The Characterisation of Trypanosomal Type 1 DnaJ-like Proteins

THE DEPARTMENT OF BIOCHEMISTRY, MICROBIOLOGY AND BIOTECHNOLOGY

FACULTY OF SCIENCE

RHODES UNIVERSITY

By

Michael Hans Ludewig

2009

The Characterisation of Trypanosomal Type 1 DnaJ-like Proteins

Thesis

Submitted in Fulfilment of the Requirements of the Degree of

Doctor of Philosophy
(Biochemistry)

IN THE DEPARTMENT OF BIOCHEMISTRY, MICROBIOLOGY AND BIOTECHNOLOGY

FACULTY OF SCIENCE

RHODES UNIVERSITY

By

Michael Hans Ludewig

2009

Declaration

I declare that this thesis is my own, unaided work. It is being submitted for the degree of Doctor of Philosophy (Biochemistry) at Rhodes University. Where the work of other researchers has been included in the thesis, the person responsible for producing the work has been appropriately acknowledged. It has not been submitted for any other degree or examination at any other university.

20 day of DECEMBER 2009

ABSTRACT

Trypanosomes are protozoans, of which many are parasitic, and possess complex life-cycles which alternate between mammalian and arthropod hosts. As is the case with most organisms, molecular chaperones and heat shock proteins are encoded within the genomes of these protozoans. These proteins are an integral part of maintaining the structural integrity of proteins during normal and stress conditions. Heat shock protein 40 (Hsp40) is a co-chaperone of heat shock protein 70 (Hsp70) and in some cases can act as a chaperone. These proteins work together to bind non-native polypeptide structures to prevent unfolded protein aggregrate formation in times of stress, translocate proteins across organelle membranes, and transport unsalvageable proteins to proteolytic degradation by the cellular proteasome. Hsp40s are divided into four types based on their domain structure.

Analysis of the nuclear genomes of eight trypanosomatid species revealed that less than 10 of the approximate 70 Hsp40 sequences per genome were Type I Hsp40s, many of which contained putative orthologues in the other seven trypanosomatid genomes. One of these Type I Hsp40s from *T. b. brucei*, *Trypanosoma brucei* DnaJ 2 (Tbj2), was functionally characterised in *T. brucei brucei*. RNA interference knockdown of expression in *T. brucei brucei* showed that cells deficient in Tbj2 displayed a severe inhibition of the growth of the cell population. The levels of the Tbj2 protein population in *T. brucei brucei* cells increases after exposure to 42 °C and the protein was found to have a generalized cytoplasmic subcellular localization at 37 °C. These findings provide evidence that Tbj2 is an orthologue of Yeast DnaJ 1 (Ydj1), an essential *S. cerevisiae* protein.

Hsp40s interact with their partner Hsp70s through their J-domain. The amino acids of the J-domain important for a functional interaction with Hsp70 were examined in *Trypanosoma cruzi* DnaJ 2 (Tcj2) (the orthologue of Tbj2) and *T. cruzi* DnaJ protein 3 (Tcj3) by testing their ability to substitute for Ydj1 in *Saccharomyces cerevisae* and for DnaJ in *Escherichia coli*. In both systems, the positively charged amino acids of Helix II

and III of the J-domain disrupted the functional interaction of these Hsp40s with their partner Hsp70s. Substitutions in Helix I and IV of the J-domains of Tcj2 and Tcj3 produced varied results in the two different systems, possibly suggesting that these helices serve to define with which Hsp70s a given Hsp40 can interact.

The inability of an Hsp40 and an Hsp70 to interact functionally does not necessarily mean a total absence of physical interaction between these proteins. The amino acid substitution of the histidine in the HPD motif (H34Q) of the J-domain of Tcj2 and Tcj3 removed the ability of these proteins to interact functionally with *S. cerevisiae* Hsp70 (Ssa1) *in vivo*. However, preliminary binding studies using the quartz crystal microbalance with dissipation monitoring (QCM-D) show that Tcj2 and Tcj2(H34Q) both physically interact with *M. sativa* Hsp70 *in vitro*.

This study is the first report to provide evidence that certain trypanosomal Type I Hsp40s are essential proteins. Futhermore, the interaction of these Hsp40s with Hsp70 identified important features of the functional interface of this chaperone machinery.

This thesis is dedicated to my family

My parents

Werner Ludewig

and

Raylene Joy Ludewig

my sister

Helga Els

and her husband

Shane Mark Els

without whom this thesis would not have been possible

Acknowledgements

I would firstly like to thank my parents for their support along this journey. Thank you for your love, support and everything you have done for me to get me to this stage.

A very big thank you to my supervisor Prof. Gregory L. Blatch, for his patience over the past number of years and for his guidance and support at every stage. I am also appreciative of his dedication and enthusiasm to his subject, which is very much infectious and creates a fantastic working environment.

I would also like to thank Dr David Horn (London School of Tropical Hygiene and Medicine, London, UK) for accommodating me in his laboratory for two months, for training me in *T. brucei* cell culture and RNAi and for his generosity with his time. Furthermore, I thank the members of Dr Horn's research group for being happy to share their laboratory and expertise with me. I am also most deeply grateful to Dr Horn for generating the fluorescence microscopy images of the subcellular localization of GFP-Tbj2 as I was unable to complete this during the time of my study visit.

I owe a very big thank you to Cassandra Alexandrovna Louw, Dr. Eva-Rachele Pesce and Dr Earl Prinsloo for kindly proofing the drafts of this thesis. Many thanks for all the time. You guys are great friends who have managed to keep me sane for the last few years.

Many thanks to Ronen Fogel for his sharing his expertise of the QCM-D.

To the members (past and present) of the Biomedical and Biotechnology Research Unit at Rhodes University for sharing the journey and thank you for the memories

Many thanks to the following people for strains, plasmids and proteins:

- Dr Elizabeth Craig (University of Wisconsin Medical School, USA) for the JJ160 *S. cerevisiae* strain and for the pRS317-Ydj1 control plasmid.
- Dr Michael E. Cheetham, University College London Institute of Opthalmology, London, UK - *Bos taurus* brain Hsc70.
- Dr Oliver Deloche (Département de Biochimie Médicale, Centre Médical Universitaire, Geneva, Switzerland) - *E. coli* OD259.
- Dr David Engman (Northwestern University Medical School, Chicago, USA) pET28aTcj2.
- Dr Petra Gentz (Department of Biochemistry, Microbiology and Biotechnology, Rhodes University) - pKG6.
- Professor Richard Zimmerman (Department of Medical Biochemistry, University of the Saarland Medical School, Homburg Germany) – manufacturing anti-Tbj3 antibody.

I would like to thank the National Research Foundation of South Africa and the Deutscher Akademischer Austauch Dienst (DAAD) for providing funding for this project. In addition I am very grateful to the International Union of Biochemistry and Molecular Biology (IUBMB) for granting me a Wood-Whelan travel award to fund my traveling expenses.

"Dulcius ex asperis" Clan Fergusson
"Aut viam inveniam aut faciam" Hannibal

Table of Contents:

List of Figures

Chapter 3:

Chapter 4:

Chapter 5:

Appendices:

List of Tables:

List of Abbreviations

1N:1K	1 nucleus: 1 kinetoplast
1N:2K	1 nucleus: 2 kinetoplasts
2N:2K	2 nucleus: 2 kinetoplasts
~	Approximately
µg	Micrograms
µM	Micromolar
°C	Degrees Celsius
A	Absorbance
Aa	Amino acids
ADP	Adenosine diphosphate
APS	Ammonium persulphate
ATP	Adenosine triphosphate
ATPase	Adenosine triphosphatase
Bag-1	Bcl2-associated anthanogene
bp	Base pairs
BSA	Bovine Serum Albumin
Cbpa	Curved DNA binding protein
CDS	Coding sequence
DNA	Deoxyribonucleic acid
DAPI	4,6-Diamino-2-phenylindole
DnaJ	Prokaryotic form of Hsp40
DnaK	Prokaryotic form of Hsp70
DMSO	Dimethyl Sulphoxide
DOGS-NTA	1,2-Dioleoyl-*sn*-Glycero-3-[(N-(5-amino-1-carboxypentyl)iminodiacetic acid) succinyl]
DTT	Dithiothreitol
E. coli	*Escherichia coli*
EDTA	Ethylene diamine tetra-acetic acid
FPLC	Fast Protein Liquid Chromatography
FTase	Farnesyl Transferase
xg	Measure of acceleration relative to standard gravity or the acceleration produced by the earths gravitational force at sea level
GGTase	Geranylgeranyl transferase
GPI	Glycophosphatidylinositol
gRNA	Guide RNA
GrpE	GroP-like protein
His Tag	6 x Histidine Tag
HSE	Heat shock element
HSF	Heat shock transcription factor
HSP	Heat Shock Protein
Hsp70	Heat Shock Protein 70
Hsp40	Heat shock protein 40
Hz	Hertz
IPTG	Isopropyl β D-thiogalactoside
kDa	Kilo Dalton
kDNA	Kinetoplast DNA
M	Molar (Moles/litre)
mg	Milligrams
ml	Millilitres

mRNA	Messenger RNA
NEF	Nucleotide exchange factor
NTA	Nitriloacetic acid
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
PDB file	Protein Data Bank file
pKa	Dissociation constant of an acid
POPC	1-Palmitoyl-2-Oleoyl-*sn*-Glycero-3-Phosphocoline
QCMD	Quartz Crystal Microbalance with Dissipation monitoring
RNA	Ribonucleic acid
RNAi	RNA interference
RPM	Revolutions per minute
S. cereviseae	*Saccharomyces cerevisiae*
SDS	Sodium Dodecyl Sulphate
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
S_k	S phase of the trypanosome kinetoplast
SmHsp	Small Heat Shock Protein
S_n	S phase of the trypanosome nucleus
SPB	Supported Phospholipid Bilayer
SPR	Surface Plasmon Resonance
SRA	Serum resistance associated protein
T. brucei	*Trypanosoma brucei*
T. b. rhodesiense	*Trypanosoma brucei rhodesiense*
T. b. gambiense	*Trypanosoma brucei gambiense*
Tbj2	*Trypanosoma brucei* J protein 2
Tbj3	*Trypanosoma brucei* J protein 3
TAE	Tris-Acetate-EDTA
TBS	Tris buffered saline
TBS-Tween	Tris buffered Saline with Tween 20
TcHsp70	*Trypanosoma cruzi* Hsp70
Tcj2	*Trypanosoma cruzi* J protein 2
Tcj3	*Trypanosoma cruzi* J protein 3
TE Buffer	Tris-EDTA
TEMED	N,N,N'N'-tetramethylethylenediamine
Tris	Tris-2-amino-2-(hydroxymethyl)-1,3,propandiol
U	Units (of enzyme-a measure of enzyme used in molecular biology reactions)
VSG	Variable Surface Glycoprotein
Ydj1	Yeast DnaJ 1
YMM	Yeast Minimal Medium
YMM galactose	Yeast minimal medium containing galactose
YMM glucose	Yeast minimal medium containing glucose
YT broth	Yeast-Tryptone Broth
ZBD	Zinc binding domain
ZFLR	Zinc finger-like region.

Publications

Publications:

Louw, C.A., **Ludewig, M.H**. and Blatch, G.L. Overproduction, purification and characterisation of Tbj1, a novel Type III Hsp40 from *Trypanosoma brucei*, the African Sleeping Sickness parasite. *Protein Expression and Purification*. 69: 168-177.

Nicoll, W.S., Botha, M., McNamara, C., Schlange, M., Pesce, E.-R., Boshoff, A., **Ludewig, M.H.**, Zimmermann, R., Cheetham, M.E., Chapple, J.P. and Blatch, G.L. (2007) Cytosolic and ER J-Domains of mammalian and parasitic origin can functionally interact with DnaK. *The International Journal of Biochemistry and Cell Biology*. 39: 736-751.

Longshaw, V.M., Nicoll, W.S., Botha, M., **Ludewig, M.H.**, Shonhai, A., Stephens, L.L. and Blatch, G.L. (2006) Getting practical with molecular chaperones. *Biotech International* 18: 24-27.

Nicoll, W.S., Boshoff, A., **Ludewig, M.H.**, Hennessy, F., Jung, M. and Blatch, G.L. (2006) Approaches to the isolation and characterisation of molecular chaperones. *Protein Expression and Purification* 46: 1-15.

Boshoff, A., Nicoll, W.S., Hennessy, F., **Ludewig, M.H.**, Daniel, S., Modisakeng, K.W., Shonhai, A., McNamara, C., Bradley, G. and Blatch. G.L. (2004) Molecular chaperones in biology, medicine and protein biotechnology. *Special Rhodes University Centenary Issue of S. Afr. J. Sci.* 100: 665-677.

Edkins, A.L.; **Ludewig, M.H.** and Blatch, G.L. (2004) A *Trypanosoma cruzi* heat shock protein 40 is able to stimulate the adenosine triphosphate hydrolysis activity of heat shock protein 70 and can substitute for a yeast heat shock protein 40. *The International Journal of Biochemistry & Cell Biology* 36: 1585-1598.

Publications in preparation:

Louw, C.A., **Ludewig, M.H**. and Blatch, G. L. The Hsp70 complement of the Tritryps: Facilitating survival of ancient organisms. *In preparation.*

Ludewig, M.H., Horn, D. and Blatch, G.L. The *in vivo* characterisation of the *T. brucei brucei* Hsp40s Tbj2 and Tbj3. *In prepration.*

Ludewig, M.H., Louw, C.A. and Blatch, G.L. The *in silico* characterisation of the kinetoplastid Hsp40 protein family. *In prepration.*

Ludewig, M.H. and Blatch, G.L. The *in vitro* characterisation of the interaction between *Trypanosoma cruzi* Hsp40 (Tcj2) and its partner Hsp70. *In prepration.*

Conference Presentations:

Ludewig, M.H. and Blatch, G.L. (2010) The characterization of the interaction of trypanosomal Type I Hsp40s with Hsp70. *South African Society of Biochemistry and Molecular Biology, Bloemfontein, South Africa.*

Ludewig, M.H., Horn, D. and Blatch, G.L. (2009) The Characterization of Type 1 and Type I-like Hsp40s of Trypanosomatid Species. *European Science Foundation conference on infectious Diseases, Somerset West, Cape Town, South Africa.*

Ludewig, M.H., Nicoll, W.S. and Blatch, G.L. (2005) The Characterisation of the J-domains of Trypanosomal Type I DnaJ-like Proteins. *South African Society of Biochemistry and Molecular Biology, Stellenbosch, South Africa.*

Chapter 1:
Literature Review

CHAPTER 1
Literature Review

1.1) The Kinetoplastids

Kinetoplastids (Domain: **Eukaryota**; Kingdom: **Excavata**; Phylum: **Euglenoza**) are a group of free-living and parasitic flagellate protozoa (Simpson *et al.*, 2002). This class of organisms has most often been subdivided into the Orders Trypanosomatida and Bodonida. However, later classification studies have proposed a new division of the class into Prokinetoplastina (consisting of *Icthyobodo* and *Perkinsiella* bodonid groups) and Metakinetoplastina (containing the remaining three clades of bodonids [Neobodonida, Parabodonida and Eubodonida] and the whole of the previous Trypanosomatida) (Moreira *et al.*, 2004). These organisms have a unique mitochondrial DNA architecture which is organized into an organelle called the kinetoplast that forms part of the single mitochondrion (Donelson *et al.*, 1999; Simpson *et al.*, 2002). Other distinguishing features of some or all kinetoplastids include: 1) complex and energy-consuming mitochondrial RNA editing, 2) trans-splicing of all RNA transcripts, 3) the arrangement of genes into giant polycistronic clusters; 4) unprecedented modifications of nucleotides, 6) compartmentalization of glycolysis, 6) evasion of the host cellular immune response using a variable surface protein coat, 7) the ability to escape destruction by the immune system by migrating out of phagocytic vacuoles (Simpson *et al.*, 2006; Donelson *et al.*, 1999).

Kinetoplastids are ancient Eukaryotes that have rRNA lineages older than animals, plants and fungi (Donelson *et al.*, 1999). It is believed that parasitism evolved on four separate occasions within kinetoplastids to produce: 1) the *Icthyobodo-Perkinsiella* clade, 2) Fish infecting *Cryptobia* (trypanoplasma) species, 3) the true *Cryptobia* and 4) trypanosomatids (Simpson *et al.*, 2006). A disproportionate amount of data has been generated for the trypanosomatids as this group contains a number of species that are of great importance to human medicine and the economy of developing countries. The most notable examples of such trypanosomatids are *Trypanosoma cruzi*, *Trypanosoma brucei* and various *Leishmania* species (Simpson *et al.*, 2006).

1.2) The Trypanosomatida

The trypanosomatids are a group of protozoa that are all parasitic. They have been classified into nine genera that consist of those that infect only insects and those genera in which the parasites cycle between invertebrates and vertebrates or plants (Simpson *et al.*, 2006). Of particular interest to humanity are the genera that infect humans or domestic animals, of which *Trypanosoma brucei* and *Trypanosoma cruzi* are two examples. *T. brucei* is subdivided into three sub-species: *T. brucei brucei*, *T. brucei gambiense* and *T. brucei rhodesiense*. Of these, only *T. b. gambiense* and *T. b. rhodesiense* can cause human disease, but all three are known to cause disease in wild and domestic animals (Vanhamme and Pays, 2004). Apolipoprotein L-1 (Apo L-1) is a tryapanolytic factor bound to high density lipoprotein (HDL) particles that is able to kill *T. b. brucei*, but the human infective trypanosomes have mechanisms of resistance (Barrett *et al.*, 2003). The *T. b. rhodesiense* serum resistance associated protein (SRA) is a variable surface glycoprotein (VSG)-like protein that is able to neutralize ApoL-1, while *T. b. gambiense* appears to have resistance by another mechanism. *T. b. gambiense* group 1 has constitutive resistance to human serum, while *T. b. rhodesiense* is reversible if the parasites are passaged in animals for a long period (Vanhamme and Pays, 2004). *Trypanosoma cruzi* has been divided into two subgroups. Lineage 1 (*T. cruzi* I) is more prevalent in the sylvatic infection cycle (affecting wild rather than domestic animals), while lineage 2 (*T. cruzi* II) is found in the domestic infection cycle between humans and domestic animals (Barret *et al*,. 2003; El-Sayed *et al.*, 2005). Lineage 2 has 5 subgroups designated IIa, IIb, IIc, IId and IIe (El-Sayed *et al.*, 2005). *T. brucei* have been classified into the salivaria as their transmission is through the saliva of their insect vector, while *T. cruzi* have been classified into the stercoraria as they are transmitted through the vector faeces (Barrett *et al.*, 2003).

1.2.1) *Trypanosoma brucei* and Human African Trypanosomiasis

The human infective subspecies of *T. brucei* cause Human African Trypanosomiasis (HAT, also known as African Sleeping Sickness). *T. b. gambiense* generates a chronic version of Human African Trypanosomiasis in Western and Central Africa, while *T. b. rhodesiense* causes a more acute version of the disease in Eastern and Southern

Africa. Both organisms cause a fatal infection if left untreated, but *T. b. rhodesiense* infection is often fatal within weeks as opposed to years in the case of *T. b. gambiense* (Barrett *et al.*, 2007).

HAT has two distinct stages: Stage 1 (hemolymphatic) and Stage 2 (encephalitic). The site of inoculation may develop a chancre caused by proliferation of parasites in that region. The parasites migrate to the draining lymph node and the bloodstream, thus initiating the hemolymphatic stage of the disease. Periods of fever occur that last from 1 to 7 days, general malaise, headache, weakness and weight loss include some of the non-specific symptoms of this stage (Barrett *et al.*, 2003; Kennedy, 2006). The acute *T. b. rhodesiense* infection can lead to fatalities from pancarditis with congestive heart failure, pulmonary oedema and pericardial effusion (Barrett *et al.*, 2003). During stage 2 of the infection, the parasites invade the central nervous system (CNS) and other internal organs. This happens within a few weeks in *T. b. rhodesiense* and months to years in *T. b. gambiense*. Symptoms resulting from CNS invasion have been characterized as sleep disturbances as well as psychiatric, motor and sensory abnormalities. The sleep disturbances resulted in the name African Sleeping Sickness, due to upset of the circadian rythms resulting in nocturnal insomnia and diurnal sleepiness. Demyelination may also occur, which may leave permanent damage even if the infection is cured (Barrett *et al.*, 2003; Kennedy, 2006).

HAT is endemic to portions of 36 African countries in which it is estimated that 60 million people are at risk of the disease and approximately 50 000 to 70 000 fatalities occur per annum (Kennedy, 2006; Matthews, 2005). The exact number of cases is uncertain due to inadequate case reporting in rural areas (Barret *et al.*, 2003). In the later part of the 20th century, there has been a resurgence of HAT due to a relaxation of vector control (Garcia *et al.*, 2006). In certain provinces of Angola, Congo and southern Sudan, it has surpassed the number of deaths attributed to HIV/AIDS (Matthews, 2005). As HAT occurs in some of the poorest populations in the world, effective treatment does not exist for these diseases due to the financial disincentive for novel drug development and rising resistance to current drug treatments (Barrett, 2003; Barrett *et al.*, 2007).

1.2.2) *Trypanosoma cruzi* and Chagas disease

Chagas Disease (also known as American Trypanosomiasis) is caused by *Trypanosoma cruzi*, which is endemic to regions of Central and South America (El-Sayed *et al.*, 2005). *T. cruzi* has two distinct stages of infection designated as the acute and the chronic stages. The acute stage occurs 6 to 10 days after infection and can last up to 2 months. Inflammation of the site of infection often produces an oedematous swelling called a chagoma. Symptoms that occur include general malaise, fever, rash, hepatosplenomegaly, lymphadenopathy, facial oedema and changes in heart rhythm. Death can occur due to complications such as myocarditis and meningoencephalitis. Parasites are easily detectable in the peripheral blood system during the acute stage and the transition to the chronic stage is recognised as the reduction of circulating parasites, which makes direct parasite diagnosis difficult (Barrett *et al.*, 2003; Massad, 2008). During the chronic stage of infection, patients typically become asymptomatic and approximately three quarters remain this way for the rest of their lives. The remaining quarter develop the cardiac, nervous and digestive forms of Chagas disease 10 to 25 years after initial infection. Chagas disease affecting the digestive system usually manifests as megaoesophagus and megacolon, where these organs swell and lose smooth muscle tone, affecting swallowing and causing constipation respectively. The cardiac form of chronic chagas disease is the most common form affecting about 94% of all symptomatic cases (Barrett *et al.*, 2003; Teixeira *et al.*, 2006). Estimates of people infected with *Trypanosoma cruzi* range from 16 to 18 million, with reports of 21 000 deaths associated with Chagas disease each year (El-Sayed *et al.*, 2005).

1.2.3) Trypanosomatid Cell Biology

Trypanosomatids have a number of cellular features that are common to the majority of the group studied to date. However, their cellular morphology differs between species and according to life cycle stage (**Figure 1.1**). Promastigotes, trypomastigotes and epimastigotes are all spindle shaped cells. However, epimastigotes have the kinetoplast situated anterior to the nucleus, while the trypomastigote kinetoplast is situated behind the nucleus with the cell flagellum running along the majority of the cell body (De Souza, 2002). Amastigotes (or spheromastigotes) are almost spherical in shape and only have a

vestigial flagellum. Promastigotes have a free flagellum that is not attached to the cell surface (Clayton *et al.*, 1995). The flagellum of promastigotes and epimastigotes emerges from the anterior of the cell, while it emerges from the posterior in trypomastigotes (McConville *et al.*, 2002).

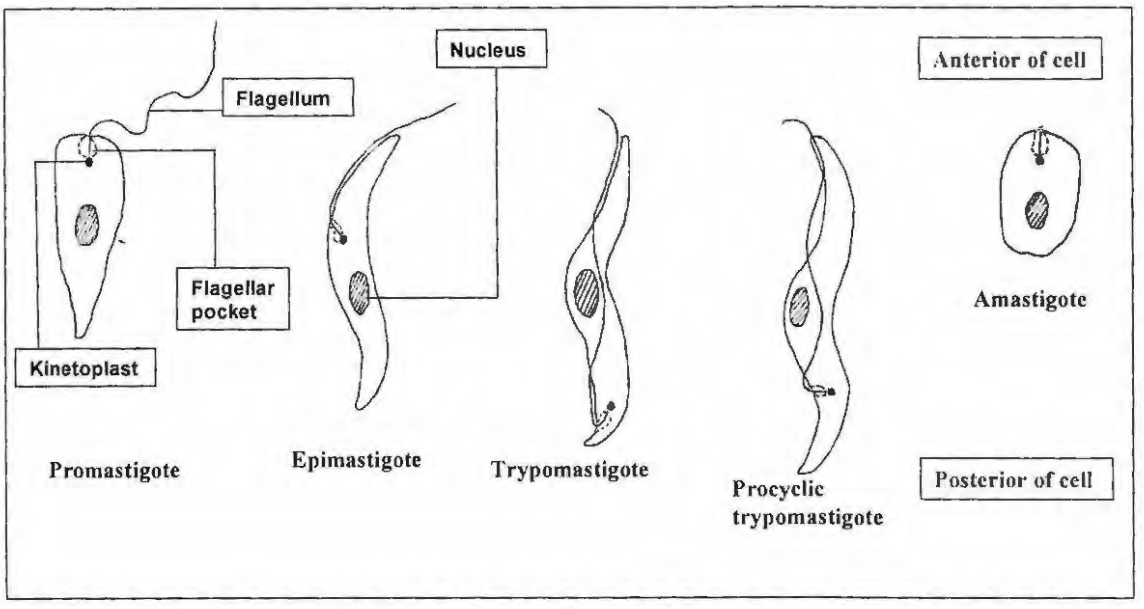


Figure 1.1: Simplified diagrams of different trypanosomatid life stages to highlight the differences in cell morphology. The organelle symbols for each diagram are the same. The cells are categorized mainly by the location of the attachment site of the flagellum relative to the nucleus. Although the epimastigote, trypomastigote and procyclic trypomastigote are all spindle shaped cells, the kinetoplast and flagellar pocket are all in different locations within the cell. Both trypomastigote forms have the majority of the flagellum running along and attached to the side of the cell. The promastigote has a free flagellum. The amastigote is almost spherical in shape and only has a vestigial flagellum that hardly extends beyond the flagellar pocket. Figure adapted from Clayton *et al.*, 1995; McConville *et al.*, 2002.

The shape of these various forms is maintained by an array of microtubules that run just below the surface of the plasma membrane of the cell **(Figure 1.2)**. The sub-pellicular microtubules run along an anterior-posterior axis and have a distinct polarity with positive ends at the point of flagellar exit (McConville *et al.*, 2002). They are cross-linked to each other and attached to the plasma membrane below which they run (Gull, 1999). This network of tubules remains intact, even during cell division (Matthews, 2005). All trypanosomatids have a flagellum which has recently been viewed as a distinct organelle (Fridberg *et al.*, 2007). It emerges from a specialized invagination of the cell surface called the flagellar pocket located at the anterior or along the side of the cell

depending on the cell morphological type (De Souza, 2002). Certain trypanosomatids have a free flagellum, but the epimastigote and trypomastigote cell morphologies have at least part of the flagellum attached along a specialized linear portion of the cell membrane called the flagellar attachment zone (FAZ) (Clayton *et al.*, 1995). This attachment appears to be mediated by desmosomes (De Souza, 2002). In addition to the conventional pattern of 9 + 2 microtubules in the flagellar axoneme, a fibrous structure called the paraflagellar rod runs along side the axoneme and is thought to play a role in cell motility (Landfear and Ignatushchenko, 2001). The flagellum is anchored into the

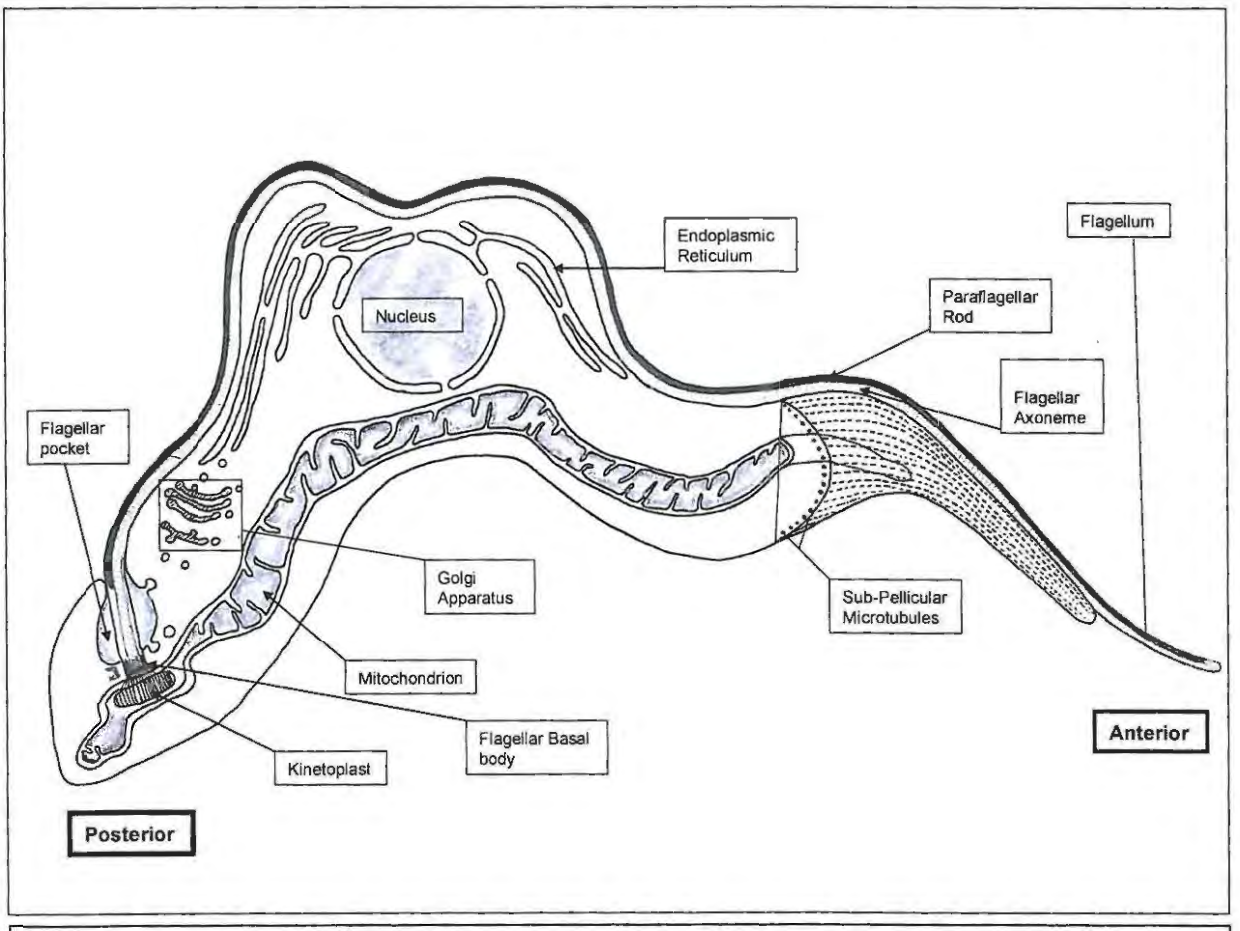


Figure 1.2: The generalized ultrastructure of a trypanosomatid trypomastigote showing the characteristic organelles of the group. The cell is shown sectioned along the axis of the flagellum, except for the anterior portion of the cell. The anterior portion of the cell shows the arrangement of sub-pellicular microtubules below the surface of the plasma membrane that maintain the cellular shape. Note the location of the golgi apparatus and the endoplasmic reticulum relative to the flagellar pocket for the secretory pathway. The flagellar basal body is connected to the kinetoplast and emerges from the opening of the flagellar pocket. It is attached along the length of the cell. Trypanosomatids typically have one mitochondrion that extends through most of the cell length. Figure adapted from Clayton *et al.*, 1995; Matthews, 2005; Field and Carrington, 2004; McConville *et al.*, 2002; Landfear and Ignatushchenko 2001.

cell below the flagellar pocket by a basal body which is directly attached to the single mitochondrion (Clayton *et al.*, 1995). In addition to motility, the flagellum also plays a role in chemotaxis, cell signaling and host cell invasion (Fridberg *et al.*, 2007). The corset of subpellicular microtubules acts as an efficient barrier to vesicular transport. Hence all vesicular traffic into and out of the trypanosome is restricted to the flagellar pocket region, which is not associated with the dense sub-pellicular microtubule corset (Gull, 2003; Field and Carrington, 2004). Only four specialized microtubules are associated with the flagellar pocket. This compartment is at least partially isolated from the extracellular medium by physical interactions between the cellular membrane at the outer rim of the flagellar pocket and the flagellar membrane by means of desmosomes (Clayton *et al.*, 1995; Landfear and Ignatuschenko, 2001). Both secreted and membrane bound proteins are delivered to the cell exterior and cell surface through this organelle, which represents 0.4 % to 3 % of the total cellular surface (Landfear and Ignatuschenko, 2001). An additional area specialized for endocytosis known as the cytopharynx occurs in some trypanosomatid species such as *T. cruzi* (epimastigote and amastigote stages), but is absent in *T. brucei*. Cellular location of this organelle is highly variable among various trypanosomatid species and cell stages (McConville *et al.*, 2002).

Trypanosomatids typically have one mitochondrion per cell, the size and shape of which varies according to the morphological cell type (De Souza, 2002). For example, the mitochondrion is diminished in size within the bloodstream form of *T. brucei* compared to the well developed mitochondrion of procyclic forms (Lukes *et al.*, 2005; Matthews, 2005). This is due to the lack of mitochondrial respiration in the bloodstream form as energy is generated through glycolysis of host blood glucose in glycosomes. However, the procyclic form, not having an abundant energy source, has a more conventional mitochondrion activity (Matthews, 2005). A specialized section of the mitochondrion called the kinetoplast contains the mitochondrial genomic DNA (referred to as Kinetoplast-DNA or kDNA because of its location in the kinetoplast) (Lukes *et al.*, 2005). This is the characteristic organelle for the Kinetoplastida (Clayton *et al.*, 1995).The kinetoplast is physically attached to the basal body of the flagellum through a structure called the tripartite attachment complex that traverses the mitochondrial

membranes (Gull, 2003). The structure of kDNA varies for different trypanosomatids. All contain varying circles of DNA that are divided into smaller circles (minicircles) and larger circles (maxicircles), which may be catenated into a network or free from each other. kDNA may be localized to a single region of the kinetoplast or spread across the entire organelle. Most of these minicircles do not supercoil and remain in a covalently closed conformation. The maxicircles code subunits for mitochondrial respiratory complexes and ribosomal RNA genes, while the minicircles are thought to encode guide RNAs (gRNAs) (Lukes *et al.*, 2005). These gRNAs act as templates to post transcriptionally modify the maxicicle transcripts (Matthews, 2005). Among other proteins associated with the kinetoplast in *T. cruzi* are mitochondrial chaperones (De Souza, 2002).

The single golgi apparatus of trypanosomatids is normally found between the nucleus and the flagellar pocket, and is well situated for post translational modification of proteins, targeted to the exocytic pathway, that have to move through the flagellar pocket (Clayton *et al.*, 1995; McConville *et al.*, 2002). The golgi apparatus of *T. brucei* bloodstream form cells is more elaborate than in the procyclic form, which possibly indicates a decrease in glycan processing requirements of the procyclic stage which performs far less complicated glycan synthesis (Field and Carrington, 2004). The endoplasmic reticulum of trypanosomatids is present throughout the cell, but can be subdivided into various subdomains that may differ slightly in function (McConville *et al.*, 2002). The flagellar pocket, flagellum, kinetoplast, mitochondrion and nucleus are precisely positioned within the cell and these positions may change between the various life cycle stages (Matthews, 2005).

1.2.4) Genome structure and RNA processing in *T. cruzi* and *T. brucei*

Trypanosomes are diploid organisms (Berriman *et al.*, 2005; El-Sayed *et al.*, 2005). The nuclear genomes of *T. brucei* and *T.cruzi* have both been sequenced (El Sayed *et al.*, 2005; Berriman *et al.*, 2005). In *T. brucei*, the nuclear genome consists of 11 megabase-sized chromosomes and over 100 mini-chromosomes (~50 kb) and intermediate chromosomes (~700kb) (Matthews, 2005; Berriman *et al.*, 2005). These

minichromosomes each contain a number of VSG genes. The *T. brucei* mitochondrial genome consists of ~50 copies of the maxicircle DNA and ~10 000 copies of the minicircle DNA catenated together into a network (Matthews, 2005).

The *T. cruzi* kDNA accounts for about 20 to 25% of total DNA in the cell and its kinetoplast consists of 20 000 to 30 000 individual minicircles of DNA (De Souza, 2002). The genome of *T. cruzi* is estimated to contain 55 megabase pairs of sequence compared to the 35 megabase pairs of *T. brucei* (El-Sayed *et al.*, 2005). Not all of the trypanosomatid genes are expressed in every life cycle stage (Atwood *et al.*, 2005; Jones *et al.*, 2006; Bridges *et al.*, 2008; Vertommen *et al.*, 2008). Gene expression in trypanosomatids is mainly regulated at the post-transcriptional level. The open reading frames are organised into arrays that are transcribed as polycistronic RNAs that are then divided into individual mRNAs by a complex RNA editing system. Differential regulation of transcription of certain gene clusters by RNA polymerase II does not appear to occur. Few regulatory transcription factors are coded by trypanosomatids and transcription appears to be globally downregulated in non-dividing forms. Exceptions to this form of gene expression control are the VSG and EP procyclin of *T. brucei*. These genes are transcribed by RNA polymerase I from dedicated promoters that appear to be developmentally regulated through chromatin mediated silencing (Clayton and Shapira, 2007).

1.2.5) Trypanosomatid life cycles

Trypanosomatid life cycles are complex and various distinct forms of the parasites exist during their life cycle, which can be identified by their morphology and location. Not all of these stages are present in all parasite species (De souza 2002). Both *T. cruzi* and *T. brucei* have life cycles that involve stages in hematophagous insects and mammalian hosts (**Figure 1.3**).

1.2.5.1) Life cycle of *Trypansoma cruzi*

T. cruzi is transmitted by insects of the family reduviidae (Class: **Insecta**, Order: **Hemiptera**; Suborder: **Heteroptera**; Superfamily: **Cimicomorpha**) such as *Triatoma*

infestans, Rhodnius prolixus and Panstrongylus megistus (De Souza, 2002). These bugs ingest a pleiomorphic mixture of bloodstream trypomastigotes and amastigotes from the peripheral circulatory system while feeding on a *T. cruzi* infected mammal. In the reduvid stomach, these parasites differentiate to epimastigotes that are able to attach to the midgut intestinal cells through hemidesmosomes (De Souza, 2002; Tyler and Engman, 2001).

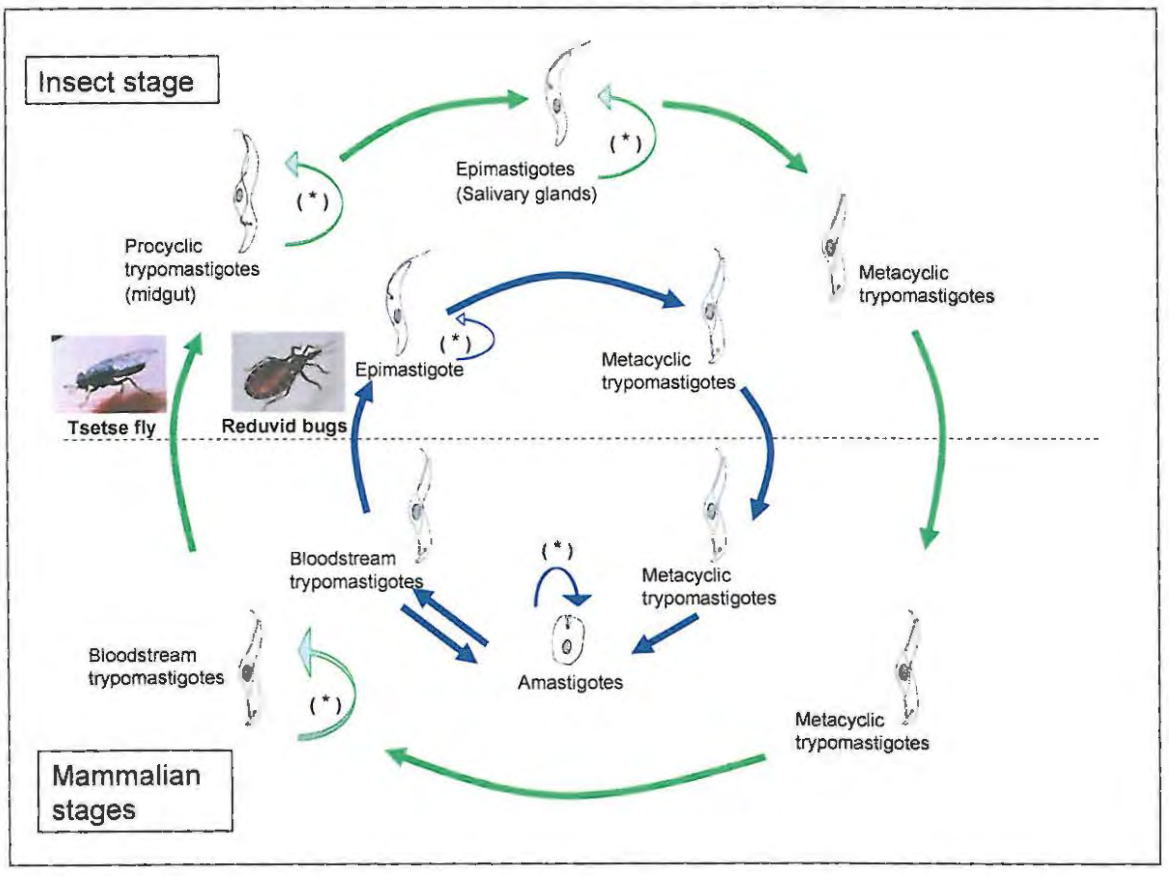


Figure 1.3: The life cycle of *Trypanosoma brucei* (green) and *Trypanosoma cruzi* (blue). Both parasites move between invertebrate and mammalian systems during their life cycles. The replicative stages (*) of *T. brucei* are procyclic trypomasitgotes, epimastigotes and bloodstream trypomastigotes, while the epimastigotes and amatigotes of *T. cruzi* are replicative. In the mammalian host, *T. cruzi* amastigotes convert to bloodstream trypomastigotes. Some of these infect other cells and convert directly into amastigotes without going through the other life cycle stages. (Adapted from Folgueira and Requena, 2007; Matthews 2005; McConville *et al.*, 2002). For more details of the individual life cycles please see **section 1.2.5.1** for the *T. cruzi* life cycle and **section 1.2.5.2** for the *T. brucei* life cycle.

The epimastigotes divide repeatedly through binary fission, some of which migrate to the rectum of the insect where they attach hydrophobically to the waxy cuticle, where they differentiate into metacyclic trypomastigotes. The mature metacyclic trypomastigotes

detach from the hindgut wall and are excreted with the faeces (Tyler and Engman, 2001). Scratching of the irritating bites of the insects leads to contamination of the bite wound, micro-abrasions of the skin or mucous membranes with the faecal matter containing the metacyclic trypomastigotes (Barrett *et al.*, 2003). The trypomastigotes invade the cells of the mammalian host and transform into amastigotes, which are able to proliferate through binary fission. The population of amastigotes then elongate into bloodstream trypomastigotes that are released into the intercellular spaces upon rupture of the infected cell. The trypomastigotes invade neighbouring cells or invade the blood stream and travel to other areas of the mammalian host's body and infect other cells. Host cells do sometimes rupture before all amastigotes have transformed into trypomastigotes. Hence, amastigotes and bloodstream trypomastigotes can both be ingested by the reduvid vector to start a new cycle in the invertebrate host. Unlike *Trypanosoma brucei*, the *T. cruzi* bloodstream trypomastigote is not known to divide in the blood stream of the mammalian host (De Souza, 2002). Various intermediate forms have been observed between the various stages of differentiation and a number of mechanisms of differentiation have been proposed (Tyler and Engman, 2001).

1.2.5.2) Life cycle of *Trypanosoma brucei*

Unlike many other major human parasitic trypanosomatids (*T. cruzi* and *Leishmania spp.*), *Trypanosoma brucei* along with most other African trypanosomes are extracellular throughout their life cycle (Pays *et al.*, 2004). *T. brucei* spp. are transmitted to their mammalian hosts by infected flies of the genus *Glossina* (Class: **Insecta**, Order: **Diptera**; Suborder: **Brachycera**; Superfamily: **Hippoboscoidea;** Family: **Muscidae**; Subfamily: **Glossinidae**) commonly known as tsetse flies (Fraumann, 2003). The parasites are injected into the bloodstream along with saliva from the parasite-colonized salivary glands in the fly. The metacyclic trypomastigotes express the variable surface glycoprotein (VSG) that aids in the protection of the parasite against the mammalian host immune response (Gull, 2002). Upon entry into the bloodstream, they convert to the proliferative long slender bloodstream trypomastigotes. Differentiation into the non-proliferative stumpy bloodstream trypomastigote is dependent on the population density in the bloodstream and is achieved by moving through a number of intermediate forms

(Hirumi and Hirumi, 1994). This helps to increase the host survival time and the chances of parasite transmission to the vector. Arrest of the stumpy bloodstream forms in the G1 phase of the cell cycle also ensures that the majority of the bloodstream population will be in the correct cell cycle phase for the successful proliferation of the tsetse midgut procyclic forms (Matthews, 2005). In the bloodstream forms, the parasite mitochondrion activity is suppressed and the parasite becomes dependent on glycolysis in the glycosome to generate energy (Gull, 2002; Matthews, 2005).

When a tsetse fly takes a blood meal from an infected mammal, bloodstream form trypomastigotes enter the fly midgut and convert to proliferative procyclic trypomastigotes. This stage replaces the VSG coat with procyclins that produce a less dense surface coat (Roditi and Liniger, 2002). Insects have a more primitive immune system, and therefore the constant changing of the surface antigen is not required (Field and Carrington, 2004). The mitochondrion becomes structurally more elaborate and plays a larger role in energy generation. The kinetoplast also moves further forward from the extreme posterior of the cell. The parasites migrate to the proventriculus and then the salivary glands of the insect and transform into epimastigotes. Here they attach to the wall of the salivary glands by their flagellum and continue to proliferate by binary fission, followed by division arrest as they convert to metacyclic trypomastigotes and release into the salivary gland lumen. The metacyclic trypomastigotes develop the VSG coat ready for survival in the mammalian host (Matthews, 2005; Hirumi & Hirumi 1994). *T. brucei* thus shows a distinct adaptation to survive in three different environments: The tsetse fly midgut, tsetse fly salivary glands and the mammalian bloodstream (Matthews, 2005). The epidemiology of both human infective subspecies indicates a use of animal reservoirs. However, this is more important in *T. b. rhodesiense* (Barrett *et al.*, 2007).

1.2.6) Parasite-host interactions

The most important area of interaction between trypanosomatid parasite and host is the surface interface between these two entities. The parasite cell surface is where the parasite has to be able to counteract the host immune system and carry out specialized tasks for that stage of its life cycle, such as host cell invasion in the case of *T. cruzi* and

Leishmania major. The cellular surface is very different for *T. cruzi* and *T. brucei*, which indicates the difference in requirements for their life cycle (McConville *et al.*, 2002). The flagellar pocket also forms part of the cell surface and as the main portal for traffic into and out of the parasite (delivering cell surface proteins and secreted proteins), it plays a large role in host-parasite interaction (Landfear and Ignatuschenko, 2001).

The surface of *T. brucei* is dominated by single protein types in both the bloodstream forms and the procyclic insect form, that are both anchored to the plasma membrane by glycophosphatidylinositol (GPI). In the bloodstream forms, the cell surface is dominated by the variable surface glycoprotein (VSG). Approximately 10^7 copies of this 58 kDa protein are evenly distributed over each bloodstream form cell surface (Berriman *et al.*, 2005). The VSG population represents ~90% of the surface proteins of the cell, which means that this is by far the most prevalent protein in the exocytic pathway (Field and Carrington, 2004; Gull, 2003). Other proteins found attached to the cell surface include polytopic transporters that are found under the VSG protective coat and type I integral membrane proteins. These proteins either have a sporadic distribution across the entire cell surface or are restricted to the flagellar pocket. The VSG serves to protect parasites from the complement-mediated lysis by the host immune system and shelters other non-variant surface proteins from the host immune system attack. Large macromolecules are excluded by this VSG layer, but small proteins and essential nutrients are still able to penetrate the cell membrane (McConville *et al.*, 2002). *T. brucei* contains a large variety of VSG encoding genes (~1000 to 2000 potential sequences) the expression of which can be rotated in individual parasites (Berriman *et al.*, 2005; Barry *et al.*, 2005). In addition, the *T. brucei* bloodstream form cells have the ability to recycle VSGs that have been bound by antibodies, through the endocytic pathway and then re-export these VSGs to the cell surface (Field and Carrington, 2004). Both of these mechanisms assist the parasite in establishing a long term infection in the host.

The predominant surface glycoprotein of the *T. brucei* procyclic stage found in the insect vector is procyclin of which two forms exist (GPEET and EP, named after the amino acid repeats found within them) (Matthews, 2005). This coat of procyclin (also referred to as

the invariant surface glycoprotein) is thought to not display antigenic variation due to a more primitive immune system in the tsetse fly (Field and Carrington, 2004). The main function of this coat is hypothesized to prevent direct attack on the parasite cell membrane by tsetse fly gut hydrolases or the immune system. Procyclins are small, acidic and highly glycosylated and there are approximately 5 $x10^6$ copies per cell (McConville *et al.*, 2002).

Major life cycle stages of *T. cruzi* have a cell surface dominated by two sets of GPI linked glycoproteins: mucin-like glycoproteins and trans-sialidase glycoproteins. The mucins are modified with O-linked glycans. Trans-sialidases transfer sialic acid from host cell glycoproteins to the O-linked glycan regions. These host sialic acid groups are thought to protect the extracellular parasite against complement-mediated lysis and anticarbohydrate antibodies in host blood. Several additional surface proteins are also produced to counteract other components of the host complement system. Free GPI components are produced in all stages of the *T. cruzi* life cycle and form a densely packed layer under the mucin coat (McConville *et al.*, 2002). *T. cruzi* also evades the host immune system in certain stages by invading and hiding in host cells, providing a possibility of a very close interaction. The invasion of host cells is a complex process involving Ca^{2+} signal transduction in both host cells and parasites that activates a number of cellular components for uptake of the parasite by the cell into a parasitophorous vacuole, which is then lysed by components secreted by the parasite to release it into the cytoplasm (Yoshida, 2006). While inside the host cell, parasites must derive their nutrients from the cytoplasm of the cell or from material assimilated by the cell.

1.2.7) Consequences of a complex life cycle

There are two important factors that result from the parasite life cycle being divided between two different hosts. These are important to remember when studying parasite molecular biology. The first is that cellular homeostasis is not the same in every life cycle stage. For example, in *T. brucei* the rate of endocytosis in the bloodstream form parasites is far higher than in the procyclic forms of the parasite. This is possibly due to mechanisms of immune evasion like the recycling of VSG (Field and Carrington, 2004).

The other factor to remember is that not all genes are expressed in every stage of the life cycle. This is clearly illustrated with VSG and procyclin in *T. brucei*. In *T. cruzi* it is estimated that only 30% of the genes identified in the annotated genome (El-Sayed *et al.*, 2005) are expressed at all stages of the life cycle (Atwood, 2005).

1.3) Protein folding

Proteins are one of the most abundant molecules in biology and are responsible for controlling most of the chemical processes on which cells depend (Dobson, 2004). The central dogma of molecular biology states that DNA is transcribed into mRNA, which then gets translated by ribosomes into protein. However, to be functional, most proteins need this sequence of amino acids to be folded into their compact three dimensional structure. Christian Anfinsen determined that proteins are able to self assemble into their native three dimensional state utilizing only their primary structure (i.e. the sequence of amino acids) (Anfinsen, 1973). He and his coworkers also developed the hypothesis that a protein's native conformation has the lowest Gibb's Free Energy of every potential conformation (Anfinsen, 1973).

Although the mechanism has not been completely elucidated to date, certain theories have been proposed to explain aspects of this process. Folding of an unfolded polypeptide chain starts with the interaction of a small number of amino acids referred to as a folding nucleus. The remainder of the polypeptide then rapidly condenses around this nucleus, shielding the hydrophobic residues from the aqueous environment and exposing the hydrophilic residues on the surface of the structure (Fersht, 2000). The final native structure is then folded by localized conformational changes of this compact folding intermediate (Dobson, 2003). This folding process indicates a stochastic search for the native conformation as opposed to a systematic sampling of a protein's conformational space (Dobson, 2004). Levinthal's Paradox indicated that a systematic search for the native conformation of a protein would, in many cases, exceed the life of the organism by many orders of magnitude (Zwanzig *et al.*, 1992; Ruddon and Bedows 1997). The stochastic conformational search is best conceptualized by linking the Gibb's Free Energy of a polypeptide chain to its conformational properties. Correct or native-like

interactions of the amino acid side chains in the polypeptide are more stable than non-native ones and hence lower the free energy of the polypeptide structure. These native-like bonds restrict the conformational space accessible to the polypeptide chain, thus allowing the protein to fold more efficiently. Thus, the energy landscape of a polypeptide has been compared to a funnel, where there are multiple pathways a polypeptide can follow to its native state in which some are more probable and efficient than others (Wolynes *et al.*, 1995; Dill and Chan, 1997; Jahn, and Radford, 2005). Proteins larger than 100 residues are also able to use this mechanism of folding, but it is hypothesized that individual domains of these proteins can fold independently of each other in a modular fashion, until they assume their final native conformation in relation to each other (Dobson, 2003).

1.4) Protein aggregation, misfolding and the cellular environment

Protein folding can go wrong, especially in the larger multi-domain proteins that form a greater number of partially folded intermediates. These intermediates may have hydrophobic amino acid side chains or unstructured polypeptide backbone segments exposed to the aqueous environment that could lead to non-native interactions within a protein molecule, which is termed misfolding. These features also drive the interaction of these partially folded proteins through hydrophobic forces and inter-chain hydrogen bonding. The folding of these aggregated proteins into their functional native state is halted and in some cases these aggregates become insoluble and precipitate (Dobson and Karplus, 1999; Radford, 2000). To reduce these effects, *in vitro* protein folding experiments utilize very dilute protein concentrations (0.01-0.02 mg/ml) (Goldberger *et al.*, 1963; Huth *et al.*, 1993), which is very low compared to the concentration of proteins and other macromolecules within the cellular interior (300-400 mg/ml) (Ellis and Minton, 2003). It has also been found that *in vitro* protein refolding reactions yielded active protein at too slow a rate to fulfill biological requirements (Anfinsen, 1973; Hartl and Hayer-Hartl, 2002), implying that folding needs to take place far faster within the cellular environment. However, under optimized *in vitro* conditions, the renaturation of ribonuclease has been shown to take 20 minutes, despite its relatively simple monomeric structure (Ruddon and Beddows, 1997).

The intracellular environment adds a number of further challenges to the folding process. In order to produce many proteins as quickly as possible, cells form polysomes, consisting of a number of ribosomes translating the same mRNA sequence into protein at the same time. This brings many partially folded polypeptide chains into close contact with each other and may result in protein aggregation (Hartl and Hayer-Hartl, 2002). In addition, certain polypeptides are co-translationally translocated into other cellular compartments such as the Endoplasmic Reticulum (ER) or mitochondria before being able to fold correctly (Dobson, 2003). Macromolecular crowding of the intracellular environment results in 20 to 30% of the total cellular volume being occupied by macromolecules, which affects cellular reaction rates due to the excluded volume effect (Ellis, 2001). This drives macromolecular compaction in the form of protein oligomerisation and association of partially folded or misassembled proteins into aggregates (Hartl and Hayer Hartl, 2002; Minton, 2000). These processes are therefore in direct competition with correct protein folding and hence decrease the efficiency of the process within the cell (Ruddon and Beddows, 1997). Mature or native proteins can also become unfolded due to stresses such as heat or toxic chemicals, resulting in their loss of function and potentially the formation of aggregates (Outeiro and Tetzlaff, 2007).

Large scale failure of proteins to fold correctly or to remain folded, results in the malfunctioning of cellular activity and the manifestation of disease (Dobson, 2003). These misfolding diseases can be categorized into three general molecular phenotypes: 1) protein misfolding that causes loss of protein function, 2) protein misfolding that leads to incorrect localization and 3) misfolding that results in toxic folding of proteins (often in the form of misassembled/aggregated protein complexes (Thomas *et al.*, 1995). Human diseases such as some cancers (Zylicz *et al.*, 2001) and cystic fibrosis (Loo *et al.*, 1998) are examples of diseases caused by protein misfolding and loss of function. Alzheimer's disease, Huntington's disease and Parkinson's disease are examples of disorders caused by deposition of misassembled/aggregated protein. In these protein aggregation disorders, the protein aggregates are arranged into ordered fibrils referred to as amyloid fibrils (Dobson, 2004). The similar structure of these fibrils in amyloid associated disorders has been attributed to the protein-protein aggregate interactions being between the

polypeptide chain back-bone through hydrogen bonds. The polypeptides are folded into β-sheets that are organized perpendicular to the long axis of the fibril (Dobson, 2003). **Table 1.1** shows examples of the growing number of protein misfolding diseases and categorises them according to the type of molecular phenotype they produce. It has, however, been recently shown that not all mammalian amyloid structures are cytotoxic (Fowler *et al.*, 2006). Some researchers therefore prefer to use the term misassembly instead of aggregation to describe the non-functional interaction of two polypeptide chains in order to avoid confusion (Ellis, 2001).

Table 1.1: Examples of misfolding diseases found in humans that result from protein misfolding and toxic folding.

Disease	**Affected Protein**	**Phenotypic effect**	**Site of folding**
Misfolding diseases			
α-Antitrypsin deficiency (Pike *et al.*, 2002)	A-Antitrypsin	Mislocalisation	ER
Cancer (Zylicz *et al.*, 2001)	P53	Misfolding	Cytosol
Cystic fibrosis (Loo *et al.*, 1998)	Cystic fibrosis transmembrane conductance regulator	Misfolding	ER
Familial Hypercholesterolaemia (Hobbs *et al.*, 1990)	Low-density lipoprotein Receptor protein	Mislocalisation	ER
Marfan Syndrome (Wu *et al.*, 1995)	Fibrillin	Misfolding	ER
Osteogenesis Imperfecta (Lamande *et al.*, 1995)	Procollagen	Misfolding	ER
Phenylketoneuria (Gamez *et al.*, 2000)	Phenylalanine hydroxylase	Misfolding	Cytosol
Retinitis Pigmentosa (Clarke *et al.*, 2000)	Rhodopsin	Mislocalisation	ER
Tay-Sachs (Lau *et al.*, 1989)	β-Hexosaminidase	Mislocalisation	ER
Toxic Folding diseases			
Alzheimer's Disease (Katayama *et al.*, 1999)	β-Amyloid	Aggregation/misassembly	ER
Cataracts (Litt *et al.*, 1998)	α-crystallins	Aggregation/misassembly	Cytosol
Creutzfeld-Jacob Disease (Masters *et al.*, 1979)	Prions	Aggregation/misassembly	ER
Familial Amyloidosis (Colon and Kelley, 1992)	Transthyretin	Aggregation/misassembly	ER
Huntington's Disease (Carmichael *et al.*, 2000)	Huntington	Aggregation/misassembly	Cytosol
Parkinson's Disease (Auluk *et al.*, 2002)	PARK2, α-synuclein	Aggregation/misassembly	Cytosol

Cells have developed a number of complex and integrated systems to regulate the quality of protein folding by minimizing populations of potentially damaging misfolded and misassembled proteins. The system of interest in this thesis is the system of molecular chaperones.

1.5) Molecular Chaperones and the cellular stress response

The term 'Molecular Chaperone' was first proposed by Laskey and his co-workers when they described the requirement of nucleoplasmin proteins for the *in vitro* assembly of nucleosomes (Laskey *et al.*, 1978). Ellis expanded this concept to a generalized type of cellular function that incorporates all proteins that mediate the correct folding and oligomeric assembly of other cellular proteins (Ellis, 1987). The molecular chaperone concept therefore altered the view of protein folding from self assembly to chaperone assisted self assembly (Ellis *et al.*, 1989; Dobson, 2004), which is required especially for longer multi-domain polypeptide chains (Hartl and Hayer-Hartl, 2002). Many of these proteins have a highly conserved sequence and are found in most organisms from prokaryotes to mammals, which indicates the importance of their cellular function (Kiang and Tsokos, 1998; Ben-Zvi and Goloubinoff, 2001). They also exist in most sub-cellular compartments of eukaryotes (Kiang and Tsokos, 1998). It is important to note that there are also non-proteinaceous molecular chaperones. Chemical chaperones such as trehalose (a disaccharide) help to stabilize non-native proteins *in vivo* (Singer and Lindquist, 1998).

Molecular chaperones recognize unfolded polypeptide chains by binding exposed hydrophobic residues (Fink 1999; Borges and Ramos, 2005). They therefore do not increase the rate of protein folding, but rather increase the efficiency of the folding process by minimizing traffic of polypeptides into misfolding and aggregation pathways within the cell, and in some cases reversing protein aggregation/misassembly (Dobson, 2003). They do not offer any steric information for the protein folding process and are not part of the final folded macromolecular structure (Hendrik and Hartl, 1993). In their function of assisted protein folding, they are involved in processes as diverse as protein translation, unfolding, translocation, degradation and protein aggregate disassembly (Qiu *et al.*, 2006 and Dobson 2004; Fink 1999).

The discovery of the biological heat shock response by Ritossa through the observation of changes in polytene chromosome puffs in *Drosophila spp.* at different temperatures Ritossa, 1962; Ritossa, 1996) led to the discovery of a group of molecular chaperones called heat shock proteins (Hsp). Expression of these proteins is now known to be

increased during various cellular stresses including heat, free radicals, amino acid analogues, cold shock, virus infection and oxygen or glucose deprivation (Hendrik and Hartl, 1993; Lelivelt and Kawula, 1995; Liang and MacRae, 1997; Santoro, 2000). However, production of some stress inducible molecular chaperones is not induced by every kind of cellular stress (Lelivelt and Kawula, 1995). The majority of molecular chaperones are expressed constitutively and are involved in the general cellular chaperoning tasks, even if they are stress-inducible (Gething and Sambrook, 1992). Constitutively expressed molecular chaperones are referred to as heat shock cognate (Hsc) proteins (Ingolia and Craig, 1982). Eukaryotic stress inducible Hsp induction is controlled by regulatory proteins called heat shock transcription factors (HSF's) which bind to heat shock elements (HSE's) (Santoro, 2000; Kiang and Tsokos, 1998). HSE's consist of a series of repeats of the sequence 5'- nGAAn – 3' found upstream of the stress inducible Hsp gene (Santoro, 2000).

Heat shock proteins have been most commonly classified according to their molecular weight in kilodaltons: Hsp100 (~100 kDa), Hsp90 (~90 kDa), Hsp70 (~70 kDa), Hsp60 (~60 kDa), Hsp40 (~40 kDa) and the small heat shock proteins (~15-30 kDa) (Fink, 1999; Frydman, 2001; Borges and Ramos, 2005). **Table 1.2** shows some examples of each chaperone family in both prokaryotes and eukaryotes including their location and functions.

These proteins have also been functionally categorized into holding, folding and unfolding chaperones (Stirling *et al.*, 2003). Chaperones that serve a holding function bind and stabilize non-native proteins until they are passed on to be folded by other chaperones. They do not have an ATPase function to drive energy dependent processes. Hsp40s (Mayer *et al.*, 2000a) and the small Hsp's (van Montfort *et al.*, 2001a) are holding chaperones. Hsp40 can prevent protein aggregation/misassembly (Langer *et al.*, 1992) and then pass the protein to a selected Hsp70 to be folded (Mayer *et al.*, 2000a). Folding chaperones are able to capture and fold non-native proteins in an ATP-dependent process. Hsp70 is an example of a folding protein that stabilizes non-native proteins by binding to them and then releases them through an ATP dependent conformational

Table 1.2: The major families of heat shock proteins and some examples of members in these groups as well as their function and cellular localization. (Data from Hendrik and Hartl, 1993; Fink,1999).

Family	Examples of protein members	Location	Main functions of the group
Hsp100	**Prokaryote (*E. coli*)** Clp **Eukaryote** Hsp104	Cytosol Cytosol	ATP-dependent proteolysis; protein aggregate disaggregation, oligomer assembly/disassembly
Hsp90	**Prokaryote (*E. coli*)** HtpG **Eukaryote** Hsp90	Cytosol Cytosol	ATP-dependent chaperone; prevention of protein aggregation and signalling protein control.
Hsp70	**Prokaryote (*E. coli*)** DnaK **Eukaryote** Ssa, Ssc (Yeast) Hsc70, Hsp70 (Higher eukaryote)	Cytosol Cytosol Cytosol and nucleus	ATP-dependent chaperone; Stabilisation of polypeptides by binding to exposed hydrophobic amino acid regions, thus assisting with folding of misfolded and nascent proteins; translocation of proteins in eukaryotes
Hsp60 (Chaperonins)	**Prokaryote (*E. coli*)** GroEL **Eukaryote** Hsp60	Cytosol Mitochondria	ATP-dependent chaperone; Creates an environment separate from the cytosol to promote conditions for stabilisation of proteins into the native state; protein translocation.
Hsp40	**Prokaryote (*E.coli*)** DnaJ **Eukaryote** Hsp40	Cytosol Cytosol	ATP-independent Binds misfolded polypeptides and acts as a co-chaperone for Hsp70 and DnaK
Small Hsps	**Prokaryote (*E.coli*)** IbpA, IbpB **Eukaryote** Hsp25	Cytosol Cytosol	ATP-independent chaperone; prevention of protein aggregation

change to allow them to fold correctly (Mayer *et al.*, 2000a). Unfolding chaperones unfold proteins and disassemble protein aggregates using ATP. Hsp100 is an example of an unfolding chaperone that utilizes ATP dependent conformational changes to free individual polypeptides from protein aggregates, which it then passes on to Hsp70 for

folding (Glover and Tkach, 2001; Goloubinoff *et al.*, 1999). Hsp100 is also involved in the disassembly of quarternary polypeptide complexes (Pak and Wickner, 1997). As these examples show, the different groups of chaperones often work in a co-ordinated manner (sequentially or simultaneously) to maintain the protein integrity and function within cells (Stirling *et al.*, 2003; Hartl and Hayer-Hartl, 2002).

It is becoming increasingly evident that molecular chaperones are important in every stage of the existence and quality-control of a polypeptide. They ensure the correct folding of a nascent polypeptide into its functional native state, maintain mature proteins in this state and then sort damaged or non-recoverable misfolded/misassembled polypeptides for degradation in the ubiquitin proteasome system (Goldberg, 2003; Jahn and Radford, 2005; Höhfeld *et al.*, 2001). Estimates indicate that approximately 30% of all nascent polypeptides fail to fold correctly and are then degraded in the proteasome system (Schubert *et al.*, 2000). **Figure 1.4** illustrates the wide variety of states accessible to a protein within the cellular environment and the areas in which molecular chaperones are known to maintain protein functional integrity within cells. The solid black lines indicate how various molecular chaperones work to drive protein folding to a native conformation. The dashed lines show how molecular crowding and other factors drive proteins away from folding into their native states. The red lines indicate the various points at which polypeptides are believed to be targeted to the ubiquitin proteasome system for degradation.

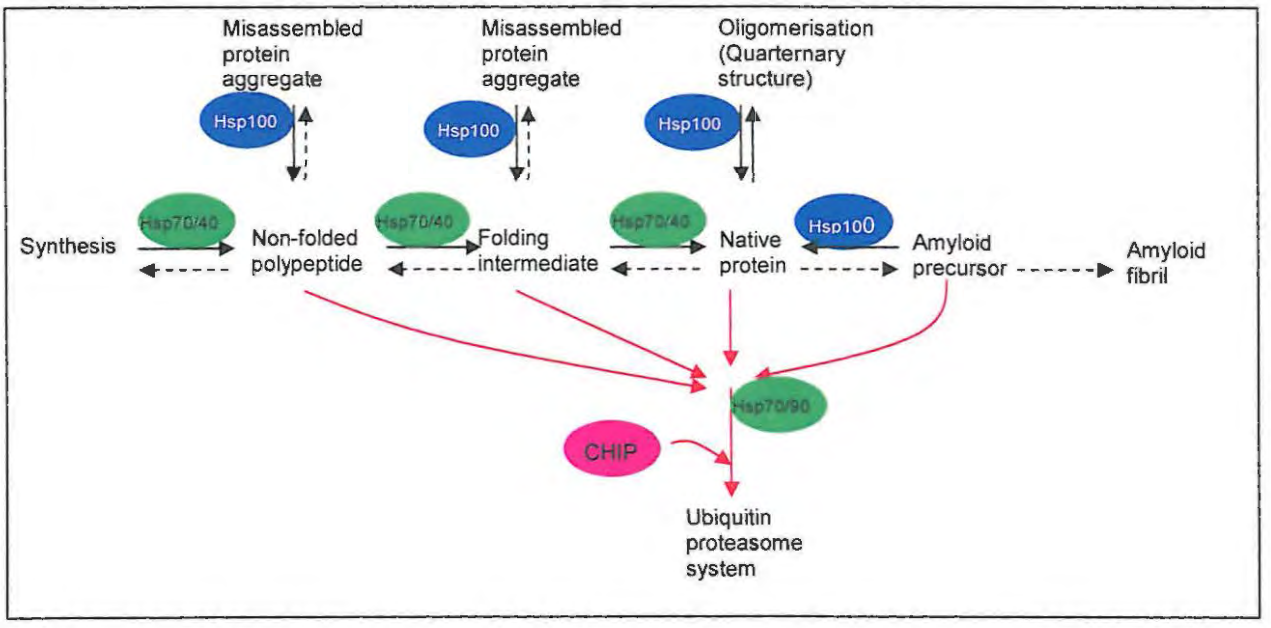


Figure 1.4: Shows some of the structures that can be formed by polypeptide chains *in vivo*. The solid black arrows show the movement of polypeptides into a more native state driven by various molecular chaperones that act in competition with forces that drive protein unfolding, misassembly and aggregation (shown in dashed arrows). The red arrows indicate the targeting of non-recoverable non-native polypeptides to the ubiquitin proteasome system. Hsp100 is given as an example of molecular chaperones used to dismantle protein misassemblies. The Hsp70/Hsp40 system is given as an example of chaperones involved in generalized protein folding and sorting of proteins for the ubiquitin proteasome pathway. CHIP (carboxy terminus of Hsp70-interacting protein) recognizes unfolded proteins through cooperation with Hsp70 and Hsp90 and ubiquitylates them so that they are recognized by the proteasome system for degradation. (Adapted from Dobson, 2004; Jahn and Radford, 2005; Goldberg, 2003).

Intracellular parasites have evolved ways of exploiting host molecular chaperones for their own purposes. Cellular Hsp70s are recruited for virion assembly and genome replication (Cripe *et al.*, 1995; Kelley, 1998; Garimella *et al.*, 2006) while others regulate the expression of cellular Hsp70s (Phillips *et al.*, 1991; Aranda *et al.*, 1996). It has further been found that plant closteroviruses have their own (non-cellular) Hsp70s (Agranovsky *et al.*,1991; Karasev *et al.*, 1996; Zhu *et al.*, 1998) that they utilize for cell- to-cell movement in infected plants (Peremyslov *et al.*, 1999). The malaria parasite, *Plasmodium falciparum*, has Hsp40s that it exports into the infected red blood cell and are hypothesized to interact with host Hsp70s (Botha *et al.*, 2007). As this thesis focuses on the co-chaperone Hsp40 in the Hsp70/Hsp40 molecular chaperone system, only this system will be mentioned in detail.

1.6) Hsp70/Hsc70

1.6.1) Introduction

Hsp70s have an approximate size of 70 kDa (Borges and Ramos, 2005) and exist in two main forms, the constitutively expressed form (Hsc70 or Heat shock cognate protein 70) which is involved in the cellular house keeping duties, and the stress inducible form (Hsp70 or Heat shock protein 70) which is induced under conditions of stress (Ingolia and Craig, 1982; Folgueira and Requena, 2007). These are one of the most evolutionary conserved protein families across species (Hunt and Morimoto, 1985; Lindquist and Craig, 1988). This conservation in the Hsp70 sequence is visible in the functionally conserved nature of Hsp70 across species. This is seen in the ability of a *Drosophila spp.* Hsp70 to protect mammalian cells against heat stress (Pelham, 1984) and human Hsp70 to functionally complement for rodent Hsp70 *in vitro* (Li *et al.*, 1991; Li *et al.*, 1992) and in transgenic mice (Angelidis *et al.*, 1996; Radford *et al.*, 1996). The *E. coli* genome encodes three distinct Hsp70s: DnaK, HscA and HscC. DnaK is by far the best studied of the three proteins (Genevaux *et al.*, 2007). The eukaryotic system is more complicated (there are 14 Hsp70s coded on the *S. cerevisiae* genome [James *et al.*, 1997]), where every cellular compartment has its own Hsp70, with multiple distinct Hsp70s found in the cytoplasm that have specialized functions (Folgueira and Requena, 2007).

1.6.2) Domain structure of Hsp70

Hsp70s have a conserved domain structure consisting of a 44 kDa ATPase domain at the N-terminus and a 25 kDa domain at the C-terminus (Flaherty *et al.*, 1990 and Wang *et al.*, 2000) (**Figure 1.5**). These two domains are linked by a conserved linker region (Jiang *et al.*, 2005; Goloubinoff and De Los Rios, 2007) (**Figure 1.6**).

1.6.2.1) The ATPase domain of Hsp70

ATP hydrolysis is essential for the chaperone activity of all Hsp70s (Mayer and Bukau, 2005). The ATPase domain consists of two large sub-domains, designated I and II, that are both further divided into smaller sub-domains (IA, IB, IIA and IIB) (**Figure 1.5B**). The two large sub-domains are separated from each other by a large cleft in which an adenosine nucleotide binding site is located at the base. Two K^+ and one Mg^{2+} ions are

bound in complex with this nucleotide and contact all four of the sub-domains (Flaherty *et al.*, 1990). The two large sub-domains can move closer together or further apart depending on the nucleotide that is bound (Zhang and Zuiderweg, 2004). The nucleotide binding cleft is open at the highest frequency in its nucleotide free state and decreasing from nucleotide free > ADP > ADP + inorganic phosphate > ATP (Gässler *et al.*, 2001). ATP hydrolysis is the rate limiting step in the ATPase cycle for most Hsp70s (McCarty *et al.*, 1995; Theyssen *et al.*, 1996). Hsp70 homologues have large differences in nucleotide dissociation rates, in spite of their almost superimposable structures. There are however, subtle differences in the structures that have led to the classification of Hsp70s into three subfamilies with *E. coli* DnaK, *E. coli* HscA and Human Hsc70 as the archetypal members (Gamer *et al.*, 1996). Certain Hsp70s interact with co-chaperone proteins called nucleotide exchange factors to stimulate the dissociation of ADP from the ATPase domain. Two examples of nucleotide exchange factors are GrpE in *E. coli* and Bag1 (Bcl2-associated athanogene) in eukaryotes (Brehmer *et al.*, 2001; Kabani *et al* 2003).

1.6.2.2) The peptide/substrate binding domain of Hsp70

The peptide binding domain is subdivided into a sandwich of two twisted β-Sheets with four antiparallel strands and a lid region consisting of a series of α-helices (A to E) and a structurally unresolved segment at the C-terminus of the protein (Gragerov *et al.*, 1994). The substrate binding pocket is embedded in the β-sandwich, with four upward protruding loops (**Figure 1.5C**). Loops $L_{1,2}$ and $L_{3,4}$ form the inner loops of the peptide binding pocket, and are stabilized by $L_{4,5}$ and $L_{5,6}$ which form the outer pair (Zhu *et al.*, 1996). α-Helix B contains a flexible hinge structure that allows the movement of the rest of the lid structure (helices B-E and the unresolved c-terminal region) from an open state to a closed state over the substrate binding cleft. This closed state is secured by two hydrogen bonds and a salt bridge that connect it to the substrate binding pocket loops $L_{3,4}$ and $L_{5,6}$ (Mayer and Bukau, 2005). ATP binding is hypothesized to open this lid to allow substrate release.

The optimal binding of the peptide substrate requires the α-helical lid, the deep hydrophobic pocket (located between the loops $L_{1,2}$ and $L_{3,4}$) and an arch enclosing the

substrate within the binding pocket (Mayer *et al.*, 2000b; Pellechia *et al.*, 2000; Buczynski, 2001; Slepenkov and Witt, 2002). This arch is formed by amino acid side chains from M404 (on loop $L_{1,2}$) and A429 (on loop $L_{3,4}$) (Mayer *et al.*, 2000b). However, of these three components, the hydrophobic pocket is the most important for substrate binding (Mayer *et al.*, 2000b). The peptide substrate (sequence NRTLLLTG) co-crystallised with the peptide binding domain (**Figure 1.5C**) indicated Hsp70 interacts with the peptide substrate in an extended conformation through van der Waals and hydrogen bonding (Zhu *et al.*, 1996). Hydrogen bonds were formed between the polypeptide backbone of the substrate binding cavity forming loops ($L_{1,2}$ and $L_{3,4}$) and the polypeptide backbone of the substrate. This provides a binding specificity for polypeptides composed of L-amino acids and not D-amino acids (Rüdiger *et al.*, 2001). Hydrophobic side chains line the substrate binding cavity and interact with the exposed hydrophobic side chains (e.g. leucine) of unfolded polypeptides (Zhu *et al.*, 1996). In the substrate binding domain structure of DnaK (**Figure 1.5C**) along with a peptide substrate bound, the central leucine is shown bound into the hydrophobic pocket, while the two flanking leucine residues are shown bound to M404 and A429 that form the hydrophobic arch and partially protecting them from the hydrophilic solvent environment (Zhu *et al.*, 1996). The substrate binding domain constantly opens and closes in the ATP and ADP bound conformations of Hsp70, but in the ATP bound conformation the substrate binding cavity is more frequently in the open conformation to allow substrate release (Mayer *et al.*, 2000b). *E. coli* DnaK interacts strongly with peptides containing a large proportion of phenylalanine, isoleucine, leucine and valine (Richarme and Kohiyama, 1993; Gragerov *et al.*, 1994). However, 90% of all regions of a polypeptide to which Hsp70 binds contain leucines. Such binding sites on an unfolded polypeptide have an average occurrence of once every 36 amino acid residues (Rüdiger *et al.*, 1997).

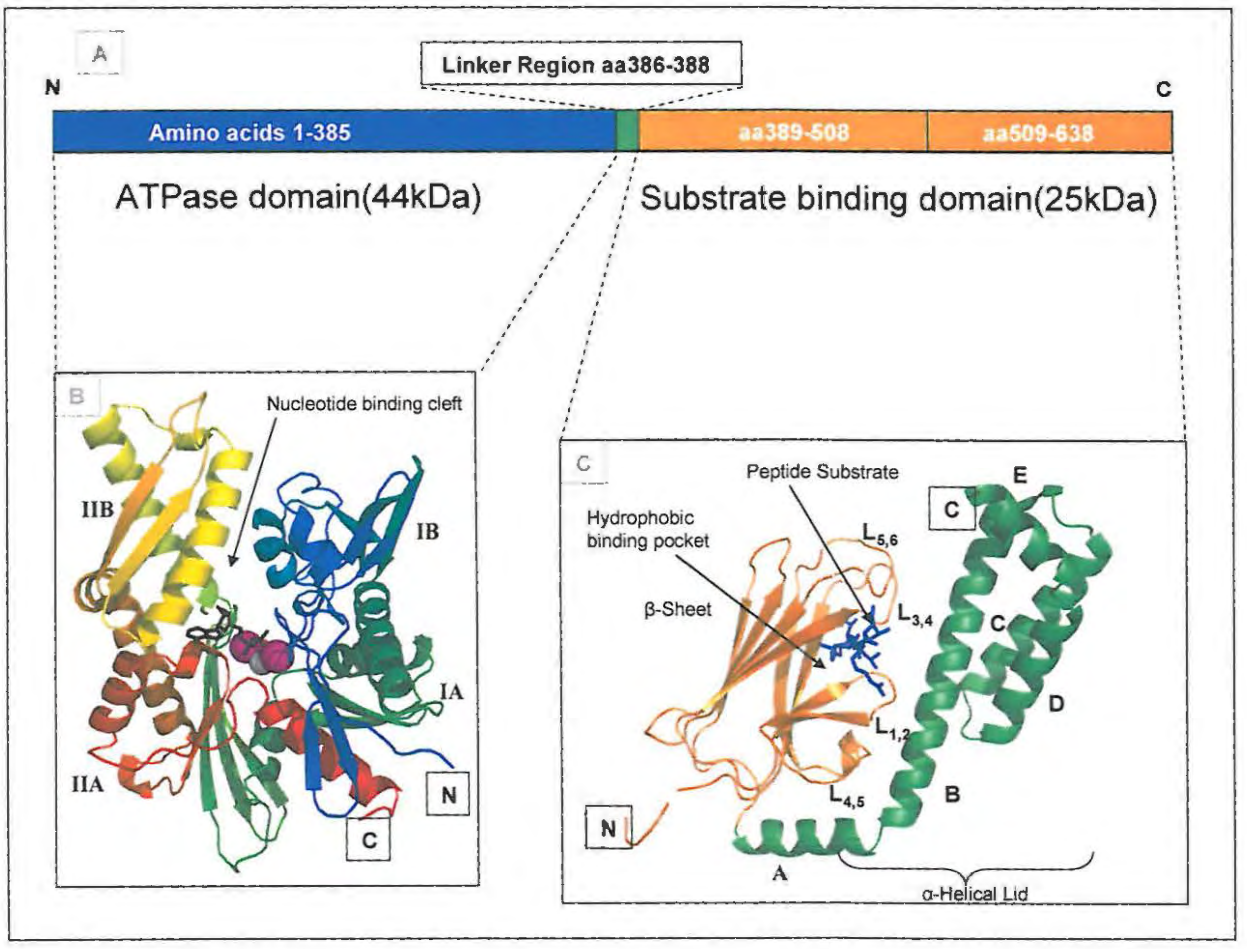


Figure 1.5 The domain structure of Hsp70. (A) a schematic view of the domain structure of Hsp70 showing the number of amino acids of which each section of *E. coli* DnaK consists. A 44 kDa ATPase domain at the N-terminus and a 25 kDa peptide binding domain at the C-terminus, which are connected by a short linker region. The peptide binding domain is further subdivided into two subdomains: the β-sandwich domain (amino acids 389-508) and the α-helical lid domain (amino acids 509-638) that clamps over the hydrophobic substrate binding pocket of the β-sandwich domain. **(B)** A three dimensional representation of the ATPase domain of *Bos Taurus* Hsc70 (3HSC.pdb; Flaherty *et al.*, 1990). The structure id divided into two subdomains (I and II), which form two the two lobes of the structure. These two lobes are further divided into subdomains A and B. The nucleotide binding cleft is found between subdomains IB and IIB. Buried within this nucleotide binding cleft are 2 K^+ atoms (purple spheres) and a Zn^{2+} atom (grey sphere). ADP is shown bound into the active site as a black stick structure (Flaherty *et al.*, 1990) The N and C termini are indicated with and N and C within a black square. **(C)** A three dimensional representation of the peptide binding domain of *E. coli* DnaK (1DKX.pdb; Zhu *et al.*, 1996). The β-Sandwich subdomain is shown in orange, with the four loops that form the hydrophobic binding pocket are labeled (inner lobes L1,2 and L3,4; outer lobes L4,5 and L4,6). The peptide substrate is shown in a stick configuration, coloured blue. The lid assembly of the peptide binding domain is coloured green and the various helices (A-E) are labeled. The domain 3-dimensional molecular structure images were generated using Pymol (DeLano, 2002).

1.6.3) Functions of Hsp70

Hsp70s perform a range of cellular functions which all appear to require the ability of Hsp70 to interact with hydrophobic peptide segments of proteins in an ATP controlled manner. These functions include: membrane translocation of proteins targeted to other cellular compartments and the secretory pathway, folding and assembly of nascent proteins, prevention of protein aggregation, refolding of unfolded or aggregated proteins and regulation of signal transduction pathways (Mayer and Bukau, 2005; Borges and Ramos, 2005). Hsp70 therefore serves standard house-keeping functions in the cell as well as protein quality control functions, which become more important during periods of cellular stress (Mayer and Bukau, 2005). An example of a general house keeping function of Hsp70 is the manipulation of nascent polypeptides. In addition to monitoring the correct folding of these proteins, Hsp70 sometimes delays protein folding to the native state by shielding hydrophobic residues from the aqueous environment. This is required for translocation of certain proteins to other cellular compartments that need to pass through membrane pores in an unfolded state (Young *et al.*, 2003). Other proteins only fold correctly if translation is complete before folding commences (Leone *et al.*, 1996), and certain proteins are damaging to the cell and are maintained in an unfolded state until they are required or inserted into their safe location (Mayer and Bukau, 2005).

Although protein quality control functions of Hsp70 are also required in the house keeping functions of the cell, they are even more important in times of cellular stress. Hsp70, along with its co-chaperones, help to prevent protein aggregation and promote the refolding of non-native proteins in an ATP dependent manner (**see section 1.8.1** explaining the Hsp70 chaperone cycle) (Mayer and Bukau, 2005). Two mechanisms have been proposed to explain the refolding of unfolded polypeptides by Hsp70. The first is passive involving repetitive binding and release of substrates to lower the unfolded protein concentration to a level that prevents aggregation, allowing the free unfolded substrates to fold on their own. In the second proposed mechanism, the multiple bind and release cycles of Hsp70 induce a localized unfolding of the misfolded substrate that results in the removal of energy barriers preventing correct folding to the native state (Mayer and Bukau, 2005).

Hsp70 plays a central role in the molecular chaperone system of cells as it is able to coordinate the movement of unfolded proteins from one chaperone pathway to another (Borges and Ramos, 2005). **Figure 1.4** illustrates some of the ways in which Hsp70 interacts with Hsp100, Hsp40 and Hsp90. Hsp100 interacts with Hsp70 in protein disaggregation processes and regulation of protein quarternary structure (Pak and Wickner, 1997; Glover and Tkach, 2001). Hsp40 is able to deliver unfolded protein substrates to Hsp70 and stimulate its ATP hydrolysis activity that is central to its role as a chaperone. Hsp90 works with Hsp70 in signal transduction pathways and targeting of proteins to the ubiquitin proteasome system (Goldberg, 2003; Mayer & Bukau, 2005). Hsp70 acts with Hsp90 to regulate signaling protein activity and therefore regulate sections of cellular signal transduction pathways (Mayer and Bukau, 2005). Hsp70 also transports nascent polypeptides from ribosomes to Hsp60 for folding (Langer *et al.*, 1992; Frydman *et al.*, 1994). Small Hsps are also known to transfer heat dentured proteins to Hsp70 for folding to their native state (Lee and Vierling, 2000; van Montfort *et al.*, 2001b).

A number of Hsp70s and Hsc70s are known to self-associate into dimers and polymers (Benhouradj *et al.*, 1996; Chou *et al.*, 2003; Guidon and Hightower, 1986; Kim *et al.*, 1992; Bhattacharyya *et al.*, 1995; Schönfeld *et al.*, 1995). Data suggests that the C-terminal 10-kDa lid domain of Hsp70 is important for this association in rat Hsc70 (Chou *et al.*, 2003). Benharoudj *et al.*, 1996 proposed a model where Hsp70 exists in an equilibrium between a polymeric and a monomeric state, which is regulated by the associated nucleotides, unfolded protein substrates and co-chaperones. The binding of ATP, unfolded polypeptide substrates and nucleotide exchange factors drives the equilibrium to the monomeric state, while ADP and Hsp40s stabilize the polymeric form. In this way oligomerisation is seen as a mechanism of regulation of the Hsp70 chaperone activity as Hsp70 is not bound to unfolded substrate in its polymeric form (Benharoudj *et al.*, 1996).

1.6.4) Interdomain communication of Hsp70

During the execution of its cellular functions, the conformational changes of Hsp70 move across both the ATPase domain and the peptide binding domain, which results in substantial relaying of signals between the two domains (Montgomery *et al.*, 1999; Jiang *et al.*, 2005; Jiang *et al.*, 2007). For example, ATP hydrolysis in the ATPase domain is known to alter the substrate binding domain conformation and its affinity for unfolded protein substrates (Jiang *et al.*, 2005).

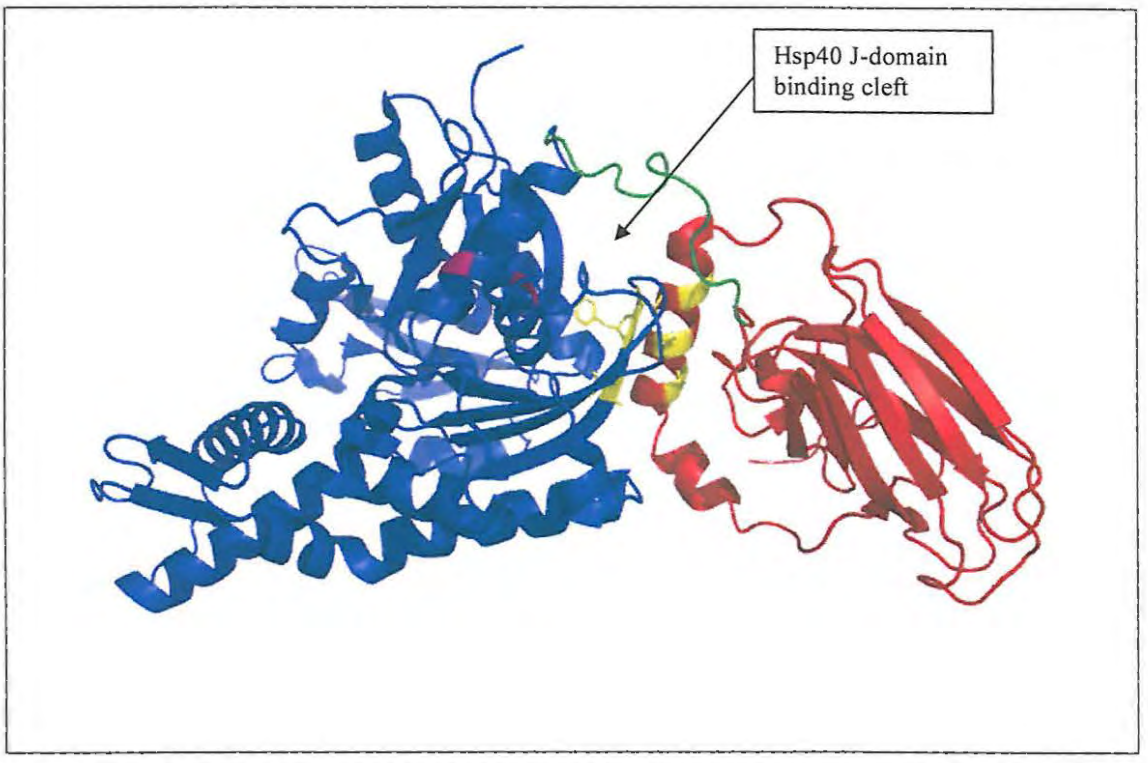


Figure 1.6: The structure of *Bos taurus* Hsc70 excluding part of the lid assembly of the peptide binding domain, showing the way in which the ATPase domain and substrate binding domain are connected (1YUW.pdb; Jiang *et al.*, 2005). A conserved linker region (amino acid residues 384-394 coloured in green) connects the ATPase domain (amino acid residues 1-383 in blue) to the unfolded peptide substrate binding domain amino acid residues 395-554, in red). The amino acids involved in the hydrophobic interactions between the two domains that are disrupted by the linker region during ATP binding are shown in yellow. The site of the binding of the Hsp40 J-domain is also indicated. The structure was visualized using PyMol (DeLano, 2002)

The inter-domain interface of both domains is well conserved and includes the groove between the lobes IA and IIA of the ATPase domain (Jiang *et al.*, 2005; Revington *et al.*, 2005; Gässler *et al.*, 1998) and the amino acids 414-417 and helix A of the substrate binding domain (Jiang *et al.*, 2005). The amino acids that appear to be specifically

involved in inter-domain communication of *Bos taurus* Hsc70 include: salt bridges formed between the two domains are K325/E530, K524/D152, E218/R426/K526 and R171/D513, and a hydrophobic cluster formed by I216, F217, V219, I515, V519 and A522. Five Hydrogen bond interactions also contribute to the interaction (N417/E192; Q520/N174, K220/K415, E516/Q376 and R376/F217) (Jiang *et al.*, 2005). The linker region (**Figure 1.6**) of Hsp70 is also believed to play an important role in the inter-domain communication mechanism. ATP-binding produces a conformational change in the linker region that results in movement from a solvent-exposed to a partially buried state (Rist *et al.*, 2006). The linker is believed to become buried between the two domains and disrupting the hydrophobic interaction between the domains, shifting their position relative to each other. The interaction of the Hsp40 J-domain with Hsp70 has a marked effect on the inter-domain communication in Hsp70 (Jiang *et al.*, 2007). See section **(1.8.2.1)** on the interaction of Hsp70 and the J-domain of Hsp40.

1.7) Hsp40

Hsp40s are approximately 40 kDa in size and act as co-chaperones of Hsp70 (Fan *et al.*, 2003). In addition, some Hsp40s are classed as holding molecular chaperones as they are able to bind and prevent the aggregation of unfolded proteins but are unable to fold them back to their native state (Stirling *et al.* 2003). Six Hsp40s have been found in the *E. coli* genome (Genevaux *et al.*, 2007) and far more in Eukaryotic systems (20 in *S. cerevisiae* [Walsh *et al.*, 2004] and 49 in humans [Kampinga *et al.*, 2009]). The Hsp40s in eukaryotes are localized to various of the cellular compartments (Walsh *et al.*, 2004; Cyr *et al.*, 1992). The large number of Hsp40s in eukaryotes relative to the number of Hsp70s is indicative of many Hsp40s interacting with the same Hsp70 to fulfill their function (Fan *et al.*, 2005; Walsh *et al.*, 2004).

1.7.1) Domain structure

There are a number of protein domains that are commonly found in Hsp40s. Their domain structure has allowed them to be divided into four different types (**see section 1.7.2**) (Cheetham and Caplan 1998; Fan *et al.*, 2003). The J-domain is common to all Hsp40s and they can contain a combination of the other domains described below. A

complete structure for Hsp40 has not been determined, but individual domains have had their structures resolved (**Figure 1.7**).

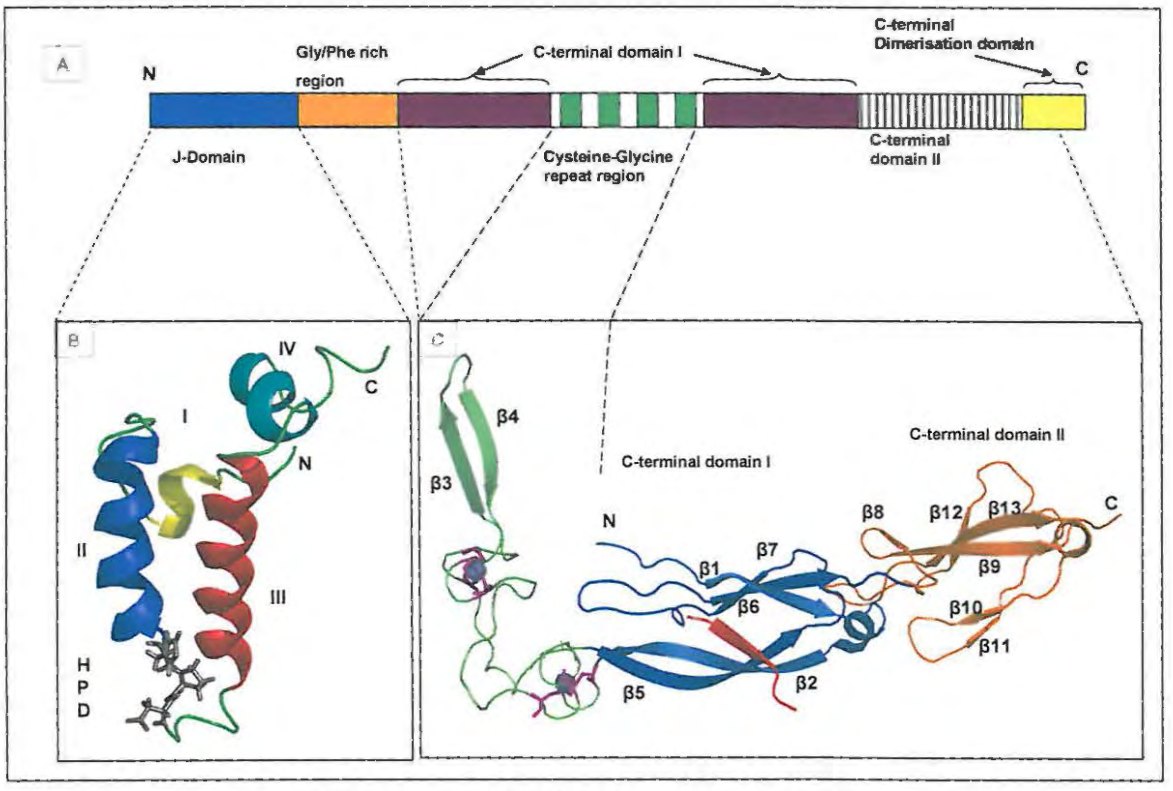


Figure 1.7: The domain structure of an Hsp40. A) A schematic diagram of the generalized domain structure of an Hsp40. The J-domain is the signature domain of Hsp40s and is connected to the rest of the protein by the glycine/phenylalanine rich region (G/F region). The zinc finger like (Cysteine/Glycine repeat) region contains four repeats of the motif CxxCxGxG that are involved in coordination of two zinc atoms and is found between two sections of the C-terminal sub-domain 1 (purple). The C-terminal sub-domain II (region with vertical grey stripes) is located just before the dimerisation domain which is located at the extreme C-terminus. **B)** The three dimensional structure of the J-domain of *E. coli* DnaJ, showing the general features of all Hsp40 J-domains (PDB file Identity no. 1XBL) (Pellechia *et al.*, 1996). It consists of four α-helices highlighted in yellow (Helix 1), dark blue (Helix II), red (helix III) and cyan (Helix IV). The HPD motif important for stimulation of the Hsp70 ATPase activity is displayed in grey as part of the loop between helices II and III, with the side chains visible. **C)** Shows the three dimensional structure of the C-terminal domains of the yeast type I Hsp40 YdJ1 (PDB file identity: 1NLT) (Li *et al.*, 2003). The Zinc finger like region (amino acids 143-208 in YdJ1) is shown in green with the two zinc atoms co-ordinated by the domain shown as grey spheres. The cysteine side chains involved in binding the zinc atoms are shown in magenta. The C-terminal peptide binding domain is divided into two domains shown as blue (amino acids 110-142 and 209-256 in YdJ1) and orange (amino acids 257-337 in YdJ1). An unfolded peptide being bound by the C-terminal region is shown in red. The complete C-terminal dimerisation domain is not shown in this structure (Li *et al.*, 2003). The C-terminus of the J-domain is connected to the N-terminus of the zinc finger like region by the flexible Glycine/Phenylalanine rich region, for which no structure has been determined (Borges *et al.*, 2005). The three dimensional structures were produced using the Pymol PDB visualization computer programme (Delano, 2002).

1.7.1.1) J-domain

The J-domain is the signature domain of Hsp40s and is known to increase the ATP hydrolyzing (ATPase) activity of Hsp70. The approximately 73 amino acid domain has long been hypothesized as the major site of interaction between Hsp40s and Hsp70s (Hennessy *et al.*, 2000). The J-domains of various Hsp40s have had their structures determined: *E. coli* DnaJ (Pellechia *et al.*, 1996; Huang *et al.*, 1998), Human DnaJ protein 1 (Hdj1) (Qian *et al.*, 1996), murine polyomavirus T antigen (PDB id: 1faf) (Berjanskii *et al.*, 2000), *E. coli* Hsc20 (PDB id:1fpo) (Cupp-Vickery and Vickery, 2000), Simian Virus 40 (SV40) T antigen (PDB id: 1GH6) (Kim *et al.*, 2001) and Bovine Auxilin (PDB id: 1nz6) (Jiang *et al.*, 2003). Other J-domain structures have been resolved and published on the PDB database (http://www.pdb.org/) only. The J-domain contains four α-helices connected by loop regions (**Figure 1.7B**). Helices II and III are antiparallel relative to each other and a hydrophobic core of amino acids are on the interior faces of helices I, II and III to maintain the structure (Pellechia *et al.*, 1996). The loop between α-helices II and III contains a conserved tripeptide motif of Histidine, Proline and Aspartic acid (HPD) which is required for the functional interaction of Hsp40 and Hsp70 in many cases (Qian *et al.*, 1996). The HPD motif has been shown to be important in the stimulation of the ATPase activity of Hsp70 (Cheetham and Caplan, 1998). The type I and type II Hsp40 J-domains are more conserved than the type III Hsp40 J-domain amino acid sequence. It has been hypothesized that the more conserved amino acids of the J-domain could be involved in maintaining structural integrity of the domain or interaction with the partner Hsp70. Less conserved residues have been suggested to play a role in the specificity of the interaction between the Hsp70 and Hsp40 partners (Hennessy *et al.*, 2000; Hennessy *et al.*, 2005a).

1.7.1.2) Glycine/Phenylalanine rich region

The Glycine/Phenylalaninine rich region (G/F region) is an extended, non-structured flexible region that has one conserved sequence motif consisting of aspartic acid-Isoleucine/valine- Phenylalanine triplets referred to as the DIF motif (Wall *et al.*, 1995). This region is believed to reduce local flexibility and allow rapid switching between favoured conformations (Pellechia *et al.*, 1996). Any alteration to the DIF motif results in

the same phenotype as complete deletion of the G/F region. This motif does not appear to be involved in substrate binding or direct stimulation of the ATPase activity of Hsp70, despite the fact that amino acid substitutions in the DIF motif slow the steady state ATPase cycle of the DnaK/DnaJ/GrpE complex and its chaperone activity (Cajo *et al.*, 2006).

The G/F regions of Type I and Type II Hsp40s are different. The yeast type II Hsp40 Sis1 G/F region has an addition 10 amino acids (GHAFSNEDAF) that is not found in the type I Hsp40 Ydj1 (Lopez *et al.*, 2003). Swapping the Sis1 G/F region in place of the Ydj1 G/F region allows the Ydj1/Sis1 chimera to recover from the lethal phenotype of Sis1Δ, but intact Ydj1 does not. This indicates that the two G/F regions help to specify the different functions of type 1 and Type II Hsp40s (Fan *et al.*, 2004).

1.7.1.3) Cysteine/Glycine repeat region

This is also known as the zinc finger-like region (ZFLR) and typically consists of four repeated Cyteine-X-X-Cysteine-X-Glycine-X-Glycine (CxxCxGxG) motifs, where X is any amino acid. The structure of this region has been determined for *E. coli* DnaJ by nuclear magnetic resonance spectroscopy (Martinez-Yamout *et al.*, 2000) and revealed a V-shaped groove with an extended β-hairpin topology that has been found to be conserved in Ydj1 (**Figure 1.7C**) (Li *et al.*, 2003). These CxxCxGxG motifs function in pairs to bind two zinc atoms that form zinc binding domain (ZBD) I (repeats one and four) and ZBD II (repeats two and three). The structure has the potential to undergo protein-protein interactions (Martinez-Yamout *et al.*, 2000); however, it does not seem to be essential for peptide substrate binding (Banecki *et al.*, 1996). The ATPase activity of DnaK was also not stimulated by the region (Lu and Cyr, 1998a). The ZFLR has been shown to be necessary in the cooperation of type I Hsp40 with Hsp70 in protein folding reactions (Fan, *et al.*, 2005; Linke *et al.*, 2003). The two zinc binding domains appear to have different functions. ZBDII is important in Hsp40 binding to DnaK/Hsp70 to suppress thermally induced protein aggregation. Both zinc binding domains were required for protein folding (Fan *et al.*, 2005; Linke *et al.*, 2003). It has been proposed that the activity of Type I Hsp40s could be partially regulated by the oxidation or

reduction of the cysteines of the ZFLR, resulting in binding or release of the zinc atoms (Choi *et al.*, 2006). Some eukaryotic type I Hsp40s have a ZFLR region in which a given CxxCxGxG repeat has another amino acid substituted for the last glycine (Martinez-Yamout *et al.*, 2000). Two examples of this phenomenon are the *S. cerevisiae* Hsp40s Ydj1 (Caplan and Douglas, 1991) and Scj1 (Blumberg and Silver, 1991). Only one such substitution appears to be tolerated by the structure to ZBDI and ZBDII respectively (Martinez- Yamout, *et al.*, 2000). Instead of a Zinc finger binding region, certain type II Hsp40s have what appears to be an extension of the G/F region that is termed the glycine/methionine rich region (G/M region) (Cheetham and Caplan, 1998; Fan *et al.*, 2003).

1.7.1.4) C-terminal peptide binding region

Type I and type II Hsp40s both have a peptide binding domain at their C-terminal ends, but Type III Hsp40s appear to have a peptide-binding region that is evolved for more specialized functions (Cheetham and Caplan, 1998). The structure of the C-terminal regions of representative members of both Type I (Ydj1- 1NLT; Li *et al.*, 2003) and type II Hsp40s (Sis1- 1C3G; Sha *et al.*, 2000 and Hdj1- 2QLD; Hu *et al.*, 2008) have been elucidated. The peptide binding fragments of both these protein Hsp40 types have two sub-domains. Some researchers have included the zinc finger binding domain of Type I Hsp40s as part of the C-terminal binding fragment of these Hsp40s as it is located between segments of the first C-terminal sub-domain in the primary amino acid sequence (Sha *et al.*, 2000; Li *et al.*, 2003). The C-terminal fragment of Ydj1 (**Figure 1.7C**) sub-domain I stretches from amino acids 110 to 142 and 209 to 256, with the zinc finger binding domain in between (amino acids 143 to 208). The C-terminal sub-domain II is located from amino acids 257 to 337 (Li *et al.*, 2003). These two sub-domains have a similar structure consisting of a major and a minor β-sheet connected by a short stretch of α-helix. The major β-sheet of subdomain I consists of the β-sheets B1, B6 and B7 and a minor β-sheet of B2 and B5 (**Figure 1.7C**). In subdomain II the major β-sheet consists of B8, B9, B12 and B13 and the minor β-sheet contains B10 and B11. One of the few major differences in the structure between these two domains is that domain II contains a larger major β-sheet with an extra antiparallel strand (B8) (Li *et al.*, 2003). The two sub-

domains (particularly sub-domain I) of the peptide binding region of the *S. cerevisiae* type II Hsp40 Sis1 have a similar fold structure with the main difference being the lack of the zinc finger like region interrupting sub-domain I in the primary amino acid sequence (Hu *et al.*, 2008; Li *et al.*, 2003).

Both type I and type II Hsp40s have a peptide binding region in their C-terminal domain I that binds unfolded polypeptides (Sha *et al.*, 2000; Li *et al.*, 2003). In Ydj1, the peptide substrate forms an anti-parallel β-strand with B2 strand of sub-domain I, binding with normal β-sheet hydrogen bond networks with the side chains of the peptide n- GWLYEIS - c. Residues in the Ydj1 hydrophobic pocket important for binding (I116, L135, L137, L216 and F249). A further hydrophobic pocket region was found on sub-domain II of Ydj1, hut it is thought not to be important in substrate binding as it is occupied by β-strand 8 (Li *et al.*, 2003). Hydrophobic residues important in the hydrophobic pocket of Sis1 are K199, F201 and F251 (Lee *et al.*, 2002).

As the hydrophobic cleft suggests, both Ydj1 and Sis1 select for peptides enriched with aromatic and bulky hydrophobic amino acids. However, it appears that Ydj1 prefers peptide substrates with stretches of 3 to 4 hydrophobic residues, while peptides selected for by Sis1 have a random distribution of hydrophobic residues (Fan *et al.*, 2004). Type I Hsp40s have a greater affinity for peptides rich in the aromatic amino acids phenylalanine, tryptophan and tyrosine, the large hydrophobic amino acids isoleucine and leucine and the histidine polar residue. Proline and lysine rich peptides are mostly excluded from preferred peptide substrates (Rüdiger *et al.*, 2001; Fan *et al.*, 2004). As type I Hsp40s bind only the side chains of the substrate peptides, they are able to recognize both L- and D- amino acids, which is in contrast to Hsp70 which recognizes the peptide backbone and side-chain contacts (Rüdiger *et al.*, 2001; Bischofberger *et al.*, 2003).

1.7.1.5) Hsp40 Quarternary Structure

Both Type I and Type II Hsp40s have been shown to dimerise and that this is important for their chaperone activity (Borges *et al.*, 2005; Shi *et al.*, 2005; Langer *et al.*, 1992).

Experimentation with these dimers has revealed that they maintain their structural integrity when exposed to temperatures as high as 50 °C, while a truncated monomer version of HdJ2/DjaI (amino acids 1-332) lost substantial structural intergrity with temperatures above 38 °C. Small angle X-ray scattering studies have shown great difference in the quarternary structure of Type I and Type II Hsp40s. Type I Hsp40s were shown to dimerise into a compact bullet shaped molecule with the N- and C-termini of each monomer in mirrored positions. The J-domains of the monomers are in the widest part of the bullet shape, while the C-termini are in the narrowest part. The Cys/Gly repeat regions form zinc finger motifs that face each other in the respective monomers and are hypothesized to keep the two J-domains at an appropriate distance from one another and to stabilize the polypeptide binding groove (Borges *et al.*, 2005).

The structure of type II Hsp40s such as the human DjB4 (also known as Hdj1, or Human DnaJ1) was interpreted as a more extended U-shaped structure (**Figure 1.8B**) with only the C-termini of the monomers interacting with each other and each monomer bending away from the other (**Figure 1.8C**)(Li *et al.*, 2003; Li *et al.*, 2006). Both protein classes dimerise through a dimerisation motif in the C-terminus of the protein (Sha *et al.*, 2000; Wu *et al.*, 2005). However, the dimerisation mechanism appears to be different as possibly indicated by the longer C-terminus of the yeast Type I Hsp40 Ydj1 (**Figure 1.8A**) compared to the yeast Type II Hsp40 Sis1 in yeast. Ydj1 dimerisation appears to occur through two mechanisms. The first involves a hydrophobic cluster of amino acids (L274, I278, L346, L349, I352, L353, F335, P336 and F340). The second mechanism involves the formation of a β-sheet between the C-terminal dimerisation motif of one monomer and the upstream sub-domain of the other monomer (**Figure 1.8A(1)**) (Wu, *et al.*, 2005). The Type II Hsp40 Sis1 has also been found to dimerise using a hydrophobic cluster of amino acids (Sha *et al.*, 2000). The β-sheet dimerisation mechanism has not been found in Type II Hsp40s (Wu *et al.*, 2005). This difference in dimer structure between type I and type II Hsp40s has been implicated in differences of substrate selection and has distinct ramifications for type I and type II Hsp40 interaction with Hsp70 (**refer to section 1.7.2**) (Borges *et al.*, 2005). However, there do appear to be

differences in the interpretation of the structure for the C-terminal domains of Type II Hsp40s (Sha *et al.*, 2003; Borges *et al.*, 2005).

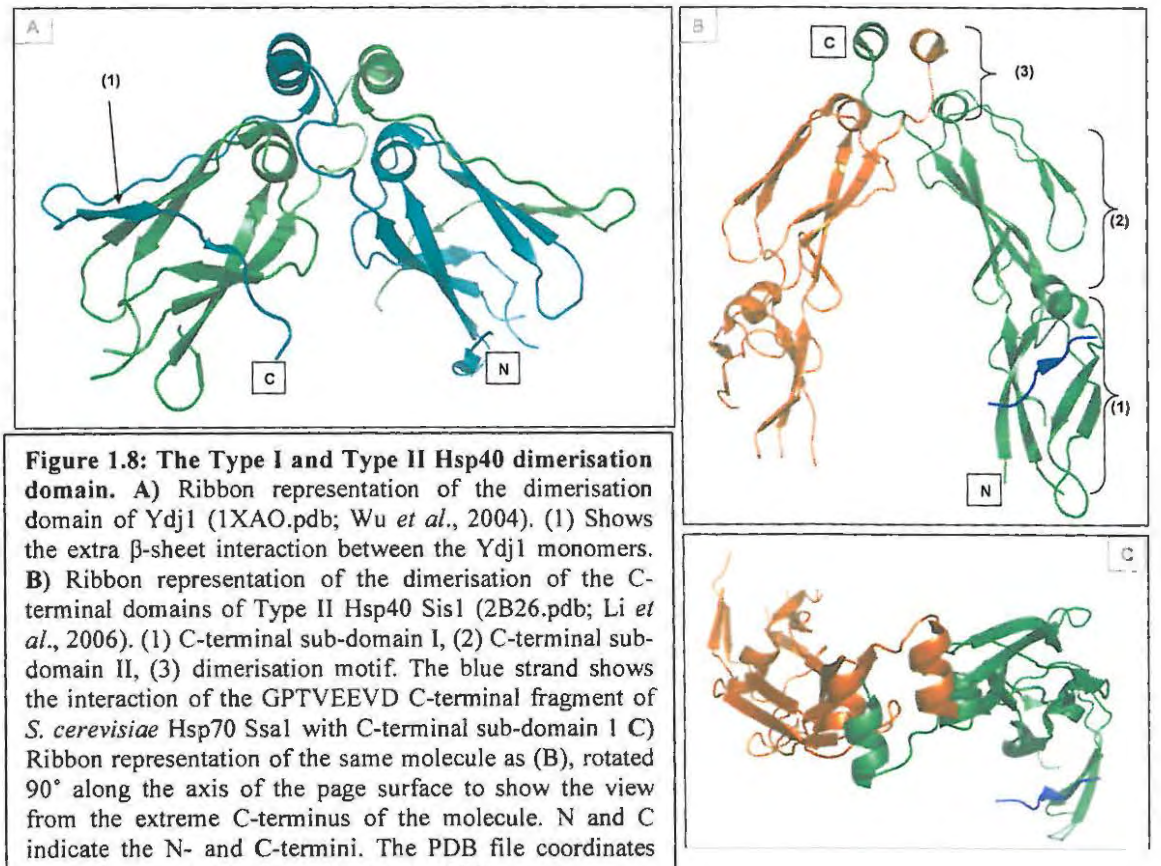


Figure 1.8: The Type I and Type II Hsp40 dimerisation domain. A) Ribbon representation of the dimerisation domain of Ydj1 (1XAO.pdb; Wu *et al.*, 2004). (1) Shows the extra β-sheet interaction between the Ydj1 monomers. **B)** Ribbon representation of the dimerisation of the C-terminal domains of Type II Hsp40 Sis1 (2B26.pdb; Li *et al.*, 2006). (1) C-terminal sub-domain I, (2) C-terminal sub-domain II, (3) dimerisation motif. The blue strand shows the interaction of the GPTVEEVD C-terminal fragment of *S. cerevisiae* Hsp70 Ssa1 with C-terminal sub-domain 1 **C)** Ribbon representation of the same molecule as (B), rotated 90° along the axis of the page surface to show the view from the extreme C-terminus of the molecule. N and C indicate the N- and C-termini. The PDB file coordinates were visualized using Pymol (Delano, 2002).

1.7.2) Classification of Hsp40 proteins

The classification of the Hsp40 proteins is based on the structural motifs that they possess, in addition to the J-domain (**Figure 1.9**). Type 1 Hsp40s have an N-terminal J-domain, in addition to a region rich in glycine and phenylalanine residues (Gly/Phe-rich region) and a cysteine/glycine repeat (CxxCxGxG) region. Type II Hsp40s contain an N-terminal J-domain and a Gly/Phe-rich region but lack the cysteine repeat region (Cheetham and Caplan, 1998). It has been proposed that Type I and Type II Hsp40s serve to bind to non-native substrates (Kelley, 1998). Type III Hsp40s appear to be more specialized in their function (Kelley, 1998) and contain only the J-domain, which, unlike that of the Type I and II Hsp40s can occur anywhere along the protein (Cheetham and Caplan, 1998). Type IV Hsp40s have been proposed to be a further classification of

Hsp40s in which the HPD motif of the J-domain is partially or completely absent (Botha *et al.*, 2007). The other domain structures within the type IV group remain the same as the type I, II and III categories. Therefore, it is possible to get type IV Hsp40 that is type I-like and another type IV Hsp40 that is type III-like (Botha *et al.*, 2007). Further characterization of Hsp40s has shown that Type II Hsp40s have a Glycine/Methionine rich region after their Glycine/ Phenylalanine rich region. The C-terminal domains of Type I and Type II proteins appear to be similar (Li *et al.*, 2003; Sha *et al.*, 2000) and both contain a C-terminal dimerisation domain (Wu *et al.*, 2004; Sha *et al.*, 2000).

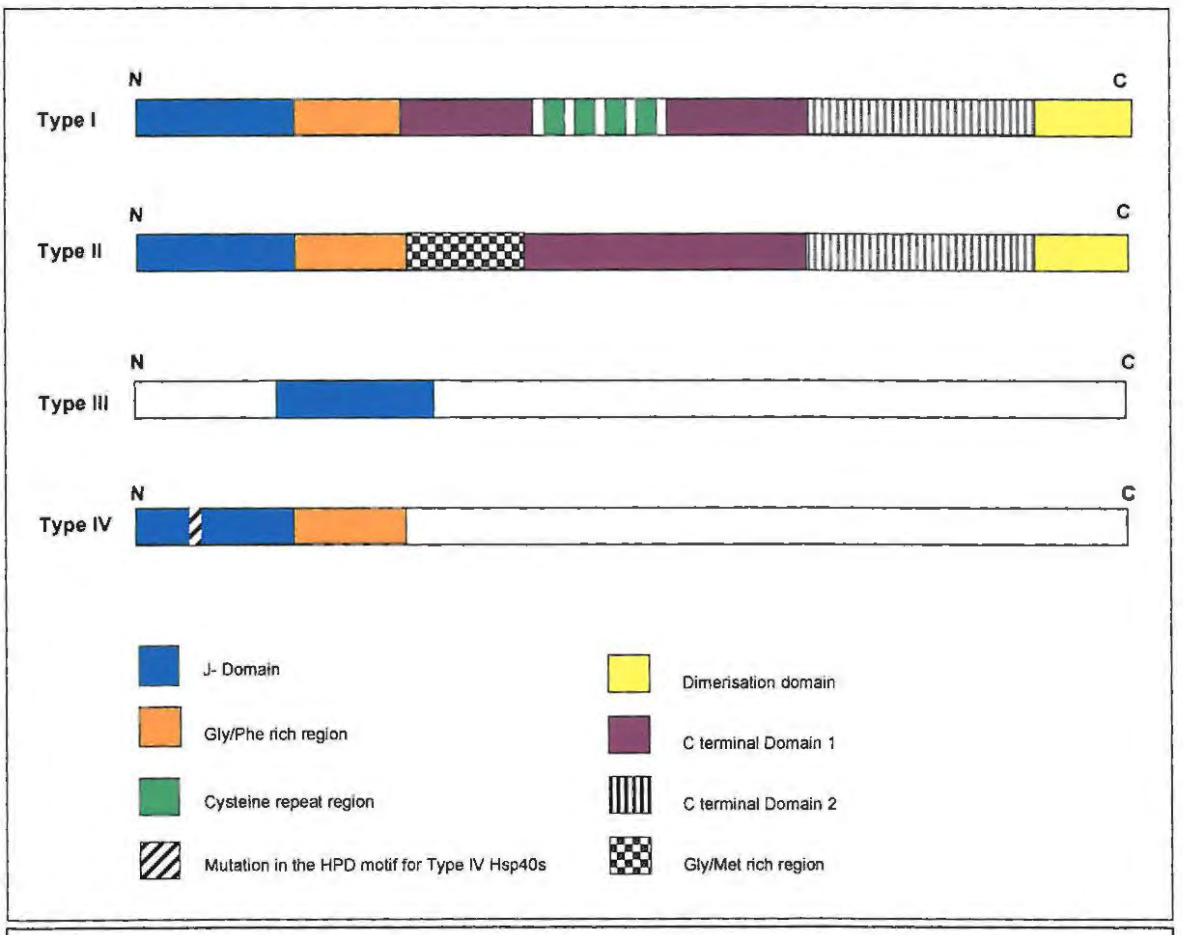


Figure 1.9: Hsp40 classification. Hsp40s have traditionally been classified according to their similarity to *E. coli* DnaJ. This includes the possession of a J-domain, G/F region and Cys/Gly repeat region (Cheetham and Caplan, 1998). Better characterization of the C-terminus of Hsp40s have led to a further level of categorization. Type I and II Hsp40s have C-terminal region divided into two subdomains: C-terminal domain 1 (CTDI) and C-terminal domain 2 (CTDII). In Type I Hsp40s the CTDI is interrupted by the Cys/Gly repeat region. It has also been found that type II Hsp40s have a Gly/Met rich region after the Gly/Phe rich region (Fan *et al.*, 2004). A further category of Hsp40 is the Type IV, in which amino acid substitutions have occurred in the HPD motif of the J-domain (Botha *et al.*, 2007).

Hsp40s are important in a multitude of cellular processes which include: protein translocation across membranes, protein degradation, disassembly of multi-protein complexes, protection from cellular stress, refolding of denatured proteins, suppression of protein aggregation, folding of nascent polypeptides, signal transduction and endocytosis (Walsh *et al.*, 2004; Fink, 1999; Terada *et al.*, 1997; Ngosuwan *et al.*, 2003; Abdul *et al.*, 2002; Prip-Buus *et al.*, 1996; Cyr *et al.*, 1994; Hartl, 1996; Sondheimer *et al.*, 2006). Although most of these functions involve Hsp40 functioning as a co-chaperone of Hsp70, it has been discovered that some Hsp40s have an intrinsic chaperone activity of their own that appears to be important for the correct functioning of Hsp70 (Wickner *et al.*, 1991; Langer *et al.*, 1992; Szabo *et al.*, 1994). However, Hsp40s do not hydrolyse ATP to perform their function (Stirling *et al.*, 2003).

1.7.3) Hsp40 as chaperone

Some Hsp40s are defined as having independent chaperone activity as they are able to bind non-native polypeptide substrates and prevent their aggregation with other such species. This phenomenon was first observed in *E. coli* DnaJ (Wickner *et al.*, 1991; Langer *et al.*, 1992) and has been subsequently observed in other type I Hsp40s such as YdJ1 in eukaryotes (Cyr, 1995; Lu and Cyr, 1998a). Although Type II Hsp40s have been shown to bind unfolded polypeptides through their peptide binding domain (Lee *et al.*, 2002; Mohler, *et al.*, 2004), data indicates that they do so with less efficiency than Type I proteins. Studies have shown that Sis1 and Hdj-1 are unable to efficiently bind and prevent the aggregation of thermally denatured luciferase (Lu and Cyr 1998b; Minami *et al.*, 1996; Borges *et al.*, 2005). This could be due to possible substrate selectivity between the two protein groups as they have been discovered to bind and select for slightly different polypeptide substrates (Fan *et al.*, 2004) (**Section 1.7.1.4**). Type III Hsp40s do not appear to be capable of suppressing protein aggregation and have not traditionally been classed as having an intrinsic molecular chaperone activity (Cheetham and Caplan, 1998). Hsp40 has been proposed to scan polypeptide substrates for hydrophobic amino acid side chains and make the initial contacts with them before targeting them to Hsp70 (Rudiger *et al.*, 2001). Another paper suggested that Hsp40s keep the polypeptide in an

elongated form between the two arms of the dimer to enable Hsp70 to bind the polypeptide (Li *et al.*, 2006).

1.7.4) Hsp40 as co-chaperone

Co-chaperones are proteins that bind to chaperones and assist in their function through regulation of their chaperone activity (Caplan, 2003). It is now well known that Hsp40s regulate the chaperone activity of Hsp70 through binding and delivering unfolded polypeptide substrates to Hsp70, and by stimulating the ATP hydrolysis by Hsp70 which results in the stabilizing of the Hsp70/substrate interaction. In addition, specialized Hsp40s are localized to different sites within the same cellular compartments and can facilitate specific processes of Hsp70 function at different locations. As with the chaperone activity of Hsp40s, data on the co-chaperone activity of these proteins indicates that Type I Hsp40s are more efficient at cooperating with Hsp70 than type II Hsp40s (Terada *et al.*, 1997; Lu and Cyr, 1998b). However, the type II Hsp40 Sis1 is required for *S. cerevisae* viability while the Type I Hsp40 YdJ1 causes less severe growth defects if it is not expressed (Caplan and Douglas, 1991). As Ydj1 cannot substitute for the absence of Sis1, it is evident that it has other important functions to fulfill (Luke *et al.*, 1991). The chaperone specificity of Hsp70 is determined by the Hsp40 co-chaperone as it selects which substrates to convey to Hsp70 (Caplan, 2003). Type III Hsp40s appear to have a more specialized co-chaperone roll and may contain specialized domains for binding of specific proteins or nucleic acids (Cheetham and Caplan, 1998; Walsh *et al.*, 2004; Fan *et al.*, 2003). An example of the different substrate specificities of Type I and Type II Hsp40s is demonstrated by the ability of the Type II Hsp40 Sis1 to maintain the prion state of RNQ+, but the Type I Hsp40 Ydj1 is unable to do so (Fan *et al.*, 2003).

1.7.5) Prenylation of Hsp40s and proteins in parasitic systems

Prenylation of proteins is the post-translational covalent addition of an isoprenoid group to these proteins, allowing them to be reversibly anchored into biological lipid membranes. These isoprenoids are either 20 carbon (Geranylgeranyl pyrophosphate) or 15 carbon (Farnesyl pyrophosphate) chains (McTaggart, 2006). Three enzymes, farnesyl transferase (FTase), geranylgeranyl transferase I (GGTase I) and geranylgeranyl

trasferase II (GGTase II) transfer isoprenoid groups to proteins with a CaaX amino acid sequence motif. This motif consists of a conserved cysteine amino acid residue (C), two usually aliphatic amino acids (aa) and any amino acid (X). Farnesylation or geranylgeranylation is determined according to the amino acid in position X and according to the sequences upstream of the CaaX box (Magee and Seabra, 2005). If X is a serine, methionine, alanine or glutamine the protein is likely to be farnesylated, while leucine or phenylalanine as X normally indicates geranylgeranylation (Kinsella *et al.*, 1991; Casey *et al.*, 1991; Eastman, *et al.*, 2006). The isoprenoid is added to the thiol of the Cysteine. In CaaX boxes at the C-terminus the aaX is cleaved off and the Cysteine carboxymethylated. GGTase II transfers two geranylgeranyl groups to one proteins ending in CC, CCXX or CXC (Magee and Seabra, 2005). Not all proteins containing a CaaX box are prenylated (McTaggart, 2006). The localization of these prenylated proteins is determined partly by their being farnesylated or geranylgeranylated (Magee and Seabra, 2005). Hsp40s such as YdJ1 (Yeast DnaJ 1) from *S. cerevisiae* and HdJ2 (Human DnaJ2, also known as DnaJA1) are farnesylated, which appears to be important for their function (Caplan *et al.*, 1992a; Kanazawa *et al.*, 1997, Davis *et al.*, 1998). Farnesylation of some Hsp40s is required for their function in peptide binding and aggregation suppression (Flom *et al.*, 2008; Summers *et al.*, 2009). Some proteins of *T. brucei* are known to be prenylated (Yokoyama *et al.*, 1997; Eastman *et al.*, 2006) and inhibitors of FTase have been investigated as chemotherapeutic targets for cancer cells and protozoan parasitic infections. Both of these cell types are more sensitive to farnesylation inhibition than are healthy mammalian cells. The FTase in *T. brucei* shows a preference to methionine or glutamine at postion X (Eastman *et al.*, 2006). Farnesylation of Type I Hsp40s has been found to be important for the normal growth of cells at elevated temperatures (Caplan and Douglas, 1991). Although farnesylation of Hsp40s appears to be more common, the human Hsj1b (also known as DnaJB2) Hsp40 is geranylgeranylated (Chapple and Cheetham, 2003).

1.8) Hsp40/Hsp70 interaction

The interaction between Hsp40 and Hsp70 is complex and transient (Genevaux *et al.*, 2007; Bukau and Horwich, 1998; Misselwitz *et al.*, 1999; Jiang *et al.*, 2007). Their

interaction is important for the regulation and function of the Hsp40/Hsp70 chaperone cycle (Mayer and Bukau, 2005; Genevaux *et al.*, 2007). Although the J-domain of Hsp40 is known as the site of minimal interaction with Hsp70, other possible sites of interaction between Hsp40 and Hsp70 have been proposed.

1.8.1) Interactions during the Hsp40/Hsp70 chaperone cycle

Hsp70 fulfills a number of chaperone functions within the cell (**See section 1.6.4**) for damaged and unfolded nascent proteins. It binds proteins displaying areas of non-native structure by utilizing the energy derived from ATP hydrolysis. The ATP hydrolysis of Hsp70 is regulated by a number of co-chaperones including Hsp40 (Palleros *et al.*, 1993; Bukau and Horwich, 1998; Mayer *et al.*, 2000b). The nucleotide bound state of Hsp70 also determines the affinity and exchange rate for these substrates. ATP bound Hsp70 has a low affinity for unfolded protein substrate, while ADP-bound Hsp70 displays high affinity for substrates and increases the relative time of contact between Hsp70 and the substrate (Genevaux *et al.*, 2007; Mayer and Bukau, 2005).

Figure 1.10 shows a generalized scheme for the Hsp40/Hsp70 chaperone cycle, highlighting the system for Type I and Type II Hsp40s. An Hsp40 dimer first recognizes and binds a non-native or unfolded protein, which it then transfers to a complex of Hsp70 and ATP (Genevaux *et al.*, 2007). Many recent models of Hsp70/Hsp40 interaction propose that the Hsp70 moves between the two arms of the U-shaped dimer Hsp40 with the Hsp70 substrate binding domain positioned deepest within the Hsp40 structure (Sha *et al.*, 2000; Li *et al.*, 2008; Genevaux *et al.*, 2007). Through interaction with the Hsp70 ATPase domain and linker region (Jiang *et al.*, 2007), the Hsp40 J-domain causes a conformational change resulting in ATP hydrolysis to ADP, causing transfer of the unfolded polypeptide into the substrate binding pocket of Hsp70 and the closing of the Hsp70 substrate binding domain lid. This produces a high affinity and low exchange rates for the unfolded protein substrates by Hsp70, producing longer periods of hydrophobic patch shielding from the aqueous environment. Hsp40 is released and can screen further proteins and interact with more unfolded polypeptides (Fan *et al.*, 2003). At this point Hsp70 can transport proteins to sites of translocation across membranes, to sites of

proteasome degradation or to other chaperones for folding. The ADP nucleotide is replaced by ATP through a nucleotide exchange factor (NEF) resulting in a lower affinity of Hsp70 for its unfolded protein substrate and the opening of the lid of the substrate binding domain. The unfolded protein is therefore released from Hsp70 and given the opportunity to refold into its native state. If this unfolded protein does not refold correctly, it can aggregate (through hydrophobic patch interaction) with other unfolded polypeptides or it can re-enter the chaperone cycle and interact with Hsp40 and Hsp70 (Fan *et al.*, 2003; Mayer and Bukau, 2005).

It has become evident that not all Hsp70s require a nucleotide exchange factor to replace ADP with ATP and NEF dependence appears to be dependent on structural differences within the ATPase domain of individual Hsp70s (Brehmer *et al.*, 2001; Kluck *et al.*, 2002). Differences in the conformational changes of individual Hsp70s have also been reported. In *E. coli* DnaK, the ATPase and substrate binding domains do not interact in the nucleotide free or ADP bound forms, while interaction occurs during the ATP bound state (Swain *et al.*, 2007). In contrast, the SBD and ATPase domains of *Bos Taurus* Hsc70 were found to interact during both the ATP and ADP bound states (Jiang *et al.*, 2007). The ability of the J-domain to stimulate Hsp70 ATPase activity is enhanced by the presence of peptides bound in the polypeptide binding site of Hsp70 (Bukau and Horwich, 1998). Unfolded peptide binding appears to move the arms of the Hsp40 dimer apart, which has been proposed as a mechanism to stretch the polypeptide into a completely extended conformation. It has also been proposed that this widening of the cleft between the Hsp40 monomers makes space for Hsp70 docking between them (Li *et al.*, 2006; Hu *et al.*, 2008).

This is a generalized cycle for Type I and Type II Hsp40s. Type III Hsp40s do not appear to be general chaperones and have evolved to contain peptide binding domains that recognize specialised substrates (Cheetham and Caplan, 1998). The eukaryotic Hsp70/Hsp40 chaperone cycle appears to use more cochaperones than in prokaryotes. For example, Hip (Hsp70 interacting protein) was found to stabilize the Hsp70-ADP complex (Höhfeld *et al.*, 2001).

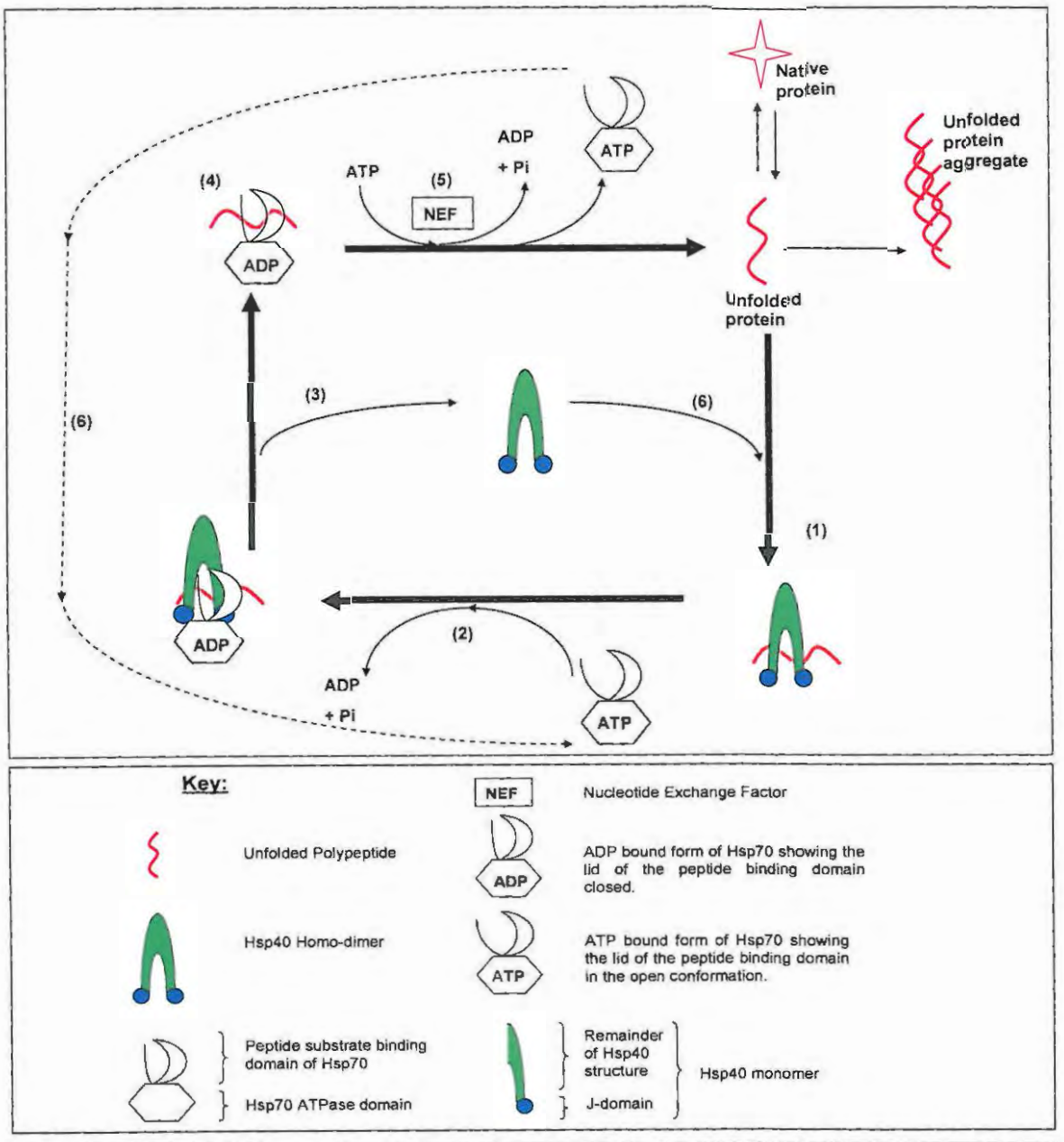


Figure 1.10: A schematic representation of the Hsp70/Hsp40 chaperone cycle. 1) Non-native proteins are recognised and bound by Hsp40. Hsp40 is depicted here in its dimer conformation. 2) Hsp40 delivers the unfolded polypeptide to Hsp70 and stimulates the ATP hydrolysis by Hsp70 to close the lid of the Hsp70 Peptide binding domain, increasing the affinity of Hsp70 for unfolded peptides. 3) Hsp40 is released to bind more non-native proteins. 4) Hsp70 maintains a high affinity for substrate until the exchange of ADP for ATP (5) in the active site of the ATPase domain, which in some cases requires a nucleotide exchange factor. The ADP bound form of Hsp70 has a lower affinity for substrate. This allows the non-native protein more time not being bound with Hsp70, allowing it to find its native state. If it does not find its native state, this protein could become aggregated or it is rebound by Hsp40/Hsp70 in a new cycle. Hsp40 and Hsp70 are recycled (6). Hsp70 has been frequently portrayed with the substrate binding domain moving between the arms of the U-shaped Hsp40 dimer, so that the J-domain of Hsp40 can make contact with the ATPase domain of Hsp70 (Hu *et al.*, 2008; Li *et al.*, 2006; Sha *et al.*, 2000; Genevaux *et al.*, 2007). (Adapted from Fan *et al.*, 2003; Goloubinoff and De Los Rios, 2007)

1.8.2) The sites of interaction between Hsp70/Hsc70 and Hsp40

1.8.2.1) Interaction between the J-domain and Hsc70/Hsp70

The crystal structure of the neuronal specific auxilin J-domain bound to *Bos Taurus* Hsc70 has been solved (Jiang *et al.*, 2007). The structure showed the J-domain to be bound into the cleft between the sub-domains IA and IIA of the Hsc70 ATPase domain. The inter-domain linker region and the portion of this cleft closest to the inter-domain linker region form the Hsc70 side of the intermolecular interface. The interface on the J-domain side involves most of the helix III and the C-terminal portion of helix II and the loop connecting the two helices (Jiang *et al.*, 2007).

Evidence suggests that binding of the J-domain alters the *Bos Taurus* Hsc70 linker region conformation inducing an ordering of the structure of most of the linker region and it is believed to be directed into a hydrophobic patch on the ATPase domain surface. This binding was found to stimulate amino acids Y371 and I181 (corresponding to DnaK I373 and L177) of the hydrophobic patch to convey the ATPase stimulatory signal within the ATPase domain. The signal is then believed to be passed on to E175 (E171 for DnaK) which is believed to transfer the signal to the nucleotide in the active site. In addition, it was found that decoupling of the ATPase domain and substrate binding domain of Hsc70 is required for the binding of the J-domain and access to the hydrophobic patch by the linker. This hydrophobic patch is normally covered by the substrate binding domain and is involved in communication of signals between the two domains (Jiang *et al.*, 2007; Jiang *et al.*, 2005). It is suggested that the nucleotide bound state of Hsc70 affects the interaction between the ATPase domain and the substrate binding domain of Hsc70. The two domains are connected in *Bos Taurus* Hsc70 in the ADP bound form (Jiang *et al.*, 2007), while DnaK-ADP shows a lack of interaction between the two domains (Swain *et al.*, 2007).

The conformation of the J-domain, and especially the HelixII/III loop of the structure, constantly changes in solution until binding to Hsp70 stablilises the structure (Landry *et al.*, 2003).The amino acids of the J-domain important for interaction with the Hsp70 can be divided into charged and hydrophobic residues. Hennessy *et al.*, 2005b summarized

the highly conserved charged residues as the positively charged residues in Helix II, the negatively charged residues in Helix IV, the KFK motif of Helix III and the QKRAA motif of Helix IV (Hennessy *et al.*, 2005b; Hennessy *et al.*, 2000; Auger and Roudier, 1997). The hydrophobic residues are believed to be involved in maintaining the structural integrity, and hence the function of the J-domain. These have been identified as conserved leucines in Helices I and III, an alanine residue in Helix III and the aromatic residues of Helix I (Hennessy *et al.*, 2005b). The HPD motif is the most well characterized region of the J-domain in terms of its necessity for a functional interaction with Hsp70 (Hennessy *et al.*, 2005b; Nicoll *et al.*, 2007; Genevaux *et al.*, 2003). Substitution of any one of these amino acids in the tripeptide motif results in the inability of the Hsp40 to stimulate the ATPase activity of Hsp70 and functionally interact during *in vivo* studies (Hennessy *et al.*, 2005b; Nicoll *et al.*, 2007)

1.8.2.2) Additional sites of interaction between Hsp40 and Hsp70

In addition to the J-domain interaction with Hsp70, there is data that suggests a four amino acid motif (EEVD) on the extreme C-terminus of some Hsp70s is involved in the regulation of intermolecular interactions of Hsp70 and some Hsp40s and the associated function of this interaction. Removal of this motif inhibited the ability of human Hsp70 to refold proteins in coordination with the Hsp40 Hdj1 (Freeman *et al.*, 1995; Qian *et al.*, 2002). Li *et al.* (2006) solved a crystal structure of a complex between the *S. cerevisiae* Type II Hsp40 Sis1 with the 8 amino acids ($G^{634}PTVEEVD^{641}$) from the extreme C-terminus of the *S. cerevisiae* Hsp70 Ssa1 lid domain. The structure showed that the peptide formed a β-sheet with the peptide binding domain of Hsp70, suggesting that such an interaction may occur between Sis1 and the full length protein. The peptide formed an anti-parallel β-strand with the B2 strand of the minor β-sheet in C-terminal domain 1 of Sis1. In addition, hydrophobic interactions (with the Sis1 peptide binding groove in domain 1) and charge-charge interactions occur between the amino acid side chains of the Ssa1 peptide and Sis1 (Li *et al.*, 2006). Sequence comparisons of Sis1 with Hdj1 and other Type II Hsp40s indicate that they could also interact with the EEVD motif of Hsp70 (Li *et al.*, 2006; Hu *et al.*, 2008). The Type I Hsp40 Ydj1 was found to have a very much weakened area of localized positive charge at the site homologous to the site of

charge-charge interactions found in Sis1. This could possibly indicate a different interaction with Hsp70 and could reflect the divergent functions of Type I and Type II Hsp40s (Li *et al.*, 2006). A model has been proposed to explain the interaction of Sis1 with the Ssa1 motif (Li *et al.*, 2006; Hu *et al.*, 2008; Qian *et al.*, 2002). Li *et al.* (2006) suggest that the interaction with the Ssa1 EEVD motif in close proximity to the peptide binding pocket of Sis1 allows the motif to break the interaction between Sis1 and the unfolded substrate and thus facilitate transfer of the substrate to Hsp70. In so doing, Hsp70 replaces some of the interactions with the Sis1 hydrophobic pocket, but with much less affinity than the unfolded substrate (Li *et al.*, 2006).

1.9) Trypanosomal Hsp40s

Since the release of the genome sequences for *Trypanosoma brucei*, *Trypanosoma cruzi* and *Leishmania major* (Berriman *et al.*, 2005; El-Sayed *et al.*, 2005; Ivens *et al.*, 2005), a large number of distinct Hsp40s have been found: 65 in *T. brucei*, 67 in *T. cruzi* and 66 in *L. Major* (Folgueira and Requena, 2007). These have yet to be characterized into Types I to IV. Fourteen different Hsp70 gene sequences have been found for *L. Major*, 12 for *T. brucei* and 28 for *T. cruzi*. However, 16 of the *T. cruzi* genes are partial sequences (Folgueira and Requena, 2007). The other molecular chaperone families of these three trypanosomatids are also reviewed in Folguiera and Requena 2007.

Relative to *E. coli* (Genevaux *et al.*, 2007) and *S. cerevisiae* (Walsh *et al.*, 2004) Hsp40s, very little work has been done to characterize the Hsp40s of trypanosomatids. Only Hsp40s from *T. cruzi* have been studied so far. Tibbetts *et al.* (1998) isolated four distinct Hsp40 genes and named them Tcj1, Tcj2, Tcj3 and Tcj4 (Tcj – *Trypanosoma cruzi* J protein). Tcj2, Tcj3 and Tcj4 displayed evidence of being Type I Hsp40s. Tcj2 and Tcj4 also contain a C-terminal CAAX prenylation motif to indicate membrane association (Tibbetts *et al.*, 1998). Tcj2 was also shown to be heat inducible and contained a conserved sequence in the upstream untranslated region of the gene that has been found in most other heat inducible mRNAs in *T. cruzi* (Tibbets *et al.*, 1998). A further Hsp40 (TcDJ1) was found to localize to the mitochondrion. Both the mRNA and protein populations were found to be greater in the epimastigote forms of the than the metacyclic

forms (Carreira *et al.*, 1998). Salmon *et al.* 2001 reported another distinct Hsp40 sequence from *T. cruzi*, which they termed Tcj6. It was found to have a high sequence similarity to *S. cerevisiae* Sis1 and was discovered to associate with ribosomes and 80S monosomes, just like Sis1 (Salmon *et al.*, 2001).

1.10) Problem statement and motivation

Little work has been published on the trypansomal Hsp40 protein family (Section 1.8; Tibbets *et al.*, 1998; Salmon *et al.*, 2001; Edkins *et al.*, 2004, Folgueira and Requena, 2007) and therefore there is very little known about this protein family. The sequencing of a number of trypanosomatid genomes (Berriman *et al.*, 2005; El-Sayed *et al.*, 2005; Ivens *et al.*, 2005; Peacock *et al.*, 2007) has produced a large amount of sequence data that requires complete annotation and functional characterisation. The categorization of the Hsp40 sequences in these various trypanosomatid genomes into Types I-IV is still to be done. It is possible that some of these trypanosomal Hsp40 sequences could have novel functions in organelles not found in other eukaryotic cells or in the development of parasitaemia in the host.

From a broader perspective, the mechanism of the interaction of Hsp70 with Hsp40 is complex and the dynamics and determinants of this interaction are still being elucidated for individual systems. There are two categories of molecular determinants that govern the ability of proteins to interact functionally: 1) those amino acids necessary and sufficient for structural integrity and physical binding of the two proteins and 2) those amino acids that are necessary and sufficient for functional (catalytic) stimulation. The disruption of the former will disrupt physical binding, and thus functional interaction, while the disruption of the latter disrupts functional (catalytic) stimulation, but may not affect binding. In *E. coli* DnaJ (prokaryotic Hsp40) the D35N mutation is known to increase binding affinity to DnaK (prokaryotic Hsp70) (Landry, 2003).

1.11) Hypothesis

Despite the apparent redundancy, certain trypanosomal Type I Hsp40s interact functionally with Hsp70s and this interaction is essential to growth and survival of trypanosomes.

1.12) Aims and Objectives

Chapter 2:

Bioinformatics analysis of various trypanosomatid genome projects

1) The identification of the Type I and Type I-like Hsp40 sequences in the genome sequences of various trypanosomatid species.
2) Identify potential orthologues of the Type I and Type I-like Hsp40s in the different trypanosomatid species.
3) Make *in silico* predictions for these proteins regarding subcellular localization and farnesylation.
4) Identify putative orthologues of trypanosomal Type I and Type I-like Hsp40s in other organisms (*Arabidopsis thaliana*, *Plasmodium falciparum*, *Homo sapiens* and *Saccharomyces cerevisiae*).
5) Compare the Type I and Type I-like Hsp40s of different trypanosomatid species by multiple sequence alignment.

Chapter 3

Investigation of the amino acid residues important for the interaction of selected trypanosomal type I Hsp40s with Hsp70 using an *in vivo* assay system

1) Test the ability of Tcj2 (*Trypanosoma cruzi* J protein 2), Tcj3 (*Trypanosoma cruzi* J protein 3) and mutants of these proteins to functionally substitute for Ydj1 in a eukaryotic system (*S. cerevisiae*), and to investigate the amino acids important in their interaction with Ssa1 (yeast cytosolic Hsp70).
2) Examine the ability of a eukaryotic J-domain to functionally replace a prokaryotic J-domain in the prokaryotic system (*E. coli*), and to investigate the amino acids important for their interaction with DnaK.
3) Investigate the effect of prenylation on the *in vivo* function of Tcj2 using a Ydj1 deficient *S. cerevisiae* system.

Chapter 4

Investigation of the interaction of the *T. cruzi* type I Hsp40 Tcj2 with Hsp70 using *in vitro* assays

1) Test the ability of His-Tcj2 to stimulate the ATPase activity of multiple Hsp70s.
2) Investigate the binding of His-Tcj2 and His-Tcj2(H34Q) to Hsp70s using quartz crystal microbalance with dissipation monitoring.

Chapter 5

The *in vivo* characterization of the function and subcellular localization of the *T. b. brucei* type I Hsp40 Tbj2 (*Trypanosoma brucei brucei* J protein 2) in *T. b. brucei* laboratory culture

1) Determination of the ability of Tbj2 to be heat stress inducible.
2) Determination of the phenotypic effects of Tbj2 deficiency.
3) Assessment of the subcellular localization of Tbj2 in *T. b. brucei*.

Chapter 2:
Bioinformatic Comparison of the Trypanosomal Type I Hsp40s

CHAPTER 2
Bioinformatic Comparison of the Trypanosomal Type I Hsp40s

2.1) Introduction

As with many other disciplines, the biological sciences have benefited from *in silico* applications. Computers have become integral in the storage, management, analysis and interpretation of biological data. An example of this is in genome sequence data storage and interpretation. The increased rate of genome sequence data collection has made it difficult to manage and draw meaningful conclusions using manual techniques. New developments in computing have provided an ideal platform for managing the large volumes of data that genome sequences contain (Moore, 2007). In addition, characterization of every predicted protein in a genome within the laboratory would be extremely costly in terms of time and resources. As many organisms utilize similar proteins for the same function, there would be an inherent degree of redundancy in this approach. The *in silico* comparison of the sequence data of different genomes is able to predict protein characteristics based on the similarity of amino acid sequences of hypothetical proteins with the amino acid sequences of proteins with known functions. The development of increasingly sophisticated computer programmes to characterize proteins based on their amino acid primary sequence has provided biological scientists with an overview of the differences and similarities of organisms and has allowed them to prioritise the focus of their research to areas that are likely to provide new knowledge and solve important problems. In the case of parasites, such bioinformatics analyses can lead to prioritizing the characterization of proteins that are thought to be important for pathogenicity or that are absent in the host and therefore potential drug targets.

Eight trypanosomatid nuclear genomes have recently been sequenced and annotated, or are in the process of being annotated. In 2005, the annotated genome sequences for *T. brucei brucei*, *T. cruzi* and *L. major* were published (Berriman *et al.*, 2005; El-Sayed *et al.*, 2005b; Ivens *et al.*, 2005). The annotated nuclear genomes of *Leishmania braziliensis* and *Leishmania infantum* have also been published (Peacock *et al.*, 2007). The GeneDB

database (http://www.genedb.org/; Hertz-Fowler *et al.*, 2004) also contains the nuclear genome sequences for *Trypanosoma vivax* (*T. vivax* sequencing project (Sanger institute) http://www.sanger.ac.uk/Projects/T_vivax/), *Trypanosoma brucei gambiense* (*T. b. gambiense* sequencing (Sanger institute) http://www.sanger.ac.uk/Projects/T_b_gambiense/) and *Trypanosoma congolense* (*T. congolense* project (Sanger institute) http://www.sanger.ac.uk/Projects/T_congolense/) that are in the process of being annotated. *T. cruzi*, *L. Major* and *T. brucei* were discovered to have a highly conserved synteny between their genes, despite a predicted divergence of over 200 to 500 million years ago (El-Sayed *et al.*, 2005b; Ghedin *et al.*, 2004). The conservation of the synteny between *L. major*, *L. infantum* and *L. braziliensis* is even higher (up to 99%). Of the three *Leishmania* species whose genomes have been sequenced, *L. major* and *L. infantum* are the most closely related, sharing an average of 92% identity in their amino acid sequences. The genome of *L. braziliensis* is less similar to that of *L. infantum* and *L. major*, showing only a 77% amino acid identity with both genomes (Peacock *et al.*, 2007). **Table 2.1** shows a comparison of the genome size, number of chromosomes and number of genes encoded in the genomes of different trypanosomatid species. *T. brucei brucei* has the smallest nuclear genome, while *T. cruzi* has the largest, with the three *Leishmania* species each having equal genome sizes and an approximate equal number of chromosomes. It has been reported that over 50% of the *T. cruzi* genome consists of repeated sequences (El-Sayed *et al.*, 2005a; Arner *et al.*, 2007). This may explain why *T. cruzi* has a genome size that is almost double that of the other trypanosomatids and contains a third greater number of genes coded by this genome. In the trypanosomatid genomes that have been sequenced, a core of approximately 6200 genes are found in each (Peacock *et al.*, 2007; El-Sayed *et al.*, 2005b). Approximately 1700 of the genes in *T. brucei* (Berriman *et al.*, 2005) and 910 in *L. major* (Ivens *et al.*, 2005) are specific to these organisms when the genomes of *T. cruzi*, *L. major* and *T. brucei brucei* are compared. Comparison of the three *Leishmania* genome sequences (*L. major*, *L. infantum* and *L. braziliensis*) revealed the presence of only 78 species-specific genes (Peacock *et al.*, 2007).

Table 2.1: a comparison of the genome sizes, number of chromosomes and number of genes coded of various trypanosomatid genomes.

	T. brucei	*T. cruzi*	*L. major*	*L. infantum*	*L. braziliensis*
Chromosome number (per haploid genome)	11	~28	36	36	35
Genome size (Mbp)	25	55	33	32	32
Number of genes coded per haploid genome	9068	~12000	8298	8154	8159

(Adapted from El-Sayed *et al.*, 2005b and Peacock *et al.*, 2007)

A large number of Hsp40s have been found in the *T. cruzi* (67), *L. major* (66) and *T. brucei* (65) genomes (Folgueira and Requena, 2007). This number is far greater than the number found in *S. cerevisiae* (22) (Walsh *et al.*, 2004), *E. coli* (6) (Genevaux *et al.*, 2007) and *Homo sapiens* (49) (Kampinga *et al.*, 2009; Qiu *et al.*, 2006). The trypanosomatid Hsp40s identified by Folgueira and Requena (2007) have yet to be classified into the different Hsp40 subtypes (1-IV) that are characterized by their amino acid sequence and domain compositions.

Hsp40s are categorized according to certain domains they contain or lack. Type I Hsp40s are most highly conserved with regard to *E. coli* DnaJ and contain a J- domain located near the N-terminus, a Glycine/Phenylalanine rich region and a Zinc finger-like region consisting of four repeats of CxxCxGxG, where x is any amino acid. Type II Hsp40s lack a Zinc finger-like region, while Type III Hsp40s only contain a J-domain that may be located anywhere within the amino acid sequence (Cheetham and Caplan, 1998). Type IV Hsp40s may possess the same domain structure as any of the other three categories but do not contain a complete HPD motif within the J-domain. There are certain Type IV Hsp40s that can therefore be Type I-like, but do not contain a complete HPD tripeptide (Botha *et al.*, 2007). This subgroup of Type IV Hsp40s will henceforth be referred to as Type IV/I Hsp40s.

Many cellular functions require the reversible localization of certain proteins to membranes. This reversible membrane localization is often achieved by the addition of a farnesyl (15 carbon) or a geranylgeranyl (20 carbon) isoprenoid group through three types of enzymes called prenyl transferases (Maurer-Stroh *et al.*, 2007). Farnesyltransferases and Geranylgeranyl transferase I (GGTase I) both recognize a CaaX

motif at the C- terminus of the protein sequence, unlike GGTase II (**Section 1.7.5; Chapter 1**). Recent developments in the understanding of protein prenylation signals have provided a greater resolution in the distinction of the type of prenylation that a CaaX motif containing protein will receive. Although aliphatic amino acids are generally located at the position of the 'a's in the CaaX box, other amino acids are also allowed, especially in the position of the first 'a' (Maurer-Stroh and Eisenhaber, 2005). The amino acid at the 'X' position of the motif is the most important in determining the type of isoprenoid that is added. The residue volume and charge properties at this position are important for access to the prenyltransferase active site. Leucine and Methionine are preferred at the 'X' position for geranylgeranylation, while farnesylation tolerates a far greater range of amino acids at this position, including these two residues (Roskoski and Ritchie, 1998). The final C-terminal portion of the protein sequence is required to be in an extended conformation, rather than in a helical or coiled conformation, in order to fit into the active site of the prenyl transferases (Reid *et al.*, 2004).

In addition to the CaaX box itself, physicochemical constraints present in the amino acid sequence up to 11 residues upstream from the cysteine of the CaaX box have been found to distinguish the type of isoprenoid that is attached to a protein (Maurer-Stroh and Eisenhaber, 2005). Maurer-Stroh and Eisenhaber (2005) therefore concluded that the prenylation signal is in fact 15 residues long, including the CaaX box and the upstream 11 amino acids. An *in silico* predictor has been developed to predict the kind of prenyl group a protein will receive based on this 15 amino acid motif (Maurer-Stroh and Eisenhaber, 2005). From these advances in prenylation motif recognition, a new prenylation prediction computer programme (The Prenylation Prediction Suite [PrePS]) has been developed, which is able to distinguish between the farnesylation and geranylgeranylation based on the amino acid sequence of a protein alone. The likelihood of this programme generating false predictions is estimated at 0.11% for farnesylation and 0.02% for geranylgeranylation (Maurer-Stroh and Eisenhaber, 2005).

The process of protein prenylation has potential medical significance in the treatment of certain cancers and the treatment of some parasitic infections (Eastman *et al.*, 2006;

Maurer-Stroh *et al.*, 2003; Gelb *et al.*, 2003). Protozoan parasites have been shown to experience greatly reduced growth, relative to mammalian cells, when the farnesylation of their proteins is inhibited. This is due to different substrate specificities between the farnesyltransferases of the two groups (Eastman *et al.*, 2006). Farnesyltransferase knockdown in *T. brucei* bloodstream form cells using RNA interference showed a 74% decrease in cell proliferation and led to the same cellular deformities that were seen in cells treated with farnesyltransferase inhibitors (Gillespie *et al.*, 2007). The application of these inhibitors as potential therapeutic agents for the treatment of *T. brucei* and *P. falciparum* infections is being investigated (Eastman *et al.*, 2006). Some farnesyltransferase inhibitors cause death to the *T. brucei* parasites 48 hours after exposure (Yokoyama *et al.*, 1998). The Type I Hsp40 Ydj1 (derived from *S. cerevisiae*), which is localized to the cytoplasm (Caplan and Douglas, 1991), require farnesylation in order for it to function in the suppression of the effects of heat stress in *S. cerevisiae* (Caplan *et al.*, 1992a). The *T. cruzi* Type I Hsp40s, Tcj2 and Tcj4 have been found to have a C-terminal CaaX motif, strongly suggesting that they undergo farnesylation (Tibbetts *et al.*, 1998). Prenylation is known to be part of *T. cruzi* and *Leishmania mexicana* cellular processes (Yokoyama *et al.*, 1998).

The subcellular localization of a protein can provide clues to its function and assist in the identification of functional orthologues in other related species. Many of the signals that are responsible for addressing the protein to its correct cellular organelle destination have been discovered to be in the form of a peptide in the nascent polypeptide chain (Nakai and Horton, 2007). These signal peptides are well conserved among many eukaryotic organisms (Alberts *et al.*, 1997). A number of computer programmes have been developed to predict the organelle to which a protein is localized based on these signal peptides found in the primary amino acid sequence (Guda and Subramaniam, 2005; Guda, 2006; Horton *et al.*, 2007; Nakai and Horton, 1999). However, those signals that consist of various sections of the primary amino acid sequence coming together once the polypeptide has been folded into its three dimensional structure are more difficult to predict from the primary amino acid sequence, even when using *in silico* algorithms (Emanuelsson and von Heijne, 2001).

Despite the conservation of the peptide signals across eukaryotes, there are cases of divergence between species. This complicates the prediction of the localization of a protein using a generalized *in silico* method which relies on signal peptides. An example of such divergence is the endoplasmic reticulum retention signal. The canonical eukaryotic signal peptide for ER retention is –XDEL at the C-terminus, where the amino acid at position X is species dependent. This signal is -KDEL in mammals, -HDEL in *S. cerevisiae* and -SDEL in *P. falciparum* (Bangs *et al.*, 1996; Teasdale and Jackson, 1996; Kumar and Zheng, 1992). In trypanosomes, a wide variation in the sequence of the retention signal appears to be tolerated. The C-terminal sequences –MDDL and –KQDL have been determined as the retention signals of BiP (Bangs *et al.*, 1993) and a protein disulphide isomerase (Hsu *et al.*, 1989) in *T. brucei*. In addition, the mammalian signal –KDEL also functioned as a valid ER retention signal in *T. brucei* (Bangs *et al.*, 1996).

Mitochondrial import appears to be more conserved among all Eukaryotes. Even the trypanosomes with their divergent mitochondrion use cleavable N-terminal peptides for mitochondrial import (Pena-Daiz *et al.*, 2004). Two groups of mitochondrial targeting signals are thought to be functional in trypanosomes. The first contains a sequence of 15 to 20 amino acids, which is very similar to those found in other organisms, and a second shorter sequence of 7 to 9 amino acids. This second sequence is also found in *S. cerevisiae*, but has been found to work very inefficiently (Hausler *et al.*, 1997).

Apart from signal peptide identification, three other groups of *in silico* methods have been developed for the prediction of protein subcellular localization. **(I)** The amino acid composition methods rely on differences in the amino acid properties and amino acid composition of proteins from different subcellular locations, such as hydrophobicity (Feng and Zhang 2001; Cedano *et al.*, 1997; Cui *et al.*, 2004). These methods have a low accuracy. **(II)** Methods based on the identification of keywords in the annotations of functions of proteins (Nair and Rost, 2002). This method relies on the consistency of descriptions between databases. **(III)** Prediction of localization using phylogenetic profiles, domain projection or structural information (Marcotte *et al.*, 2000; Mott *et al.*,

2002, Nair and Rost, 2003). Most contemporary computer programs designed for protein subcellular localization prediction use a combination of these various methods.

The WoLF PSORT protein localization prediction system is able to predict more than 10 distinct destinations of proteins within a eukaryotic cell (Horton *et al.*, 2007). It was also specifically designed to predict the dual localization of proteins, where proteins shuttle between organelles (Horton *et al.*, 2006; Horton *et al.*, 2007). The system was trained on three separate protein datasets for fungi, plant and animal proteins. The accuracy of the prediction depends on the destination organelle, with the nucleus, mitochondria, cytosol, plasma membrane, extracellular and chloroplast destinations estimated at 70% sensitivity and specificity. Prediction sensitivity for the golgi and peroxisome are predicted to be much lower (Horton *et al.*, 2007). This programme is an extension of and uses features from PSORT II (Nakai and Horton, 1999), PSORT (Nakai and Kanehisa, 1992) and iPSORT (Bannai *et al.*, 2002) and uses the same *k* nearest classifier algorithm as PSORTII (Horton *et al.*, 2006). The parameters considered include amino acid composition, signal peptide identification and orthologue identification with BLAST (Horton *et al.*, 2007; Guda *et al.*, 2006).

pTARGET (Guda, 2006) is able to predict nine different subcellular locations in fungi and metazoans but not plants. It does this by targeting proteins based on domains that are location specific (using data from the Protein Family (PFAM) database [http://pfam.wustl.edu; Bateman *et al.*, 2004]) and differences in the amino acid composition of proteins from different locations (Guda and Subramaniam, 2005; Guda, 2006). If there is no PFAM annotation for a protein sequence or if the PFAM domains contained in the sequence are found in multiple locations, prediction of localization is based on the amino acid composition alone. Therefore, unlike WoLF PSORT, this system is currently unable to correctly predict dual localization of proteins (Guda, 2006). The authors of this prediction method, however, claim that there is a predicted maximum false positive rate of approximately 4% (Guda and Subramaniam, 2005).

As Type I Hsp40s are the focus of this thesis, this chapter aims to identify the Type I and Type I – like Type IV Hsp40s in the different trypanosomatid genomes that have been sequenced. In addition, this chapter intends to identify Hsp40 orthologues in the different genomes and identify interesting features and characteristics of these proteins based on their primary amino acid sequence.

2.2) Materials and Methods

2.2.1) The identification of Hsp40s from various trypanosomatid species whose genomes have been sequenced

In order to find the Hsp40 protein sequences contained within the various trypanosomatid genome sequences on the GeneDB database (http://www.genedb.org/; Hertz-Fowler *et al.*, 2004), the keywords DnaJ and Hsp40 were used to search the database for proteins annotated as DnaJ homologues/Hsp40 proteins. In addition the genome databases for the various trypanosomatid species on GeneDB were subjected to a BLAST search using the Tbj2 (Tb927.2.5160) and Tcj2 (Tc00.1047053511627.110) amino acid sequences. The amino acid sequences of the resultant hits were submitted to PROSITE (http://au.expasy.org/prosite/; Hulo *et al.*, 2007; Sigrist *et al.*, 2002) to identify domains in these sequences. PROSITE was chosen to identify the domains instead of a manual process, to aid efficiency in dealing with a large family of protein sequences from eight different genomes. Only sequences containing a J-domain identified in PROSITE were selected as putative Hsp40s. Type I Hsp40s were identified as proteins containing both the J-domain and the Zinc finger-Like region (CxxCxGxG repeats) as identified by PROSITE. The presence of a glycine/phenylalanine rich region was confirmed by inspection.

2.2.2) The identification of putative Type I Hsp40 orthologues in the different trypanosomatid genomes

Putative homologues of the Type I Hsp40s of the different trypanosomatid species were identified by comparison of amino acid sequence similarities using a phylogenetic comparison of the protein sequences. Pairwise alignments using the Clustal W alignment package on AlignX (part of the Vector NTI software suite) were used to confirm that the

proteins identified as orthologues in the phylogenetic analysis had the highest percentage identity among the Hsp40s of the two species being compared. In addition, the BLAST search function on the GeneDB website was used to confirm orthologues. If the most similar protein in a genome being searched is the predicted orthologue of the query protein, the two proteins are confirmed as being orthologues. Type I proteins of the *T. brucei* genome were used as query proteins for the other genomes. *L. major* was used for orthologue prediction of the J4 grouping (**please refer to Figure 2.2**) as there is no orthologous sequence in *T. brucei*.

The cladogram for the phlyogenetic comparison was generated by aligning the sequences of those proteins identified as Type I or Type I-like Hsp40s in Clustal X (Larkin *et al.*, 2007). This same programme was used to generate the cladogram, which was then visualized using Treeview (Page, 1996).

2.2.3) Prediction of Prenylation in Type I and Type I – like Hsp40s

To predict the farnesylation or geranylgeranylation of the various type I Hsp40s, the Prenylation Prediction Suite (PrePS) (http://mendel.imp.ac.at/sat/PrePS/; Maurer-Stroh and Eisenhaber, 2005) was used.

2.2.4) Prediction of the subcellular localization of the various Type I and Type I-like Hsp40s

Two online software applications were used to predict the possible subcellular localization of the Type I and Type I-like Hsp40s: Target P (http://bioapps.rit.albany.edu/pTARGET/; Guda and Subramaniam, 2005; Guda, 2006) and WolfPsort (http://wolfpsort.org/; Horton *et al.*, 2007).

2.2.5) Homology Modeling of various domains

2.2.5.1) Template identification

The J-domain amino acid sequence for each protein to be modeled was submitted to the Phyre (http://www.sbg.bio.ic.ac.uk/~phyre/; Bennet-Lovsey *et al.*, 2008; Kelley *et al.*, 2000) and Fugue (http://tardis.nibio.go.jp/fugue/prfsearch.html; Shi *et al.*, 2001) servers

for identification of the most suitable template structures. Up to five different structural templates were selected for modeling of the selected Type I Hsp40 J-domains. Templates were chosen based on the similarity of their secondary structure to the predicted secondary structure of the target and the percentage identity of the J-domain amino acid sequences with the target molecules.

2.2.5.2) Model Generation

The selected structural template coordinate files were retrieved from the Research Colaboratory for Structural Bioinformatics (RCSB) Protein Data Bank (PDB) (http://www.pdb.org/pdb/home/home.do). The amino acid sequences of the template structures were retrieved in FASTA format from the same database where it was also verified that there was structural data for all of the amino acid sequence present in the FASTA file. The template amino acid sequences and the query domain amino acids sequence were aligned using the European Bioinformatics Institute (EBI) of the European Molecular Biology Laboratory (EMBL) ClustalW2 world wide web-based alignment programme (http://www.ebi.ac.uk/Tools/clustalw2/index.html; Larkin *et al.*, 2007). The resultant output sequences (possibly containing gaps) were used as input for the generation of a model of the query sequence using the Modeller 9 version 3 (Sali and Blundell, 1993; http://www.salilab.org/modeller/) protein homology modeling programme. The five templates selected in **section 2.2.5.1** were used in the multiple template based modeling setting.

2.2.5.3) Visualisation of modeled structures and generation of figures

The PDB file generated by Modeller 9v3 was visualized and figures generated using Pymol (Delano, 2002).

2.2.5.4) Verification of the accuracy of the generated homology model

The quality of the resultant homology models was assessed firstly by superimposing the model structural coordinates onto the template structures using the University of California, San Francisco (UCSF) UCSF Chimera (http://www.cgl.ucsf.edu/chimera/; Pettersen *et al.*, 2004) protein structure visualization and analysis programme. In

addition, the PROCHECK (Laskowski *et al.*, 1993; Morris *et al.*, 1992; http://www.biochem.ucl.ac.uk/~roman/procheck/procheck.html) computer programme was used to check the quality of the homology model by plotting phi-psi torsion angles and other distorted geometry parameters. The plots produced by PROCHECK included a Ramachandran plot of each of the homology models.

2.2.6) Design of Peptide polyclonal antibodies

The amino acid sequences of Tbj2, Tbj3 and Tcj3 were examined for potential antigenic regions using the protein sequence composite analysis of Generunner (available at: http://www.generunner.net/). The composite analysis uses the methods by Eisenberg and coworkers (1984), Hopp and Woods (1981), Kyte and Doolittle (1982) and Engelman and coworkers (1986) to determine hydrophobic versus hydrophilic regions of the sequence. Other factors that were assessed are chain flexibility (Karplus *et al.*, 1988) and amino acid surface exposure probability (Emini *et al.*, 1985; Janin *et al.*, 1978).The Jameson-Wolf Antigenic index (Jameson and Wolf, 1988) is also used by the protein composite analysis as an overall measure of antigenicity of a particular portion of the amino acid sequence. This method combines the secondary structure prediction methods of Chou and Fasman (1978) and Robson-Garnier (Garnier *et al.*, 1978) with results from the Hopp-Woods hydropathy, Emini Surface Probability and Karplus-Schultz chain flexibility prediction methods to produce an index of antigenicity (Jameson and Wolf, 1988). The charge density of the amino acid sequence is also assessed by Generunner and plotted as the average charge of each amino acid at pH 7.0 (calculated from their pKa) using a sliding window calculation method. The potential antigenic regions were then assessed for their potential in the manufacture of a polyclonal antibody from a peptide region that is sufficiently specific to the protein of interest to prevent binding to different proteins that may occur in the study. This included BLAST of the peptide against the *E. coli*, *S. cerevisiae* and *T. brucei* genomes as these were the systems to be used in this study. In addition the chosen Tbj2 and Tcj3 peptides were assessed for their antigenic suitability by the online antigenicity assessment provided on the Genescript Corporation (Piscataway, New Jersey, USA) website (http://www.genscript.com/) before being ordered through this company. The selected Tbj3 peptide sequence was sent to Prof.

Zimmerman (Department of Medical Biochemistry, University of the Saarland Medical School, Homburg, Germany) for manufacture and rabbit polyclonal antibody production.

2.3) Results

2.3.1) Determination of the number of Hsp40s in various Trypanosomatid species genomes that are Type I Hsp40s

Figure 2.1 shows a graphic representation of the total number of Hsp40s found in the genome database of each trypanosomatid species compared to published literature of the Hsp40 families of *S. cerevisiae* (Walsh *et al.*, 2004), *H. sapiens* (Qiu *et al.*, 2006; Kampinga *et al.*, 2009), *Arabidopsis thaliana* (Miernyk, 2001), *Plasmodium falciparum* (Botha *et al.*, 2007) and *E. coli* (Genevaux *et al.*, 2007). With the exception of *A. thaliana*, all of the trypanosomatid species were found to have a significantly larger number of Hsp40 sequences than the other species (**Figure 2.1**). *T. cruzi* has approximately double the number of sequences compared to the other trypanosomatid species, which can be attributed to 50 % of the *T. cruzi* genome consisting of copies of existing coding sequences (El-Sayed *et al.*, 2005a; Arner *et al.*, 2007). Type I Hsp40s were identified as the sequences containing a J-domain near the N-terminus of the sequence, directly followed by a Gly/Phe rich region and the zinc finger like domain. *E. coli* has by far the fewest Hsp40s in its genome (6 sequences) (Genevaux *et al.*, 2007), which is possibly due to the lack of cellular compartmentalization.

Domains within the sequences identified as Hsp40s in the various genomes were identified using PROSITE (http://au.expasy.org/prosite/; Hulo *et al.*, 2007; Sigrist *et al.*, 2002) and categorized according to their domain composition. The number of Type I and Type IV/I proteins identified in each genome are shown graphically in **Figure 2.1**, with the actual number shown in **Table 2.2.** From **Figure 2.1** it is clear that the number of Type I Hsp40s, in all of the organisms compared, make up a very small proportion of the total number of Hsp40s in each genome. All contain less than 10 and an average of

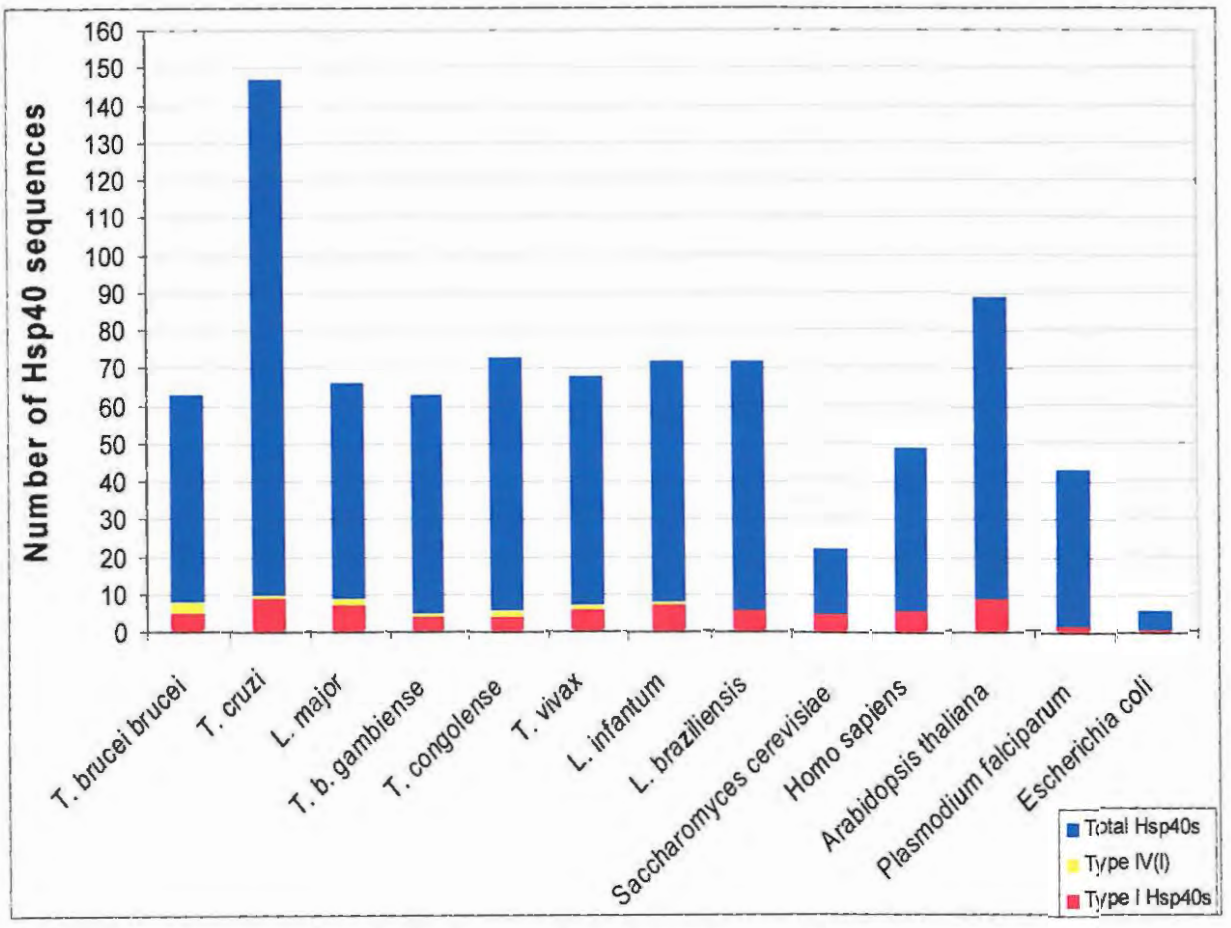


Figure 2.1: The number of Hsp40 sequences of various trypanosomatid organisms in comparison to the number of Hsp40s found in *Homo sapiens* (Qiu *et al.*, 2006; Kampinga *et al.*, 2009), *Arabidopsis thaliana* (Miernyk, 2001), *Saccharomyces cerevisiae* (Walsh *et al.*, 2004), *Plasmodium falciparum* (Botha *et al.*, 2007) and *Escherichia coli* (Genevaux *et al.*, 2007). The number of Type I Hsp40s per species is shown in red, while the total number of Hsp40s per species is shown in blue. The Type IV proteins that are Type I-like for each trypanosomatid species is shown in yellow. The number of Type IV Hsp40s that are type I-like have not been determined for the non-trypanosomatid organisms listed (*S. cerevisiae*, *H. sapiens*, *A. thaliana*, *P. falciparum* and *E. coli*). No Type I-like Type IV proteins were found for *L. braziliensis*. The GeneDB accession numbers of the protein sequences for the trypanosomatid sequences and the NCBI (http://www.ncbi.nlm.nih.gov/; Sayers *et al.*, 2008) protein accession numbers for *S. cerevisiae*, *H. sapiens*, *A. thaliana* and *E. coli* are shown in **Table 2.1**. The *P. falciparum* sequences were derived from the PlasmoDB database (http://plasmodb.org/plasmo/; Bahl *et al.*, 2003; Aurrecoechea *et al.*, 2009), while the *S. cerevisiae* sequences were derived from the Saccharomyces Genome Database (www.yeastgenome.org/; Cherry *et al.*, 1998).

6 Type I Hsp40s per genome. With the exception of *L. braziliensis*, one to two Type IV/I sequences were found in all of the trypanosomatid species contained in this study. *T. brucei brucei* and *T. cruzi* both contain two sequences of this category, but it is likely that the two *T. cruzi* sequences are duplications. *Plasmodium falciparum* has been shown to not have any of this category of Type IV Hsp40s either (Botha *et al.*, 2007). Of the

Hsp40s described as Type I for *S. cerevisiae* (Walsh *et al.*, 2004), *Homo sapiens* (Qiu *et al.*, 2006; Kampinga *et al.*, 2009), *A. Thaliana* (Miernyk, 2001) and *E. coli* (Genevaux *et al.*, 2007), none could be classified as Type IV/I as they all had an intact HPD tripeptide in the J-domain (**Appendix 2B, 2C, 2E and 2F**). This strongly indicates that these organisms do not contain this subtype of Hsp40s as the only difference between Type I and Type IV/I Hsp40s is the HPD motif. The GeneDB (http://www.genedb.org/; Hertz-Fowler *et al.*, 2004) accession numbers for the sequences of the various trypanosomatid species and the National Centre for Biotechnology Information (NCBI) sequence database (http://www.ncbi.nlm.nih.gov/; Sayers *et al.*, 2008), PlasmoDB (http://plasmodb.org/plasmo/; Bahl *et al.*, 2003; Aurrecoechea *et al.*, 2009) and Saccharomyces Genome Database (www.yeastgenome.org/; Cherry *et al.*, 1998) accession numbers for the non-trypanosomatid species of this study are shown for both the Type I and Type IV/I sequences in **Table 2.2**. Certain of the GeneDB trypanosomatid sequences did not start with a methionine amino acid residue, did not end with a stop codon or contained stop codons within the sequence. The latter two may be attributed to frame shift events and in the case of internal stop codons, GeneDB obtained the sequences by joining two sequences. The sequences containing these anomalies are highlighted in green in **Table 2.2** and **Table 2.3** and are classified as putative pseudogenes. They were even found in genomes that have been published (*T. cruzi*; [El-Sayed *et al.*, 2005b] and *L. braziliensis*; [Peacock *et al.*, 2007]). The Type IV/1 accession numbers highlighted in blue indicated amino acid sequences for which prosite was not able to identify an N-terminal J-domain. These sequences were included as Type IV/I sequences as they had a high sequence identity with other Type IV/I sequences from other trypanosomatids and a readily identifiable KDPQ motif instead of the HPD motif was visible. Nine Hsp40s were defined as Type I in *Arabidopsis thaliana* (Miernyk, 2001), but Prosite was unable to identify a Zn finger-like domain in either A63 or A36 and A36 was annotated by PROSITE as having a C-terminal J-domain. Assuming the accession numbers in Miernyk (2001) and NCBI (http://www.ncbi.nlm.nih.gov/; Sayers *et al.*, 2008) correspond to the same sequence, this would indicate a misannotation. However, A63 and A36 were not defined as Type I Hsp40s in this study.

PROSITE identified a Diphthamide (DPH) type of zinc-finger in some of the putative Hsp40 sequences. However, these were not categorized as type I Hsp40s as they did not have the canonical four repeats (CxxCxGxG) of the Type I Hsp40 zinc finger- like region and many are truncated to exclude the C-terminal peptide binding domain and dimerisation domain associated with Type I Hsp40s. The *T. brucei* sequence Tb10.389.1150 and *T. b. gambiense* sequence Tbgamb.32246 are examples of such proteins that were excluded from the Type I Hsp40 group for this reason. Likewise, the *H. sapiens* DnaJA5 (Qiu *et al.*, 2006) sequence was identified by PROSITE to contain the DPH type zinc-finger like domain. It was therefore not acknowledged as a true Type I Hsp40 in this study, even though Qiu *et al.* (2006) described it as such.

Table 2.2 (Page 69): Showing a comparison of the number of confirmed Type I Hsp40s and the total number of Hsp40s found in the genomes of 8 trypanosomatid species with the same parameters in the genomes of *Saccharomyces cerevisiae*, *Homo sapiens*, *Arabidopsis thaliana*, *Plasmodium falciparum* and *Escherichia coli*. The amino acid sequence accession numbers for the typanosomatids (GeneDB; [http://www.genedb.org/; Hertz-Fowler *et al.*, 2004]), *S. cerevisiae* (Saccharomyces Genome Database [www.yeastgenome.org/; Cherry *et al.*, 1998]), *P. falciparum* (PlasmoDB; [http://plasmodb.org/plasmo/; Bahl *et al.*, 2003; Aurrecoechea *et al.*, 2009]), *H. sapiens*, *A. thaliana* and *E. coli* (NCBI; [http://www.ncbi.nlm.nih.gov/; Sayers *et al.*, 2008]) are also shown. The accession numbers representing pseudogene sequences are highlighted in green. Those Type IV/I sequences for which PROSITE was unable to find a J-domain are highlighted in blue.

Notes:

* many of the number of sequences in the *T. cruzi* genome are believed to be allelic copies and duplicates (Folgueira and Requena, 2007; El-Sayed *et al.*, 2005a; Arner *et al.*, 2007).

\# The number of Type IV/I sequences per species was not mentioned in publications of the Hsp40s of these different organisms. However, examination of the Type I Hsp40 sequences identified in these publications, no Type IV/I sequences were found (**Appendix 2B, 2C, 2D, 2E and 2F**).

The Type IV/I accession numbers highlighted in blue indicated amino acid sequences for which prosite was not able to identify an N-terminal J-domain, but percentage identity of these sequences with other Type IV proteins in which prosite identified a J-domain indicated possible status as a type IV Hsp40. In addition, the KDPQ motif that aligns with the HPD of canonical J-domains is also found in these sequences, sometimes with some amino acid substitutions in these sequences (LmjF20.0550 – TDPV) (Tviv796e07.plk_2 – KDPK) (LinJ20_v3.0620 - KDPA). The DP of this motif is consistent throughout these orthologues (J47) (Please refer **to alignment in Figure 2.4**).

Qiu *et al.*, 2006 classifies *Homo sapiens* genome of having 6 Type I Hsp40s, but DnaJA5 (according to the PROSITE annotations) does not have four CxxCxGxG repeats like *E. coli* DnaJ to which Type I Hsp40s are supposed to be most similar. (**Please refer to Appendix 2B**).

Miernyk (2001) defines 9 Hsp40s in the *A. thaliana* genome as being type I Hsp40s, but PROSITE annotations did not define a ZN-Finger motif in either A63 or A36. In addition, a C-terminal J-domain was defined in A36, ruling it out as a Type I Hsp40, the criteria of its definition requiring an N-terminal J-domain. (**Please refer to Appendix 2E**).

Table 2.2: Confirmed Type I and Type IV/I Hsp40s per species of kinetoplastids

Species	Number of confirmed type I Hsp40s	Total number of Hsp40 sequences found	Type I Hsp40s	Putative Type I/IV
Trypanosoma brucei brucei	5	63	Tb927.5160 Tb10.70.5440 Tb927.3.1430 Tb09.211.3680 Tb11.01.8480	Tb09.211.0330 Tb927.1.1230
Leishmania major	7	66	LmjF27.2400 LmjF21.0490 LmjF25.1100 LmjF32.3300 LmjF35.2980 LmjF04.0940 LmjF15.1220	LmjF20.0550
Trypanosoma cruzi	9 (in addition to two pseudogenes)	147 *	Tc00.1047053506445.121 Tc00.1047053510575.200 Tc00.1047053511025.100 Tc00.1047053509233.80 Tc00.1047053509437.40 Tc00.1047053510743.100 Tc00.1047053510659.210 Tc00.1047053510243.30 Tc00.1047053511367.138 Tc00.1047053511627.110 Tc00.1047053507801.130	Tc00.1047053511423.170 Tc00.1047053507949.10
T. brucei gambiense	4	63	Tbgamb.42991 Tbgamb.23116 Tbgamb.2981 Tbgamb.24631	Tbgamb.0216
Trypanosoma congolense	4	73	Congo92c09.q1k_8 congo365g12.p1k_2 congo1293a06.q1k_4 congo541b10.q1k_1	congo1350g02.q1k_11 congo520e01.p1k_13 (truncated paralogue of congo1350g02.q1k_11?)
Leishmania infantum	7	72	LinJ21_V3.0550 LinJ27_V3.2350 LinJ25_V3.1140 LinJ35_V3.3030 LinJ04_V3.0940 LinJ32_V3.3500 LinJ15_V3.1220	LinJ20_V3.0620
Trypanosoma vivax	6	68	Tviv1192b08.p1k_5 tviv1100a12.q1k_0 tviv1163f03.q1k_15 tviv1189h03.q1k_1 tviv623d01.q1k_23 Tviv1689e09.p1k_6	tviv796e07.p1k_2
Leishmania brasiliensis	6 (in addition to one pseudogene)	72	LbrM25_V2.0990 LbrM32_V2.3590 LbrM04_V2.0730 LbrM27_V2.2610 LbrM21_V2.0550 LbrM34_V2.2890 LbrM15_V2.1170	None found
Saccharomyces cerevisiae (Walsh *et al.*, 2004)	5	22	Ydj1 (Ynl064C) Xdj1 (Ylr090w) Scj1 (Ymr214w) Apj1 (Ynl077w) Mdj1 (Yfl016c)	Not defined #
Homo sapiens (Qiu *et al.*, 2006; Kampinga *et al.*, 2009)	6	49	DnaJA1 (NP-001530) DnaJA2 (NP_005871) DnaJA2b (AAB69313) DnaJA3 (NP_005138) DnaJA4 (EAW99177) DnaJA5 (NP_919259)	Not defined #
Arabidopsis thaliana (Miernyk, 2001)	9	89	A2 (AAB86799) (Hdj2) A3 (AAB49030) A24 (CAB80659.1) A26 (AAD22362) A30 (BAB11067.1) A52 (AAD55483) A54 (BAB02706) A63 (AAF07843) A36 (CAB86083.1)	Not defined #
Plasmodium falciparum (Botha *et al.*, 2007)	2	43	PF11_0359 PFD0462w	Not defined #
E.coli (Genevaux et al., 2007)	1	6	DnaJ (NP_414556) or (P08622)	Not defined #

2.3.2) Identification of putative orthologues of Type I and Type I-like Hsp40s in the various trypanosomatids

Homologues are genes that are related to each other by their descent from a common ancestral sequence. Orthologues are homologous genes that are derived from a common ancestor and were separated by speciation. They normally retain the same function in their respective organisms (Koonin, 2005). Alignment of all the trypanosomatid amino acid sequences shown in **Table 2.1** and generation of a tree (**Figure 2.2**) using Clustal X (Larkin *et al.*, 2007) showed distinct clustering of the sequences from the different trypanosomatid genomes that are the most similar. These clustered sequences also had the highest percentage identity in pairwise alignments comparing the sequence identity of Hsp40s between two species. The sequences within a cluster are therefore likely to be orthologues of one another. **Figure 2.2** uses the nomenclature derived by Folgueira and Requena (2007) to categorise the sequences of Hsp40s that they found in the genomes of *T. cruzi*, *T. brucei* and *L. major*. Within each cluster, the Leishmania and Trypanosomal derived sequences are consistently found in two separate sub-clusters. This indicates that the Leishmania sequences in the main clusters are more related to others of Leishmania origin, as are the Trypanosomal sequences more related to other trypanosomal sequences of the same sequence cluster.

Table 2.3 shows the orthologues identified according to their sequence similarity for each of the trypanosomatid species. The putative pseudogenes are again highlighted in green. In this table the sequences are designated as Type I or Type IV/I with a green **I** or **IV** respectively. It is clear from this table that not all of the putative orthologues have been categorized into the same group. This is quite possible as the only difference between the Type I and Type IV/I group lies within the HPD tripeptide of the J-domain. Of all the sequences designated as J27, only the *T. brucei* sequence (GeneDB accession number: Tb09.211.0330) is categorized as a Type IV/I protein. All other trypanosomatid species appear to have only one Type IV/I sequence, while *T. brucei* has two.

Those amino acid sequences identified as J66 by Folgueira and Requena (2007) were detected in the analysis by BLAST of the *T. cruzi*, *L. major*, *T. brucei brucei* and other

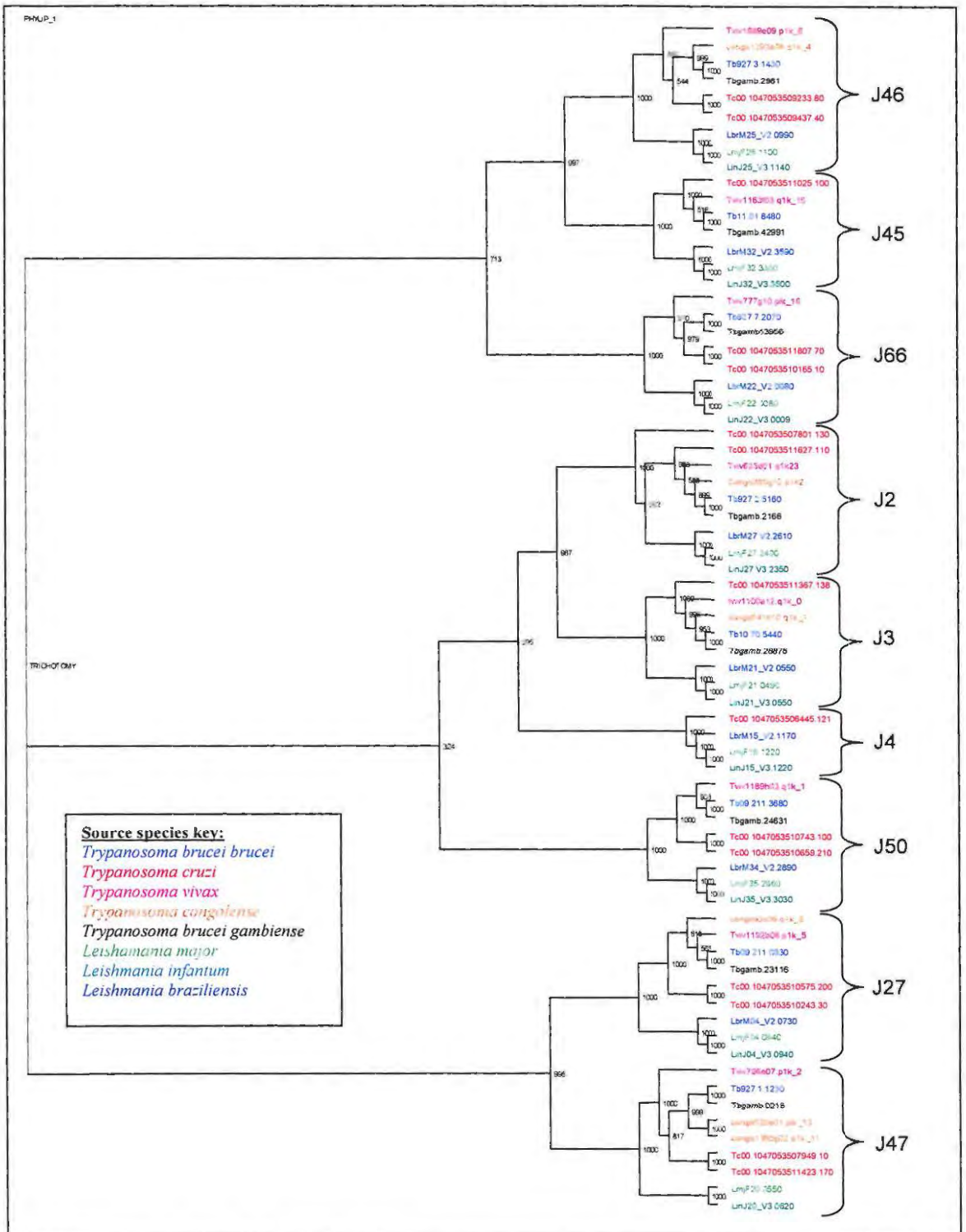


Figure 2.2: Phylogenetic tree showing the clustering of Hsp40 sequences from various trypanosomatid species. The tree was generated using the neighbour-joining method of Clustal X (Larkin *et al.*, 2007). The clustered sequences are most similar and based on this they can be called putative orthologues. The clustered sequences are named according to the nomenclature used for Folguiera and Requena (2007). The Clustal X output was visualized using Treeview (Page, 1996). The sequence accession numbers are colour coded to indicate the source organism (Please see the Source Species Key in the figure). The numbers at the nodes represent bootstrap values.

trypanosomatid genomes with Tbj2 (Tb927.2.5160) conducted during this study. However, annotation by PROSITE indicates that they are unlikely to be Type I or Type IV Hsp40s and that they may not even be Hsp40s due to the lack of detection of a J-domain. A Zinc finger-like region containing the four canonical repeats (CxxCxGxG) is found very close to the N-terminus of each of these sequences, thus indicating that if they are Hsp40s, they are unlikely to be Type I Hsp40s. However, there is a high sequence similarity of Tbj66 (Tb927.7.2070), from the *T. brucei* sequence, to Tbj2 (Tb927.2.5160) from the zinc finger-like region to the C-terminus in both sequences. As these sequences occur in all of the trypanosomatid species it is unlikely that this would be an annotation error in any of the sequence databases. Even if these sequences are not Hsp40s, it is likely that they have a legitimate function of substantial importance within each trypanosomatid, as they have been reasonably well conserved through the process of speciation.

The *T. b. gambiense* orthologue of Tbj2 (Tbgamb.2166) has a high sequence similarity to the C-terminal region of Tbj2 starting from the Zinc finger-like region but does not contain a J-domain. This may be due to the fact that the *T. b. gambiense* genome is still being annotated or this may be a partial deletion of the gene that took place during the divergence of *T. brucei brucei* and *T. brucei gambiense*. The same phenomenon is seen in the orthologue for Tbj3 (Tbgamb.26876). Both TbgambJ3 and TbgambJ2 therefore have a very similar domain structure to that identified for J66 in *T. b. gambiense* (Tbgamb.13956) and the other J66 orthologues. Although these two *T. b. gambiense* sequences were predicted as J2 and J3 orthologues, their lack of a J-domain casts serious doubt on their status as Hsp40s.

The clustering of two *T. cruzi* derived sequences into every J-protein group, with the exception of J3, J4 and J45, indicates that these sequences are gene duplications within the *T. cruzi* genome (Arner *et al.*, 2007) due to the close similarity of both of these sequences to a particular J-protein group. It is also interesting that two of these duplicated sequences (Tc00.1047053507801.130 and Tc00.1047053510165.10) are pseudogenes

indicating that these duplicated sequences could have lost their function as a result of a lack of evolutionary pressure produced by the duplication.

Only 4 of the 8 trypanosomatid genomes examined had a sequence representative in the J4 group. It is interesting to note that those species that have a putative J4 orthologue sequence (the three Leishmania species and *T. cruzi*) all have an intracellular stage to their life cycle within their mammalian hosts (De Souza, 2002; Maia *et al.*, 2007). GeneDB contains an incomplete version of Tcj4, but previous work by Tibbets and coworkers (1998) showed an intact sequence for Tcj4 which is available on the NCBI database (http://www.ncbi.nlm.nih.gov/; Sayers *et al.*, 2008) with the accession number AAC18897. This indicates that the pseudogene status of Tcj4 was either due to a sequencing error or a strain specific variation in the strain used to generate the GeneDB sequence.

T. congolense is missing any putative orthologue for J4, J45, J50 and J66, but has two sequences associated with J47. One of the J47 sequences is designated as a pseudogene. The lower number of Type I Hsp40 sequences in *T. congolense* could be due to the preliminary nature of the annotation of this genome sequence (http://www.genedb.org/genedb/tcongolense/; Hertz-Fowler *et al.*, 2004) or it could indicate a species variation from the other trypanosomatids. *L. braziliensis* is more completely annotated, as is indicated by the publication Peacock *et al.* (2007), but this genome does not appear to have an orthologous sequence for J47.

These sequences have been categorized into orthologues according to their sequence similarities. However, functional orthology implies that they have a similar function within the different species. This requires information on the subcellular localization of these proteins within the cell. Similar localisation implies a similar function within the cell.

Table 2.3 (Page 75): The categorization of the Type I and Type IV/I Hsp40 orthologues of the various trypanosomatid species. The left hand column contains the Hsp40 nomenclature for the different sequences that was adopted by Folguiera and Requena (2007). The amino acid sequences, represented by their geneDB (Hertz-Fowler *et al.*, 2004) accession numbers are categorized as Type I or Type IV/I with a green I for a Type I Hsp40 and a green IV. Due to their high sequence identity with the C-terminal domains of the Type I Hsp40s, the J66 proteins are included in the table, even though they are not identified as Hsp40s because they lack a J-domain. Sequences not containing a J-domain are indicated with a red trident (**refer to Table 2.3 Key**). As with **Table 2.2**, those sequences designated as pseudogenes are highlighted in green. The predicted localization for each of these sequences, as determined by pTARGET (http://bioapps.rit.albany.edu/pTARGET/; Guda and Subramaniam, 2005; Guda, 2006) and WOLF PSORT (http://wolfpsort.org/; Horton *et al.*, 2007) are shown for each sequence (***Please refer to* Table 2.3 Key**). Those sequences that are predicted to be farnesylated, using PrePS (http://mendel.imp.ac.at/sat/PrePS/; Maurer-Stroh and Eisenhaber, 2005) are also indicated (**Please refer to Table 2.3 key**).

Key to Table 2.3:

Tc00.1047053506445.121 sequence does not finish with a stop codon. A more complete version of the gene is displayed on the NCBI nucleotide and protein sequence database (http://www.ncbi.nlm.nih.gov/; Sayers *et al.*, 2008) with the accession number: AAC18897.

*Sequences that are designated as pseudogenes as they contain stop codons or do not start or end with the correct start and stop codons.

Θ	Nucleus localization prediction
Ω	Cytoplasmic localization prediction
Ψ	Mitochondrial localization prediction
Σ	extracellular localization prediction
P	plasma membrane localization prediction
G	Golgi apparatus localization prediction
I	**Type I Hsp40s**
IV	**Type IV Hsp40s that are Type I-like except for lacking an intact HPD motif**
Ψ	**These proteins contain no N-terminal J-domain (their Zinc Finger motif starts within a few amino acids of the N-terminus)**
Δ	proteins that contain a cysteine residue as the 4th last position in the amino acid sequence and contain a putative CAAX box, but are not predicted to be farnesylated.
F	Predicted Farnesylation

- A single localization indicates a consensus between pTARGET and WOLF PSORT in their predictions.
- When the two localization prediction programmes do not predict the same localization the two predictions are separated with a '/' in the order:
  (pTARGET prediction / WOLF PSORT prediction)
- In cases where Wolf PSORT predicts duel localization, the 2 predicted localizations are separated by an underscore e.g. Ω_ Ψ, which indicates a duel localization to the cytoplasm and to the mitochondrion.

	T.brucei brucei	*L. major*	*T. cruzi*	*T.b. gambiense*	*T. congolense*	*L. infantum*	*T. vivax*	*L. braziliensis*
J2	Tb927.1.5160 I Ω F	LmjF27.2400 I Ω F	Tc00.1047053511627.110 I Ω F Tc00.1047053507801.130 I (Ω/ Θ)	Tbgamb2166 Ψ (no J-domain) (Ω/ Ω _Θ) F	Congo365g12.p1k_2 I Ω F	LinJ27_V3.2350 I Ω F	tviv623d01.q1k_23 I Ω	LbrM27_V2.2610_I Ω F
J3	Tb10.70.5440 I Ω F	LmjF21.0490 I Ω F	Tc00.1047053511367.138 I Ω Δ	Tbgamb.26876 (no J-domain) Ψ (Ω/ Σ) F	congo541b10.q1k_1 I Ω	LinJ21_V3.0550 I Ω F	tviv1100a12.q1k_0 I Ω Δ	LbrM21_V2.0550 I Ω Δ
J4		LmjF15.1220 I (Θ/ Ω)	# Tc00.1047053506445.121 I F (Ω/ Θ)			LinJ15_V3.1220 I Ω		LbrM15_V2.1170 I Θ
J45	Tb11.01.8480 I (II /Σ)	LmjF32.3300 I (II / Σ)	Tc00.1047053511025.100 I (Ω/ Σ)	Tbgamb.42991 I (II / Σ)		LinJ32_V3.3500 I (II / Σ)	tviv1163f03.q1k_15 I (Ω/ Σ)	LbrM32_V2.3590 I (Ω/ Σ)
J46	Tb927.3.1430 I II	LmjF25.1100 I II	Tc00.1047053509233.80 I (Ω/ Σ) Tc00.1047053509437.40 I (II / Σ)	Tbgamb.2981 I II	congo1293a06.q1k_4 I (Ω/ Σ)	LinJ25_V3.1140 I (II/ G)	Tviv1689e09.p1k_6 I (II/ P)	LbrM25_V2.0990 I (II / Σ)
J50	Tb09.211.3680 I II Δ	LmjF35.2980 I II F	Tc00.1047053510659.210 I II Δ Tc00.1047053510743.100 I II Δ	Tbgamb.24631 I II Δ		LinJ35_V3.3030 I II F	tviv1189h03.q1k_1 I II Δ	LbrM34_V2.2890 I II F
J27	Tb09.211.0330 IV Ω	LmjF04.0940 I II	Tc00.1047053510575.200 I II Tc00.1047053510243.30 I (Ω/ II)	Tbgamb.23116 I II	congo92c09.q1k_8 I II	LinJ04_V3.0940 I II	Tviv1192b08.p1k_5 I II	LbrM04_V2.0730 I II
J47 (Type IV)	Tb927.1.1230 IV II	LmjF20.0550 IV (Θ/ II)	Tc00.1047053507949.10 IV (Θ/ II) Tc00.1047053511423.170 IV (Θ/ II)	Tbgamb.0216 IV II	congo1350g02.q1k_11 IV II congo520e01.plk_13 IV (Θ/ II)	LinJ20_V3.0620 IV (Θ/ II)	tviv796e07.p1k_2 IV (Θ /Ω_Θ)	
ZN finger containing proteins without N-terminal J-domain								
J66	**Tb927.7.2070 Ψ Ω Δ**	LmjF22.0080 Ψ (Ω/ Σ) F	Tc00.1047053511807.70 Ψ (Ω/ Σ) Δ Tc00.1047053510165.10 Ψ (Ω/ Σ) Δ	Tbgamb13956 Ψ (Ω/ Ω _Θ) Δ		LinJ22_V3.0009 Ψ (Ω/ Σ) Δ	Tviv777g10.plk_16 Ψ (Ω/ Σ) Δ	LbrM22_V2.0080 Ψ (Ω / II) Δ

2.3.3) Proposed nomenclature of trypanosomatid Hsp40s

Trypanosomal Hsp40s have traditionally been named according to the first letter of the genus and species followed by a J to indicate that it is a DnaJ-like protein and a number to distinguish the different Hsp40s from the same organism (Tibbets *et al.*, 1998; Folgueira and Requena, 2007). *Trypanosoma cruzi* J-protein 2 has traditionally been referred to as Tcj2. Tibbetts and coworkers (1998) used this nomenclature for *T. cruzi* Hsp40s and Folgueira and Requena (2007) proposed to maintain this nomenclature for *L. major* and *T. brucei brucei.* In the naming of the Hsp40s of the other species discussed in this chapter, it is proposed that this same nomenclature is maintained, with the protein numbering outlined by Folguiera and Requena (2007)(**Table 2.4**) However, to avoid confusion between *Trypansoma cruzi* and *Trypanosoma congolense*, the initial TcongJ should be used for *T. congolense*. Likewise, to avoid confusion between sequences from *T. brucei brucei* and *T. brucei gambiense*, TbgambJ will be used for *T. brucei gambiense.* Perhaps an extension of this nomenclature should include the distinction made between Type I (A), Type II (B) and Type III (C) Hsp40s used by Qiu and coworkers (2006) and Kampinga and coworkers (2009). For example TbjA2 to designate that Tbj2 is a Type I Hsp40.

Table 2.4: The proposed nomenclature for trypanosomatid Hsp40s

Species	**Hsp40 nomenclature**
Trypanosoma brucei brucei	TbJ
Trypanosoma cruzi	TcJ
Trypanosoma vivax	TvivJ
Trypanosoma congolense	TconJ
Leishmania major	LmJ
Trypanosoma brucei gambiense	TbgambJ
Leishmania infantum	LinJ
Leishmania braziliensis	LbrJ

2.3.4) Prediction of the sub-cellular localization of Type I Hsp40s and Type I-like Hsp40s

Owing to the absence of a computer programme designed specifically for prediction of the trypanosomatid protein subcellular localization, programmes that predict the localization of proteins in animal cells were used. Both pTARGET and WOLF PSORT

are able to process query sequences using one of three datasets: those from animals, plants or fungi. The animal setting was used for these predictions. **Table 2.3** shows the predicted subcellular localization of each protein next to the accession number with coloured symbols (See **Table 2.3 KEY**).

All of the sequences that are designated as true Type I Hsp40s in the J2 and J3 groups, that are not classified as pseudogenes were all predicted to be localized to the cytoplasm. The two sequences shown as orthologues for *T. b. gambiense* in the J2 and J3 families may not be Hsp40s due to their lack of a J-domain. The less common J4 family of sequences were predicted to have a cytoplasmic or a nuclear localization. The J50 orthologues were all predicted by both localization prediction programmes to be localized to the mitochondrion. The J27 proteins were also all predicted to localize to the mitochondrion with the exception of one of the *T. cruzi* gene repeats and the *T. brucei brucei* orthologue, which was predicted to be cytoplasmic. Other evidence for the functional divergence of Tbj27 from the other J27 proteins is that it is a Type IV/I Hsp40 while all the other J27 proteins are Type I Hsp40s. All of the sequences predicted as mitochondrial contained a stretch of amino acids rich in arginine directly at the N-terminus, before the start of the J-domain (Please Refer to Appendix 2A for the annotated amino acid sequences of Type I and Type IV/I Hsp40s of the different Trypanosomatids). The general characteristics of the mitochondrial import signal include it being rich in positively charged residues such as arginine (Emanuelsson and von Heine, 2001) and often N-terminal (Emanuelsson and von Heine, 2001; Pena-Diaz *et al.*, 2004; Claros and Vincens, 1996). This is the subcellular localization destination that is the most accurately predicted by most contemporary subcellular localization prediction computer software (Sprenger *et al.*, 2006). This is therefore the likely reason for the high level of consensus between the localization prediction programmes when a true prediction of mitochondrial localization occurs.

In the J45, J46 and J47 orthologue groups, the predicted localization is less consistent between the various member sequences. WOLF PSORT consistently predicts the J45 group of sequences to have an extracellular localization, while pTARGET designated half

as mitochondrial and half as cytoplasmic. PTARGET on the other hand predicted most of the sequences in the J46 group of orthologues to be mitochondrial, while WOLF PSORT predicted either mitochondrial or extracellular localization, with the exception of LinJ46 (LinJ25_V3.1140) and TvivJ46 (Tviv1689e09.p1k_6), which were predicted to be localized to the golgi apparatus and plasma membrane respectively. However, pTARGET has been found to be better at predicting plasma membrane localization than WOLF PSORT (Sprenger *et al.*, 2006), so it is unlikely that TvivJ46 is localized to the plasma membrane. Prediction of a protein targeting to the golgi apparatus using the primary amino acid sequence is less accurate using contemporary localization prediction programmes (Sprenger *et al.*, 2006). It is therefore questionable as to whether WOLF PSORT predicted the correct localization. With the exception of TvivJ47 WOLF PSORT predicted a mitochondrial localization for the J47 orthologues, while pTARGET predicted either mitocbondrial or nuclear localization. TvivJ 47 was predicted to have a cytoplasmic and nuclear dual localization. As pTARGET predicted a nuclear localization and does not make provision for dual localizations, it is possible that this protein is localized to both the cytoplasm and the nucleus.

In the J66 sequences, pTARGET consistently predicted a cytoplasmic localization, as it did for the other two J66 like proteins, TbgambJ2 (Tbgamb.2166) and TbgamJ3 (Tbgamb.26876). WOLF PSORT, however, predicted a wider range of localizations for the various proteins of this group.

The predicted localization of the Hsp40s, leads to the possibility of predicting the Hsp70 partners of each Hsp40 using *in silico* localization prediction. **Table 2.5** Shows the predicted *T. brucei brucei* Hsp70/Hsp40 partners based on *in silico* localization prediction. However, it is difficult to determine the Hsp40/Hsp70 partnerships in organelles that contain more than one of each of these protein families. It can only be concluded that all of the Hsp70 and Hsp40 proteins predicted to localize to a certain cellular compartment are potential partners. These will need to be confirmed not only through subcellular localization experiments, but also through "bait and prey" studies *in vivo*.

Table 2.5: The prediction of potential Hsp40/Hsp70 partners for *T. brucei brucei* based on their predicted subcellular localization. The Hsp70 localisation prediction (using PTarget [http://bioapps.rit.albany.edu/pTARGET/; Guda and Subramaniam, 2005; Guda, 2006] and WOLF PSORT [http://wolfpsort.org/; Horton *et al.*, 2007]) was derived from Louw, 2009.

Subcellular Organelle	Hsp70	Hsp40
Cytoplasm	Tb11.01.3110 Tb11.01.3080 Tb927.7.710 Tb10.389.0880	Tbj2(Tb927.1.5160) Tbj3 (Tb10.70.5440) Tbj27(Tb09.211.033)0 * Tbj66(Tb927.7.2070) Ψ
Endoplasmic Reticulum	Tb11.02.5500 Tb11.02.5450	-
Nucleus	Tb927.7.1030	-
Mitochondrion	Tb927.6.3740 Tb927.6.3750 Tb927.6.3800	Tbj45 (Tb11.01.8480) Tbj46 (Tb927.3.1430) Tbj50 (Tb09.211.3680) Tbj47 Tb927.1.1230 *
Secreted	Tb09.211.1390 Tb09.160.3090	Tbj45 (Tb11.01.8480)

* These Hsp40 sequences were identified as Type IV Hsp40s that were Type I-like.
Ψ Tbj66 contained all the Type I Hsp40 domains, except the J-domain.

2.3.5) Prediction of prenylation in trypanosomatid Type I Hsp40s

Of the sequences predicted to be prenylated (**Table 2.3**) all were predicted to be farnesylated. Geranylgeranylation is recognised as being very rare modification for Hsp40s (Chapple and Cheetham, 2003). Many of the sequences in **Table 2.3** contain a putative CaaX box with a cysteine residue at the fourth last amino acid position at their C-terminus. However, PrePS (http://mendel.imp.ac.at/sat/PrePS/; Maurer-Stroh and Eisenhaber, 2005) did not predict these sequences to be farnesylated. While an average of approximately 2 Type I-like Hsp40 sequences per species was predicted to be farnesylated, each species had an average of 2 putative CaaX containing sequences that were not predicted to be farnesylated.

Three of the *T. vivax* sequences in **Table 2.3** contain a putative CaaX sequence, but none of them were predicted to be farnesylated. None of the *L. major* sequences with a putative CaaX sequence identified were not predicted to be farnesylated. According to the sequences in **Table 2.3**, the Leishmanias showed a marginally higher rate of predicted farnesylation than their Trypanosomal counterparts.

All of the orthologues in the J2 group were predicted to be farnesylated, with the exception of the *T. vivax* sequence, while the J3 group contained the second highest

number of farnesylation predictions. Of the orthologue groups that contain sequences that are predicted to be farnesylated (J2, J3, J50 and J66) the other sequences that are not predicted to be farnesylated contain a putative CaaX sequence. This motif was not found in the J46, J27 and J47 groups at all. This could mean that farnesylation was a feature that was gained or lost by the orthologous sequences of the J2, J3, J50 and J66 groups during the process of species divergent evolution. It is also interesting to note that Leishmania sequences from J50 were all predicted to be farnesylated, while the trypanosomal sequences were all predicted to not be farnesylated. Tcj4 is the only sequence predicted to be farnesylated in the J4 group, while none of the others in this group even possess a putative CaaX motif.

2.3.6) Expression of the various Type I and Type IV Hsp40s in the different life cycle stages of the different trypanosomatid species

There is currently limited data for the differential expression of Type I and Type IV/I Hsp40s in the life-cycle stages of the trypanosomatid species. All of the *Leishmania major* proteins represented by the sequence accession numbers shown in **Table 2.3** have been found to be constitutively expressed (Leifso *et al.*, 2007). There is more variation in the expression of the different Type I Hsp40s in the different life cycle stages of *T. brucei brucei* and *Leishmania infantum*, for which the data is shown in **Table 2.6.** The data concerning the expression status for the *T. brucei brucei* Type I Hsp40s in the procyclic and bloodstream stages is incomplete. Of the 5 Type I Hsp40s in *T. brucei brucei*, only Tbj2 is known to be expressed in both the procyclic and the bloodstream stages (Vertommen *et al.*, 2008). It is not known if Tbj46 (Tb927.3.1430), Tbj45 (Tb11.01.8480) and Tbj47 (Tb927.1.1230) are expressed in the *T. brucei brucei* procyclic stages. Only Tbj2 (Jones *et al.*, 2006; Vertommen *et al.*, 2008) and Tbj66 (Subramaniam *et al.*, 2006) have been shown to be expressed in the bloodstream stages. In *Leishmania infantum*, LinJ2 (LinJ27_V3.2350) has down regulated expression during the mature amastigote stage, while expression of LinJ50 (LinJ35_V3.3030) and LinJ45 (LinJ32_V3.3500) is increased during this same life-cycle stage relative to the promastigote stages (Rosenzweig *et al.*, 2008). This sparse data appears to indicate that, despite the genetic similarities between the trypanosomatids especially between the

different *Leishmania* species, the different species have a very different system of regulating the expression of Type I Hsp40 proteins in the different life-cycle stages. However, more complete data is required to substantiate this claim before more detailed conclusions can be made.

Table 2.6: The differential expression of Type I and Type IV Hsp40s in the different life-cycle stages of *T. brucei brucei* and *L. infantum*. The names follow the same nomenclature derived from Folgueira and Requena, 2007). Tbj - *T. brucei brucei* J protein. LinJ - *Leishmania infantum* J protein. The names in blue indicate Type IV/I Hsp40s and the names highlighted in red indicate those proteins that lack an N-terminal J-domain, despite their sequence identity with other Hsp40s. The expression of *T. brucei brucei* proteins is divided into expression of the proteins in the procyclic stages within the insect and expression in he bloodstream stages within the mammalian host. However, the expression of proteins in *Leishmania infantum* are divided into expression in the insect vector or the amastigote stage inside the mammalian host.

Name	Sequence accession no.	Expressed in insect stages	Expressed in mammalian host stages
Trypanosoma brucei brucei			
		Expressed in insect procyclic stages	**Expressed in Bloodstream stages**
Tbj2	Tb927.2.5160	**Yes** (Vertommen *et al.*, 2008)	**Yes** (Jones *et al.*, 2006; Vertommen *et al.*, 2008)
Tbj3	Tb10.70.5440	**Yes** (Vertommen *et al.*, 2008)	**No data**
Tbj45	Tb11.01.8480	**No data**	**No data**
Tbj46	Tb927.3.1430	**No data**	**No data**
Tbj50	Tb09.211.3680	**Yes** (Vertommen *et al* 2008; Jones *et al.*, 2006)	**No data**
Tbj27	Tb09.211.0330	**Yes** (Vertommen *et al.*, 2008)	**No data**
Tbj47	Tb927.1.1230	**No data**	**Yes** (Subramaniam *et al.*, 2006)
Tbj66	Tb927.7.2070	**Yes** (Jones *et al.*, 2006)	**No data**
Leishmania infantum			
		Expressed in the insect stages	**Expression in the mature amastigotes in the mammalian host relative to the insect stages**(Rosenzweig *et al.*, 2008)
LinJ2	LinJ27_V3.2350	**Yes**	Downregulated expression
LinJ3	LinJ21_V3.0550	**No data**	**No data**
LinJ4	LinJ15_V3.1220	**No data**	**No data**
LinJ45	LinJ32_V3.3500	**Yes**	Upregulated expression
LinJ46	LinJ25_V3.1140	**No data**	**No data**
LinJ50	LinJ35_V3.3030	**Yes**	Upregulated expression
LinJ27	LinJ04_V3.0940	**No data**	**No data**
LinJ47	LinJ20_V3.0620	**No data**	**No data**
LinJ66	LinJ22_V3.0009	**No data**	**No data**
Leishmania major			
All *L. major* sequences mentioned in this chapter are constitutively expressed (Leifso *et al.*, 2007)			

2.3.7) Putative orthologues of *T. brucei* Type I Hsp40s in the *Homo sapiens*, *Arabidopsis thaliana*, *Saccharomyces cerevisiae* and *Plasmodium falciparum* genomes

The *T. brucei brucei* Type I Hsp40 and Type I-like Hsp40 sequences were compared to the Type I protein sequences of *S. cerevisiae* (according to Walsh *et al.*, 2004), *H. sapiens* (according to Qiu *et al.*, 2006), *P. falciparum* (according to Botha *et al.*, 2007) and *A. thaliana* (according to Miernyk, 2001). Although the *H. sapiens* sequence DnaJA5 and A36 and A63 from *A. thaliana* have been reported as Type I Hsp40s (Qiu *et al.*, 2006; Miernyk, 2001), these sequences were excluded from the dataset for the same reason outlined in **section 2.3.1**. The sequences from other trypanosomatid species were also excluded to simplify the comparison and the putative orthologues with the non-trypanosomatid species could be inferred by their orthology with the *T. brucei brucei* sequences. The cladogram generated to compare the similarity of the sequences (**Figure 2.3**) showed a generalized clustering of those sequences predicted to be localized to the mitochondria or chloroplasts (**Figure 2.3A**) and those predicted to be localized to the cytoplasm (**Figure 2.3B**). The notable exceptions being Tbj27 in **Group A**, which is localized to the cytoplasm, and A2 and A3 from *A. thaliana*, which are predicted to be localized to the nucleus. With the exception of the *S. cerevisiae* sequence Apj1 and the *H. sapiens* sequences DnaJA2 and DnaJA2b, all of the sequences in the **group B** cluster were predicted by PrePS (http://mendel.imp.ac.at/sat/PrePS/; Maurer-Stroh and Eisenhaber, 2005) to be farnesylated. In addition, DnaJA2 and DnaJA2b both contained a putative CaaX box with a cysteine at the fourth last position before the C-terminus. Assuming that PrePS is correct in its prediction of a lack of farnesylation of these two proteins, it could imply a subtle evolutionary divergence from the other sequences of **group B**. The lack of farnesylation of DnaJA2 and DnaJA2b could have profound implications for the function of these two proteins relative to DnaJA1 and DnaJA4, despite their high level of sequence similarity. Scj1 from *S. cerevisiae* showed the most divergence from the *T. brucei brucei* sequences as is indicated by its separation into a third branch of the cladogram on its own. It is also the only Type I Hsp40 in **Figure 2.3** that is predicted to be localized to the endoplasmic reticulum.

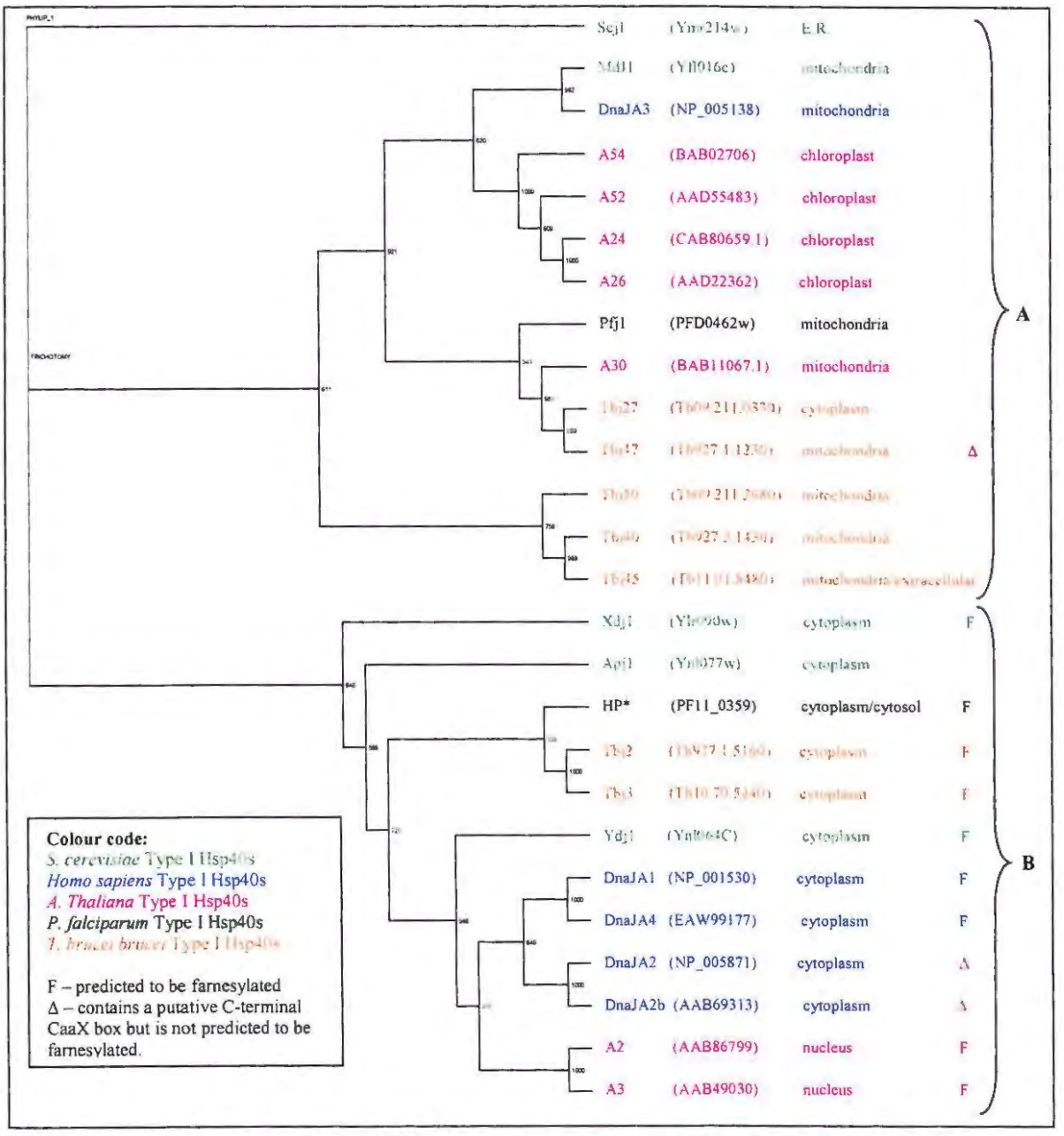


Figure 2.3: Cladogram displaying the similarities of *T. brucei* Type I and Type I like Hsp40s with those from *H. sapiens* (Qiu *et al.*, 2006), *S. cerevisiae* (Walsh *et al.*, 2004), *A. thaliana* (Miernyk, 2001) and *P. falciparum* (Botha *et al.*, 2007). *S. cerevisiae* sequences were obtained from Saccharomyces Genome Database (www.yeastgenome.org; Cherry *et al.*, 1998), *P. falciparum* sequences were obtained from PlasmoDB ((http://plasmodb.org/plasmo/; Bahl *et al.*, 2003; Aurrecoechea *et al.*, 2009) and sequences for *H. sapiens* and *A. thaliana* were obtained from NCBI (http://www.ncbi.nlm.nih.gov/; Sayers *et al.*, 2008). The sequences are colour coded according to the organism from which they are derived (Inset Key). The names of the proteins are accompanied by the amino acid sequence accession numbers for the respective sequence databases. Predictions of subcellular localization using WOLFPSORT (Horton *et al.*, 2007) and pTarget (Guda and Subramaniam, 2005; Guda, 2006) are also shown. Farnesylation prediction using PrePS (http://mendel.imp.ac.at/sat/PrePS/; Maurer-Stroh and Eisenhaber, 2005) for each of the sequences uses the same nomenclature as for **Table 2.3.** **F** indicates farnesylation, while **Δ** indicates that the protein contains a putative C-terminal CaaX box but is not predicted to be farnesylated. *The *P. falciparum* sequence (PF11_0359) was named Hypothetical Protein (HP) after the Botha *et al.*, 2007 nomenclature.

The large evolutionary distance between the organisms from which the sequences in **Figure 2.3** are derived, is evident in the fact that these sequences don't cluster very well with sequences from other organisms. This is clearly shown by the tendency of the *T. brucei brucei* sequences to be clustered as a group partially separated from the other sequences (**Figure 2.3**). This is in direct contrast to **Figure 2.2,** where sequences from different trypanosomatid species tend to cluster as opposed to clustering with sequences from the same species. The other organisms being compared in **Figure 2.3** also have the same tendency to cluster with other sequences derived from the same organism. The four Hsp40 sequences from *H. sapiens* that form part of **group B** in **figure 2.3** are shown as more related to one another than to sequences from other organisms in this group. The same is seen in the clustering of the sequences A2 and A3 from *A. thaliana.* Tbj2 and Tbj3 are shown to be the most similar to the *P. falciparum* protein HP (PF_11_0359) in **Figure 2.3 group B** and form a sub-cluster on the cladogram. Likewise, Ydj1, the Human Hsp40s and the *A. thaliana* sequences of **group B** form a cluster separate from the Tbj2, Tbj3 and HP cluster. The *S. cerevisiae* sequences in group B show a large divergence in their sequences, especially in Ydj1, which shows more similarity to the *H. sapiens* and *A. thaliana* sequences.

Despite the evolutionary distances between the source organisms in **Figure 2.3,** tentative orthologues can be predicted using this comparison along with information from pairwise sequence alignments. These putative orthologues are displayed in **Table 2.7**. Pairwise alignment of HP (PF_11_0359) from *P. falciparum* with Tbj2 and Tbj3 both resulted in approximately 41% identity. Therefore, it is possible to label both Tbj2 and Tbj3 as tentative orthologues of HP (PF11_0359). A2 and A3 of *A. thaliana* are shown as equally distant from Tbj2 and Tbj3 in **Figure 2.3**. Pairwise alignment shows the Tbj2 has a higher percentage identity to A3 (40.2%) than A2 (39.8%), while Tbj3 has a slightly higher percentage identity to A2 (36.3%) relative to A3 (35.9%). However, these differences in the percentage identities could easily be due to slight differences in the lengths of the different proteins (Tbj2 is 404 aa long, Tbj3 416aa, A2 419aa and A3 420aa). It is therefore prudent to label Tbj2 and Tbj3 as putative orthologues of both A2 and A3. The four human Hsp40s in group B were equally difficult to distinguish as

orthologues of Tbj2 and Tbj3, with each showing a marginally higher percentage identity (~2-3%) to Tbj2 than Tbj3. DnaJA1 showed the highest percentage identity with Tbj2 (44%) and Tbj3 (40%). Tbj2 and Tbj3 are found between Apj1 and Ydj1 in **Figure 2.3** and this could indicate similar levels of sequence similarity relative to both of these *S. cerevisiae* sequences. However, from pairwise alignments and size comparisons, it appears that Ydj1 is most similar to both Tbj proteins. Its length of 409 aa is similar to Tbj2 (404aa) and Tbj3 (416aa), while Apj1 is significantly longer (528aa). Due to its similar length with Tbj2 and Tbj3, Ydj1 shows a higher percentage identity (~40%) with these sequences than does Apj1 (~20%).

Table 2.7: Putative orthologues of Type I and Type I like Hsp40s in *Homo sapiens*, *Saccharomyces cerevisiae*, *Arabidopsis thaliana* and *Plasmodium falciparum*. The classifications in this table were derived from **Figure 2.3**

T. brucei	*S. cerevisiae*	*Homo sapiens*	*A. thaliana*	*P. falciparum*
J2/J3	Ydj1	DnaJA1 DnaJA2 DnaJA2b DnaJA4	A2/A3	PF11_0359
J45	-	-	-	-
J46	-	-	-	-
J50	-	-	-	-
J27	-	-	-	-
J47	-	-	-	-
J27/J47	Mdj1 *	DnaJA3*	BAB11067.1	PFD0462w

*pairwise alignment of Mdj1 and DnaJA3 with J27 and J47 of *T. brucei brucei* did not produce a sequence identity close to 40%. It is therefore speculation that these sequences may be orthologues.

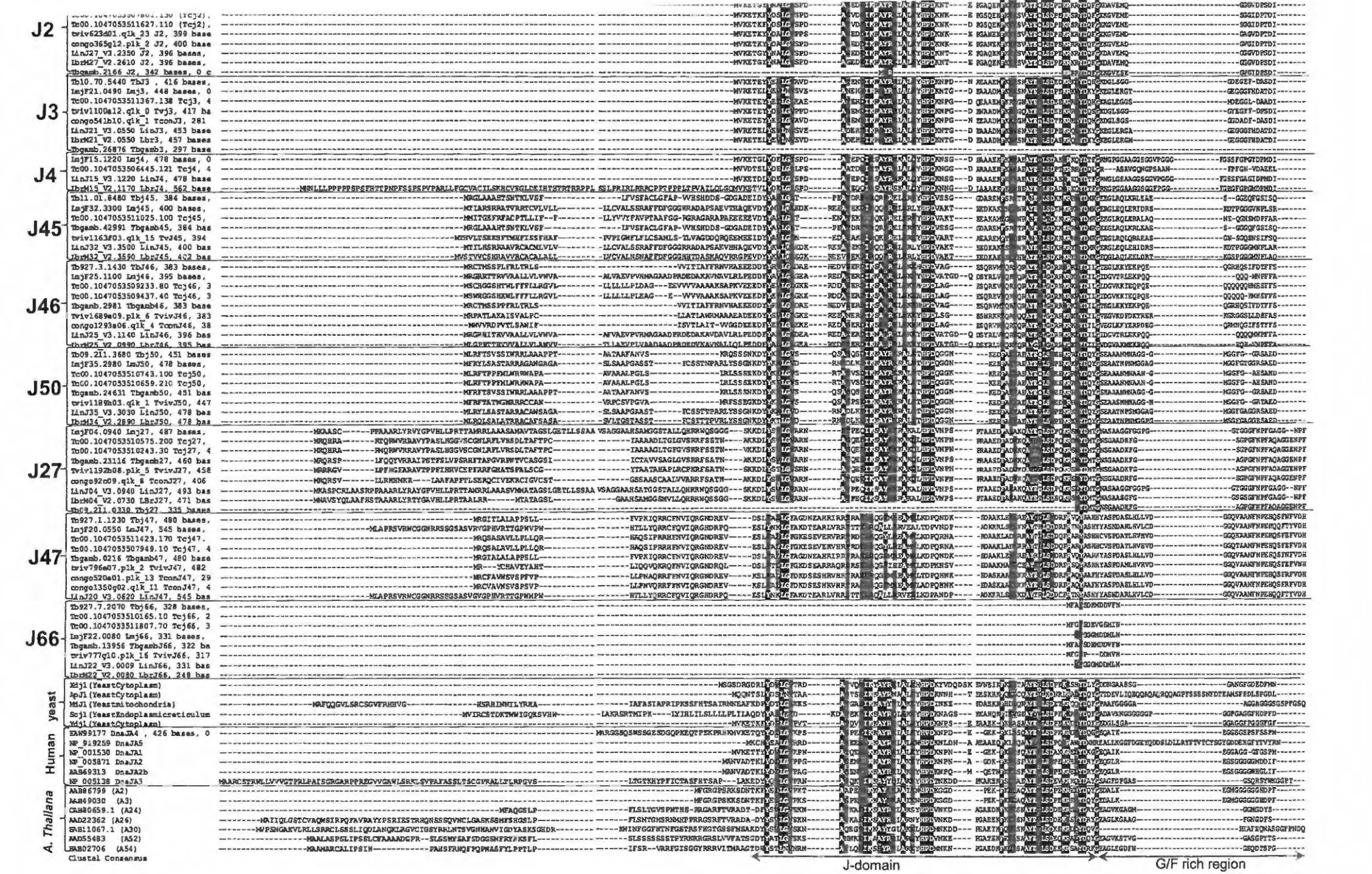
In the **group A** of **Figure 2.3,** Tbj50, Tbj46, and Tbj45 are shown to cluster on their own, while Tbj27 and Tbj47 appear to be more related to the sequences in **group A** from other species. The cladogram shows Pfj1 (from *P. falciparum*) and A30 (from *A. thaliana*) to be most related to the two *T. brucei brucei* Type IV/I Hsp40 sequences (Tbj27 and Tbj47). It is interesting that Mdj1 and DnaJA3, the only *S. cerevisiae* and human sequences in group A respectively, showed greater similarity to the *A. thaliana* genes predicted to be localized to the chloroplast (A54, A52, A24, A26) than to other sequences predicted to be localized to the mitochondria. All of these sequences in **Group A Figure 2.3**, even Pfj1 and A30 had a very low percentage identity to Tbj27 and Tbj47 (~20% or

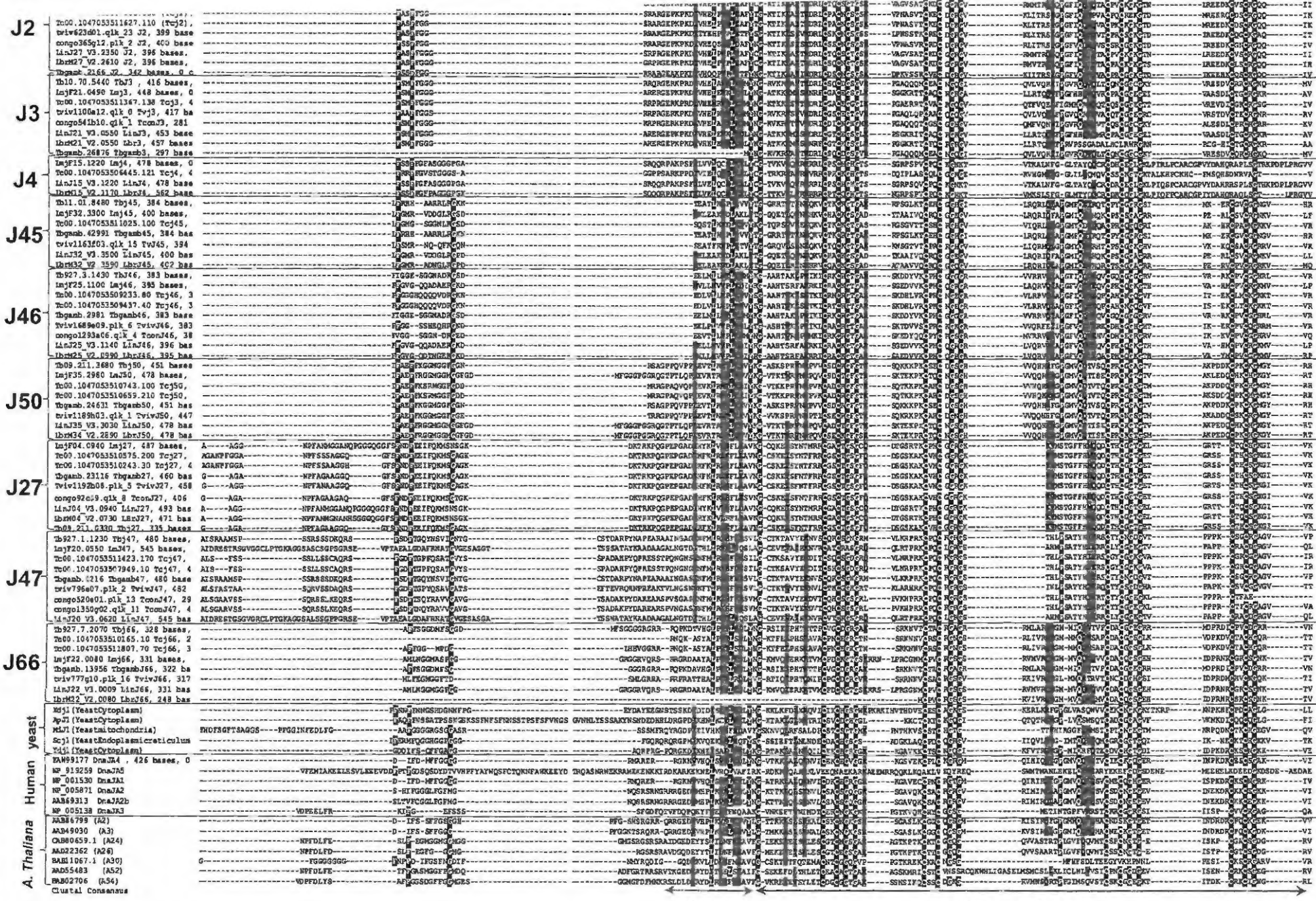
less). This weakens the argument for their status as orthologues, but the fact that they have a similar domain structure and are predicted to be localized to the same locations adds strength to this argument. It should therefore be noted that the orthologue identification for TbJ27 and TbJ47 in the non-trypanosomatid species shown in **Table 2.7** is very speculative.

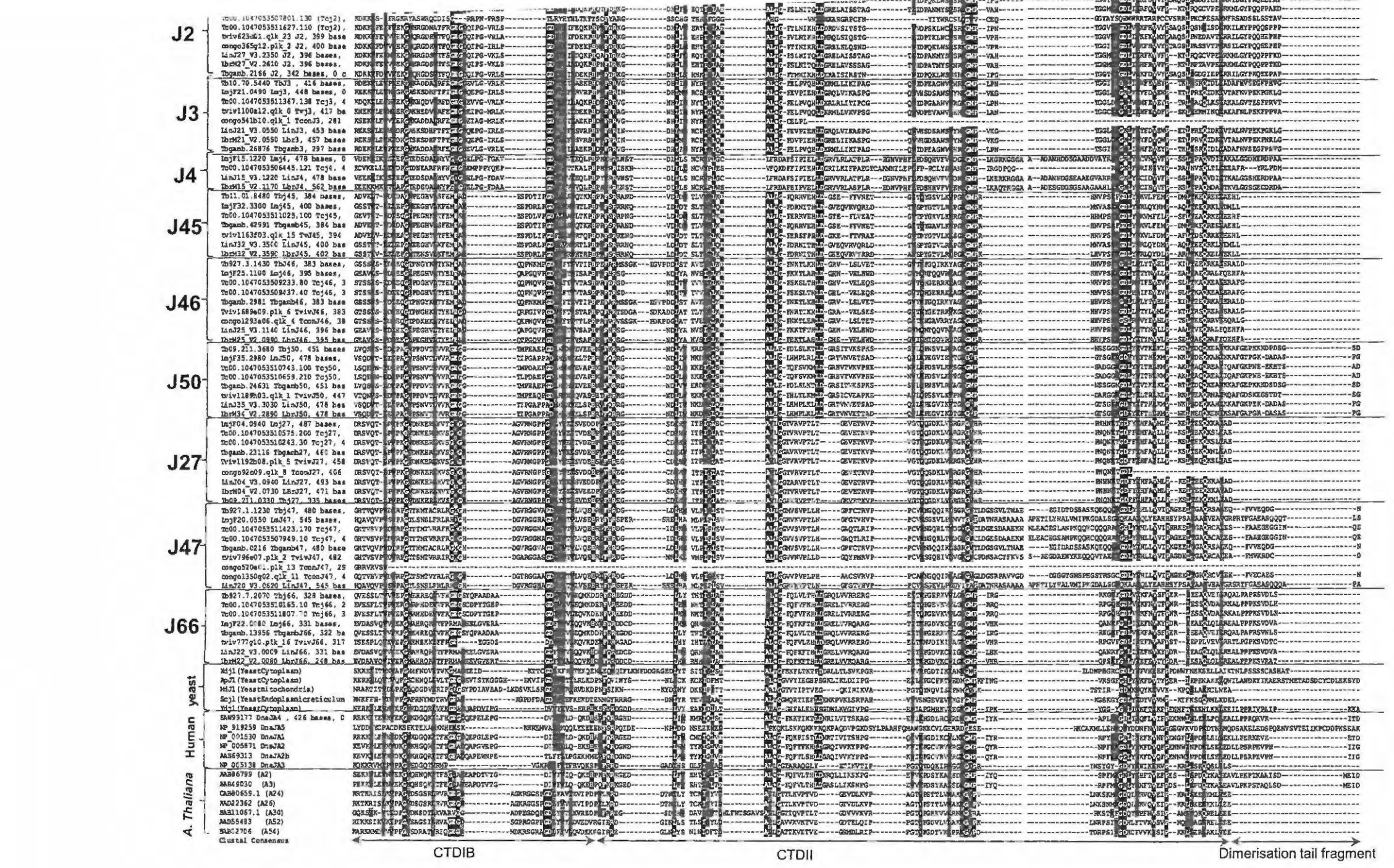
2.3.8) Comparison of features of the different Type I and Type IV classifications through multiple sequence alignment

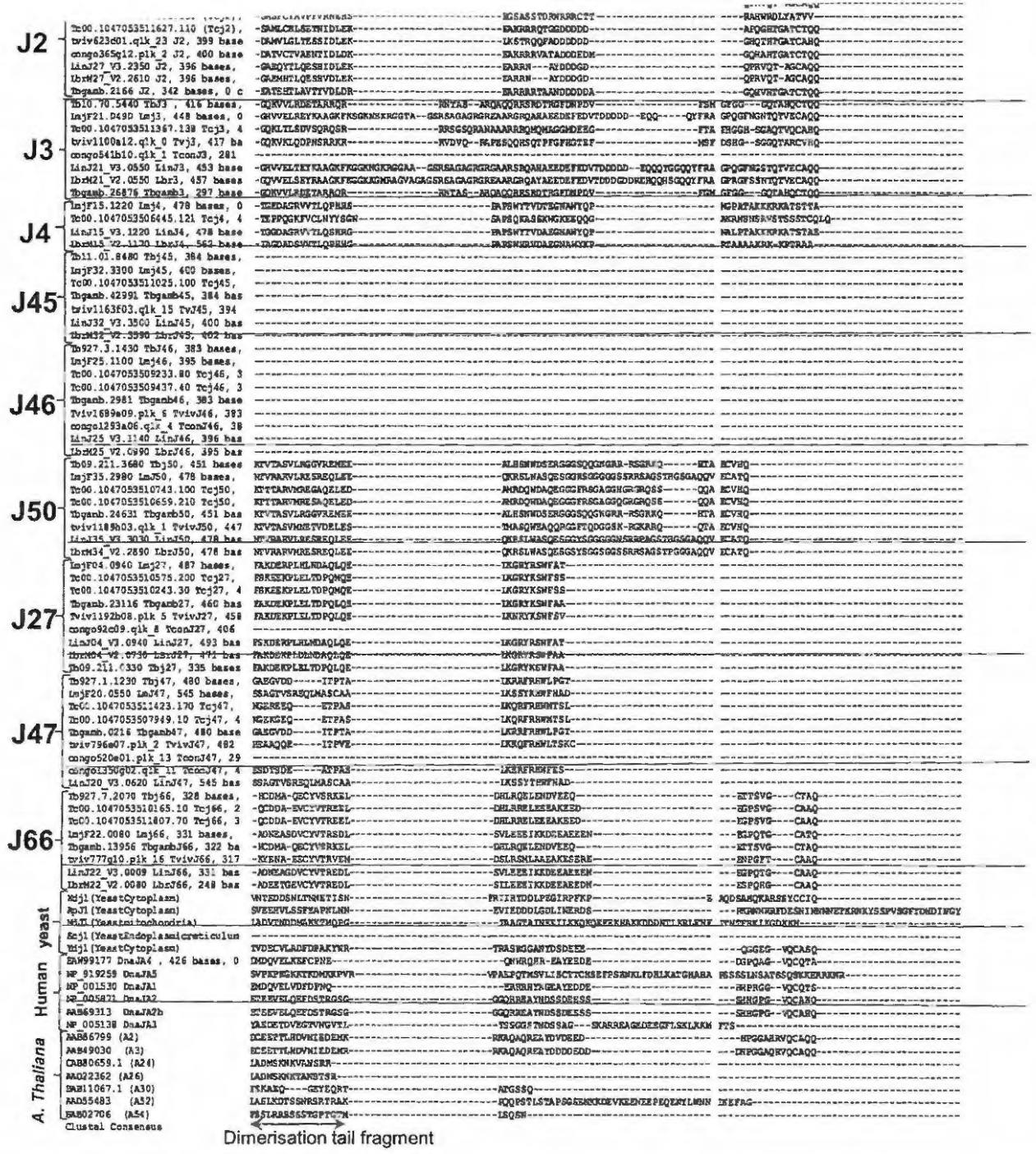
A comparison of the Type I and Type I/IV sequences of all the trypanosomatids was prepared using a multiple sequence alignment using the Clustal W algorithm. The Type I and Type I/IV sequences from *H. sapiens*, *S. cerevisiae*, *A .thaliana* and *P. falciparum* were also included **(Figure 2.4 – A PDF copy of the alignment figures is included on a CD attached at the back of the thesis).** This alignment confirmed the finding of **Figure 2.2** that the trypanosomatid sequences within a particular J group (e.g. J2 consists of Tbj2, Tcj2, Lmj2, TvJ2, TconJ2, LinJ2, LbrJ2 and TbgambJ2) have a high level of similarity.

Figure 2.4: (Page 87-90) A multiple sequence alignment of all of the trypanosomatid Type I and Type IV/I Hsp40 sequences compared to the Type I Hsp40s of *S. cerevisiae* (Yeast), *H. sapiens* (Human) and *A. thaliana*. **A)** displays the alignment from position 0 to 240. **B)** displays the alignment from positions 241 to 480. **C)** alignment from position 481 to 720. **D)** alignment from position 741 to 841. The alignment used a Clustal W algorithm and was produced using Bioedit (Hall, 1999). The residues highlighted with a dark background indicate sequence identity with other sequences at the position. The lighter (grey) background indicates amino acid similarity at that position between sequences highlighted in this way. An electronic version of this figure is attached at the rear of this thesis to provide a better resolution.


J2
J3
J4
J45
J46
J50
J27
J47
J66
yeast
Human
A. Thallana
Clustal Consensus
J-domain
G/F rich region


J2
J3
J4
J45
J46
J50
J27
J47
J66
A. Thaliana
Human
yeast
CTDIA
Zinc Finger domain


J2
J3
J4
J45
J46
J50
J27
J47
J66
Human yeast
A. Thaliana
CTDIB
CTDII
Dimerisation tail fragment


J2
J3
J4
J45
J46
J50
J27
J47
J66
A. Thaliana
Human
yeast
Clustal Consensus
Dimerisation tail fragment

The J2 and J3 groups show very similar sequences, except that the G/F region of the J3 sequences is longer by 3 to 4 amino acids (**Figure 2.4A**) and that they have a longer C-terminal region subsequent to the dimerisation domain (**Figure 2.4D**). Notable exceptions in these two groups are the TbgambJ2 and TbgambJ3 that are missing part (TbgambJ2) or all (TbgambJ3) of the J-domain. TconJ3 is missing the sequence from halfway through its C-terminal Domain II to the C-terminus of its J3 counterparts (**Figure 2.4C**). The J4 sequences are the rarest, showing only 4 examples out of the 8 trypansomatid species investigated. Their sequences are very similar to J2 and J3, with the exception that they have an even longer G/F region than the J3's (**Figure 2.4A**) and a disruption of the fourth CxxCxGxG repeat of the Zinc finger-like region (**Figure 2.4B**). Their extreme C-terminal region, subsequent to the tail end of the dimerisation domain (shown to overlap the CTDII of the partner monomer **Chapter 1 Figure 1.8**) is the same length as the J2 proteins (**Figure 2.4C/D**). The J2, J3 and J4 groups all typically have a very short sequence (~5aa) before the start of the J-domain (**Figure 2.4A**). The exception to this general observation being LbrJ4, which has a long N-terminal sequence preceding the J-domain. J45, J46, J47, J50 and J27 all have a length of sequence up to 85 amino acids long before the start of the J-domain (**Figure 2.4A**). Although these sequences can differ in length substantially (~30aa) they are likely to contain the N-terminal mitochondrial localization sequences responsible for the predicted localization of these sequences to the mitochondria. (**See Section 2.3.4**). All of the trypanosomatid J-groups have a J-domain with the exception of J66 that was designated as an Hsp40 by Folgueira and Requena (2007). Sequences in this group are missing a J-domain, G/F region and part of their C terminal Domain IA (CTDIA). Their status as Hsp40s is therefore questionable. TbgambJ2, TbgambJ3 and Tbj27 also lack a J-domain. In general, J2, J3, J4, J45, J46, J50 and J27 sequences all contain an HPD motif. This is absent in the J47 sequences (**Figure 2.4A**) as these are designated as Type I/IV Hsp40s and have a (K/T/A)DP(Q/V/K/A) motif instead of the HPD motif. The J-domain shows the greatest level of sequence conservation in the alignment (**Figure 2.4A**).

All of the J2 and J50 sequences and almost all of the J3 and J4 sequences contain a DIF motif (Found in the region of position 231/232 in the alignment) in their G/F rich region.

This is absent in the other five J protein groupings (J45, J46, J27, J47 and J66). This motif is not a consistent feature in all of the Type I Hsp40s of *H. sapiens*, *S. cerevisiae* and *A. thaliana* either. One *A. thaliana* sequence (A30) contains two DIF motifs in the G/F rich region.

The C-terminal domain designations were taken according to the known domain boundaries for Ydj1 (Li *et al.*, 2003) (Last of the Yeast sequences in **Figure 2.4**) and thus it was assumed that the domains are the same for each J-protein group. Assuming that these domain designations are correct, J27 and J47 both show a C terminal domain I section A (CTDIA) that is about two and a half times as long as those in the other J protein groups and J50 shows a slightly longer domain than J2, J3, J4 J45 and J46. However, J47 and J27 may just have extended G/F rich region.

The zinc finger-like region is of similar size in all the trypanosomatid proteins aligned (**Figure 2.4B**). The first three CxxCxGxG repeats are conserved in all the tryapanosmatid sequences with the exception of LbrJ66 which is missing the first cysteine/glycine repeat and LbrJ47, where the first glycine in the first repeat is replaced with an alanine. Three of the four J4 group of sequences (LmJ4, LinJ4 and LbrJ4) have the last glycine of the second cysteine/glycine repeat replaced with a lysine. The only trypanosomatid Type I/Type IV/I sequence without an intact third cysteine/glycine repeat is LbrJ3 in which only the last glycine of this repeat is evident. The cysteine/glycine repeat shows a much larger amount of variation between the different J protein groupings (J2, J3 etc). All the sequences of J2, J46 and J66 and most of the sequences in J3 and J45 did not have the final glycine of the fourth cysteine/glycine repeat. The last glycine of this repeat is present in all the sequences of the J50, J27 and J47 groups. The J4 sequences lack most of this repeat, with only the first glycine synchronizing with the fourth repeats of the other groups. The CxxC is visible, but is ~12 amino acids upstream of where it should be in relation to the first glycine. The J4 sequences also show an extended sequence between the third and fourth cysteine/glycine repeats.

The second portion of the C-terminal Domain I (CTDIB), found after the Zn finger-like region, is of a similar length in all of the sequences of the alignment (**Figure 2.4C**). Conserved amino acids in the CTDIB are G (position 493) (numbered according to the above the alignment in **Figure 2.4**) and the G(D/N) (position 527/528), which are found in almost all sequences, even in the non trypanosmatid sequences in the alignment.

The structure of the Ydj1 fragment (amino acids 110 to 337) complexed with a peptide substrate (1NLT.pdb; Li *et al.*, 2003) revealed a hydrophobic pocket in CTD1 that was thought to be involved in peptide substrate binding (Li *et al.*, 2003). These hydrophobic amino acids (I116, L135 , L137, L216, V247 and F249 – according to the YDJ1 amino acid numbering) were investigated for corresponding hydrophobic residues in the other sequences in the alignment (**Figure 2.4**). Results revealed that although the hydrophobic residues are not the same for each sequence, there is a high conservation of hydrophobic residues at these positions (**Figure 2.4**). Position 347 of the alignment, corresponding to Ydj1's Ile116, showed a conservation of a hydrophobic amino acids for J2, J3, J4, J50, J27 and J47. J46 and J45 only showed hydrophobic amino acids at this position for the LmJ, LbrJ and the LinJ sequences. J66 showed some sequences with the weakly hydrophobic alanine, or a non-hydrophobic residue. Position 367 corresponds to L135 of Ydj1 displayed hydrophobic amino acids for all trypanosomatid sequences wit the exception of LmJ46, LinJ46 and LbrJ46 that had serine at this position. L137 of Ydj1 (position 369) again shows a hydrophobic amino acid for almost all sequences at this position, with the exception of most sequences in J66 that had a histidine. L216 of Ydj1, corresponding to position 486 in the alignment, shows less hydrophobic conservation among the trypanosomatid sequences. The J47, J50 and J27 families showed a complete absence of hydrophobic amino acids at this position, while only some sequences in J45 contained hydrophobic amino acids at this position. This may just be a manifestation of the divergence of these sequences from Ydj1 as there are hydrophobic amino acids just to the left of this position that could possibly form part of the hydrophobic pocket in these sequences. Position 530 (F249 in Ydj1) shows hydrophobic amino acid conservation in all the trypanosomatid sequences except for the TconJ47 pseudogene (congo520e01.plk_13), which is a partial copy of the full length version

(congo1350g02.q1k_11). All of the *S. cerevisae* sequences show hydrophobic residues at all of these positions, with the exception of Scj1, which is missing a hydrophobic amino acid at position 486 in the alignment. The *H. sapiens* and *A. thaliana* sequences also mostly have hydrophobic amino acids at these positions.

Wu and coworkers (2005) describe Ydj1 dimerisation as occurring by two mechanisms. The first involves interaction of a number of hydrophobic amino acids of the C – terminal domain II (L274 [position 566 on the alignment], I278 [position 570], L346 [position 684], L349 [position 687], I 352 [position 690], L353 [position 691], F335 [position 673], P336 [position 674] and F340 [position 640]) between the two monomers. The second is suggested to be unique to Type I Hsp40s and consists of a loop with an extra two β-sheets that rests over the CTDII of the other monomer. **Figure 2.4 C** shows that the first portion of the CTDII domain (region 539 to 687 of the alignment Figure 2.4), has an almost uniform size across the various sequences, with the exception of the J4 and J47 groups which have an extended sequence from position 631 to 654 of the alignment. The extra dimerisation loop and sheet region described by Wu *et al.* (2005) (Region 350 to 390 using the Ydj1 sequence as a reference) **(See section 1.7.1.5; Figure 1.8A)** shows a significant variation in length. This whole section appears to be absent in the J45 and J46 sequences, indicating that not all trypanosomatid Type I Hsp40s use the same mechanism to dimerise as that described for Ydj1 (Wu *et al.*, 2005).

Of the hydrophobic amino acids involved in the first mechanism of dimerisation of Ydj1, hydrophobic amino acids are consistently found at the positions corresponding to L274, F335, P336 and L349. Positions corresponding to I278 of Ydj1 show conservation of a hydrophobic residue at this position in most trypanosomatid sequences with the exception of J27 and J47. This may be due to the insertion of gaps in the alignment for these two sequence families. If these gaps were removed they would mostly have a hydrophobic amino acid at this position. Hydrophobic amino acids are seen at positions corresponding to I352 and L353 for all trypanosomatid Type I-like sequence families except J27 and J47. Positions corresponding to F340 and L346 show less conservation of the hydrophobic character of these amino acids, but this may be due to a gap insertion in

most trypanosomatid sequences just prior to this position. The sequence (containing the L274 and I278 hydrophobic amino acids) at position 566 to 572 is well conserved, with the exception of the J27, J47 and J4 sequence families. This corresponds to the First α-helix in the CTDII region of YdJ1. The second α-helix of the Ydj1 CTDII region (alignment position 677 to 687) shows less sequence conservation across the sequences. Other areas of high sequence conservation in CTDII are 538-HxxFxR-543, L558, 591-LD-592 and 627-GMP-629. J27 and J47 show a much lower consensus with the other J group sequences in the latter four regions of conserved sequence.

The C-terminal region (position 688 to 823) of J2, J3, J4 and J66 have a length most similar to Ydj1, most of the *H. sapiens* sequences and A2 and A3 from *A. thaliana*. Although this C-terminal section of the J50 family is of a similar length to Ydj1, these sequences do not align well with J2, J3, J4 and J66 in this area. The J27 and J47 sequence families are significantly truncated in this region relative to Ydj1.

2.3.9) Detailed analysis of Tcj2,Tcj3, Tbj2, Tbj3 and their homologues in other species

As the remainder of this thesis deals with the proteins Tcj2, Tcj3, Tbj2 and Tbj3, a detailed analysis of their characteristics was undertaken. Tbj2 and Tcj2 were shown to be orthologues in **section 2.3.2**, as were Tbj3 and Tcj3. The pairwise alignment of Tcj2/Tbj2 (74%) and Tcj3/Tbj3 (65%) are therefore not surprisingly very high. However, the pairwise alignment of Tcj2 with Tcj3 and Tbj2 with Tbj3 is also high. This finding was also recorded for Tcj2 and Tcj3 by Tibbetts and coworkers (1998). All four of these trypanosomal proteins have a similar length, except for a slight lengthening of the extreme C-terminus of the two J3 proteins (**Figure 2.5 Section 390 to 435 of the alignment**). In addition, these proteins are all predicted to be localized to the cytoplasm (**Section 2.3.4; Table 2.3**) and are all predicted to be farnesylated (**Section 2.3.5; Table 2.3**) and have a very similar domain structure.

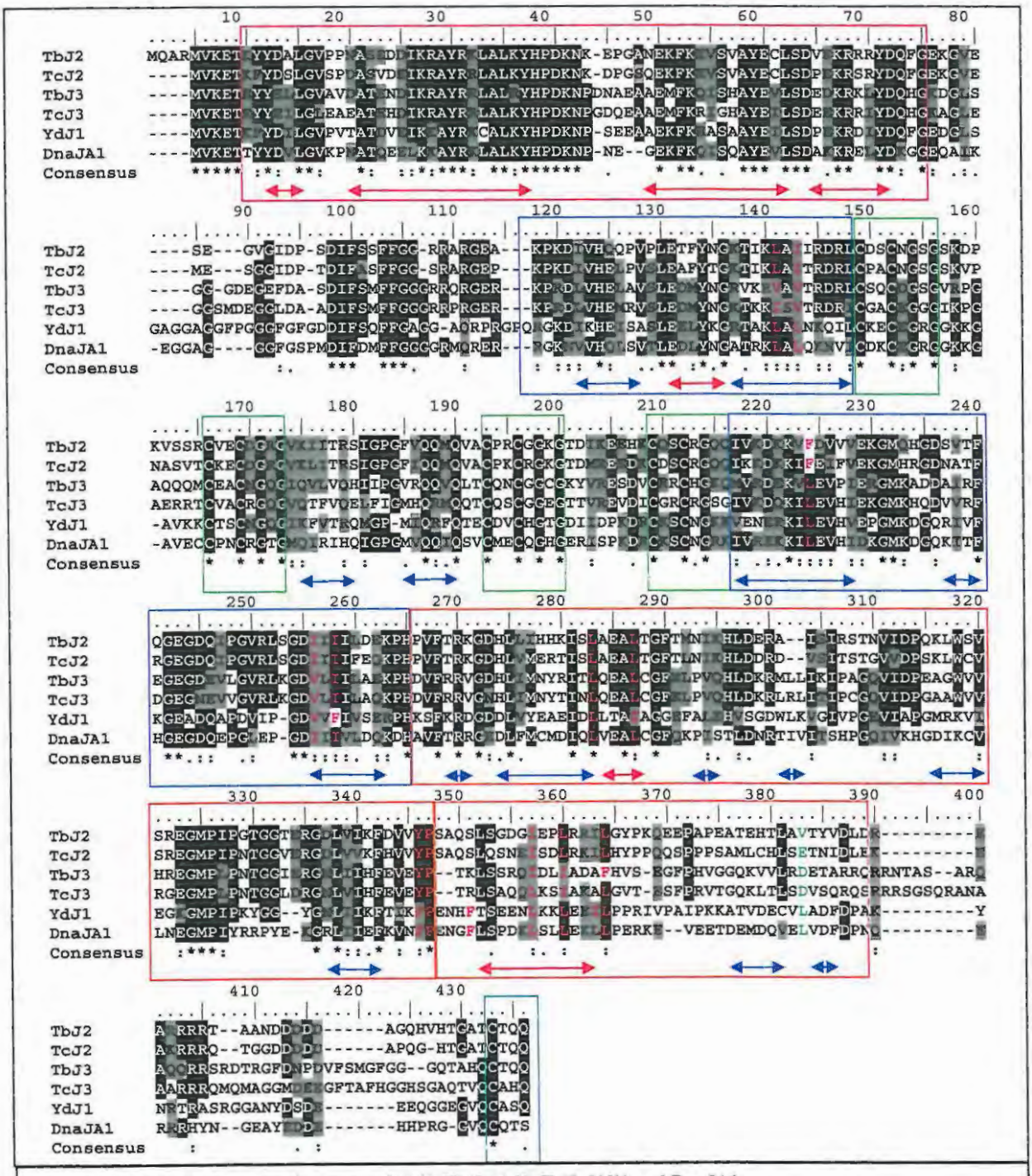


Figure 2.5: Multiple sequence alignment of Tbj2, Tbj3, Tcj2, Tcj3, Ydj1 and DnaJA1.
The red box indicates the J-domain. The green boxes indicate the Cys/Gly repeats of the Zn finger-like region. The C terminal domain I (CTDI) is shown with a dark blue box. CTDIA is before the An finger-like region, while CTDIB is after this domain. The C-terminal Domain II (CTDII) is indicated by an orange box. The extra dimerisation loop and β-sheet associated with some Type I Hsp40s is shown as a separate orange box after the more conserved portion of the CTDII. The C-terminal CaaX box of each sequence is highlighted with a light blue box. The α-helices are shown as red arrows, while the β-sheet regions are indicated with blue arrows. The CTDI, CTDII, α-helices and β-sheet regions were designated according to the Ydj1 structural data (Li *et al.* 2003; Wu *et al.*, 2005). The amino acids at positions highlighted in purple have been described as part of the hydrophobic peptide binding pocket in YdJ1 (Li *et al.*, 2003; Li and Sha, 2005). The hydrophobic amino acids at positions highlighted in red are indicated as important for dimerisation of Ydj1. The amino acid position highlighted in green is part of the extra dimerisation loop and β-sheet found to wrap around the CTDII of the other monomer. This amino acid interacts with a hydrophobic pocket present on the CTDII of the other monomer (Wu *et al.*, 2005).

The multiple sequence alignment in **Figure 2.5** reveals the significant sequence similarity of the J2 and J3 proteins of *T. cruzi* and *T. brucei brucei* to Ydj1 and DnaJA1. All 6 proteins were revealed to have the domain structure very typical of Type I Hsp40s. They all have a highly conserved sequence in their N-terminal J-domains, with the exception of portions of the loop regions. The cysteine/glycine repeat region is a well known feature of the zinc finger-like domain of Type I Hsp40s (Cheetham and Caplan, 1998). The C-terminal domain I (CTD I) and portions of the C-terminal domain II (CTD II) are also well conserved in the sequences shown in **Figure 2.5**. This can lead to possible inferences about the C-terminal domain three-dimensional structure of the J2, J3 and DnaJA1 proteins from the known structure of the C-terminal domains of Ydj1 (Li *et al.*, 2003). This could possibly help with the determination of surface exposed regions of Tbj2, Tbj3 and Tcj3 for the design of an antibody that would detect the protein in its native conformation.

The likely structural similarity of these proteins and the conservation of hydrophobic amino acid residues at positions 122, 141, 143 in CTD IA and 224, and 258 in CTD IB (**Figure 2.5**) indicate that these proteins all have a hydrophobic patch similar to Ydj1 based on the structural information of the C-terminal domains of Ydj1 (Li *et al.*, 2003). The three hydrophobic residues in CTD IA show similar properties in terms of spatial volume, with leucine, valine or isoluecine at these positions. The hydrophobic residues at the aforementioned positions in CTD IB, however, are more divergent in size, containing a leucine, isoleucine or phenylalanine. Comparison of the predicted secondary structure of Tbj2, Tbj3, Tcj2 and Tcj3 with that of YdjI using the PSIPRED (McGuffin *et al.*, 2000; http://bioinf.cs.ucl.ac.uk/psipred/psiform.html) algorithm revealed a very similar secondary structure shared by all 5 proteins.

2.3.10) Generation of Homology models of Tcj2,Tcj3, Ydj1, *A. tumefaciens* DnaJ, Tbj2

The homology models that were generated for Tcj2, Tbj2, Tcj3, Tbj3, Ydj1 and *A. tumefaciens* DnaJ (Agt DnaJ) are depicted in **Figure 2.6**. The potential template structures that were identified by Fugue and Phyre (Bennet-Lovsey *et al.*, 2007; Kelley *et al.*, 2000; Shi *et al.*, 2001) that were used to generate the models were only selected if they were from Type I or Type II Hsp40s (2CTP.pdb and 2O37.pdb are from Type II Hsp40s while 1XBL.pdb, 2OCH.pdb and 2DN9.pdb are all J-domain structures derived from Type I Hsp40s) that have more conserved J-domains between (Hennessy *et al.*, 2000) and if their amino acid sequences had a high similarity to the target sequence. The Ramachandran plots generated by PROCHECK (Laskowski *et al.*, 1993; Morris *et al.*, 1992; http://www.biochem.ucl.ac.uk/~roman/procheck/procheck.html) all described >96% of the amino acids being in the allowable or core regions of the plot for each model. The other plots generated for PROCHECK also indicated a high level of accuracy for the models (data not shown). In addition, the models superimposed well onto the structures of the templates that were used, with only a small amount of divergence in position of the loop region between helix II and helix III of the template structures and the model. This is not surprising as the loop has been described as being a flexible and structurally dynamic section of the J-domain (Greene *et al.*, 1998; Hennessy *et al.*, 2000).

In **Figure 2.6** It is clear that all the structures of the modeled J-domains, especially of the eukaryotic J-domains, are very similar. This is likely a consequence of the same template structures being used for each model, with the exception of the Agt DnaJ model For which only the *E. coli* DnaJ J-domain structure (1XBL.pdb; Pellechia *et al.*, 1996) was used as a template. The Tbj2 and Tcj2 structures appear the most similar, which is not entirely unexpected as they were found to be orthologue proteins from *T. brucei brucei* and *T. cruzi* **(Section 2.3.2)**. It has also been previously reported that the J-domains of Type I Hsp40 have the highest level of conservation relative to members of the Type II or Type III Hsp40 category (Hennessy *et al.*, 2000). The four helices and the loops between them are also of a very similar length to their counterparts in the 6 models shown in **Figure 2.6**, each region differing by a maximum of 1 to 3 amino acids. The most

divergent structure is that of Agt DnaJ which is of prokaryotic origin These models therefore support the conclusions of Hennessy *et al.* (2000), that Type I Hsp40 J-domains have a high level of conservation. It is therefore likely that these models are a close approximation to the actual structure of the J-domains in these proteins.

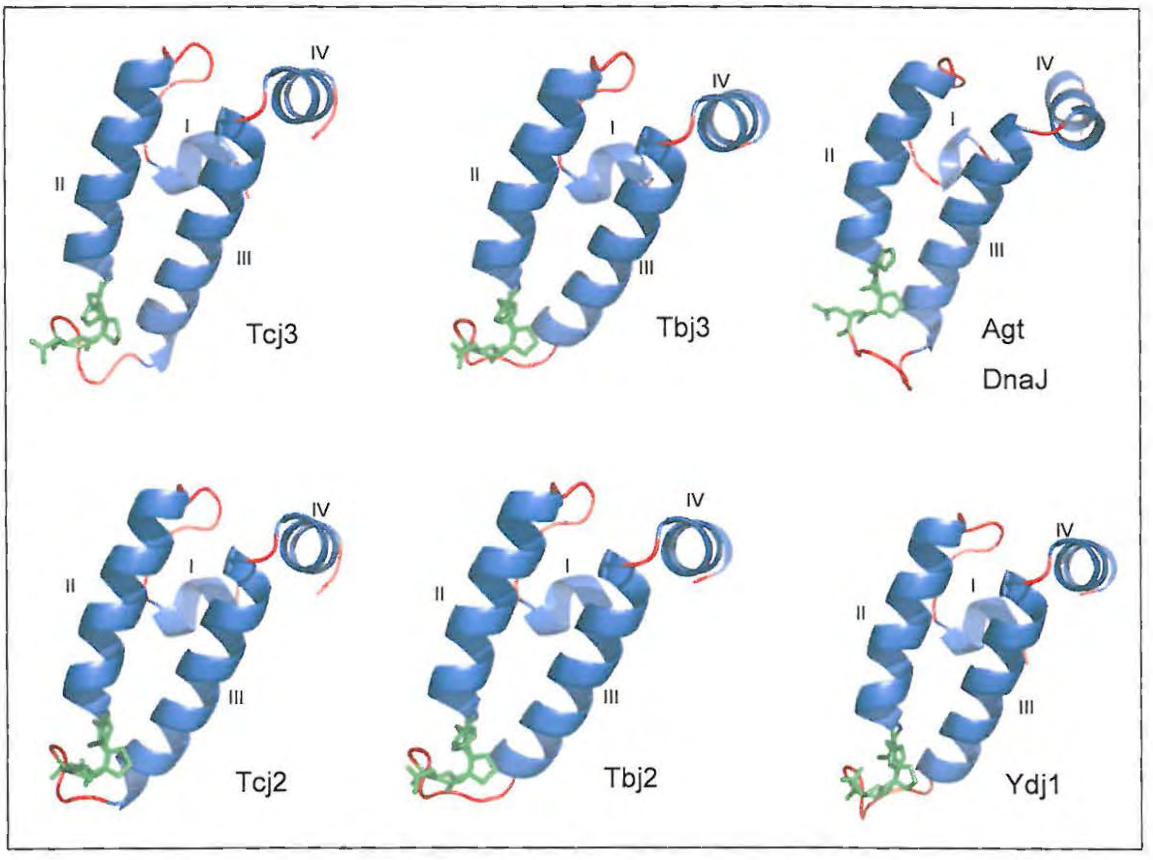


Figure 2.6: Homology models generated for the J-domains of *T. cruzi* Type I Hsp40s Tcj3 and Tcj2, *T. brucei brucei* Type I Hsp40s Tbj3 and Tbj2, the *S. cerevisiae* Type I Hsp40 Ydj1 and *A. tumefaciens DnaJ* (Agt DnaJ). The helices of each model are shown in blue and labeled I to IV. The loop regions are shown in red. The HPD motif in the loop between Helix II and III is highlighted in green. Multiple templates (1xbl.pdb, 2DN9.pdb, 2OCH.pdb, 2O37.pdb and 2CTP.pdb) were used to generate the Tcj2, Tcj3, Tbj2, Tbj3 and Ydj1 homology models using Modeller (Śali and Blundell, 1993). The crystal structure for *E. coli* DnaJ (1xbl.pdb, Pellechia *et al.*, 1996) was used as a single template for the production of a homology model of *A. tumefaciens* DnaJ. All structures were visualized using Pymol (Delano, 2002).

2.3.11) Design of peptide polyclonal antibodies for Tbj2, Tbj3 and Tcj3

The Tbj2, Tbj3 and Tcj3 primary amino acid sequences were examined for antigenic regions for the synthesis of a peptide to make polyclonal antibodies specific to each individual protein. The design of Tcj2 was not included in this section as an existing Anti-Tcj2 peptide polyclonal antibody was available from previous work (Edkins *et al.*,

2004).To find sections of a protein that have potential antigenic and immunogenic potential, a number of generalized antigenic characteristics can be exploited. Regions that are surface exposed, have a high level of peptide chain flexibility, contain a high number of hydrophilic amino acids and a high charge density are believed to make good antigenic regions for the manufacture of antibodies (Jameson and Wolf, 1988; Karplus and Schultz, 1988; Hopp and Woods, 1981; Emini *et al.*, 1985). The surface exposure of the antigenic sequence is particularly important if the polyclonal antibody to be manufactured from this peptide sequence is to be used to detect non-denatured proteins.

As hydrophilic amino acid side chains prefer an aqueous environment, regional concentrations of these amino acids have a greater probability of being surface exposed. The methods by Engelman *et al.* (1986), Hopp and Woods (1981), Kyte and Doolittle (1984) predict the regions of the primary amino acid sequence that are hydrophilic or hydrophobic (**Figure 2.7, Figure 2.8 and Figure 2.9**). However, in each, the method used to assign hydrophobic versus hydrophilic status is different. These methods all use a sliding window calculation in which the average measurements of a window size of N amino acids in the sequence is plotted for a particular position on the graphical output. The window is then shifted by one amino acid to the left to obtain the next plotted value. Hopp-woods uses vapour free energy of transfer between water and ethanol of the amino acids in its prediction of hydropathic and hydrophilic sections of the protein (Hopp and Woods, 1981), while the Kyte-Doolittle method uses water vapour free energies of transfer for the amino acids in plotting the hydropathy of the primary sequence. Kyte-Doolittle also uses the preference of amino acids for the inside or outside environments of proteins in its calculation (Kyte and Doolittle, 1982). The method by Engelman and coworkers (1986) plots the sum of the hydrophobic and hydrophilic characteristics of each amino acid. The hydrophilic elements are calculated from the free energy for the insertion of charged groups into a membrane lipid bilayer, while the hydrophobic elements are determined by the free energy of transfer from water to oil of the amino acid side chains (Engelman *et al.*, 1986). The hydrophobic moment calculation by Eisenberg and coworkers (1984) uses the asymmetrical arrangement of polar and apolar residues in the solvent exposed relative to the buried surfaces of α-helices and β-sheets of proteins.

The method calculates the hydrophobic moment assuming the primary amino acid sequence is all α-helical or all β-sheet on two separate graphs. Predicted α-helical regions correspond to regions with a high hydrophobic moment on the plot assuming α-helical structure of the protein, while predicted β-sheet has a high hydrophobic moment on the plot assuming a β-sheet structure (Eisenberg *et al.*, 1984). These hydropathy and hydrophobic moment plots are useful in the prediction of transmembrane regions, of which none were found in Tbj2, Tbj3 or Tcj3. The Emini Surface Probability plot in the Generunner protein composite analysis is also able to predict the probability of a regions surface exposure from the primary amino acid sequence (Emini *et al.*, 1985). The method is based on the calculation by Janin and coworkers (1978) who calculated the solvent accessible area of each amino acid side chain in a sample of 28 proteins with a known three dimensional structure. Charge density is also a factor in predicting the surface exposure of a protein from its amino acid sequence. The Generunner composite analysis plots the charge density of a protein sequence based on the pKas of each amino acid at a pH of 7.0.

Karplus and Schultz provided a method for predicting peptide chain flexibility, which is a known characteristic of some highly antigenic regions of a protein, from the amino acid sequence. The plotted flexibility values are produced from a calculation based on the assignment of amino acids as flexible or rigid using a sample of 31 proteins that have a known three dimensional structure (Karplus and Schultz, 1988). The Karplus-Schultz and Emini Surface Probability methods both use a version of the sliding window method described earlier. The Jameson-Wolf Antigenic Index provides an overall measure of the antigenic potential of areas along the amino acid sequence of a protein. It does this by combining the Chou-Fasman (Chou and Fasman, 1978) and Garnier-Robson (Garnier *et al.*, 1978) secondary structure prediction methods with results from the Hopp-Woods hydropathy, Emini Surface Probability and Karplus-Schultz chain flexibility to determine an index of antigenic potential (Jameson and Wolf, 1988). Peaks on the plot are predicted to have a greater probability of being strongly antigenic.

Although reports have been made that, in general, the N-terminal regions of proteins have better antigenicity (Jameson and Wolf, 1988), this region was avoided due to the presence of conserved features including the J-domain (Hennessy *et al.*, 2000), the repeats of the zinc finger-like region (Martinez-Yamout *et al.*, 2000) and the DIF motif (Wall *et al.*, 1995). This is particularly important for Tbj2 and Tbj3, due to their high sequence identity (~ 46%) and their presence in the same organism, which could lead to cross reactivity problems during *in vivo* or whole cell lysate experiments. Regions of the C-terminus were therefore chosen due to the sequence disparity in these regions. Likewise, a peptide near the C-terminus was chosen for Tcj3. A peptide of at least 14 amino acids was used to try to increase its antigenicity and the specificity of the antibody produced.

The composite analysis of the Tbj2 amino acid sequence (**Figure 2.7**) shows that the Goldman, Engelman and Steitz plot (**Figure 2.7A**) reveals a largely hydrophilic protein. As this method plots the sum of the hydrophilic and hydrophobic components of each amino acid in the sequence, the most hydrophilic amino acids correspond to the most negative values on the plot. These most hydrophilic regions correspond, with minor differences, to the hydrophilic regions of the sequence predicted by the Hopp-Woods and Kyte-Doolittle methods (**Figure 2.7F** and **Figure 2.7G** respectively). Owing to the more conserved nature of the N-terminal region of Type I Hsp40s, the region from amino acid 250 to amino acid 404 was selected as the preferred region to find an immunogenic peptide. Sections of the amino acid sequence that are strongly hydrophilic are categorized as having a greater probability of surface exposure and hence immunogenic potential in the native structure (Jameson and Wolf, 1988). In the C-terminal half of Tbj2, the most hydrophilic regions were predicted to be in the region from 350 aa to 404 aa. This same region is shown to have the highest probability of being surface exposed, according to the Emini surface probability plot (**Figure 2.7H**), with the exception of the J-domain (amino acids 1 to 75). The Karplus-Schultz plot (**Figure 2.7E**) predicts that there are regular regions of flexibility that is recognised as significant (above the dashed line) along the length of Tbj2, including the region containing the peptide that was chosen (**Figure 2.7**). A region of high flexibility with a net negative or positive charge (**Figure 2.7D**) and is surface exposed is considered a potentially good immunogenic region (Jameson and

Wolf, 1988). Although not as strong as other regions of the protein the selected peptide region is indicated by the Jameson-Wolf Antigenic index plot (**Figure 2.7I**) as having a good probability of being antigenic and immunogenic. Although the region from amino acids 370 to 390 display similar favourable characteristics for an antigenic/immunogenic peptide and a slightly higher hydrophilicity prediction, the peptide from amino acids 349 to 362 was chosen as indirect structural information is available for this region. A pairwise alignment of Tbj2 with Ydj1 showed that the Tbj2 peptide region –LGYPKQEEPAPEAT- corresponds to a region of the dimerisation domain of Ydj1 (**Figure 1.8A**). This is the loop region located before the extra β-sheet (designated as **(1)** in the figure that is unique to Type I Hsp40 dimerisation domains relative to Type II Hsp40s. This correlation provides evidence that this peptide sequence is indeed surface exposed. No such structural data was available for the potential peptide at the downstream position. As the peptide region is likely to be a loop region, the Eisenberg plots (**Figure 2.7B and Figure 2.7C**) were regarded as less important as they assume the protein consists of only α-helix or β-sheet respectively. Heavier weighting was given to the probability of surface exposure, flexibility, hydrophilicity, antigenic index and charge properties during the selection of the peptide regions.

BLAST of the selected peptide against the *T. brucei brucei* genome database in GeneDB (http://www.genedb.org/; Hertz-Fowler *et al.*, 2004) revealed no significant amino acid similarities other than to the Tbj2 sequence itself. In addition, the proteins designated as containing the closest sequences to the peptide were significantly smaller than Tbj2 (44.8 kDa) and were less than 32 kDa. These could easily be distinguished from Tbj2 on an SDS-PAGE gel.

The composite analysis of the Tbj3 amino acid sequence (**Figure 2.8**) revealed only one obvious region for the peptide to make the Tbj3 peptide based polyclonal antibody (amino acids 365-395). It is the region of the C-terminal half of the sequence to show the highest surface probability and hydrophilicity and has a strong net positive charge. This region also shows a high flexibility according to the Karplus-Schultz plot (**Figure 2.8E**). Pairwise alignment of Tbj3 with Ydj1 led to a putative correlation in which the forward portion of the peptide (369 to 394 aa in Tbj3) is likely to straddle the extra β-sheet found

in the dimerisation domain of Type I Hsp40s (**Figure 1.8A (1)**). The same region (amino acids 366 to 395) was chosen as the immunogenic peptide from the Tcj3 amino acid sequence (**Figure 2.8**) analysis. As with the Tbj3 sequence, this region of Tcj3 was the only region in the C-terminal half of the protein to have a significantly high surface probability prediction, good hydrophilicity properties, and a net positive charge over part of the region. The Jameson-Wolf Antigenic index also predicted this region to have good antigenic properties. This peptide region of Tcj3 is predicted by pairwise alignment to be likely to straddle the same β-sheet of the dimerisation domain as Tbj3. The similarity of the profiles generated in the composite analysis for Tcj3 and Tbj3 add further support to their categorization as orthologues. BLAST of theTbj3 peptide in the *T. brucei brucei* genome database (http://www.genedb.org/; Hertz-Fowler *et al.*, 2004) and BLAST of the Tcj3 peptide in the *T. cruzi* genome database (http://www.genedb.org/; Hertz-Fowler *et al.*, 2004). In addition, BLAST of the *E. coli* (K12) genome (Blattner *et al.*, 1997; Cummings *et al.*, 2002; http://blast.ncbi.nlm.nih.gov/Blast.cgi) and *S. cerevisiae* genome (Cherry *et al.*, 1998) revealed no significant hits. This indicates the lower probability of proteins in these organisms being recognised by the antibodies raised against these peptides. The composite analysis plots for Tbj2, Tcj3 and Tbj3 all showed that the only other region of these three amino acid sequences that had a comparably high predicted surface exposure and hydrophilicity was in the first 100 amino acids at the N-terminus. This region contains the J-domain of each of the Hsp40s and was predicted to have a strong antigenic potential. All three sequences are predicted to have regions of significant flexibility and net charge across their whole length.

All of the selected peptides had a cysteine added to the N-terminal side in order to facilitate the cross-linking to a carrier protein, such as limpet keyhole haemocyanin (KLH) prior to injection into rabbits, which were used to generate the antisera for all three peptides. The Tbj2 and Tcj3 peptides and antibodies were synthesized by the Genscript Corporation (http://www.genscript.com/; Piscataway, New Jersey, USA), while the Tbj3 peptide and antibody was synthesized by Professor R. Zimmermann's laboratory (Department of Medical Biochemistry, University of the Saarland Medical School, Homburg, Germany).

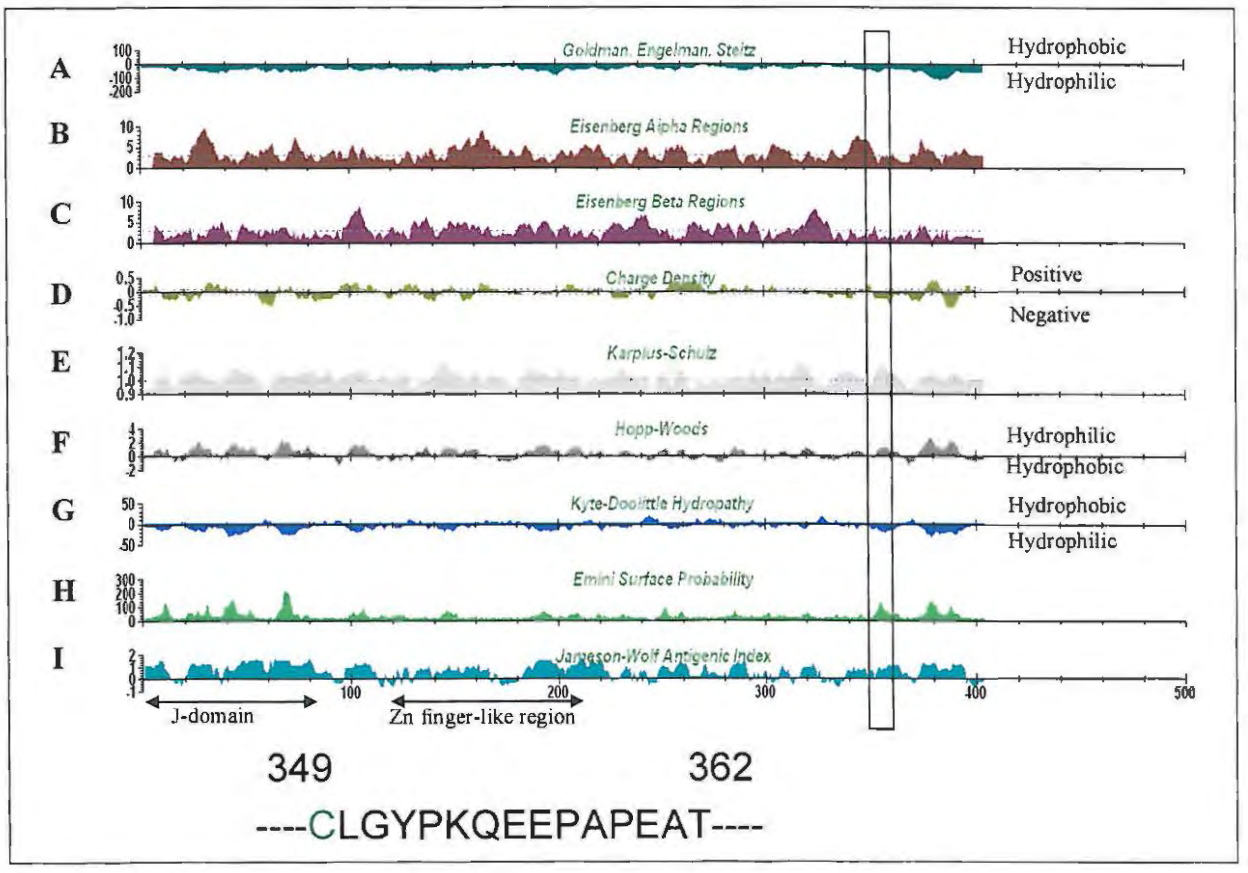


Figure 2.7: Composite analysis of Tbj2 primary amino acid sequence to determine suitable unique epitope regions for the design of a peptide polyclonal antibody specific to Tbj2.
The analysis was performed using the protein composite analysis function in Generunner (available at: http://www.generunner.net/). **(A)** The Goldman, Engelman and Steiz method plots hydrophobic regions as positive and hydrophilic regions as negative (Engelman *et al.*, 1986). **(B,C)** The Eisenberg method was applied to peptide bond angles assuming that the protein folded as an α-helix **(B)** or a β-sheet **(C)**. This method plots the hydrophobic moment which is a measure of the asymmetry in the arrangement of polar and non-polar residues (Eisenberg *et al.*, 1984). **(D)** Charge density plots the average charge for each amino acid based on the sliding window calculation for each amino acid. Overall positive charge is plotted as positive and overall negative charge is plotted as negative. **(E)** The Karplus-Schultz method measures chain flexibility and plots regions of increased flexibility as a peak (Karplus and Schultz, 1988). **(F)** Hopp-Woods plots the average local hydropathy (hydrophilic and hydrophobic regions), with positive plots indicating hydrophilic regions and negative plots indicating hydrophobic regions (Hopp and Woods, 1981). **(G)** The Kyte-Doolittle method also plots hydropathy (hydrophobicity is positive and hydrophilicity is negative) but using a slightly different calculation (Kyte and Doolittle, 1984). **(H)** The Emini surface probability plots the probability of a region of the primary amino acid sequence being on the surface of the folded protein is plotted with the peaks indicating the areas that have the greatest likelihood of surface exposure (Emini *et al.*, 1985; Janin *et al.*, 1978). **(I)** The Jameson-Wolf Antigenic index uses a combination of the data from the methods **(E)**,**(F)** and **(H)** along with the secondary structure prediction methods of Chou-Fasman (Chou and Fasman, 1978) and Robson Garnier (Garnier *et al.*, 1978) to plot an antigenicity index in which peaks show the greatest antigenic probability (Jameson and Wolf, 1988). The dashed lines on the plots in **(B)**, **(C)**, **(D)**, **(E)**, **(H)** and **(I)** indicate a threshold above (positive plot) and below (negative plot) which the values are considered significant. The peptide region that was selected is shown below the plots with the numbering indicating the position in the amino acid sequence. The cysteine coloured in green is not part of the original sequence, but was added to assist the cross-linking of the peptide to the carrier protein for antibody manufacture. (In the case of these antibodies, Keyhole Limpet Haemocyanin [KLH] was used as the carrier protein (Genscript, http://www.genscript.com/antibody.html). The region on the plots **(A-I)** highlighted with the black rectangle is the region at which the selected peptide is located in the sequence.

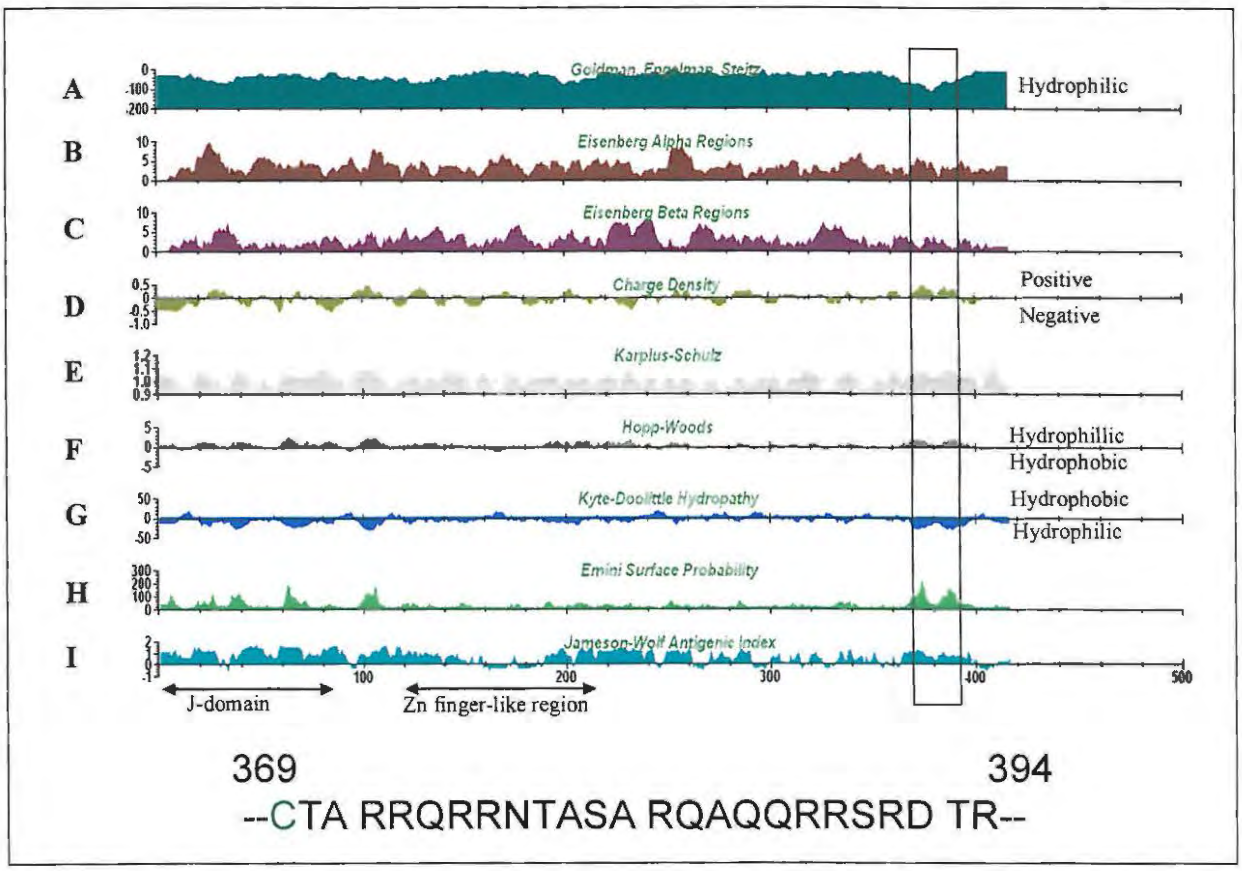


Figure 2.8: Composite analysis of Tbj3 primary amino acid sequence to determine suitable unique epitope regions for the design of a peptide polyclonal antibody specific to Tbj3.
The analysis was performed using the protein composite analysis function in Generunner (available at: http://www.generunner.net/). **(A)** The Goldman, Engelman and Steiz method plots hydrophobic regions as positive and hydrophilic regions as negative (Engelman *et al.*, 1986). **(B,C)** The Eisenberg method was applied to peptide bond angles assuming that the protein folded as an α-helix **(B)** or a β-sheet **(C)**. This method plots the hydrophobic moment which is a measure of the asymmetry in the arrangement of polar and non-polar residues (Eisenberg *et al.*, 1984). **(D)** Charge density plots the average charge for each amino acid based on the sliding window calculation for each amino acid. Overall positive charge is plotted as positive and overall negative charge is plotted as negative. **(E)** The Karplus-Schultz method measures chain flexibility and plots regions of increased flexibility as a peak (Karplus and Schultz, 1988). **(F)** Hopp-Woods plots the average local hydropathy (hydrophilic and hydrophobic regions), with positive plots indicating hydrophilic regions and negative plots indicating hydrophobic regions (Hopp and Woods, 1981). **(G)** The Kyte-Doolittle method also plots hydropathy (hydrophobicity is positive and hydrophilicity is negative) but using a slightly different calculation (Kyte and Doolittle, 1984). **(H)** The Emini surface probability plots the probability of a region of the primary amino acid sequence being on the surface of the folded protein is plotted with the peaks indicating the areas that have the greatest likelihood of surface exposure (Emini *et al.*, 1985; Janin *et al.*, 1978). **(I)** The Jameson-Wolf Antigenic index uses a combination of the data from the methods **(E),(F)** and **(H)** along with the secondary structure prediction methods of Chou-Fasman (Chou and Fasman, 1978) and Robson Garnier (Garnier *et al.*, 1978) to plot an antigenicity index in which peaks show the greatest antigenic probability (Jameson and Wolf, 1988). The dashed lines on the plots in **(B)**, **(C)**, **(D)**, **(E)**, **(H)** and **(I)** indicate a threshold above (positive plot) and below (negative plot) which the values are considered significant. The peptide region that was selected is shown below the plots with the numbering indicating the position in the amino acid sequence. The cysteine coloured in green is not part of the original sequence, but was added to assist the cross-linking of the peptide to the carrier protein for antibody manufacture. (In the case of these antibodies, Keyhole Limpet Haemocyanin [KLH] was used as the carrier protein. The region on the plots **(A-I)** highlighted with the black rectangle is the region at which the selected peptide is located in the sequence.

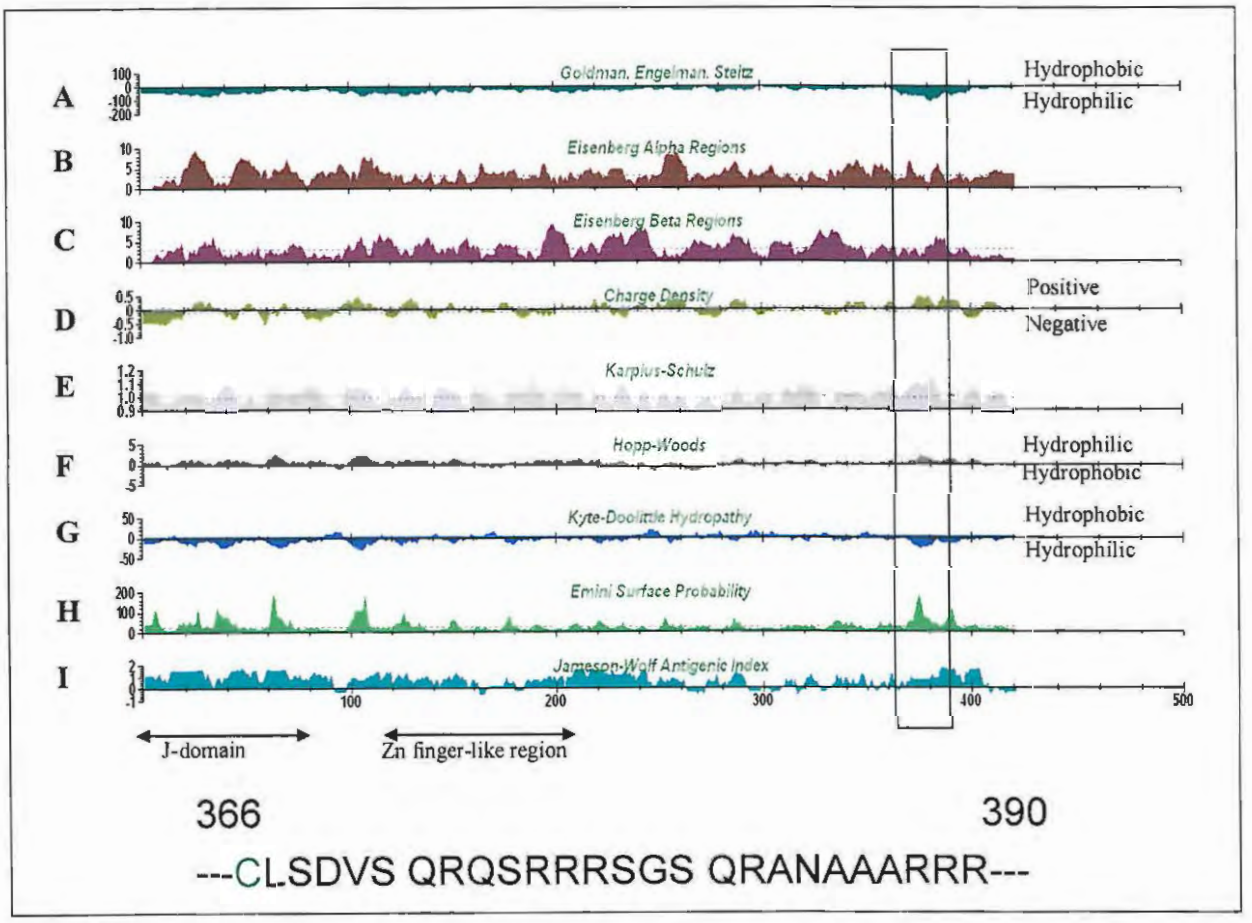


Figure 2.9: Composite analysis of Tcj3 primary amino acid sequence to determine suitable unique epitope regions for the design of a peptide polyclonal antibody specific to Tcj3.
The analysis was performed using the protein composite analysis function in Generunner (available at: http://www.generunner.net/). **(A)** The Goldman, Engelman and Steiz method plots hydrophobic regions as positive and hydrophilic regions as negative (Engelman *et al.*, 1986). **(B,C)** The Eisenberg method was applied to peptide bond angles assuming that the protein folded as an α-helix **(B)** or a β-sheet **(C)**. This method plots the hydrophobic moment which is a measure of the asymmetry in the arrangement of polar and non-polar residues (Eisenberg *et al.*, 1984). **(D)** Charge density plots the average charge for each amino acid based on the sliding window calculation for each amino acid. Overall positive charge is plotted as positive and overall negative charge is plotted as negative. **(E)** The Karplus-Schultz method measures chain flexibility and plots regions of increased flexibility as a peak (Karplus and Schultz, 1988). **(F)** Hopp-Woods plots the average local hydropathy (hydrophilic and hydrophobic regions), with positive plots indicating hydrophilic regions and negative plots indicating hydrophobic regions (Hopp and Woods, 1981). **(G)** The Kyte-Doolittle method also plots hydropathy (hydrophobicity is positive and hydrophilicity is negative) but using a slightly different calculation (Kyte and Doolittle, 1984). **(H)** The Emini surface probability plots the probability of a region of the primary amino acid sequence being on the surface of the folded protein is plotted with the peaks indicating the areas that have the greatest likelihood of surface exposure (Emini *et al.*, 1985; Janin *et al.*, 1978). **(I)** The Jameson-Wolf Antigenic index uses a combination of the data from the methods **(E)**,**(F)** and **(H)** along with the secondary structure prediction methods of Chou-Fasman (Chou and Fasman, 1978) and Robson Garnier (Garnier *et al.*, 1978) to plot an antigenicity index in which peaks show the greatest antigenic probability (Jameson and Wolf, 1988). The dashed lines on the plots in **(B)**, **(C)**, **(D)**, **(E)**, **(H)** and **(I)** indicate a threshold above (positive plot) and below (negative plot) which the values are considered significant. The peptide region that was selected is shown below the plots with the numbering indicating the position in the amino acid sequence. The cysteine coloured in green is not part of the original sequence, but was added to assist the cross-linking of the peptide to the carrier protein for antibody manufacture. (In the case of these antibodies, Keyhole Limpet Haemocyanin [KLH] was used as the carrier protein (Genscript, http://www.genscript.com/antibody.html). The region on the plots **(A-I)** highlighted with the black rectangle is the region at which the selected peptide is located in the sequence.

2.4) Discussion

The non-trypanosomatid organisms compared in **Figure 2.1** show that there is a vast difference in the possible number of Hsp40s in different organisms. It is tempting to hypothesise that the large number of Hsp40s in the trypanosomatids compared in this chapter is due to an adaptation for the pathogenicity of these organisms. An example of this would be the use of these proteins in protection against the stresses experienced by moving between the poikilotherimic insect vector, to the endothermic mammalian host. However, *P. falciparum* is adapted to similar stresses and has two thirds of the number of Hsp40 sequences (Botha *et al.*, 2007). *P. falciparum* exports some of the Hsp40 proteins into the infected erythrocyte to interact with host Hsp70s and assist in the remodeling of the erythrocyte by the parasite (Botha *et al.*, 2007). This is a possible scenario for those trypanosomatids that have an intracellular stage to their life-cycle in the mammalian host (e.g. *T. cruzi* or *L. major*), however, this is not likely to be the case for extracellular parasites such as *T. brucei*, which has almost the same number of Hsp40s in its genome as other trypanosomatids.

In *T. brucei*, the greatest number of Hsp40s have been categorized into the Type III category (data not shown) which is therefore believed to be the main source of Hsp40 proliferation in trypanosomatids. This is likely to be the case for the other trypanosomatid species too, due to the high genetic similarity between them, especially for the *Leishmanias* (El-Sayed *et al.*, 2005; Peacock *et al.*, 2007). The Type III Hsp40s are also the ones believed to have more specialized functions (Cheetham and Caplan, 1998; Walsh *et al.*, 2004; Fan *et al.*, 2003). It is therefore tempting to speculate that the specialization of function in this group is necessary for adaptations in a complex parasite life-cycle in functions other than adaptation to temperature stress.

Another possible explanation for the large number of Hsp40s found in the trypanosomatid species is the method used to find them. The assumption used is that all sequences with a J-domain (the defining feature of the Hsp40 family) are an Hsp40 and that no other proteins use or contain this domain. However, the BLAST of the genome databases with Tbj2 and Tcj2 was confirmed by the search of the annotated databases using the Keywords Hsp40 and DnaJ, which revealed slightly fewer, but almost the same

number as the BLAST search. Folgueira and Requena (2007) found a similar number of Hsp40s in the genomes of *T. cruzi*, *T. brucei* and *L. major*. However, it could be questioned as to whether some of these are actually Hsp40s. J66, for example does not appear to have a J-domain, despite the sequence identity it shares with the C-terminal domains of the J2 protein family. It is therefore safer to refer to these sequences (particularly the J66 group) as putative Hsp40s, until function has been determined. These sequences were found in the annotated nuclear genomes of a number of trypanosomatid species (El-Sayed *et al.*, 2005b; Berriman *et al.*, 2005; Ivens *et al.*, 2005; Peacock *et al.*, 2007).

The absence of a J-domain in TbgambJ2 and TbgambJ3 calls into question whether these sequences are the functional equivalent of the other J2 and J3 sequence members. Despite a high sequence identity with the J2 and J3 sequences from the Zinc finger-like region to the C-terminus, they share a more common domain structure with the J66 sequence family, which also all lack a J-domain and possibly a G/F rich region and CTDIA. Assuming that this is not a sequencing annotation error, due to the preliminary nature of the *T. b. gambiense* genome sequence, it is surprising that these sequences are so different from Tbj2 and Tbj3, considering that *T. brucei brucei* and *T. brucei gambiense* are subspecies. None of the full length Type I Hsp40s of *T. b. gambiense* are predicted to localize to the cytoplasm and therefore are unable to perform the J2 or J3 function, assuming that the localization predictions are correct.

The absence of the J-domain and other N-terminal domains in the J66 protein family, TbgambJ2 and TbgambJ3 raises the question as to whether these proteins can still be categorized as Hsp40s. The J-domain has been described as the major area of interaction between Hsp40s and Hsp70s (Hennessy *et al.*, 2005; Hennessy *et al.*, 2000). It is therefore unlikely that these sequences will have a conventional interaction with and regulation of a partner Hsp70. However, assuming the intact nature of the hydrophobic peptide binding pocket in CTDI, it may be possible for these sequences to bind non-native polypeptides as the other Type I Hsp40s do in order to suppress protein aggregation. This then raises the interesting academic question as to whether an Hsp40

requires a J-domain to fulfill all of its functions, especially if it can act as a chaperone in its own right, as is the case for Type I and Type II Hsp40s. An alternative hypothesis regarding these proteins proposes that the C-terminal regions of these proteins and a J-domain that are transcribed and translated separately interact subsequent to translation as and form a functional Hsp40 unit (Bursać and Lithgow, 2007). Some putative Hsp40s found in the *T. brucei brucei* genome database (Hertz-Fowler *et al.*, 1999) are short sequences that seem to contain only a J-domain. Any one of these may be the other half of Tbj66. Possible examples of these J-domain only sequences in the *T. brucei brucei* genome are Tb11.01.0135; Tb927.8.7010 and Tb927.5.2880 (protein sequence nomenclature according to the GeneDB database (http://www.genedb.org/; Hertz-Fowler *et al.*, 2004).

The exclusive presence of the J4 sequences in those trypansomatids that have an intracellular stage to their life-cycle in their mammalian hosts may indicate their importance in parasites evolved to be intracellular. Their localization was predicted to be cytoplasmic or nuclear. However, was this to the cytoplasm/nucleus of the parasite or the host cell? *P. falciparum* is known to export proteins into the infected erythrocyte using a *P. falciparum* specific export signal. These exported proteins interact with the host cell machinery to remodel the cell to the needs of the parasite (Botha *et al.*, 2007). The J4 proteins could be exported to the host cell, by an unknown signal, for similar adjustment to the host cell.

The prediction of the subcellular localization of a protein from its primary amino acid sequence is not always correct (Sprenger *et al.*, 2006). The prediction is very much dependent on the programme and the method it uses for the prediction. This is shown by the different predictions of localization by WOLF PSORT and pTARGET for some sequences. In addition, the use of computer programmes that predict the protein localization for animal cells assumes that the targeting signal is the same for each animal and its cells. The endoplasmic reticulum retention signal has been found to be very different between trypanosomatids and other animal cells such as *S. cerevisiae* and *H. sapiens* (Bangs *et al.*, 1993, Bangs *et al.*, 1996; Teasdale and Jackson, 1996). This is

the likely reason that no endoplasmic reticulum retention was predicted for any of the Type I/ Type IV/I sequences. The possibility of different localization and targeting mechanisms also exists. One of the *T. brucei* specific protein targeting domains has been found to be important in the localization of proteins to the flagellar pocket (Hill *et al.*, 1999). In addition, trypanosomatids have certain organelles that are not found in many other animal cells such as flagellae and glycosomes that are likely to have different targeting signals. The pseudogene sequences have an even greater chance of incorrect targeting prediction due to errors in the sequence that could result from frame shifts and missing parts of the sequence.

It is intriguing that so many Type I/TypeIV/I proteins should be localized to the mitochondrion, while none of the trypanosomatid sequences were predicted to localize to the endoplasmic reticulum. This is possibly due to the divergence of the endoplasmic reticulum retention signals between trypanosomatids and other animal cells (Bangs *et al.*, 2006) and may explain the inconsistency in the localization prediction made by WOLF PSORT and pTARGET for sequences in the J45, J46 and J47 families. The more consistent prediction of mitochondrial localization for the J27 (except TbJ27) and J50 could indicate that these are the true mitochondrial Type I/Type IV/I Hsp40s in these organisms. Even if some of these localization predictions are incorrect it is interesting that the localization prediction for a particular group is often the same.

The High percentage identity found for Tcj2 and Tcj3 (57%) (Tibbetts *et al.*, 1998) and Tbj2 and Tbj3 (44%ID) suggest that the J2 and J3 could be paralogues. This is further indicated by the clustering of Tbj2 and Tbj3 relative to *H. sapiens*, *S. cerevisiae* and *A. thaliana* sequences (**Figure 2.3**). The identification of Ydj1 as a putative orthologue of J2 and J3 (**Figure 2.3**) is further supported by the predicted localization of J2/J3 to the cytoplasm (**Table 2.3**), the predicted farnesylation of these proteins (**Table 2.3**) and the high percentage identity of these trypanosomatid protein families with Ydj1 (~ 40%). Ydj1 is known to have a cytoplasmic subcellular localization (Caplan and Douglas, 1991) and is known to be farnesylated (Caplan *et al.*, 1992a). Both Ydj1 (Caplan and Douglas, 1992; Atencio and Yaffe, 1992) and Tcj2 (Tibbetts *et al.*, 1998) are stress inducible. As

Ydj1 is better characterized than its trypanosomatid protein orthologues, it is possible to make some putative extrapolations about the function and characteristics of the J2/J3 proteins. Ydj1 is involved in the translocation of proteins into the endoplasmic reticulum and mitochondria in *S. cerevisiae* (Caplan and Douglas, 1991; Caplan *et al.*, 1992b; Atencio and Yaffe, 1992; Becker *et al.*, 1996). Deletion of Ydj1 produces a severe growth defect at all incubation temperatures in *S. cerevisiae*, even though these cells are still viable (Sahi and Craig, 2007; Johnson and Craig, 2000; Caplan and Douglas, 1991). Data indicates that farnesylation of Ydj1 is required for the correct function of Ydj1 at elevated temperatures Caplan *et al.*, 1992a). Farnesylation of Ydj1 also allows non-membrane associated Ydj1 to interact with cytosolic substrates such as prion proteins and suppress their toxicity (Summers *et al.*, 2009). In addition, Ydj1 is the most abundant Hsp40 found in the yeast cytosol (~119000 molecules per cell). It is by far the most abundant cytosolic Type I Hsp40 in *S. cerevisiae*, with Apj1 and Xdj1 measured at 125 and 1210 molecules per cell, respectively (Ghaemmaghami *et al.*, 2003; Sahi and Craig, 2007). YdJ1 generates and functions as a homodimer (Wu *et al.*, 2005; Ramos *et al.*, 2008). It is therefore possible that J2 and J3 would show these similar characteristics in the trypanosomatid species.

Chapter 3:
Analysis of the Amino Acids of the Tcj2 and Tcj3 J-domains Important for the Interaction with Hsp70

CHAPTER 3
Analysis of the amino acids of the Tcj2 and Tcj3 J-domains important for the interaction with Hsp70

3.1) Introduction

Hsp40 is known to regulate the Hsp70 chaperone cycle by physical interaction (Greene *et al.*, 1998; Hennessy *et al.*, 2005b). However, not all Hsp40s can form a functional interaction with all Hsp70s (Hennessy *et al.*, 2005b; Genevaux *et al.*, 2007). This is clearly illustrated by the ability of *E. coli* DnaJ to stimulate the ATPase activity of mammalian Hsc70, and by the inability of the human Hsp40 Hdj1 to stimulate the ATPase activity of DnaK (Minami *et al.*, 1996). In addition, it is known that *E. coli* DnaK can interact with the Hsp40 proteins DnaJ, CbpA and DjlA but not with HscB and other Hsp40s of the *E. coli* system (Genevaux *et al.*, 2007).

Although other areas of interaction are thought to exist, the J-domain appears to be the predominant site of functional interaction between Hsp40 and Hsp70 (Kelley, 1998; Kelly 1999; Hennessy *et al.*, 2005b). As is the case for the full length proteins, not all J-domains are interchangeable. DjlC, a Type III Hsp40 of *E. coli* normally interacts with HscC. The J-domain of this Hsp40 was unable to functionally replace the J-domain of *E. coli* DnaJ (Kluck *et al.*, 2002). Simian virus 40 (SV40) T antigen does not function when its J-domain is substituted for the J-domains of *E. coli* DnaJ or Ydj1 (Sullivan *et al.*, 2000). In *S. cerevisiae*, the J-domains of the yeast Hsp40s Sis1 (Type II Hsp40) and Mdj1 (Type I Hsp40) were unable to substitute for the J-domain of Sec63 (Type III Hsp40). However, Scj1 (Type I Hsp40) was able to functionally substitute for the J-domain of Sec63 (Schlenstedt *et al.*, 1995). As both Sec63 and Scj1 are Hsp40s of the endoplasmic reticulum, while Sis1 and Mdj1 are localised to the cytosol of *S. cerevisiae* (Schlenstedt *et al.*, 1995), this implies that the determinants for the functional interaction between Hsp40s and Hsp70s are specialized within some of the subcellular compartments of eukaryotes. However, many J-domains can functionally substitute for each other. *E. coli* DnaJ J-domain, for example, can be substituted by that of Mdj1 (Deloche *et al.*, 1997), DjlA (Genevaux *et al.*, 2001), Hdj1 and mammalian papovavirus T antigens (Kelley and Georgopoulos, 1997; Berjanskii *et al.*, 2000). In addition, simian virus 40 Large T antigen J-

domain has produced a functional chimera with the C-terminal region of Ydj1 (Fewell *et al.*, 2002). In general it appears that J-domains of Hsp40s that interact with Hsp70s that are functionally equivalent are able to substitute for one another (Deloche *et al.*, 1997; Genevaux *et al.*, 2001).

The amino acids important for Hsp40 J-domain interaction with Hsp70 can be subdivided into two categories: 1) those amino acids that are important for maintaining the structural integrity of the J-domain; and 2) those amino acids that are important for the physical interaction with Hsp70 (Hennessy *et al.*, 2000). Some of those J-domain amino acids that have been implicated in the interaction between Hsp70 and Hsp40 are believed to be involved in defining which of the Hsp70s the Hsp40 is able to interact with (i.e. defines specificity), while others are believed to be involved in more standard contacts between the two molecules (Hennessy *et al.*, 2005a; Hennessy *et al.*, 2005b; Hennessy *et al.*, 2000). A growing amount of literature has identified key regions and amino acid residues that are required for a functional interaction with the partner Hsp70. It was predicted that Helix II, Helix III and the loop region between these helices of Hsp40 form the major surface of interaction between Hsp40 and Hsp70 (Greene *et al*, 1998; Landry *et al.*, 2003; Hennessy *et al.*, 2005b; Garimella *et al.*, 2006). The net positive charge of Helix II has been hypothesized to interact with a negatively charged region on the ATPase domain of Hsp70 (Greene *et al.*, 1998; Gässler *et al.*, 1998). The most well characterized amino acids of Hsp40 required for a functional interaction with Hsp70 are located in the HPD tripeptide motif that is conserved among the large majority of Hsp40s (100% across the Type I and II proteins in a sample of over 200 Hsp40s)(Hennessy *et al.*, 2000; Hennessy *et al.*, 2005b). Mutation of these amino acids abolishes the functional interaction of Hsp40 and Hsp70. (Wild *et al.*, 1992; Szyperski *et al.*, 1994). Mutation of the histidine to a glutamine in this motif (H33 in DnaJ and H34 in Ydj1) resulted in the disruption of Hsp40 function (Kelley and Georgopoulos, 1997; Genevaux *et al.*, 2002 [*E. coli* DnaJ H33Q]; Hennessy *et al.*, 2005a [*A. tumefaciens* DnaJ H33Q]; Tsai and Douglas, 1996 [Ydj1 H34Q]; Genevaux *et al.*, 2001 [DjlA-DnaJ chimera, H233Q]; Stubdal *et al.*, 1997; Zalvide *et al.*, 1998 [Simian virus 40 Large T antigen H42Q]; Fewell *et al.*, 2002 [Sv40 T-antigen-Ydj1 chimera H42Q). Likewise mutation of the proline and the aspartic acid of the HPD motif results in a lack of functional interaction with Hsp70 (Genevaux *et al.*, 2002 [P34F and

P34/D35 deletion]; Genevaux *et al.*, 2001 DjlA-DnaJ chimera [D235N]; Sullivan *et al.*, 2000; Zalvide *et al.*, 1998; Stubdal *et al.*, 1997 [SV40 Large T-Antigen D44N]; Hennessy *et al.*, 2005a [*A. tumefaciens* DnaJ D35E]). The HPD motif was also found to be important in the stimulation of the ATPase domain of Hsp70 (Tsai and Douglas, 1996 (Ydj1); Chamberlain and Burgoyne, 1997 (Csp); Yan *et al.*, 2002 [P58IPK]). This conserved region in the Helix II/III loop that is essential for functional interaction with Hsp70/DnaK was proposed, by Genevaux and co workers, to include HPD(R/K)N based on their experiments with the Hdj1 and *E. coli* DnaJ J-domains (Genevaux *et al.*, 2002)

Helix II of the J-domain has a predominantly positively charged solvent exposed face rich in lysine amino acids and the sequence of this domain is well conserved between organisms (Hennessy *et al.*, 2000). Amino acids of this helix that have been mapped as important for functional interaction with Hsp70 are Y25, K26 in *E. coli* DnaJ (Genevaux *et al.*, 2002) and R26 in *A. tumefaciens* DnaJ (Hennessy *et al.*, 2005a). A synergistic effect was observed with the double mutant KR26,27AA in *E. coli* DnaJ which produced a greater inability to complement than K26 on its own, despite the phenotypically silent mutation of R27A (Genevaux *et al.*, 2002). Hennessy and colleagues proposed that R26/K26 makes a network of interactions with DnaK/Hsp70 (Hennessy *et al.*, 2005a). The same amino acids (Y24, R25, R26) showed the same importance in Hdj1J-domain-*E. coli* DnaJ chimera (Genevaux *et al.*, 2002). SV40 T antigen, a type III Hsp40, showed a requirement of other amino acids (L29, Y34 that are not equivalent positions to those of *E. coli* DnaJ) of the Helix II for restoration from heat sensitivity. The equivalent residues found to be important in *E. coli* and *A. tumefaciens* DnaJs were not tested (Fewell *et al.*, 2002).

In contrast to Helix II, Helix III contains an approximately equal number of negatively and positively charged amino acids and showed slightly lower sequence conservation, with the exception of the KFK tripeptide motif (Hennessy *et al.*, 2000). The phenylalanine of this tripeptide is especially well conserved (Hennessy *et al.*, 2000). Substitution of this amino acid (F47) in *A. tumefaciens* DnaJ to a leucine prevented the proper functioning of this protein (Hennessy *et al.*, 2005a). However, the F47A substitution in *E. coli* DnaJ, F45A in the Hdj1-DnaJ chimera and F47L substitution in *S. cerevisiae* Ydj1 removed the ability of these Hsp40s

to function (Genevaux *et al.*, 2002; Johnson and Craig, 2000). Substitutions of the lysines flanking F47; (K48T,K46T) in *A. tumefaciens* DnaJ (Hennessy *et al.*, 2005a) and *E. coli* DnaJ (Genevaux *et al.*, 2002) caused little if any detectable effect on the ability of these proteins to interact with *E. coli* DnaK. Other alanine scanning mutations of the Helix III also produced no functional difference from the wild type DnaJ (Genevaux *et al.*, 2002). In contrast, a more crucial role for Helix III has been demonstrated for the SV40 and Polyomavirus T-antigens (Garimella *et al.*, 2006 and Genevaux *et al.*, 2003; Fewell *et al.*, 2002). The lysines of the KMK motif (in place of the KFK motif), in addition to other amino acids of Helix III, were shown to be crucial in large T antigen function (Genevaux *et al.*, 2003; Fewell *et al.*, 2002). Garimella and colleagues confirmed that Helix III of the SV40 T antigen is important for its interaction with mammalian Hsc70 (Garimella *et al.*, 2006).

According to structural data, Helix I of the J-domain has a rigid and stable position relative to the other helices (Berjanskii *et al.*, 2000; Cupp-Vickery and Vickery, 1997; Huang *et al.*, 1998; Pellechia *et al.*, 1996; Qian *et al.*, 1996), and interacts with the back face of the helices II and III (Pellechia *et al* 1996; Qian *et al.*, 1996). This helix is therefore believed to have a role in maintaining the structural integrity of the J-domain through interaction of conserved hydrophobic and aromatic amino acid residues with amino acids in Helix II and III (Hennessy *et al.*, 2005a). Substitution of Y7A and L10A of *A. tumefaciens* DnaJ resulted in a loss of protein function (Hennessy *et al.*, 2005a). However, substitution of leucine residues to methionine, isoleucine and valine in SV40 large T antigen had no effect on its function (Li *et al.*, 2001). This was due to possible differences between the Helix I of the Type I and Type III J-domains.

Helix IV of the J-domain is not as well conserved as the others and is even completely absent in some J-domains such as the SV40 T antigens (Kim *et al.*, 2001). Here a longer Helix III and extended loop region is thought to substitute for Helix IV (Genevaux *et al.*, 2002). This helix is believed to be highly mobile and solvent exposed due to its low level of hydrophobic amino acid side chains (Hennessy *et al.*, 2005b). However, its function is not very well understood. The QKRAA motif, found to be conserved in many prokaryotic J-domains, makes up the bulk of Helix IV in these organisms (Auger and Roudier, 1997). Genevaux *et al* found that the

amino acid residues of Helix IV did not appear to be essential for the functional interaction of *E. coli* DnaJ J-domain with Hsp70 (Genevaux *et al.*, 2002). However, mutation of residue R63 to an alanine resulted in a lower ability of *A. tumefaciens* DnaJ to complement for DnaJ deficiency in *E. coli* (Hennessy *et al.*, 2005a). Helix IV of the J-domain has been proposed to specify which types of Hsp70s are bound by the Hsp40 (Hennessy *et al.*, 2005b). D59A of the Helix III/IV loop also disrupted the functional interaction of *A. tumefaciens* DnaJ with *E. coli* DnaK, but no effect was observed with the conservative substitution D59N (Hennessy *et al.*, 2005a) or in the double substitution of TD58,59AA in *E. coli* DnaJ (Genevaux *et al.*, 2002).

The structural stabilization of the J-domain helices in their relative positions is likely to be important for proper functional interaction of Hsp40 with Hsp70 (Hennessy *et al.*, 2000). I9, L10, V12, I21, A53 and L57 have been implicated in the structural stabilization of the *E. coli* DnaJ J-domain by forming a core of hydrophobic residues buried between the helices (Hill *et al.*, 1995; Pellechia *et al.*, 1996; Szyperski *et al.*, 1994). Other conserved residues L57, L10 and Y7 (Hennessy *et al.*, 2000) have been suggested to be important in the structural integrity of the J-domain and are necessary for proper function of *A. tumefaciens* DnaJ (Hennessy *et al.*, 2005a). In *A. tumefaciens* DnaJ, according to computer homology modeling of *A. tumefaciens* DnaJ, L10 interacts with Helix II amino acids and it is suggested that this is important in maintaining the stability of Helix II for its interaction with Hsp70. Y7 projects away from the J-domain but it was proposed that it could possibly interact with residues in helices I, II and III and possibly L57. L57 is thought to be a key residue in holding Helices II and III together (Hennessy *et al.*, 2005a). F47 has also been implicated in a structural role. By interacting with the histidine of the HPD motif, it is thought to stabilize helices II and III (Hennessy *et al.*, 2000).

An investigation of the amino acids of the J-domain important for the interaction between Hsp40 and Hsp70 has not been conducted on trypanosomal Hsp40s. This chapter makes use of alanine scanning mutagenesis to determine if selected amino acids of the Tcj3 and Tcj2 J-domains that are important for interaction with a partner Hsp70. The amino acids chosen for testing (**Figure 3.3**) are well conserved within J-domains across Hsp40s from different organisms and within the Hsp40 types (1 to IV) (Hennessy *et al.*, 2000; **chapter 2 section**

2.3.8 and 2.3.9). Conservation of amino acids in a sequence implies an importance in the maintenance of structural and/or functional integrity (Hennessy *et al.*, 2005b). The selected amino acids did not only show high levels of conservation within various J-domains, but have also been shown to be of functional importance in other Hsp40s (Hennessy *et al.*, 2005a; Genevaux *et al.*, 2002). The lack of a readily accessible laboratory culture of Type I Hsp40 deficient *T. cruzi* or *T. brucei* prompted experiments to be conducted in *S. cerevisiae* JJ160 (Johnson and Craig, 2000), which is a Ydj1 deficient strain. The absence of Ydj1 within *S. cerevisiae* cells results in a strain that is viable, but sensitive to high temperatures, unlike Sis1 deficiency that causes inviability (Johnson and Craig, 2001). Ydj1 deficient strains grow poorly at 30 °C (which is classed as the optimal growth temperature for *S. cerevisiae*) and do not grow well in liquid media (Caplan and Douglas, 1991; Atencio and Yaffe, 1992; Meacham, *et al.*, 1999; Johnson and Craig, 2000). This strain is therefore frequently incubated at temperatures between 23 °C and 25 °C (Edkins *et al.*, 2004; Johnson and Craig, 2000). The importance of prenylation in Tcj2 and the loss of the terminal glycine in its zinc finger like region for Tcj2 function was also assessed in *S. cerevisiae* JJ160 through amino acid substitutions. As was mentioned in chapter 1 (section 1.6.1.3), some Type I Hsp40s lack the terminal glycine in some of the CxxCxGxG repeats. A lysine or arginine has most often been found at this position (Martinez-Yamout *et al.*, 2000). Although, the glycines of the ZFLR repeats do not seem to be involved in the coordination of the zinc atoms, they do appear to be important for the overall structure of the ZFLR and it appears that only one of the glycines per Zinc Binding Domain can be substituted and maintain the correct structure of the domain (Martinez-Yamout *et al.*, 2000).

It was questioned if the same amino acids required for interaction of a type I Hsp40 J-domain with the *S. cerevisiae* (eukaryotic) Hsp70 Ssa1 would also be required for interaction with *E. coli* DnaK of the prokaryotic system. In order to assess this, a DnaJ/CbpA deficient strain of *E. coli* (OD259) (Deloche *et al.*, 1997) was used to test the effect of these same alanine point substitutions in the J-domain of Tcj3 on the interaction with DnaK. *E. coli* OD259 is a DnaJ and CbpA deficient strain of *E. coli* that is sensitive to temperatures above 37 °C and below 16 °C. It is therefore often cultured at 30 °C, unlike the wild type *E. coli* that grows optimally at 37 °C (Deloche *et al.*, 1997; Hennessy *et al.*, 2005a). The use of *E. coli* and *S. cerevisiae*

reporter systems to assess the importance of various amino acids in the J-domains of heterologous organisms has been successful in other studies (Genevaux *et al.*, 2002; Genevaux *et al.*, 2003; Fewell *et al.*, 2002). In addition, the culture of these organisms is less complicated and potentially less hazardous than *T. cruzi* or *T. brucei*.

The objectives of this chapter were to:

1) Test the ability of Tcj2, Tcj3 and mutants of these proteins to functionally substitute for Ydj1.
2) Examine the ability of a eukaryotic J-domain to functionally replace a prokaryotic J-domain in the prokaryotic system.
3) Investigate the effect of prenylation on the *in vivo* function of Tcj2 using a Ydj1 deficient *S. cerevisiae* system.

3.2) Materials and Methods

3.2.1) Construction of plasmids used during the course of this study

3.2.1.1) Construction of pKG6Tcj2 (Figure 3.2)

Tcj2 was PCR amplified from the original pET28aTcj2 construct kindly supplied by Dr David Engman (North Western University Medical School, Chicago USA) using the Expand High Fidelity PCR kit (Roche) (**Appendix C.1.15**), with oligonucleotide primers Tcj2F and Tcj2R (**Appendix A table A2**) with *Eco*RI and *Xho*I restriction endonuclease recognition sites attached to the forward and reverse primers respectively. The fragment was ligated into pGEM-T Easy (**Appendix C.1.14**) shuttle vector. Putative pGEM-T Easy-Tcj2 plasmids were isolated from *E. coli* transformants (**Appendix C.1.8**) and screened by restriction endonuclease digestion to release the Tcj2 insert with *Eco*RI and *Xho*I. A bulk digestion of pGEM-T Easy Tcj2 was then used to excise the Tcj2 coding sequence with *Eco*RI and *Xho*I, which was subsequently resolved on a 1% agarose gel (**Appendix C.1.9**). The band corresponding to the Tcj2 coding sequence was excised and purified from the agarose (**Appendix C.1.10**). The Tcj2 fragment was ligated into *Eco*RI and *Xho*I linearised pKG6 (A kind donation from Dr Petra Gentz, Department of Biochemistry, Microbiology and Biotechnology, Rhodes University, South Africa) that had been treated with shrimp alkaline phosphatase (Roche) (**Appendix C.1.12**). Putative pKG6Tcj2 plasmids were isolated from *E. coli* JM109 transformants and

screened for the Tcj2 insert with *Eco*RI and *Xho*I. The identity and nucleotide sequence of the coding region of Tcj2 was confirmed by chain-termination based cycle sequencing (**Appendix C1.16**).

3.2.1.2) Construction of pKG6Tcj3 (Figure 3.2)

Tcj3 was cloned into pKG6 in the same manner as Tcj2, but the restriction enzymes *Bam*HI and *Eco*RI were attached to the PCR primers (Tcj3F and Tcj3R) (**Appendix A Table A2**) and the coding sequence was amplified from pET23bTcj3 kindly supplied by Dr David Engman (North Western University Medical School, Chicago USA).

3.2.2) Generation of mutants by site directed mutagenesis

Point mutations were generated in the J-domains of Tcj2 and Tcj3 according to the protocol for rational site directed mutagenesis outlined in **Appendix C.1.16**.

3.2.3) *In vivo* complementation with *Escherichia coli* OD259

3.2.3.1) pQE30Tcj3-Agt expression vector

This plasmid (**Figure3.1**) consists of the pQE30 expression vector (Qiagen, USA) containing the coding region for the chimera of the Tcj3 J-domain and the remaining domains of *A. tumefaciens* DnaJ (excluding its J-domain). The Tcj3 J-domain coding region is located between the *Bam*HI and *Bst*BI restriction sites, while the coding region for the *A. tumefaciens* DnaJ portion of the chimera is between the *Bst*BI and *Hind*III restriction enzyme sites. The expression of the chimera is controlled by the *Lac* promoter and is inducible with IPTG. This vector contains an ampicillin resistance selection marker. This plasmid was kindly donated by Dr Aileen Boshoff (Dept. of Biochemistry, Microbiology and Biotechnology, Rhodes University, South Africa).

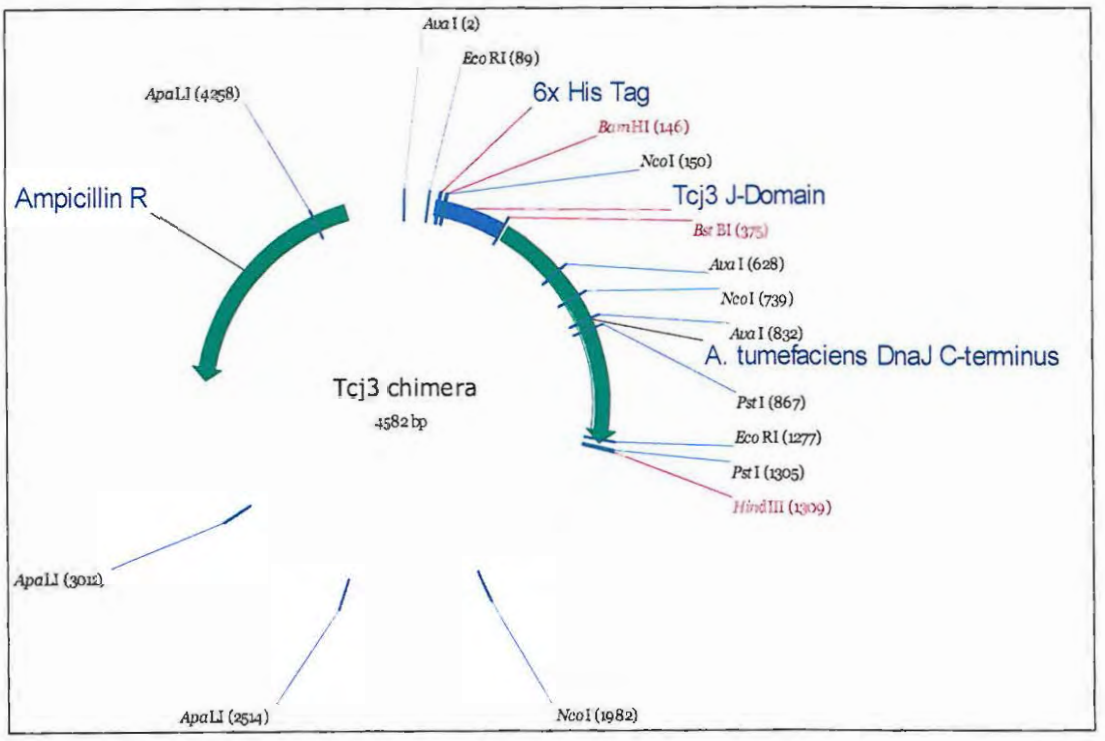


Figure 3.1: Map of the pQETcj3-Agt prokaryotic expression vector. The Tcj3-J-domain coding region, shown in blue, forms a chimera with the coding region for the remaining domains of *A. tumefaciens* DnaJ (excluding the J-domain). A His-tag attached to the 5' end allows for easy detection of protein by Western blot using anti-His antibodies. The backbone of the plasmid comes from pQE30 (Qiagen, USA) expression vector system. A coding sequence for β-lactamase (Ampicillin R) acts as the selective marker for the plasmid.

3.2.3.2) *In vivo* complementation of Tcj3-*A. tumefaciens* DnaJ Chimera (Tcj3-Agt) *E. coli* OD259

E. coli OD259 (MC4100 *araD139* Δ*ara714* Δ*CbpA::kan dnaJ*::Tn10-42) (Deloche *et al.*, 1997) was kindly donated by Dr Olivier Deloche (Dèpartement de Biochemie Mèdicale, Centre Mèdicale Universitaire, Facultè de Medicinè, Universitè de Genève, Geneva, Switzerland). This strain of *E. coli* is temperature sensitive and is unable to grow above 37 °C and below 16 °C. Fresh transformants (for preparation of transformation competent cells, see **Appendix C.1.1**) of *E. coli* OD259 were inoculated into 5 ml of YT broth containing kanamycin (50 μg/ml final concentration) for selection of the DnaJ and CbpA deficient strain (OD259) and Ampicillin (100 μg/ml final concentration) selection for the pQE30 expression vector. These starter cultures were incubated at 30 °C with shaking overnight and inoculated into 50 ml YT broth containing the same antibiotics and incubated at 30 °C until the cultures reach an A600

of 1.8 to 2 absorbance units. A portion of these cultures was diluted using sterile YT broth to an A_{600} of 0.2 absorbance units in a final volume of 1 ml. Five 10-fold dilutions (10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}) of the culture were made. A 2 µl sample of each of these dilutions was spotted onto kanamycin (50 µg/ml), ampicillin (100 µg/ml) and IPTG (50 µM) containing YT agar plates in increasing order of dilution. The spots were allowed to dry and the plates sealed with parafilm, prior to incubating replicate plates at 15 °C, 30 °C, 40 °C and 42 °C for 12 to 24 hours. Data was collected by scanning the agar plates on a digital image scanner.

3.2.3.3) Verification of expression of the various proteins in *E. coli* OD259

The cells from 1 ml of the complementation flask culture were harvested by centrifugation. The cells were resuspended in a volume of phosphate buffered saline (PBS) (pH 7.4, 137 mm NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 2 mM KH_2PO_4) to equalize the concentration of cells in each culture (the formula: A_{600nm}/0.5 x 150 µl = volume of PBS, was used to determine the volume of PBS to add). SDS-PAGE loading buffer (20 µl) (50 mM Tris-HCl pH 6.8, 25% glycerol, 2% SDS 0.001% bromophenol blue) was added to 40 µl of the PBS resuspended cells. These samples were boiled in a boiling water bath for 10 minutes and loaded on a discontinuous SDS-PAGE gel (**Appendix C.1.3**). Western blot analysis (**Appendix C.1.4**) was used to detect the His-Tcj3-Agt protein chimera in these *E. coli* OD259 cultures, using mouse anti-his antiserum as the primary antibody (1:5000) and a horse radish peroxidase conjugated goat anti-mouse IgG secondary antibody (1:5000) (**Appendix C.1.5**).

3.2.4) *In vivo* complementation with *Saccharomyces cerevisiae* Ydj1 deficient strain (JJ160)

S. cerevisiae JJ160 (*mat a trp1-1 ura3-1 leu2-3,112 his3-11,15 ade2-1 can1-100 GAL2+ met2-Δ1 lys2-Δ2 ydj1::HIS3*) (Johnson and Craig, 2000) was a kind donation of Dr Elizabeth Craig (University of Wisconsin Medical School, USA). These cells are able to grow at 23 °C, but are sensitive to higher temperatures. They were used to determine the ability of Tcj2 and Tcj3 to complement for the lack of the Type I Hsp40 Ydj1 in a heterologous eukaryotic system. The JJ160 strain was grown in yeast minimal medium (YMM) supplemented with 0.2% Adenine, 0.2% Tryptophan, 0.2% Methionine, 1% Leucine and 0.2% Uracil. Omission of Histidine from

the growth medium selects for the gene insertion that inactivates the Ydj1 gene in *S. cerevisiae* JJ160, while the carbon source for maintenance of the strain was 2 % glucose.

3.2.4.1) Vectors for expression of Tcj2 and Tcj3 in *Saccharomyces cerevisiae*

The positive control plasmid pRS317-Ydj1 (a kind donation of Dr Elizabeth Craig, University of Wisconsin Medical School, USA) contains the gene for Ydj1, thus restoring *S. cerevisiae* to a non-temperature sensitive strain. It contains a gene for synthesizing lysine, allowing *S. cerevisiae* JJ160 to grow on yeast minimal medium (YMM) lacking the amino acid (**See section 3.2.4.2**). The pKG6Tcj2 and pKG6Tcj3 expression vectors contain the *Ura3* gene and *S. cerevisiae* cells containing them were selected on YMM lacking uracil. Expression of Ydj1, Tcj2 and Tcj3 are regulated by the *gal* promoter which is induced by galactose in the growth medium. Therefore, growth of the cells in a medium containing glucose did not induce expression of the Ydj1, Tcj3 and Tcj2 coding sequences.

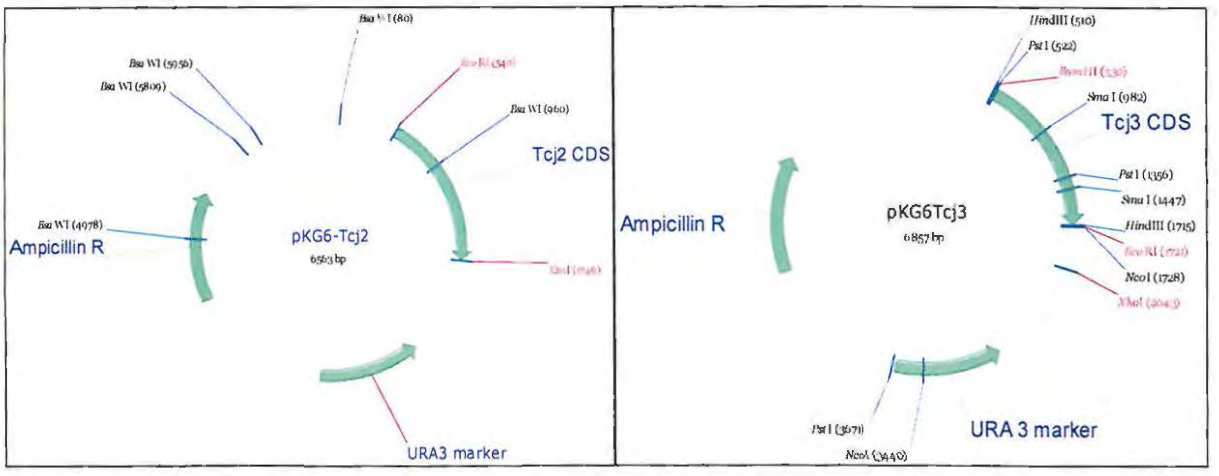


Figure 3.2: The plasmid maps of pKG6Tcj2 and pKG6Tcj3. These plasmids contain the coding sequence for Tcj2 (cloned into the plasmid with *Eco*RI and *Xho*I restriction enzymes) and Tcj3 (cloned into the plasmid with *Bam*HI and *Eco*RI restriction enzymes) respectively. pKG6 is a yeast expression vector that contains the coding sequence for Uracil (URA3) which acts as a selective marker in yeast cells. In addition, the plasmid can be propagated in prokaryotes and contains a coding sequence for β-lactamase (Ampicillin R) which acts as a selection marker in prokaryotic cells. Expression of Tcj2 and Tcj3 is induced by the presence of galactose in the growth medium.

3.2.4.2) Generation of transformation competent cells of *Saccharomyces cerevisiae* JJ160 and transformation of these cells

Transformation competent *S. cerevisiae* JJ160 were produced using the Frozen-EZ Yeast Transformation II^{TM} kit (Zymo Research, USA) as per the manufacturers instructions. A 10 ml culture was prepared on YMM [Glucose, HIS^-, Ade^+, TRP^+, MET^+, LEU^+, LYS^+] by incubating with shaking at 23 °C until an $A600_{nm}$ of 0.8-1.0 absorbance units. The cells were harvested by centrifugation at 500 x g for 5 minutes and resuspended in 10 ml of EZ 1 solution. After a second centrifugation of 500 x g for 5 minutes, the cells were resuspended in 1 ml of EZ 2 solution and aliqouted into 60 µl aliquots and allowed to freeze gradually in a -80 °C freezer.

To transform these competent cells, 15 µl of competent cells were added to 300 ng plasmid DNA. EZ 3 solution (150 µl) was then thoroughly mixed with the cells and DNA and the mixture was incubated at 23 °C for 1 hour mixing by brief vortexing every 15 minutes. The cells (150 µl) were spread onto the relevant selective YMM glucose for each plasmid, allowed to dry and incubated for 2 to 4 days.

3.2.4.3) *In vivo* complementation of Tcj3 and Tcj2 in *Saccharomyces cerevisiae* (JJ160)

pRS317-Ydj1 (the control plasmid containing the coding sequence for Ydj1), pKG6Tcj2, pKG6Tcj3 and their respective mutations were transformed into *S. cerevisiae* JJ160 and incubated on glucose yeast minimal medium (YMM) agar plates at 23 °C for 2 to 4 days. Colonies were picked and patched onto fresh glucose YMM agar plates and incubated at 23 °C for 2 to 3 days. These patches were then harvested into 1 ml of liquid YMM and vortex mixed until a homogenous suspension was generated. A 1 in 100 dilution of each was measured at A_{600nm} to estimate the volume required to inoculate an equivalent number of cells into 100 ml of liquid YMM containing galactose. These cultures were incubated at 23 °C with shaking until they reached mid log phase (A_{600nm} of 3 to 4 absorbance units). Culture cell densities were diluted to an $A600_{nm}$ of 0.2 absorbance units in a final volume of 1ml in microcentrifuge tubes. Five serial dilutions (10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}) each in a final volume of 1 ml were made, and 10 µl of each (including the 100 dilution solution) were spotted onto galactose YMM agar plates. After the spots had dried, replicate plates were incubated at 23 °C, 28 °C,

30 °C, 32 °C, 34 °C and 37 °C for 3 to 5 days. Data was recorded by scanning the plates on a digital image scanner.

3.2.5) Verification of expression of the parasitic Hsp40s and their mutants in *S. cerevisiae*

The cell densities of the *S. cerevisiae* cultures expressing Tcj3, Tcj2 and their mutants were equilibrated to an A_{600nm} of 4.0 absorbance units. The cells from 90 ml of each culture were harvested by centrifugation and resuspended in 5 ml of 100 mM Tris-SO_4^{2-} (pH 9.4) with 10 mM DTT prior to incubation in a 30 °C water bath for 10 minutes with gentle agitation. These cells were subsequently separated from the solution by centrifugation at 2000 x *g* for 5 minutes at 4 °C. The cell pellets were resuspended in 2.5 ml of PBS/Sorbitol buffer (155 mM NaCl, 25 mM Na-phosphate buffer [pH 7.3], 1.1 M sorbitol). The yeast cell walls were removed by addition of 50 µl (10 mg/ml) zymolyase (100 000 U/g; MP Biomedicals, LLC.) and incubated at 30 °C for 2 hours with gentle agitation. Protease inhibitor cocktail (200 µl) (Roche complete EDTA-free protease inhibitor cocktail [25 x stock solution made up in PBS/Sorbitol buffer]) was also added with the zymolyase. Spheroplasting of the yeast cells was confirmed using light microscopy and then Tween 20 (0.2% final concentration) was added to the spheroplasts and incubated on ice for 1 hour. Each cell lysate (200 µl) was added to 40 µl SDS-PAGE loading buffer (0.0625 M Tris-HCl pH 6.8, 10% glycerol, 2% SDS, 2% β-mercaptoethanol, 0.05% Bromophenol blue) and boiled for 10 minutes. The whole cell lysates (20 µl) were resolved on an SDS-PAGE gel (**Appendix C.1.3**) and blotted onto a nitrocellulose membrane (**Appendix C.1.4**). The anti-Tcj3 peptide polyclonal rabbit antibody (**Chapter 2 section 2.3.11**) was used to detect Tcj3 proteins within the lysates (1:5000 dilution) using the chemiluminescent based immunodetection technique (**Appendix C.1.5**). Tcj2 protein was detected in the lysates using anti-Tcj2 peptide polyclonal rabbit antibody (1:5000 dilution), raised from a region of the C-terminus of Tcj2 (Edkins *et al.*, 2004), also using chemiluminescent based immunodetection (**Appendix C.1.5**).

3.3) Results

3.3.1) Comparison of the Tcj2, Tcj3, Ydj1 and *Agrobacterium tumefaciens* J-domains

Homology models of the Tcj2, Tcj3 and Ydj1 J-domains were prepared in **chapter 2** (**Section 2.3.10**). In **Figure 3.3 A, B, C and D** the predicted positions and orientations of the amino acids substituted by alanine or glutamine in the J-domain are shown (Y8A, L11A, R27A, H34Q, K37A, F47A, L57A, D59A and R63A in Tcj2. Y8A, L11A, R27A, H34Q, K37A, F48A, L58A, D60A and R64A in Tcj3). The amino acids that were chosen for substitution in this study all displayed similar locations and orientations relative to the rest of the J-domain in all three models, including the crystal structure of *E. coli* DnaJ (Pellechia *et al.*, 1996; 1XBL.Pdb).The hydrophobic amino acid residue L11A (L10A in *E. coli* DnaJ) projects out of Helix I towards the rear face of Helices II and III. Y8A, though partially solvent exposed, is postulated to make contacts with other residues of Helix I, II III or L57 (Hennessy *et al.*, 2005a). The selected amino acids of helices II and III and the interhelical loop region (R27A, H34Q, K37A and F47A) all project out of the face of the J-domain and are likely to be at least partially solvent exposed. F47A is potentially partially shielded from the aqueous environment by the adjacent lysine residues of the KFK motif. The hydrophobic L57 amino acid residue projects from Helix III towards Helix II and has been previously suggested to make contacts with amino acids of this helix (Hennessy *et al.*, 2005a). In the Helix III/IV loop, aspartic acid 59 (*E. coli* numbering) is negatively charged and in the Helix IV, arginine 63 is positively charged. Both appear solvent exposed according to the various homology models, but R63 projects over the helices I, II and III.

A comparison of the primary amino acid sequence of the J-domains of Tcj2, Tcj3, Ydj1, *A. tumefaciens* DnaJ and *E. coli* DnaJ (**Figure 3.3E**) showed that the amino acids chosen for substitution with alanine were completely conserved across these proteins. The *E. coli* sequence has a lysine instead of an arginine at position 26 (position 27 for Tcj2, Tcj3 and Ydj1) and an arginine instead of a lysine at position 36 (position 37 for Tcj2, Tcj3 and Ydj1). Arginine and lysine are both basic and positively charged amino acids of similar size and structure. It is therefore likely that these differences in the *E. coli* sequence to the other sequences should produce minimal difference in the overall structure and function of these J-domains.

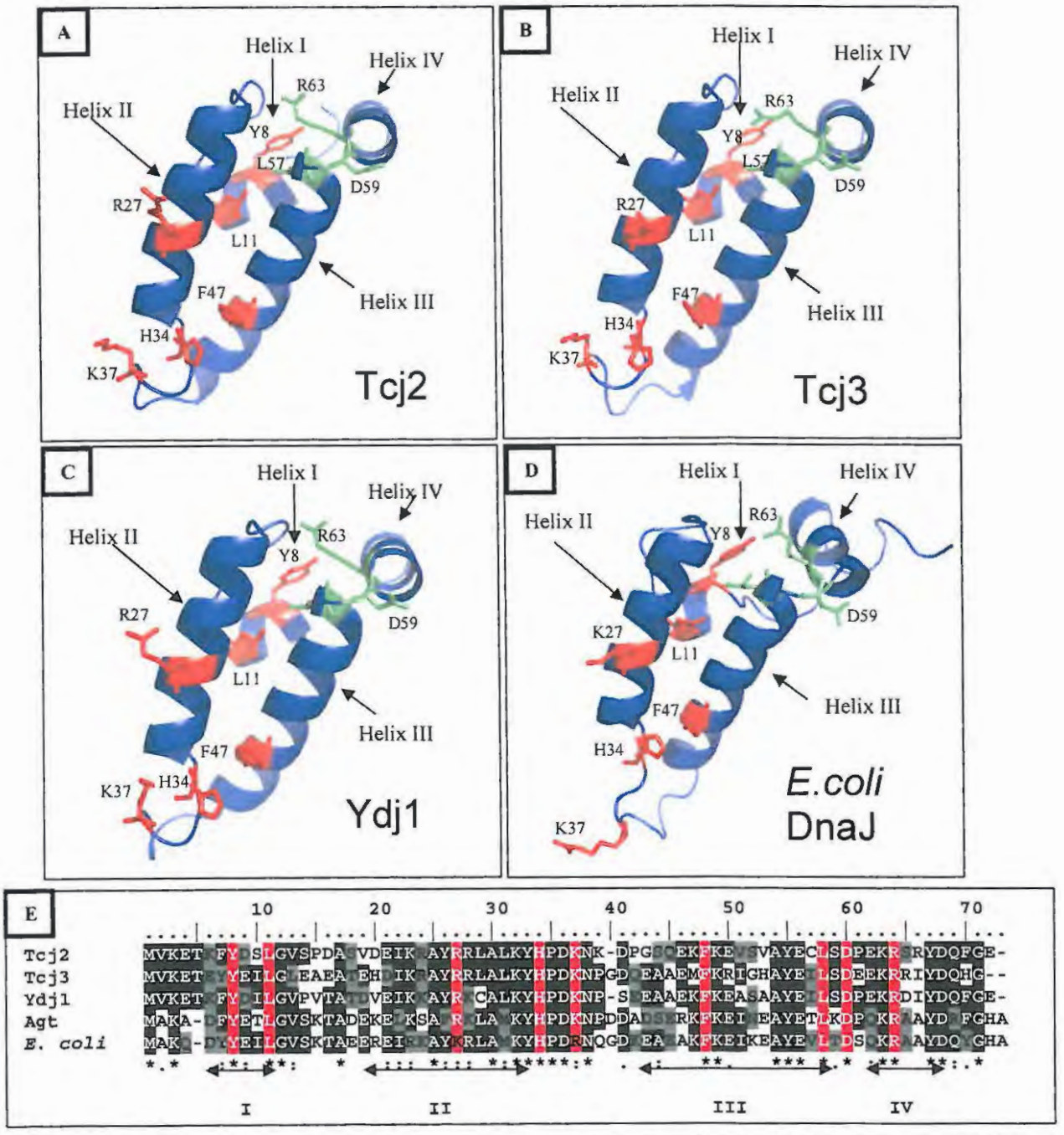


Figure 3.3: Ribbon representations of homology models of the J-domains of Tcj2 (A), Tcj3 (B) and Ydj1 (C) and crystal structure of *E. coli* DnaJ (D) (1XBL.pdb; Pellechia *et al.*, 1996). The amino acids that were mutated in this study are highlighted in red and green. All except H34 were mutated to an alanine (H34 was mutated to glutamine). Modeling produced using modeler (See chapter 2 section 2.3.10) and the model structure was visualized using Pymol (Delano, 2002).**E)** Alignment of the Tcj2, Tcj3, Ydj1, *A. tumefaciens* DnaJ (Agt) and *E. coli* DnaJ (*E. coli*). A consensus sequence is shown below the alignment. '* 'indicate 100% identity across the alignment, while ': 'and '. ' indicate partial matches across the sequences. The amino acids that were mutated in Tcj2 and Tcj3 are indicated with red background shading. The black background shading indicates identical amino acids at that position in the various sequences, while grey shading shows amino acids that are similar at the same position in the various sequences. The amino acids that form the α-helices I-IV are indicated below the alignment. The sequence alignment was produced using Bioedit (Hall, 1999).

These sequences showed a high level of sequence identity at the C-terminus of Helix II and N-terminal side of the loop between Helix II and III. Although, Helix II of the J-domain showed a low level of identity across all the sequences, a high level of amino acid sequence similarity was observed. The sequences of the Helix IV in the J-domains of eukaryotic origin (Tcj2, Tcj3 and Ydj1) do not contain the QKRAA motif that is found in Helix IV of most prokaryotic DnaJ proteins (Auger and Roudier, 1997). Instead, the eukaryotic sequences in this alignment had a conserved EKR motif. The lysine (K) and arginine (R) are conserved across all sequences, and the glutamic acid (E) is similar to glutamine (Q) in terms of size and structure. However, E will be negatively charged at pH 7, while Q will be neutral but polar.

3.3.2) Tcj2 and selected J-domain point mutants can functionally replace Ydj1 in *S. cerevisiae*

The fact that Tcj2 and Ydj1 are homologues based on their amino acid sequence similarity and identity (**Chapter 2 section 2.3.7 and 2.3.9**), presented the *S. cerevisiae* Ydj1 knockout strain (JJ160 [Johnson and Craig, 2000]) as an ideal system to study Tcj2 function *in vivo*, owing to the lack of access to a laboratory colony of *T. cruzi* for such studies. Plasmid expression vectors containing Ydj1, Tcj2 and Tcj2 mutant proteins were transformed into *S. cerevisiae* JJ160 and tested for their ability to substitute for the absence of Ydj1 in this yeast strain under normal and heat stressed conditions. These experiments were performed three times in order to assess the reproducibility of the data. The positive and negative controls for *in vivo* complementation in *S. cerevisiae* JJ160 (Figure 3.4) use different selective growth media in relation to the cells containing the pKG6 expression vector. The Ydj1 expression vector as the positive control has a lysine selective marker and the JJ160 strain has a histidine selective marker, that requires deficiency of these amino acids in the growth medium, while pKG6 selection requires a uracil deficiency in the growth medium. pKG6-Tcj2 (wild type Tcj2) and pKG6 vector alone control could therefore be better respective positive and negative controls for this study. Both showed results consistent with those observed in Edkins *et al.* (2004).

All *S. cerevisiae* cultures produced colonies even at the highest dilutions when incubated at 23 °C with the exception of the JJ160 strain on its own. This could be indicative of the different growth medium required for selection of the *S. cerevisiae* strain without expression vector

selection. When incubated at 30 °C, which is accepted as the optimal growth temperature of *S. cerevisiae* (Johnson and Craig, 2000), the cultures not expressing proteins that are able to substitute for Ydj1 show a decrease in cell viability relative to the cell line in which Ydj1 was restored (positive control) and the cells expressing Tcj2 wild type. The distinction between the ability and inability to functionally substitute for Ydj1 is a lot clearer at 32 °C. The various proteins were also expressed in *S. cerevisiae* at 34 °C and 37 °C (data not shown), but even the positive control was showing signs of failure to maintain cell viability at the higher cell dilutions. This showed that even Ydj1 was not able to completely rescue the cells from heat stress at these temperatures.

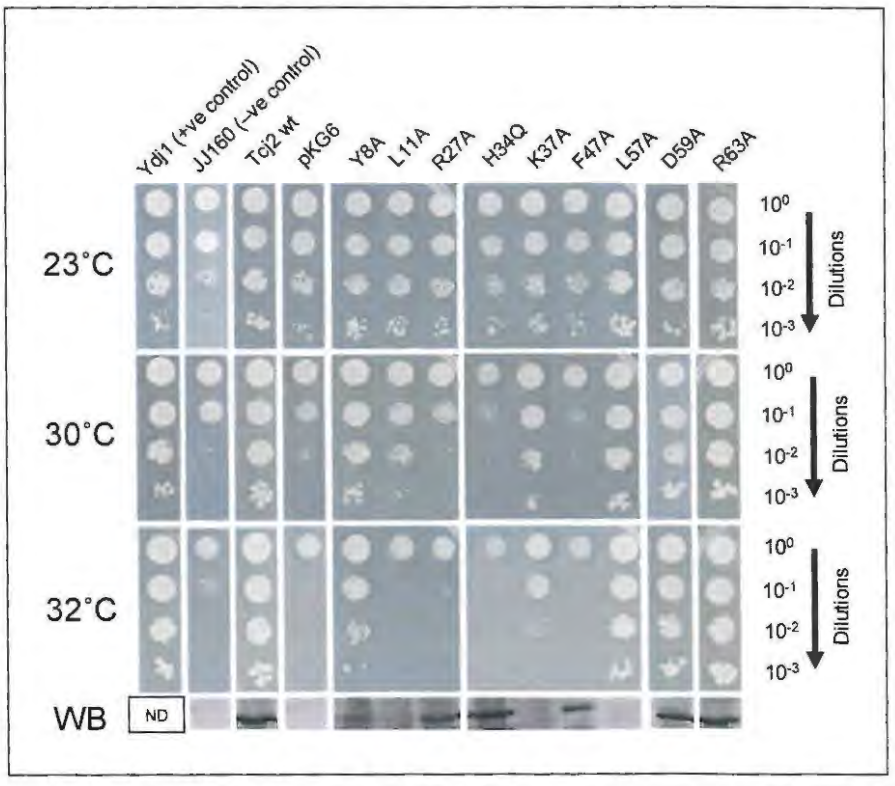


Figure 3.4: *In vivo* complementation data showing the ability of Tcj2 and Tcj2 mutants to functionally replace Ydj1 in *S. cerevisiae* JJ160. The temperatures used to incubate the yeast cells is indicated to the left of the panels. The dilutions of the yeast cultures are indicated to the right of the panels, starting from an A_{600nm} of 0.2 Absorbance units optical density (10^0). The expression of these various mutants was analysed using Western blotting (WB), with Tcj2 C-terminal peptide polyclonal antibody (1:5000 dilution) as primary antibody (Edkins *et al.*, 2004). The GE Healthcare, Amersham ECL advance kit horse radish peroxidase labeled anti-rabbit IgG (1:5000 dilution) was used as secondary antibody. The expression of Ydj1 (not detemined [ND]) was not assessed as there was no antibody for Ydj1 available. Ydj1 (Ydj1 on the figure) was used as a positive control, while the cells alone (JJ160) and vector alone (pKG6) were used as negative controls. Tcj2 wt is the wild type full length Tcj2. Y8A, L11A, R27A, H34Q, K37A, F47A, L57A, D59A and R63A indicated the point mutants of the J-domain that were used in each column on the figure. Apart from these point mutations, the mutant proteins were identical to Tcj2 wt.

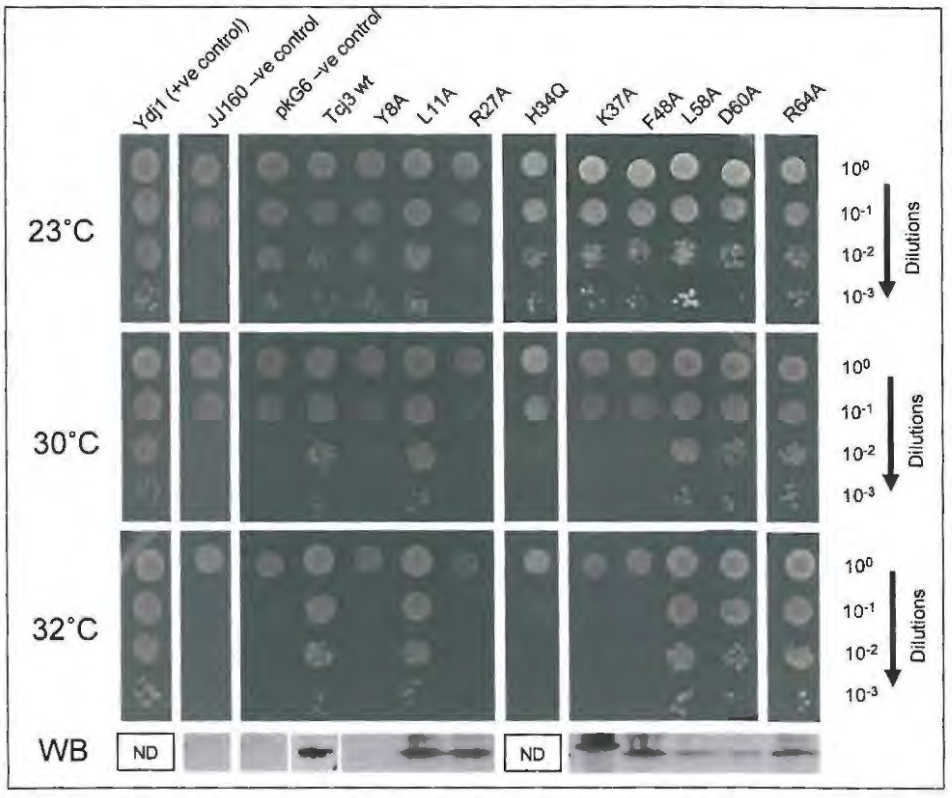
Tcj2 Y8A was able to complement for Ydj1 deficiency almost as well as the wild type Tcj2, but Tcj2 L11A showed only a partial complementation at 30 °C, which failed completely at 32 °C. Tcj2 R27A (Helix II), Tcj2 H34Q (Loop between Helix II and III) and Tcj2 F47A (Helix III) were all unable to complement for Ydj1 deficiency. K37A was able to complement at 30 °C but only showed a partial complementation at 32 °C. Tcj2 L57A, Tcj2 D59A and Tcj2 R63A all complemented as well as wild type Tcj2. Despite the lack of expression confirmation for Tcj2 L57A, its ability to complement for Ydj1 as well as wild type Tcj2 indicates that Tcj2 L57A was produced, but at undetectable levels. The same can be argued for Tcj2 K37A and its partial complementation, except that the low levels of expression could have resulted in the lower ability of this Tcj2 mutant to complement for Ydj1 deficiency.

The *S. cerevisiae* strains that did not complement showed a slightly slower growth rate even at 23 °C, which manifested as smaller individual colonies in the dilutions 10^{-2} and 10^{-3}. The 10^{0} cell culture spot of the negative controls and cultures expressing proteins that do not complement for Ydj1 deficiency still displayed growth even at 32 °C. This is possibly due to a cell density effect, where a high cell density shields cells from the effect of increased temperature. Alternatively, a high cell density could allow cells to multiply to a visible patch before the adverse effects of temperature stress are manifested in lack of growth. The cell density effect has been observed in other studies (Hennessy *et al.*, 2005a; Genevaux *et al.*, 2002; Johnson and Craig, 2001.

3.3.3) Tcj3 and selected mutants can functionally replace Ydj1 in *S. cerevisiae*

S. cerevisiae JJ160 was also used to study the J-domain of Tcj3, due to the lack of a *T. cruzi* laboratory culture and the fact that Tcj3 and Ydj1 are both type I Hsp40s. Plasmid expression vectors containing Ydj1, Tcj3 wild type and Tcj3 mutant coding sequences were transformed into *S. cerevisiae* JJ160 to assess their ability to compensate for Ydj1 deficiency (**Figure 3.5**). These experiments were repeated at least three times to ensure reproducibility of the data. The Ydj1 expressing positive control and the strain alone negative control show similar results to Tcj2 and indicate that the JJ160 strain was still functioning correctly. The various incubation temperatures show the same trend as for the Tcj2 complementation data (**Figure 3.4**). All cell

cultures were able to survive with an incubation temperature of 23 °C. The inability to complement for Ydj1 deficiency becomes evident at the normal *S. cerevisiae* growth temperature of 30 °C and becomes clearer at 32 °C. Tcj3 wild type protein was able to complement for lack of Ydj1 even under heat stress, while the vector control (pKG6) showed the same result as for the replicate shown in the Tcj2 complementation data (**Figure 3.4**).



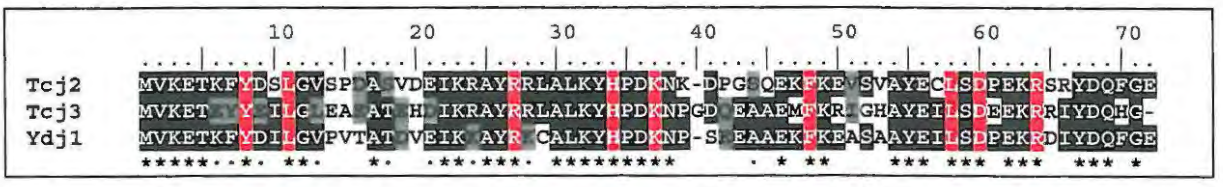


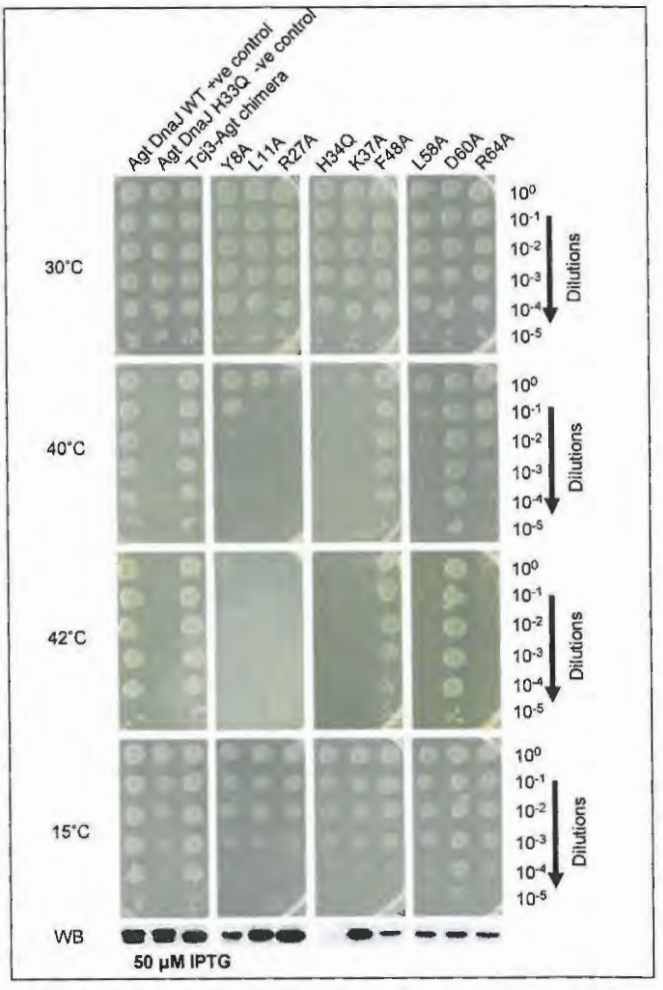
Figure 3.5: *In vivo* complementation data showing the ability of Tcj3 and Tcj3 mutants to functionally replace Ydj1 in *S. cerevisiae* JJ160. The temperatures used to incubate the yeast cells are indicated to the left of the panels. The dilutions of the yeast cultures are indicated to the right of the panels, starting from an A_{600nm} of 0.2 Absorbance units optical density (10^0). The expression of these various mutants was analysed using Western blotting (WB), with Tcj3 C-terminal peptide polyclonal antibody (1:5000 dilution) as primary antibody (See chapter 2 section 2.3.11 for the design of the peptide). The GE Healthcare, Amersham ECL advance kit horse radish peroxidase labeled anti-rabbit IgG (1:5000 dilution) was used as secondary antibody. "ND" shows that expression for these particular proteins was not determined. The complementation of Ydj1 (Ydj1 on the figure) was used as a positive control, while the cells alone (JJ160) and vector alone (pKG6) were used as negative controls. Tcj3 wt is the wild type full length Tcj3. Y8A, L11A, R27A, H34Q, K37A, F47A, L57A, D59A and R63A indicated the point mutants of the J-domain in Tcj3 that were used in each column on the figure. Apart from these mutations, the proteins were the same as Tcj3 wt protein. The alignment compares the J-domains of Tcj2, Tcj3 and Ydj1 and the amino acids substituted for alanine or glutamine are highlighted in red. '*' = 100% identity and '.' = partial match across the sequences. The alignment was produced using Bioedit (Hall, 1999).

Tcj3 Y8A showed a loss of compensation for a lack of Ydj1, while Tcj3 L11A was able to compensate for Ydj1 deficiency as was wild type Tcj3. The expression of Tcj3 Y8A was not detected using Western blot and could indicate that the lack of complementation is due to a lack of expression of the mutant protein. Tcj3 R27A, Tcj3 H34Q, Tcj3 K37A and Tcj3 F48A were all unable to compensate for Ydj1 under heat stress. The expression of Tcj3 H34Q was not determined and therefore little conclusion can be drawn from its lack of complementation. As with all the other mutations, the Tcj3 H34Q mutation was confirmed with restriction endonucleases and sequencing (data not shown). The band recognised by anti-Tcj3 in the expression confirmation of Tcj3 K37A was slightly larger than the bands of other Tcj3 proteins that were expressed. As was seen in mutations of Helix III C-terminal end and Helix IV of the J-domain in Tcj2 (L57A, D59A and R63A), the same mutations in Tcj3 (L58A, D60A and R64A) produced no loss of functional complementation for Ydj1 deficiency by Tcj3. All three of these mutations of Tcj3 were confirmed to be expressed. The complementation experiments of Tcj3 and its mutants in *S. cerevisiae* JJ160 were repeated at least three times to ensure reproducibility of the data.

3.3.4) The Tcj3 J-domain is able to functionally replace *Agrobacterium tumefaciens* DnaJ J domain in a Tcj3-J-*Agrobacterium tumefaciens* DnaJ chimera.

The Tcj3 full length protein was unable to complement for the lack of DnaJ and CbpA in *E. coli* OD259 (Data not shown). It was therefore decided to use a chimera of the Tcj3 J-domain and *A. tumefaciens* DnaJ (excluding the *A. tumefaciens* DnaJ J-domain) (Tcj3-Agt) to assess the effect of point mutations in the Tcj3 J-domain on the ability of Tcj3-Agt to functionally replace *E. coli* DnaJ and CbpA (Nicoll *et al.*, 2007).

For this set of experiments, *A. tumefaciens* DnaJ wild type and the *A. tumefaciens* DnaJ H33Q mutant were used as the positive and negative controls respectively (**Figure 3.6**). *A. tumefaciens* DnaJ has been shown to complement for the lack of *E. coli* DnaJ and CbpA, while the H33Q mutant of this protein does not (Hennessy *et al.*, 2005a). The expression of all proteins that were tested showed no inhibitory effect on the growth of *E. coli* OD259 when grown at 30 °C, which is the permissive growth temperature of this *E. coli* strain (Deloche *et al.*, 1997; Hennessy *et al.*, 2005a). Growth of these cells at 40 °C showed a distinct difference



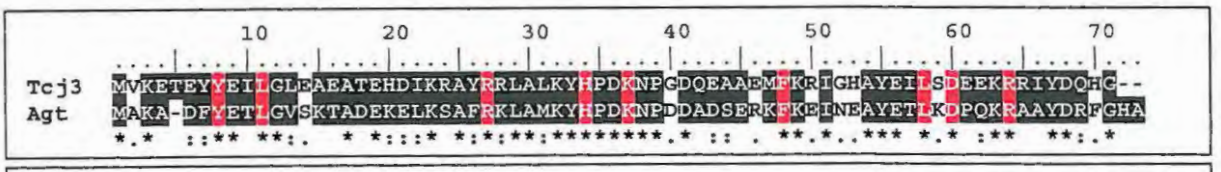


Figure 3.6: *In vivo* complementation data testing the effect of mutants in the J-domain of Tcj3 on the ability of the chimera of Tcj3-J-domain and *A. tumefaciens* DnaJ C-terminal domains (Tcj3Agt) to substitute for DnaJ/CbpA in *E. coli* OD259. The temperature of incubation of each dilution series is shown to the left of the photo panels and the dilution series, starting from an A_{600nm} of 0.2 Absorbance Units (10^0) is shown on the right. The amino acid substitutions were all in the J-domain of Tcj3Agt. All agar plates used in this experiment contained 50 µM IPTG. Expression analysis by Western blot is shown in the lowest panel of the figure (WB). The alignment compares the amino acids of the J-domains of Tcj3 and *A. tumefaciens* DnaJ. The amino acids substituted for alanine or glutamine are highlighted in red. '*' = 100% identity and ':' or '.' = partial match across the sequences. The alignment was produced using Bioedit (Hall, 1999).

between those proteins that are able to complement and those that are not, with the exception of Tcj3Agt-Y8A, Tcj3Agt-L58A and Tcj3Agt-R64A, which show partial complementation at this temperature. When the cells were incubated at 42 °C, only the positive control, Tcj3-Agt wild type protein, Tcj3-Agt F48A and Tcj3-Agt D60A were able to complement for DnaJ/CbpA deficiency. The partial complementation observed in Tcj3Agt-Y8A, Tcj3Agt-L58A and Tcj3Agt–R63A suggested that these mutant proteins were partially functional and that the substituted amino acids were not critical for the functional interaction with DnaK under these stress conditions. However, as the stress increased at 42 °C, these mutant proeins were no longer able to reverse the thermosensitivity of *E. coli* OD259, suggesting that the substituted residues were functionally and/or structurally important at this increased stress condition.

Incubation of the various cultures of *E. coli* OD259 expressing the test proteins at 15 °C, showed that those cultures that were able to complement at 42 °C grew better than those that were unable to complement at 42 °C. This possibly indicates a role of the functional Hsp40/Hsp70 interaction in suppressing the effects of cold shock in *E. coli* OD259. Other DnaJ/CbpA deficient strains of *E. coli* have displayed the same phenomenon (Genevaux *et al.*, 2003; Genevaux *et al.*, 2002). Tcj3-Agt-F48A does not show any growth rate advantage at 15 °C, which possibly indicates the slightly lower efficiency of complementation for DnaJ/CbpA relative to Tcj3-Agt wild type and Tcj3-Agt D60A. This can also be seen in the cell density colonies in the spots of the 42 °C dilution series of Tcj3-Agt F48A relative to the other proteins that complemented for the absence of DnaJ/CbpA.

Expression of all of the proteins except Tcj3-Agt H34Q was confirmed by Western blot (**Figure 3.6**) using anti-His as the primary antibody in the chemiluminescent detection system. This therefore sheds doubt on the ability to conclude whether the Tcj3-Agt H34Q protein does not complement in *E. coli* OD259 due to lack of expression or functional ability. A wealth of literature exists concerning the equivalent mutation in other Hsp40s in both eukaryotic and prokaryotic systems, all of which fail to complement (Genevaux *et al.*, 2002 [*E. coli* DnaJ H33Q]; Nicoll *et al.*, 2007 [Pfj1 H33Q, Pfj4 H33Q, Hsj1 H33Q]; Hennessy *et al.*, 2005a [*A. tumefaciens* DnaJ H33Q]; Genevaux *et al.*, 2003 [Sv40 T antigen H42Q]). This along with

the inability of Tcj2-H34Q to complement in *S. cerevisiae* (**Section 3.3.2**) implies that it is likely that Tcj3-Agt-H34Q would also not complement even if expression was high enough to be detected. The positive and negative controls show a double band pattern that was detected in the Western blot. This did not appear to affect the chimera proteins and did not appear to affect the ability of the control proteins to complement or not complement in the same way as in previous work (Hennessy *et al.*, 2005a).

3.3.5) Comparison of the eukaryotic and prokaryotic systems

Table 3.1 shows a comparison of the complementation data for the same mutations in the J-domains Tcj2 and Tcj3 and published data on *A. tumefaciens* DnaJ (Hennessy *et al.*, 2005a) and *E. coli* DnaJ (Genevaux *et al.*, 2002). Tcj2Y8A and Tcj3L11A are still able to complement for Ydj1 deficiency, while Tcj2L11A and Tcj3Y8A are not. K37A also shows a slight difference between Tcj2 and Tcj3 as Tcj2 is marginally able to complement for Ydj1 deficiency at 32 °C in *S. cerevisiae*.

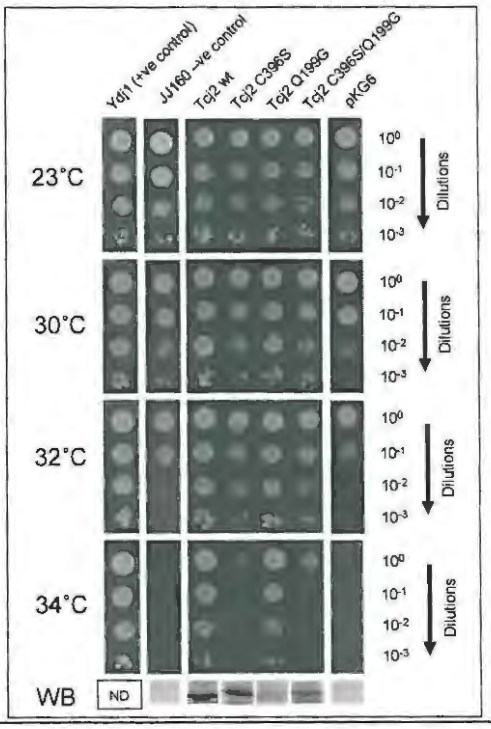
Table 3.1: A summary of the complementation data presented in this chapter compared to corresponding data within *A. tumefaciens* DnaJ (Hennessy *et al.*, 2005a) and *E. coli* DnaJ (Genevaux *et al.*, 2002). Yes indicates the ability of a mutant protein to complement for the lack of an Hsp40 in that system (*S. cerevisiae* JJ160 or *E. coli* Hsp40 knockout strain). No indicates a lack of ability of that protein to complement, while partial indicates a partial ability to complement. Those equivalent mutations that do not correspond to the sequence numbering and amino acids indicated in the first column are indicated in brackets in the relevant columns

Protein/ mutation/strain	Tcj2 Full length (Yeast system) 32°C	Tcj3 Full length (Yeast System) 32 °C	Tcj3-Agt Chimera (Prokaryotic system) 42 °C	Agt DnaJ (Prokaryotic system) (Hennessy et al., 2005)	*E. coli* DnaJ (Genevaux *et al.*, 2002) Data for 40 °C
Strain alone	No	No	No	No	
Plasmid (-ve) control (no insert)	No	No	-	-	No
Wt protein (+ve control)	Yes	Yes (not as well as Tcj2)	Yes	Yes	
Y8A	Yes	No	No	No	
L11A	No	Yes	No	No	
R27A	No	No	No	No (R26A)	Partial
H34Q	No	No	No	No (H33Q)	No (H33Q)
K37A	No (partial)	No	No	No	
F47A	No	No (F48A)	Yes (F48A)	Yes (F47L)	No (F47A)
L57A	Yes	Yes (L58A)	No (L58A)	No	
D59A	Yes	Yes (D60A)	Yes (D60A)	No	Yes (58TD59)
R63A	Yes	Yes (R63A)	No (R63A)	Yes	Yes (62KR63)

Interestingly, the full length proteins Tcj2, Tcj3 and *E. coli* DnaJ (Genevaux *et al.*, 2002) were unable to complement when containing the F47A mutation, while Tcj3AgtF48A and *A. tumefaciens* DnaJ F47L (Hennessy *et al.*, 2005a) were able to complement. Most of the differences in the ability of certain mutants to complement for Ydj1 or DnaJ/CbpA deficiency occurred between the prokaryote and eukaryote system, particularly in the region of Helix IV of the J-domain. In both full length Tcj2 and Tcj3, the equivalent mutations L57A, D59A and R63A are able to complement for Ydj1 deficiency in *S. cerevisiae*. However, of these three mutations, only Tcj3-AgtD60A/D59A and *A. tumefaciens* DnaJ R63A (Hennessy *et al.*, 2005a) were able to complement for DnaJ/CbpA deficiency in *E. coli*. The double mutations TD58,59AA and KR62,63AA both had no effect on the ability of *E. coli* DnaJ to functionally interact with DnaK (Genevaux *et al.*, 2002).

3.3.6) The completion of the fourth cysteine/glycine repeat of Tcj2 does not enhance its function, while the prevention of prenylation in Tcj2 negatively affects its function

It was found in chapter 2 (**section 2.3.8 and 2.3.9 and Figure 3.7**) that many Type I Hsp40s possess only a partial fourth cysteine/glycine repeat. The terminal glycine is often replaced by another amino acid, usually a lysine or an arginine (Martinez-Yamout *et al.*, 2000). It was hypothesized that the arginine or lysine adds a positive amino acid into an electronegative pocket in the structure, possibly enhancing structure or function. It is claimed that the substitution of only one glycine in the CxxCxGxG repeats for each zinc binding domain is tolerated in order to maintain the zinc binding domain structure (Martinez-Yamout *et al.*, 2000). Tcj2 does not have a lysine or an arginine at position 199 (the last glycine in the fourth CxxCxGxG repeat), but has a polar amino acid instead. To investigate the possible effect of this substitution, the glutamine present at this position in Tcj2 (aa 199) was swapped to the glycine found in *E. coli*/Agt DnaJ. This Tcj2 mutant was tested for its ability to substitute for Ydj1 in *S. cerevisiae* JJ160. This substitution was found to have no effect on the ability of Tcj2 to complement for Ydj1 deficiency in *S. cerevisiae* (**Figure 3.7**). Despite the inconclusive determination of expression of the Tcj2 Q199G mutant protein in *S. cerevisiae*, the ability of the protein to complement for Ydj1 as well as the wild type Tcj2 suggests that the protein is being expressed.




```
                10        20        30        40        50        60        70        80
                ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....
Tcj2  MVKETKFYDSLGVSPDASVDEIKRAYRRLALKYHPDKNKDPGSQEKFKEVSVAYECLSDPEKRSRYDQFGEKGVEME--S
Ydj1  MVKETKFYDILGVPVTATDVEIKKAYRKCALKYHPDKNPSEEAAEKFKEASAAYEILSDPEKRDIYDQFGEDGLSGAGGA
      ********* ***.  *:  ***:***: ********* .  : *****.*.*** *******. ******.*:.    :

                90       100       110       120       130       140       150       160
                ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....
Tcj2  GGIDP------TDIFASFFGG---SRARGEPKPKDIVHELPVSLEAFYTGKTIKLAITRDRLCPACNGSGSKVPNASVTC
Ydj1  GGFPGGGFGFGDDIFSQFFGAGGAQRPRGPQRGKDIKHEISASLEELYKGRTAKLALNKQILCKECEGRGGKK-GAVKKC
      **:          ***:.***.   .*.**  : *** **:..*** :*.*:* ***:.:: **  *:* *.*  .*  .*

               170       180       190       200       210       220       230       240
                ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....
Tcj2  KECDGRGVKLITRSIGPGFIQQMQVACPKCRGKGTDMREEDKCDSCRGQKIKKDKKIFEIFVEKGMHRGDNATFRGEGDQ
Ydj1  TSCNGQGIKFVTRQMGP-MIQRFQTECDVCHGTGDIIDPKDRCKSCNGKKVENERKILEVHVEPGMKDGQRIVFKGEADQ
      ..*:*:*:*::**.:** :**::*. *  *:*.*  :  :*:*.**.*::::::::**:*:.** **: *:. .*:**.**

               250       260       270       280       290       300       310       320
                ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....
Tcj2  IPGVRLSGDIIIIFEQKPHPVFTRKGDHLVMERTISLAEALTG--FTLNIKHLDDRDVSITSTGVVDPSKLWCVSREGMP
Ydj1  APDV-IPGDVVFIVSERPHKSFKRDGDDLVYEAEIDLLTAIAGGEFALEHVSGDWLKVGIVPGEVIAPGMRKVIEGKGMP
      *.* :.**:::*..::**  *.*.**.** *  *.*  *::*  *:*:    *  .*.*..  *: *.    :. :***

               330       340       350       360       370       380       390       400
                ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....
Tcj2  IPNTGGVERGDLVVKFHVVYPSAQSLQSNEISDLRKILHYPPQQSPPPSAMLCHLSETNIDLEKEAKRRRQTGGD--DDD
Ydj1  IPKYGGY--GNLIIKFTIKFPENHFTSEENLKKLEEILPPRIVPAIPKKATVDECVLADFDPAKYNRTRASRGGANYDSD
      **: **   *:*::** : :*. :  ..::::..*.:**      : * .* : .   :::*  *  : * . **   *.*

               410
                ....|....|....
Tcj2  DAPQGHTGATCTQQ
Ydj1  EEEQGGEGVQCASQ
      **  *. *:.*
```

Figure 3.7: *In vivo* complementation data showing the ability of Tcj2 C396S and Tcj2 Q199G mutants to functionally replace Ydj1 in *S. cerevisiae* JJ160. The temperatures used to incubate the yeast cells are indicated to the left of the panels. The dilutions of the yeast cultures are indicated to the right of the panels, starting from an A_{600nm} of 0.2 Absorbance units (10^0). The expression of these various mutants was checked using Western blotting (WB), with Tcj2 C-terminal peptide polyclonal antibody (1:5000 dilution) as primary antibody (Edkins *et al.*, 2004). The GE Healthcare, Amersham ECL advance kit horse radish peroxidase labeled anti-rabbit IgG (1:5000 dilution) was used as secondary antibody. The expression of Ydj1 (not detemined [ND]) was not assessed as there was no antibody for Ydj1 available.Ydj1 (Ydj1 on the figure) was used as a positive control, while the cells alone (JJ160) and vector alone (pKG6) were used as negative controls. The comparison of the full-length amino acid sequences of Tcj2 and YdJ1 is shown in the sequence alignment. The amino acids substituted for alanine or glutamine are highlighted in red. '*' = 100% identity and ':' or '.' = partial match across the sequences. The alignment was produced using Bioedit (Hall, 1999).

Tcj2 has been shown to have a C-terminal CaaX motif (**Chapter 2 section 2.3.9;** Tibbets *et al.*, 1998). In order to test its importance to the function of Tcj2 *in vivo*, the cysteine of the motif (position 396) was substituted for a serine. This removes the ability to farnesylate the protein, as the farnesyl group is bound to the cysteine of the CaaX motif (Caplan *et al.*, 1992a). The complementation data in **Figure 3.7** show subtle differences in the ability of C396S and Q199G/C396S mutants to complement for a lack of Ydj1 compared to the wild type Tcj2 protein at 30 °C and 32 °C incubation temperatures. This manifests as the smaller size of the individual colonies in the higher dilutions (10^{-2} and 10^{-3}) of the mutant spots and indicates the slower growth rate of these *S. cerevisiae* strains. A more distinct difference only occurred at 34 °C, where an almost complete lack of growth of cells expressing these mutants was observed. The C396S mutation appears to have less effect on Tcj2 function than many of the mutations shown in **Figure 3.4 Section 3.3.2.**

3.4) Discussion

As all of the J-domains compared in this chapter are from Type I Hsp40s, it is hardly surprising that they have a relatively conserved sequence identity. It has been shown that Type I and Type II Hsp40s have a higher level of conservation in their J-domain sequences than in Type III J-domains and the loop regions of Type I J-domains are shorter than in the other Hsp40 types (Hennessy *et al.*, 2000). This also confirms why the orientation of the amino acids studied in this chapter, are similar in each of the models (**Figure 3.3**).

It has been shown in this study and previously (Edkins *et al.*, 2004) that Tcj2 and Tcj3 are able to replace Ydj1 in *S. cerevisiae* JJ160 and hence interact with the yeast Ssa Hsp70s in a functional manner. It is well known that Ydj1 partners with the Ssa Hsp70 family (Caplan and Douglas, 1991; Caplan *et al.*, 1992b; Atencio and Yaffe, 1992; Becker *et al.*, 1996). The ability of Tcj2 and Tcj3 to complement for Ydj1 indicates that they are located to the same cellular compartment in *S. cerevisiae* as Ydj1. Assuming the localization signals in *T. cruzi* and *S. cerevisiae* are the same, it would indicate that both Tcj2 and Tcj3 have a cytosolic localization. The fact that Tcj2 was found to be a *T. cruzi* homologue to Ydj1 (Based on sequence identity – see **Chapter 2 section 2.3.7**) implies that Tcj2 is likely to be cytosolic as has been shown with Ydj1 (Caplan and Douglas, 1991). The Tcj2 and Tcj3 J-domains share

60.6% amino acid sequence identity, while the Ydj1 and Tcj2 J-domains have 70% amino acid identity. The J-domains of Ydj1 and Tcj3 share 66.2% amino acid identity.

The cell density effect that allows growth of non-complementing strains at low dilutions has also been observed in other studies to varying degrees (Genevaux *et al.*, 2002; Genevaux *et al.*, 2003; Hennessy *et al.*, 2005a; Johnson and Craig, 2001). In this study, the amino acid substitutions that abrogated complementation were possibly essential for a functional interaction with the partner Hsp70 in the test organism (*E. coli* or *S. cerevisiae*). It is also possible that these amino acids are important for the structural integrity of the J-domain. Those amino acids whose substitution did not affect complementation could have done so in two possible scenarios. Firstly, the amino acid at the position of the point mutation has no direct role in the functional interaction of the J-domain with Hsp70 or has no deleterious effect on the structural integrity of the J-domain. Secondly, the mutations that complement may affect the functional interaction with Hsp70 or structural integrity of the J-domain, but the effect is small enough that the Hsp40 and Hsp70 can still compensate for the mutation and complete their function. L57A, D59A and R63A of both studies of Tcj3 and Tcj2 in *S. cerevisiae* are examples of this. The partial complementation indicates a mutation that affects the functional interaction of Hsp40 with Hsp70, but is not as important for that interaction as the mutations that cause a total lack of complementation. K37A, L11A and Y8A of the Tcj2 study are examples of partial complementation. However, their partial effect is possibly due to a low level of expression.

The mutation of the histidine of the HPD motif to glutamine has almost universally eliminated Hsp40 functional interaction with its partner Hsp70 (Kelley and Georgopoulos, 1997; Genevaux *et al.*, 2002; Hennessy *et al.*, 2005a; Tsai and Douglas, 1996; Genevaux *et al.*, 2001; Fewell *et al.*, 2002; Stubdal *et al.*, 1997; Zalvide *et al.*, 1998). Therefore, despite the lack of determination of expression of Tcj3H34Q and the lack of confirmation of expression of Tcj3Agt-H34Q, it would have been very surprising if the proteins were able to complement for Ydj1 deficiency and DnaJ/CbpA deficiency respectively.

The differences in complementation between the Y8A and L11A mutations in Tcj3 and Tcj2 in the *S. cerevisiae* complementation studies is surprising due to the high level of amino acid identity between these two proteins and in comparison to Ydj1 (**Figure 3.3**). This could be due to the low levels of expression of some of these proteins relative to their counterparts in the other protein during the *S. cerevisiae* complementation studies. It may also be due to subtle differences regarding the amino acids of Helix II, Helix III or Helix IV, with which Y8 and L11 of the two proteins interact. These differences could result in different effects on the structural robustness of the respective J-domains containing these mutations relative to the wild type protein.

The ability of J-domains to be swapped from prokaryotes and eukaryotes and remain functional is a clear indication of the functionally conserved nature of this domain in Type I Hsp40s. In addition the conserved nature of most J-domains is indicated by the fact that many Type II (Hdj1 [Kelley and Georgopoulos, 1997]) and Type III, even of viral origin (SV40 T antigen [Fewell *et al.*, 2002]; mammalian papovavirus [Berjanskii *et al.*, 2000]) have also been substituted for Ydj1 or DnaJ J-domains. However, there are exceptions such as Human ERj1 (also known as DnaJC1 [Qiu *et al.*, 2006]) whose J-domain was unable to replace *A. tumefaciens* DnaJ J-domain to reverse the thermosensitivity of *E. coli* OD259 (Nicoll *et al.*, 2007). In addition, Ydj1 and *E. coli* DnaJ J-domains cannot functionally replace the J-domain of SV40 T Antigen (Sullivan *et al.*, 2000). It has been proposed that those Hsp40s whose J-domains are unable to functionally replace the J-domains of other Hsp40s have become specialized for specific functions (Nicoll *et al.*, 2007). The inablilty of Tcj3 to functionally replace *E. coli* DnaJ/CbpA indicates that the remainder of the Hsp40 structure is less interchangeable.

It is intriguing to note that despite the interchangeability of the majority of Hsp40 J-domains tested, different amino acids were highlighted as important for the interaction of Tcj3 full length protein with Ssa1 and the Tcj3Agt with DnaK. As amino acids that were substituted in this chapter are conserved across all of the J-domains studied (**Figure 3.3**), this tends to indicate that there are subtle differences in the amino acid requirements of Ssa1 and DnaK for interaction with the J-domains of their partner Hsp40s. Other differences between DnaK and

eukaryotic Hsp70s have been observed. For example, the nucleotide binding domain (NBD) and substrate binding domain (SBD) of *Bos Taurus* Hsc70 interact in both the ATP and ADP bound state of the NBD (Jiang *et al.*, 2007), while *E. coli* DnaK has been reported to interact only in the ATP bound state (Swain *et al.*, 2007).

The J-domains of the Type II Hsp40s Pfj4 (*Plasmodium falciparum* J protein 4) and the human Hsj1 were both able to functionally replace *E. coli* DnaJ J-domain with a D59A mutation, as was the case for R63A in the J-domains of Pfj4 and Pfj1 (Nicoll *et al.*, 2007). Therefore there does not appear to be a clear distinction between the amino acids required for the interaction of J-domains with prokaryotic DnaK or the eukaryotic Hsp70. Another possible explanation for the differences in the complementation capability of these mutations could be that the point mutations affect other J-domain amino acids in their vicinity (that are not necessarily the same for each J-domain) in the 3 dimensional structure of the J-domain. This could result in different abilities of J-domains to functionally interact with the same Hsp70 (e.g. DnaK) and compensate for certain point mutations. If this is the case, it does lead to certain skepticism regarding the determination of amino acids important for structure and function in a J-domain using a heterologous *in vivo* system. It is therefore preferable to use a system as similar to the wild type system as possible. Data in these heterologous systems should be interpreted carefully.

The ability to complement does not always equate to a lack of interaction. Despite the ability of R63A to complement in *E. coli* DnaJ (Genevaux *et al.*, 2002), R63A displayed a lowered interaction with DnaK relative to wild type DnaJ when measured by surface plasmon resonance spectroscopy (Laufen *et al.*, 1999). Therefore complementation data does not supply all the information with regard to the physical interaction of Hsp40 and Hsp70. It should therefore be remembered that *in vivo* complementation studies take place in a complex cellular system and that there may be compensatory factors that can buffer the mutations effect.

The structure of a stabilized interaction between Hsc70 and the Bovine Auxilin J-domain, using disulphide crosslinking of two cysteines substituted for an amino acid in Hsc70 (D876) and the Auxilin J-domain (R171), has been determined (Jiang *et al.*, 2007). Of the amino acids

determined to be important for the interaction of *Bos taurus* auxilin with Hsc70 (H784, D896, F891, H874, H875; M892; M889; E884; Q885) only H874 and F891 (Jiang *et al.*, 2007) corresponded to the amino acids examined in this study (corresponding to H34 and F47 in Tcj2). **Figure 3.8** Shows a multiple sequence alignment comparing Tcj2, Tcj3, Ydj1, *A. tumefaciens* DnaJ and *E. coli* DnaJ J-domains to the J-domain of auxilin. The only amino acid residues examined in this study that were proposed by Jiang and coworkers (2007) to be important for the physical interaction between *Bos taurus* Hsc70 and *Bos taurus* auxilin were the F47 and H34. Jiang and colleagues (2007) appear to place more emphasis on the importance of the loop between Helix II and III and Helix III for the interaction with Hsp70 than Helix II. This is in contrast to Greene and colleagues (1998) who emphasized the importance of Helix II and the HPD motif in the interaction of *E. coli* DnaJ and DnaK (Greene *et al.*, 1998). These differences could possibly be due to *Bos taurus* auxilin being a Type III Hsp40 (Jiang *et al.*, 2003), while *E. coli* DnaJ is a Type I Hsp40 (Genevaux *et al.*, 2007). The auxilin J-domain structure also has a structure divergent from the J-domain of *E. coli* DnaJ extended loop between Helices I and II and is lacking a Helix IV (Jiang *et al.*, 2003; Hennessy *et al.*, 2005b; Pellechia *et al.*, 1996). As Tcj3 and Tcj2 are both Type I Hsp40s, it is more likely that their J-domains would follow the *E. coli* DnaJ pattern of interaction with its partner Hsp70.

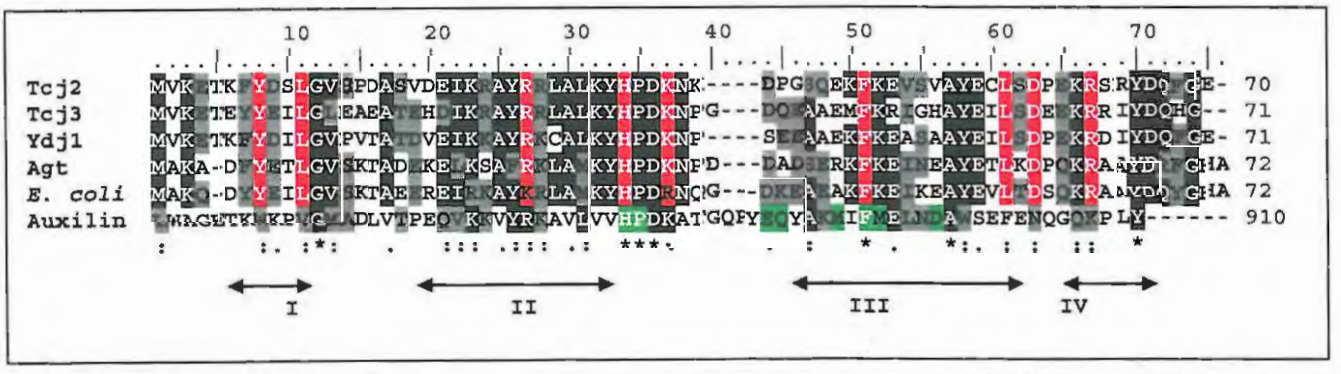


Figure 3.8: Multiple sequence alignment comparing the J-domains of Tcj2, Tcj3, Ydj1, *A. tumefaciens* DnaJ, *E. coli* DnaJ and *Bos taurus* auxilin. The amino acids that were investigated for their importance in the interaction with Hsp70 are highlighted in red, while the amino acids proposed by Jiang and colleagues to be important for *Bos taurus* auxilin interaction with its partner Hsc70 are highlighted in green (Jiang *et al.*, 2007). The multiple sequence alignment was produced using Bioedit (Hall, 1999). Identical amino acids at a position are indicated with a black background, while the grey background indicates similar amino acids. The * below the alignment indicate that amino acids are 100% conserved at this position in the alignment. ":" indicates that the amino acids are slightly less conserved at that position, while "." indicates even less conservation. The arrows below the figure indicate the Helix locations, with the exception of *Bos taurus* auxilin.

Although *in vivo* complementation does not provide conclusive evidence of physical interaction in the case of a failure to complement, data in this chapter indicates that amino acids of Helix II (R27A), the Helix II/III loop (H34Q, K37A) and Helix III (F47A) are important for the functional interaction of Tcj2 and Tcj3 J-domains with their partner Hsp70s in the heterologous complementation systems. According to the homology models of Tcj2 and Tcj3 J-domains (**Figure 3.3**), these amino acids all project forwards from the Helices II and III into the aqueous environment and hence are ideally located for physical interaction with Hsp70. The data for the R63A/R64A mutation, supports the hypothesis by Hennessy and coworkers (2005b; 2000) that Helix IV of the J-domain supports the specificity of interaction between Hsp40s and their partner Hsp70, discerning with which Hsp70 an Hsp40 is able to interact.

Due to the positions and orientation of Y8, L11 and L57 in the homology models in **Figure 3.3** it is most likely that these amino acids would affect functional interaction indirectly through disruption of the structural integrity of the J-domain; as was proposed by Hennessy and coworkers (2005a). Where complementation is still able to occur to some extent, this indicates that the partner Hsp70 is still able to interact with the J-domain to some extent, despite the structural disruption.

A similar effect to the C396S mutation has been observed in the equivalent mutation of Ydj1 (C406S) (Caplan *et al.*, 1992a; Johnson and Craig, 2001). However, the lack of complementation in the experiments reported by Johnson and Craig occurred at 30 °C due to the experiment being carried out in a yeast strain that was Ydj1 deficient and contained a truncated version of Sis1 (amino acids 1-121). The greater temperature sensitivity of the yeast cells in that experiment could be attributed to the overlapping function of Ydj1 and Sis1 (Johnson and Craig, 2001). The fact that Tcj2 has a CaaX motif (Tibbetts *et al.*, 1998) and that C396S behaves in a similar way to the Ydj1 C406S, which is known to be farnesylated, suggests that Tcj2 is also prenylated. Hdj2 and Ydj1 have both been reported to function in the translocation of proteins into the endoplasmic reticulum or mitochondria (Caplan and Douglas, 1991; Atencio and Yaffe, 1992; Kanazawa *et al.*, 1997). The membrane association provided by prenylation could be important for this process and could indicate that Tcj2 is also involved

in protein translocation across membranes, due to the likelihood that it is prenylated. Yeast cells showed an increase in the amount of membrane associated Ydj1 after heat shock at 37 °C (Caplan *et al.*, 1992a). Hdj2 (DnaJA1) has also shown a change in the sub-cellular localization of the protein from predominantly in the cytoplasm (with small amounts associated with the nuclear envelope) to a greater proportion of the protein being associated with the golgi apparatus, nuclear membrane and nucleolus during long periods of heat shock (Davis *et al.*, 1998). It has been proposed that prenylation of these proteins is at least partially responsible for this differential localization and could be due to the stabilizing of farnesyl and membrane lipid interactions at higher temperatures (Caplan *et al.*, 1992a; Davis *et al.*, 1998). This could possibly explain why Tcj2 C396S and Ydj1 C406S are unable to complement for wt Ydj1 deficiency in *S. cerevisiae* during heat stressing temperatures. The role of prenylation during heat stress is further supported by the finding that *E. coli* DnaJ (a prokaryotic homologue of Ydj1), produced without a CaaX box prenylation motif, was able to substitute for Ydj1 in *S. cerevisiae* at growth temperatures of 30 °C but not 37 °C (Caplan *et al.*, 1992a).

Recent data has shown that the Zinc binding domain 1 is less important for the suppression of aggregation function of the Zinc Finger-like Region during heat stress than Zinc binding domain 2 (Fan *et al.*, 2005). Q199 forms part of the structure of ZBDI and is therefore less likely to cause as much disruption to the suppression of aggregation function of Tcj2 that was tested through exposing *S. cerevisiae* JJ160 expressing Tcj2Q199G to heat stress. This is a possible explanation for no visible effect being observed. However, it is possible that changing the glutamine back to the glycine, that many Type I Hsp40s contain at the equivalent position of Q199 of Tcj2 (Martinez-Yamout *et al.*, 2000), is insufficient to cause a noticeable disruption of other zinc finger like region functions such as protein folding and coordination with hsp70 that require both zinc binding domains (Fan *et al.*, 2005). It would be interesting to replace Q199 with a strongly negatively charged or small hydrophobic amino acid (i.e. alanine or valine) to disrupt the interaction with the electronegative patch, described by Martinez-Yamout *et al.* (2000) and test the ability of Tcj2 to functionally replace Ydj1 in *S. cerevisiae*.

Chapter 4:

The *in vitro* Characterisation of the Interaction between Tcj2 and Hsp70

CHAPTER 4: The *in vitro* characterization of the interaction between Tcj2 and Hsp70

4.1) Introduction

It is preferable to study many aspects of proteins while they occupy their natural environment (i.e. *in vivo*) as much as possible. This, however, is not easily achieved due to the complicated mix of molecules and macromolecules in the cell, which complicates detection. *In vitro* assays, that utilize simplified systems of molecules, are therefore often required as a compromise to enhance the understanding of a given protein. Despite this simplification, *in vitro* assays can be a valuable complement to *in vivo* work, provided it is remembered that this system is not identical to that of the cell.

Intrinsic to the molecular chaperone function of Hsp70 proteins is their ability to hydrolyse ATP (Mayer *et al.*, 2000b) and this characteristic indicates a functional Hsp70 *in vitro*. This activity can be monitored by *in vitro* ATP hydrolysis assays by monitoring the amount of inorganic phosphate released from the conversion of ATP to ADP or the determination of the ratio of ATP to ADP in each solution (Mayer *et al.*, 2003; Mayer *et al.*, 2000c). The detection of phosphate is the most common, and may be detected by the separation of radiolabeled ATP from the ^{32}P released (Blond-Elguindi *et al.*, 1993) or the generation of colorimetric complexes with the inorganic phosphate (Chifflet *et al.*, 1988; Lanzetta *et al.*, 1979). Some Hsp40s are able to act as co-chaperones for Hsp70 by delivering peptide substrates to them and by stimulating their ATPase activity. This stimulation of the Hsp70 ATPase domain by Hsp40s can be used as an indirect method to detect a functional interaction between Hsp70/Hsp40 and therefore the functional integrity of both proteins *in vitro*. However, the lack of a detected functional interaction between Hsp70 and Hsp40 mutants (such as those with a H34Q/H33Q mutation in the J-domain), through the measurement of the stimulation of Hsp70 ATPase (**Chapter 4 section 4.3.4**; Hennessy *et al.*, 2005b; Tsai and Douglas, 1996; Wall *et al.*, 1994) or *in vivo* complementation interactions (**Chapter 3**; Hennessy *et al.*, 2005a; Genevaux *et al.*, 2002; Tsai and Douglas, 1996; Stubdal *et al.*, 1997) does not necessarily imply a complete lack of physical interaction between a given Hsp40 and its partner Hsp70.

The general chaperone cycle of Hsp70 has been outlined (**Chapter 1 section 1.8.1**). In its ATP-bound state, the Hsp70 substrate binding domain is in an open state exhibiting high association and dissociation rates for unfolded protein subtrates. In contrast, ATP hydrolysis (the ADP-bound state) closes the substrate binding domain and produces low association and dissociation rates for unfolded peptide substrates (Schmid *et al.*, 1994; Fan *et al.*, 2003). Little work has been done on the *in vitro* characterization of trypanosomal Hsp70 proteins. The best characterised of these proteins is the stress inducible *Trypanosoma cruzi* Hsp70 (TcHsp70) (Krautz *et al.*, 1998; Olson *et al.*, 1994; Edkins *et al.*, 2004). As it is localized to the cytoplasm it is the most likely partner of those *T. cruzi* Hsp40s, such as Tcj2, that are also predicted to localise to the cytoplasm (**Chapter 2 section 2.3.4**).

Piezoelectricity is a phenomenon whereby materials generate an electric potential in response to a mechanical stress (Curie and Curie, 1880; Ward and Buttry, 1990). In the context of crystal lattices, this occurs when separated but symmetrically distributed charges (resulting in an electrically neutral crystal) are shifted by mechanical stress, causing an imbalance in the charges and therefore an electrical potential. This process can be reversed to produce the reverse piezoelectric effect, whereby the application of an electric field to a material causes mechanical distortion of that crystal (Ward and Buttry, 1990).

The reverse piezoelectric effect has been applied in the quartz crystal microbalance (QCM), in which the application of an alternating current produces mechanical oscillations within a quartz crystal (Ward and Buttry, 1990). QCM applications most commonly make use of a thin circular wafer of AT-cut α-quartz (approximately 0.3mm thick and 14mm diameter) sandwiched between two gold electrodes (10 to 300nm thick) to produce the quartz crystal resonator. The AT-cut is obtained by cutting the quartz wafers at approximately 35° from the *z*-axis. This results in the movement of the resonant oscillations being parallel to the faces of the circular wafer (also referred to as transverse shear or thickness shear mode), when an alternating electric field is applied across these faces with the electrodes (Ward and Buttry, 1990) (**Figure 4.1**). The piezoelectric strain

and vibration of the quartz crystal is effectively restricted to the area between the two electrodes in a process referred to as energy trapping. This enables the mounting of the crystal at its edges without greatly affecting measurements (Ward and Buttry, 1990). The frequency of the quartz crystal oscillation is dependent on the thickness of the circular quartz wafer. However, as this remains constant during operation of the oscillator the addition of any mass to one of the electrodes of the crystal can be detected as a change in the oscillation frequency of the crystal (Sauerbrey, 1959). Sensitivity of certain QCM systems has been reported to be as little as 1ng/cm^2 (Rodahl *et al.*, 1996).

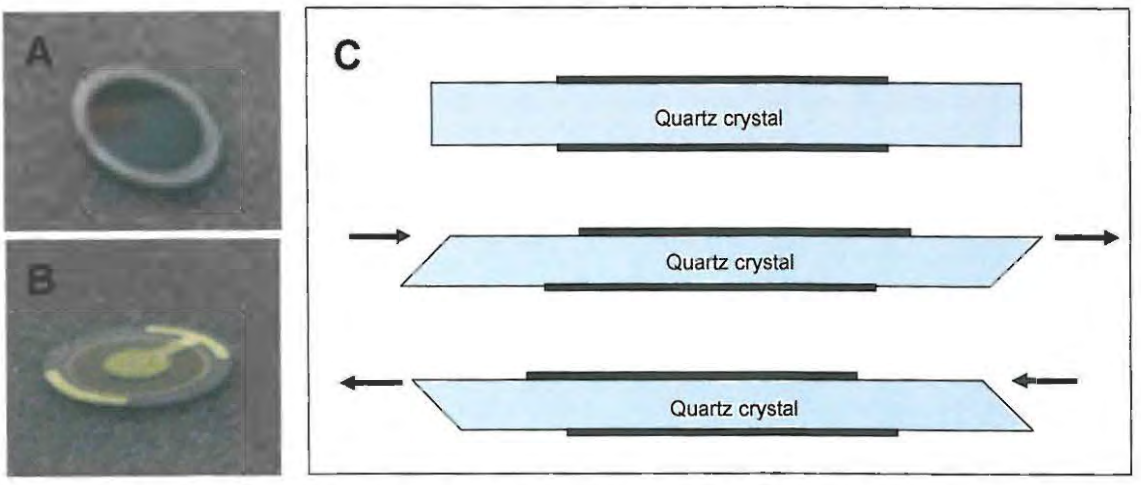


Figure 4.1: Quartz crystal resonator structure and graphic description of the thickness shear wave oscillation.
A) The liquid interface side of the SiO_2 coated quartz crystal resonator. **B)** The opposite side of the quartz crystal resonator. **C)** Schematic representation of the side view of a circular quartz crystal resonator. The quartz crystal resonator is sandwiched between two gold electrodes (thick black bars). The application of an alternating current to these electrodes generates an electric field producing a mechanical distortion of the quartz crystal. As the alternating current switches the polarity of the electrodes, the electric field polarity reverses and mechanically distorts the crystal in the opposite direction.

A number of overtones of the fundamental oscillation frequency can be measured simultaneously to increase or decrease the sensitivity of detection (Ward and Buttry, 1990). The odd overtone numbers (3rd, 5th, 7th, 9th and so on) are the most commonly used. The third overtone records a frequency three times the fundamental oscillation frequency, while the fifth overtone records a frequency five times the fundamental oscillation frequency of the crystal resonator (i.e. if the fundamental frequency is 5 mHz, the third overtone oscillation frequency would be 15 mHz and the 5th overtone oscillation frequency would be 25 mHz) (Höök *et al.*, 2001; Su *et al.*, 2005).

The quartz piezoelectric oscillations are known to be stable in a vacuum, air (Rodahl *et al.*, 1995) and liquid (Nomura and Hattori, 1980; Nomura and Okuhara, 1982). However, the density and viscosity of the medium in which the crystal oscillates is known to introduce dissipative losses in the oscillating system.This is due to the shear mode oscillation of the crystal inducing a shear wave in the surrounding medium. This damping effect is significantly higher for resonator oscillations in liquids relative to gases due to an increase in viscosity, which causes frictional losses, within the liquid relative to the gas (Rodahl *et al.*, 1995). In a similar manner, the addition of mass to the surface of an electrode of the quartz oscillator is able to introduce dissipative losses into the oscillating system (Rodahl *et al.*, 1995; Rodahl *et al.*, 1996). The absolute dissipation losses of an oscillating quartz crystal system can be measured by switching off the alternating electric current driving the system and measuring the voltage over the crystal during the decay oscillations as the stored energy in the system dissipates. The rate of decay determines the amount of dissipation in the system. A technique of simultaneous measurement of the oscillating frequency and oscillating energy dissipation in the system has been developed by Rodahl and coworkers (1995).

If the only change to a resonator system is the adsorption of a molecular film to the surface of the crystal electrode, the change in the frequency and energy dissipation of this crystal resonator can be attributed to the adsorption of the molecular film. The amount of dissipation attributed to the adsorbed layer can be used to determine the nature of the adsorbed layer. A film of small rigid molecules adsorbed to the surface of a quartz resonator will result in a lower contribution to the dissipation of oscillation than a film of larger molecules or flexible molecules (Höök *et al.*, 1998a; Höök *et al.*, 1998b). Protein conformational/structural changes can therefore be observed using the dissipation factor measurements (Höök *et al.*, 1998b).

Surface plasmon resonance (SPR) spectroscopy is an optically based technique for detection of molecular binding events using changes in the refractive index of thin films assembled on a noble-metal surface. In contrast to macromolecular binding studies using SPR spectroscopy, the QCM-D often overestimates the collective mass of these

molecules that are bound to the solid surface (Su *et al.*, 2005; Reimhult *et al.*, 2004). The frequency of oscillation shifts obtained using the QCM-D system results from the total mass of molecules coupled to the movement of the crystal resonator. This includes the mass of the protein and the mass of the water that is in the hydration shell of solvation of the protein and water that may be trapped between these biomacromolecules (Höök *et al.*, 2001; Höök *et al.*, 2002). In addition, the Sauerbrey relation (Sauerbrey, 1959) is only valid for evenly distributed, rigid and thin film layers of adsorbed material (<25 nm thickness) such as lipid bilayers (Keller *et al.*, 2000). A thicker or viscoelastic film results in the change in frequency not being directly proportional to a change in mass of the resonator as the oscillation of the resonator is not completely transferred to these layers (Ward and Buttry, 1990; Lucklum *et al.*, 1999; Rodahl *et al.*, 1997; Bandey *et al.*, 1997). Such layers require further modeling using a Voight-based viscoelastic film model to determine amounts of mass bound (Voinova *et al.*, 1999; Bandey *et al.*, 1997).

In order to conduct measurements of the interaction between two molecules using QCM-D, one of these molecules needs to be immobilized onto the sensor crystal surface. Although proteins are known to bind directly to solid surfaces at solid-liquid interfaces *in vitro* (Andrade and Hlady, 1986; Norde, 1995; Ozeki *et al.*, 2009; Dolatshahi-Pirouz *et al.*, 2008), it is preferable to have proteins soluble in the aqueous liquid medium attached to the surface in such a way that most of the protein surface is accessible to its potential partner in solution. This can be achieved through the functionalisation of phospholipids to attach a nitriloacetic acid (NTA) group, which will allow the binding of a histidine tagged protein through metal chelation in the same principle that is used for histidine tagged protein purification (Altin *et al.*, 2001; Dorn *et al.*, 1998; Kienberger *et al.*, 2001). Other functionalized lipid systems that can be used for immobilization are iminodiacetic acid (Pack *et al.*, 1997; Vogel *et al.*, 1997), maleimide thiol (Svedhem *et al.*, 2003; Elliot *et al.*, 2000) and biotin streptavidin (Fant *et al.*, 2002; Reviakine *et al.*, 2001; Larsson *et al.*, 2003). An advantage of the metal chelating systems of protein immobilization is that they are reversible through the addition of ethylenediamine tetraacetic acid (EDTA) or adjusting the pH (Wong *et al.*, 1991).

Two common ways of producing a supported phospholipid bilayer on a hydrophilic solid surface are 1) The Langmuir-Blodgett transfer technique (Naumann *et al.*, 2002) and the adsorption and spontaneous rupture of lipid vesicles (Keller and Kasemo, 1998; Glasmäster *et al.*, 2002; Richter *et al.*, 2003; Reimhult *et al.*, 2006). This second method is preferable for QCM-D as it allows the lipid layer to be formed on the SiO_2 coated quartz crystal in the flow cell, enabling the observance of the formation of a complete lipid bilayer in real time.

Phospholipid vesicles react very differently to different surfaces. Alkane thiolated gold is a hydrophobic surface that allows for the formation of a monolayer of phosphatidylcholine lipids with the hydrophobic section of the molecule attached to the gold surface. An oxidized gold surface is hydrophilic and results in the binding of intact bilipid layer vesicles, while SiO_2, another hydrophilic surface, results in the formation of an intact phospholipid bilayer supported on the SiO_2 surface (Keller and Kasemo, 1998). The formation of a phospholipid bilayer on a SiO_2 surface involves three stages in which (1) intact phospholipid bilayer vesicles are adsorbed to the surface of the SiO_2, (2) these vesicles break open, when a sufficiently high surface coverage is reached, to form phospholipid bilayer regions interspersed with intact phospholipid vesicles. (3) a complete phospholipid bilayer is spread over the whole surface (Keller and Kasemo, 1998; Keller *et al.*, 2000; Zhdanov *et al.*, 2000). An alternate mechanism has been reported in which the lipid vesicles rupture on contact with the surface irrespective of vesicle concentration (Richter *et al.*, 2003; Dimitrievski *et al.*, 2004). The nanotopography of the surface to which the lipid layer is being adsorbed has a significant influence on the kinetics and mechanism of lipid bilayer formation (Pfeiffer *et al.*, 2008). Some of the factors, other than the surface of adsorption, that affect the formation of the supported lipid bilayer are: 1) temperature (Reimhult *et al.*, 2002; Reimhult *et al.*, 2003; Seantier *et al.*, 2005), 2) buffer composition, 3) buffer pH and 4) lipid vesicle concentration and lipid vesicle constituents. A lower concentration of lipid vesicles is required for the formation of a supported phospholipid bilayer (SPB) at temperatures close to 30 °C as opposed to 5 °C, i.e. the rate of formation of this SPB is increased at the higher temperatures (Reimhult *et al.*, 2002). The presence of salts such as sodium

chloride and calcium chloride accelerate the formation of a SPB through the pathway of initial intact vesicle adsorption and rupture. The pH of the buffer has the greatest effect on the SPB formation by zwitterionic lipids. Below a pH of 7.4, these lipids tend to rupture on contact with a SiO_2 surface, while above this pH a critical density of lipid vesicles adsorbed to the surface appears to be required (Seantier *et al.*, 2005). However, a buffer pH of approximately 7 is preferable for simulation of *in vivo* conditions.

The use of a supported lipid bilayer in these experiments has a number of advantages over other protein immobilization strategies. Lipid bilayers are more representative of cellular boundaries and are therefore more closely resemble a physiological scenario in contrast to solid liquid interfaces that are not covered in a lipid bilayer. A lipid bilayer film over a solid surface prevents the spontaneous and irreversible adsorption of proteins to these solid surfaces. Proteins soluble in an aqueous medium are known to not adsorb to lipid bilayer films (Gläsmaster *et al.*, 2002). The lipid layer therefore effectively acts as a blocking agent to prevent any of the proteins binding non-specifically to the surface of the crystal and affecting measurement of the interaction between the proteins in question. In addition, the NTA group modified lipids in a lipid bilayer allow the tethering of a histidine tagged protein to the crystal surface, while still allowing an almost complete surface area of the protein to be exposed to the liquid medium for interaction with potential partner proteins. This also allows for the protein to maintain its legitimate aqueous environment structure, which is known to distort during adsorption to a solid surface (Andrade and Hlady, 1986; Norde, 1995).

QCMD was used for this study of the interaction between Hsp40 and Hsp70 as we hoped to obtain more information regarding the conformational changes associated with the interaction between Hsp70 and Hsp40, and SPR was believed to be less capable of measuring this. In addition, the protein immobilization strategies available to the QCMD system were more representative of the physiological conditions (e.g. tethered to a lipid bilayer). It is our hypothesis that less gentle protein immobilization strategies can hinder the dynamic interactions between Hsp40and Hsp70.

The aims of this chapter were to characterize aspects of the interaction of Tcj2 with partner Hsp70 *in vitro*:

1) Test the ability of His-Tcj2 to stimulate the ATPase activity of multiple Hsp70s.
2) Investigate the binding of His-Tcj2 and His-Tcj2(H34Q) to Hsp70s using the QCM-D.

4.2) Materials and Methods

4.2.1) Heterologous expression of His-Tcj2 and His-Tcj2 mutant proteins for purification

Tcj2 with a (6x)Histidine tag attached to the N-terminus was expressed from pET28aTcj2 (**FIGURE 4.2**), in *Escherichia coli* BL21(DE3) cells. pET28aTcj2 was a generous gift from Dr David M. Engman (Northwestern University Medical School, Chicago, USA).

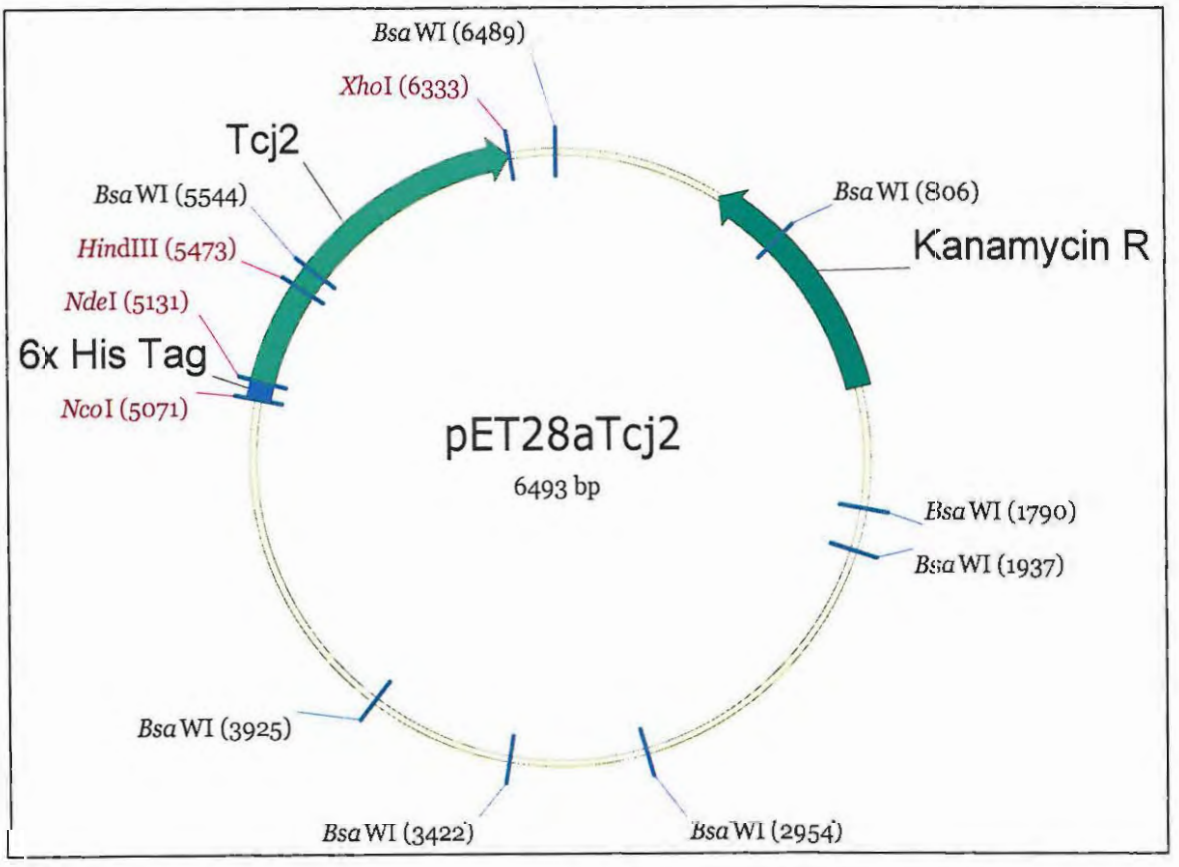


Figure 4.2: Plasmid map of the pET28aTcj2 expression vector.
The Tcj2 and kanamycin resistance gene coding sequences are highlighted with green arrows. The 6x – histidine tag attached to the N-terminus of Tcj2 is shown in blue.

4.2.2) Site-Directed mutagenesis of His Tcj2

Site directed mutagenesis of His-Tcj2 in pET28aTcj2 was performed as per **Appendix C.1.16.** Histidine 34 of Tcj2 was mutated to glutamine using the same primers used for this mutation in **Chapter 3** (Tcj2H34QF and Tcj2H34QR) **(Appendix A Table A3).**

4.2.3) Purification of His-Tcj2 and His-Tcj2 mutants

The pET28aTcj2 expression construct was transformed into *E. coli* BL21(DE3) and a single colony was inoculated into 25 ml of YT broth containing 50 µg/ml kanamycin and incubated at 37 °C with shaking (180 RPM) overnight. The starter culture was diluted 1:10 with YT broth containing the same concentration of kanamycin and incubated at 30 °C until the $A_{600\ nm}$ was between 0.8 to 1.0 absorbance units. IPTG was then added to the culture to a final concentration of 1 mM prior to incubation at 30 °C with shaking for 6 hours. A further 5 mg kanamycin was added to each 250 ml volume of culture at 2 hourly intervals to improve selection for cells containing pET28aTcj2 and prevent the survival of untransformed cells due to antibiotic resistance factors secreted by other cells in the population.

Two litres of bacterial culture were harvested for each protein purification by means of centrifugation in a Beckman JA-10 centrifuge tubes centrifuged at 5000 RPM for 15 minutes at 4 °C. The cell pellets were resuspended in a total volume of 20 ml of degassed lysis buffer (40mM Tris-HCl (pH8.0); 100mM NaCl; 10mM Imidazole; 0.1% triton X-100; lysozyme (1µg/ml); EDTA free Protease inhibitor cocktail (Roche)) and incubated at room temperature for 15 minutes before freezing at -80 °C overnight. The lysate was subsequently thawed at room temperature and the contents of each tube sonicated at 40 Hz for 30 seconds at 4 °C before being pelleted for 25 min at 16000 x*g* in a refrigerated (4°C) microcentrifuge (Eppendorf, Germany). The supernatant was collected.

Purification of the protein from the clarified lysate was achieved using nickel affinity purification with the His trap 5ml nickel sepharose columns (HiTrap™ 5ml column, GE Healthcare, USA) connected to an FPLC (AKTÄ-Basic 900, Amersham Biosciences, USA). The columns were rinsed with 5 column volumes of distilled water, and then with

3 column volumes of lysis buffer (excluding the lysozyme and triton X100), all with a flow rate of 1 ml/min. The column was disconnected from the FPLC to apply the clarified lysate through a syringe connected to a 0.22 µm filter (to prevent clogging of the Histrap column) at a flow rate of approximately 1 ml/min. The column was subsequently reattached to the FPLC and washed with 10 to 15 column volumes of wash buffer (40 mM Tris-HCl pH 8.0; 100mM NaCl; 10 mM Imidazole). The column was washed with an increasing gradient of Native Elution buffer (40 mM Tris-HCl pH 8.0; 100 mM NaCl; 1M imidazole) up to 40 % Native Elution buffer and 60% wash buffer over a period of 10 min. This was done in order to wash off any non-specific binding of proteins weakly bound to the nickel on the nickel sepharose. Elution of His-Tcj2 was effected by increasing the percentage of Elution buffer in the mixture to 100% over a period of six minutes again with a flow rate of 1 ml/min.

The fractions containing the Eluted protein were identified by placing 10 µl of each fraction into individual wells of a 96-well microtitre plate and adding 200 µl of Bradfords Reagent (Sigma) to each well. After an incubation for 5 minutes at room temperature, the absorbance at 595nm was read in a microtitre plate reader (Powerwave$_x$, Biotek Instruments Inc., USA) to determine the fractions containing protein.

4.2.4) Removal of Imidazole from solutions of purified proteins

The fractions containing protein (approximately 10 ml to 15 ml) were pooled together into dialysis tubing (Snakeskin #68100; molecular weight cut off of 10 kDa; Pierce Chemical Company, USA) and dialysed against 1 litre of buffer exchange buffer (100mM Tris-HCl pH 8.0; 100mM NaCl) with stirring, which was changed 5 times at approximately 3 hourly intervals. After this removal of the imidazole the protein solution was checked for purity using SDS-PAGE (12% resolving gel and 5 % stacking gel), quantified and its degree of aggregation determined. Typical yields of the purified protein using this method were 3 mg to 6 mg of protein per 2 litres of cell culture.

4.2.5) The cloning, expression and purification of pQETchsp70

Trypanosoma cruzi Hsp70 (TcHsp70) was amplified by PCR (**Appendix C.1.15**) using primers Tc70F and Tc70R (**Appendix A2**) from pET14bTchsp70 (a kind donation from Dr David Engman, Northwestern University Medical School, Chicago, USA). The product DNA was resolved on an agarose gel (**Appendix C.1.9**), gel purified (**Appendix C.1.10**) and ligated into pGEM-T Easy® (**Appendix C.1.14**). The pGEM-T Easy® plasmid containing the Tchsp70 coding sequence was digested with *Bam*HI and *Hind*III to release the fragment, which was subsequently ligated (**Appendix C.1.13**) into pQE30 using these restriction sites. (pQETchsp70 plasmid map **Appendix 4.1**). His-Tchsp70 was expressed in *E. coli* XL1blue (**Appendix A2 Table A1**) and purified using the same method outlined in **Section 4.2.3**, except that 0.1% Polyethyleneimine was added to the lysis buffer.

4.2.6) Protein Quantification using Bradfords Assay

Protein solutions were quantified using a variation of the Bradford's Assay (Bradford, 1976). Protein solutions (10 µl) and Bovine Serum Albumin (BSA) standard solutions were added to individual wells in a 96-well microtitre plate and 200 µl of Bradfords Reagent added to each well prior to incubating at room temperature for 5 minutes. The absorbance of each well at 595 nm was determined using a microtitre plate reader (Powerwave$_x$, Biotek Instruments Inc., USA). The concentration of the purified protein solutions was determined by comparing the absorbance$_{595nm}$ for the test solution to the regression line produced by the BSA standard curve. Protein absorbance readings were performed using 3 to 4 replicates that were averaged before the comparison with the BSA standard curve.

4.2.7) Analysis of the extent of aggregation of purified proteins by means of light scattering in a spectrofluorimeter

Small and large aggregates within the solutions of the purified proteins were detected using the light scattering that they produce within a fluorescence spectrophotometer (Buchner *et al.*, 1998). Protein solutions were diluted to 10 µg/mL and 2 mL of these solutions were used for measurements. These solutions were excited at 360 nm (for the

detection of small protein aggregates) and 500 nm (for the detection of large protein aggregates). The light emission from the solutions was measured over the spectrum of 300 nm to 600 nm. A native solution of 10 µg/mL BSA was used as a negative control, while heat denatured (by boiling for 15 minutes) 10 µg/mL BSA was used as a positive control for the detection of protein aggregates.

4.2.8) ATPase assay protocol

The ATPase enzyme reactions were prepared in a final volume of 1000 µl containing 10 mM Hepes (pH7.4), 10 mM $MgCl_2$, 20 mM KCl, 0.5 mM DTT. Hsp70s and Hsp40s were added to a final concentration of 0.4 µM or 0.08 µM. The reactions were equilibrated to 37 °C prior to initiating the ATPase reaction by the addition of ATP to a final concentration of 600 µM. Samples (50 µl) were taken from the reaction at 30 minute intervals for 300 minutes after the start of the reaction. These samples were added to 50 µl of 10 % SDS (in phosphate free water) in a flat bottomed microtitre plate to stop the reaction at these various time points. Three replicate reactions were set up for each test parameter and a sample was taken from each at each time point. A standard curve to determine phosphate concentration in the samples was set up using KH_2PO_4.

The levels of inorganic phosphate in the samples and standard curve was determined using a colourimetric based assay adapted from work published by Chifflet and coworkers (1988). A 1% solution of ammonium molybdate dissolved in 1M HCl solution (50 µl) was added to the microtitre plate wells containing the mixture of the 50 µl sample and 50 µl 10% SDS. Ascorbic acid (6%; 50 µl) in phosphate free water was subsequently added, resulting in a blue colour in the presence of inorganic phosphate. The absorbance of the solutions in the microtitre plate at 850 nm were determined using a microtitre plate reader (Powerwave$_x$, Biotek Instruments Inc., USA) microtitre plate reader. The ATPase assay was performed three times, each with a separate batch of purified proteins. The ATPase activites for each reaction were reported as the nmol of phosphate released per min per mg of Hsp70 (nmol/min/mg of Hsp70). A control containing ATP, but no Hsp70 or Hsp40 was used to estimate the amount of phosphate released into the reaction solution through spontaneous ATP degradation. In addition,

two reactions containing Hsp70 and Hsp40 respectively, excluding ATP were used to detect the presence of contaminating phosphate in the system. An Hsp70 deficient control (containing Tcj2 and ATP) was used to detect any ATP degradation resulting from Tcj2.

4.2.9) The determination of the presence of contaminating DnaK in solutions of His-Tcj2 purified from *E. coli* BL21(DE3) [pET28aTcj2]

His-Tcj2 was expressed in *E. coli* BL21 [pET28aTcj2], the cells harvested and cell lysate produced as described in **Section 4.2.3**. However, after the clarification of the lysate by centrifugation, the supernatant fraction was divided into two equal fractions (fraction A and B) and incubated with Ni^{2+} doped NTA-Sepharose beads (that selectively bind the His-tag of the His-Tcj2) for 12 hours at 4 °C with agitation. The beads were collected by centrifugation (1000 x*g* for 1 minute at 4 °C). The supernatant was removed and the fraction A beads were washed twice with 15 ml of wash buffer with ATP (40 mM Tris-HCl, pH 8.0; 100mM NaCl; 200 mM imidazole; 5 mM ATP) and fraction B was washed two times with wash buffer excluding ATP (40 mM Tris-HCl, pH 8.0; 100mM NaCl; 200 mM imidazole). Two additional washes excluding ATP were performed per fraction. The beads were recollected by centrifugation after each wash. His-Tcj2 was eluted from the NTA-Sepharose beads by the addition of 3 bead volumes of elution buffer (40 mM Tris-HCl; pH 7.5; 100 mM NaCl; 1 M imidazole). Samples of the various stages of the purification were probed for the presence of DnaK using Western blotting and detection with an Anti-DnaK antibody.

4.2.10) Analysis of Hsp40 and Hsp70 interaction using Quartz Crystal Microbalance with Dissipation monitoring (QCM-D)

4.2.10.1) QCM-D machine and crystal specifications

All measurements were conducted on the E-4 quartz crystal microbalance machine (Q-sense, Sweden), in which the temperature controlled flow cell was set to 28 °C. This model allows for the simultaneous measurement of the resonance, frequency and dissipation factor of four separate quartz crystal resonators. All quartz crystal resonators were sourced from Q-sense (Sweden). Their fundamental frequency of oscillation was 4.95 MHz (+/- 50 kHz), a diameter of 14 mm, thickness of 0.3 mm and an AT cut

orientation. Their electrode layer constituted a 100 nm thick gold coating covered in 50 nm of Silicon dioxide (catalogue number: QSX303, Q-sense, Sweden).

4.2.10.2) Quartz Crystal Preparation and conditions for measurement

The silicon dioxide (SiO_2) coated sensor crystals were immersed in a 0.4 % SDS solution for at least 2 hours and rinsed with Milli-Q water. The crystals were subsequently dried under nitrogen gas and subjected to 10 minutes of Ultraviolet light/Ozone treatment using the UV/Ozone Procleaner™ (Bioforce Nanosciences, Inc; USA) with the crystal 5 mm from the ultraviolet lamp (185 nm and 254 nm).

4.2.10.3) Preparation of lipid vesicles

The lyophilized powders of the synthetic lipids 1,2-Dioleoyl-*sn*-Glycero-3-[(N-(5-amino-1-carboxypentyl)iminodiacetic acid) succinyl] (DOGS-NTA; molecular weight 760.09 g/mol) and 1-Palmitoyl-2-Oleoyl-*sn*-Glycero-3-Phosphocoline (POPC; molecular weight 1015.40 g/mol), both supplied by Avanti Polar Lipids Incorporated (USA), were mixed in a ratio of 5 mol % (DOGS-NTA) to 95 mol % (POPC). This mixture was dissolved in chloroform in a glass beaker. The hydrogen was removed by evaporation under a nitrogen gas stream for 1 hour. The lipids were redissolved in a degassed buffer containing 100 mM Tris-HCl, 100 mM NaCl (pH 8.0) to a final concentration of 8.5 mg/ml. The solution was sonicated in a water bath sonicator for twenty to thirty minutes to provide the energy required for vesicle formation. The solution was then extruded through a 0.2 μM syringe filter. A 1:25 dilution with the same buffer was made using a portion of the 8.5 mg/ml lipid solution, which was then sonicated. The lipid solutions were stored under nitrogen in a light impermeable container at 4 °C.

4.2.10.4) Supported Lipid Bilayer preparation, nickel doping and application of the His-Tcj2 to the surface of the quartz crystal

The diluted lipid solution (**section 4.2.10.3**) was equilibrated to room temperature and degassed in a sonicating water bath. Once the sensor checks in air and milli-Q water were completed, buffer (100 mM Tris-HCl; 300 mM NaCl; pH8.0) was run through the QCM-D flow cell (50 μl/min). The lipid solution was applied at the same flow rate until the

spontaneous formation of a bilipid layer on the SiO_2 surface was accomplished. Buffer was pumped through the flow cell to remove any unbound lipid away from the flow cell until a steady baseline was obtained in the resonance frequency and dissipation values. A saturated solution of $NiCl_2$ in 100mM Tris-HCl, 300mM NaCl (pH.8.0) (10 mM) was passed over the sensor crystal to activate the nitriloacetic acid (NTA) groups of the DOGS-NTA in the supported phospholipid bilayer. The unbound Ni^{2+} was washed out of the flow cell with 100 mM Tris-HCl, 300 mM NaCl (pH8.0), until a stable baseline was once more achieved. Purified Tcj2 (~300 mg/ml) or its mutant Tcj2 H34Q (~300 mg/ml) was applied to the sensor crystal using the same flow rate. Unbound Tcj2 was washed off of the sensor chip with 100 mM Tris-HCl (pH8.0); 300 mM NaCl. All solutions used in the flow cell of the QCM-D were degassed.

4.2.10.5) Conditions for the testing of the interaction of Hsp70 with TcJ2

Varying concentrations of Hsp70 lacking a $(His)_6$ tag, in a degassed solution of 100 mM Tris-HCl; 300 mM NaCl; 10 mM ATP, was passed over the immobilized TcJ2. Between the application of each different Hsp70 concentration, the flow cell was washed with 100 mM Tris-HCl (pH8.0); 300 mM NaCl. The same concentration series of BSA contained in the same solution was utilized as a negative control.

4.3) Results

4.3.1) The native purification of TcHsp70

Purification of His-Tchsp70 produced in *E. coli* BL21 [pET14bTchsp70] required the use of a partial denaturing purification protocol involving the use of urea (8 M) in the lysis buffer (Edkins *et al.*, 2004). This raised questions about the proportion of denatured His-Tchsp70 that regained its native structural conformation after purification. It was therefore attempted to purify His-Tchsp70 produced in *E. coli* XL1blue [pQE30Tchsp70] using a native purification protocol (**Section 4.2.5**). The N-terminal peptide containing the $(His)_6$ tag is shorter in Tchsp70 expressed from pQE30Tchsp70 (13 amino acids) compared to protein expressed from pET14bTchsp70 (27 amino acids). It was thought that the longer $(His)_6$-tag peptide of proteins from the pET14bTchsp70 expression vector may have interfered with the structural or functional integrity of the protein *in vitro*.

However, the solubility of His-TcHsp70 expressed from pQE30Tchsp70 in XL1blue was still very poor (**Figure 4.3**). The majority of the His-Tchsp70 was present in the insoluble fraction. This could indicate that more polyethyleneimine should have been added to the

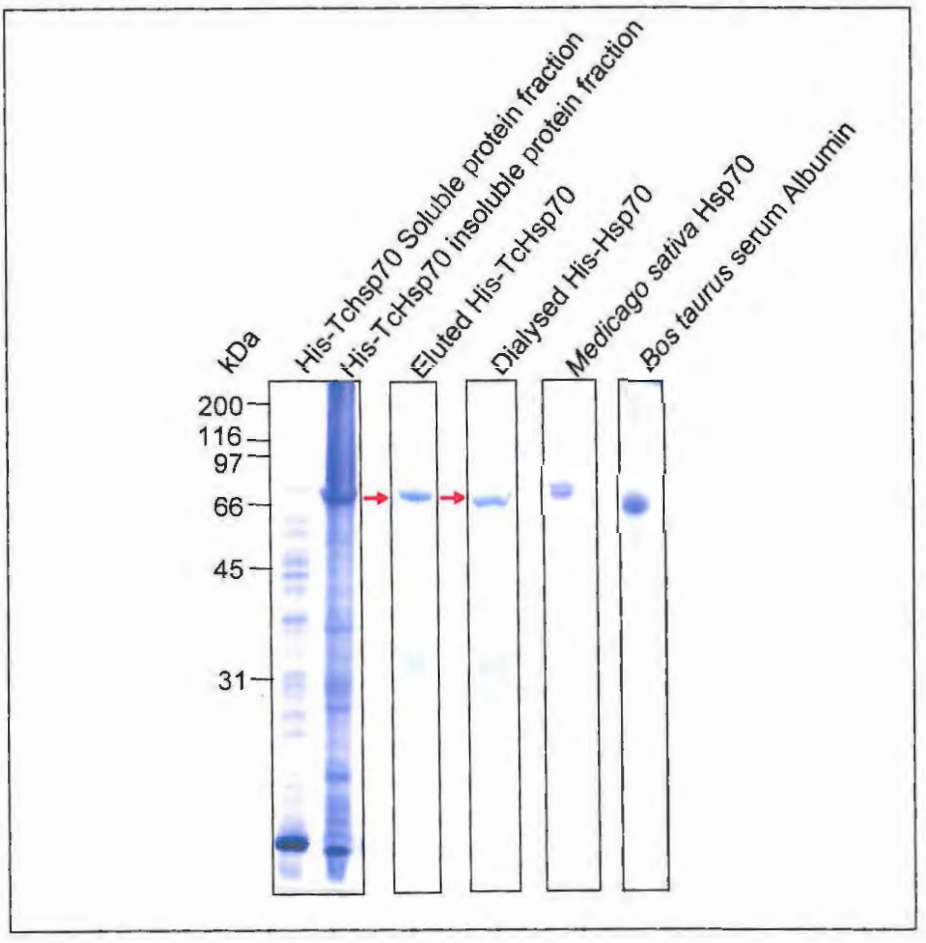


Figure 4.3: The native purification of His-TcHsp70 from *E. coli* XL1blue [pQE30Tchsp70]. The protein was purified using a native purification protocol with 0.1% polyethyleneimine in the lysis buffer. Samples were resolved using discontinuous SDS-PAGE (**Appendix C.1.3**). Soluble protein fraction indicates the fraction of the cell lysate that was loaded onto the nickel-doped NTA-sepharose matrix. The insoluble fraction is the portion of the lysate that was sedimented during centrifugation of the whole cell lysate. The red arrows indicate the purified His-Tchsp70. Samples of *Medicago sativa* Hsp70 and *Bos taurus* serum albumin were also resolved to confirm a lack of proteolytic activity. The molecular weights to the left of the figure are shown in kilodaltons (kDa).

lysis buffer or possibly that lysis was incomplete, as there appeared to be a higher proportion of all proteins in the insoluble fraction relative to the soluble fraction. Despite this, there was still sufficient soluble His-TcHsp70 to purify as is shown by samples of

the purified His-Tchsp70 eluted from the nickel-doped NTA-bead matrix and the dialysed His-Tchsp70 fractions. Yield was approximately 0.5 mg per litre of cell culture. Samples of *Medicago sativa* Hsp70 (purchased from Alpha Biogene International, The Netherlands) and Bovine serum albumin (BSA) that were used for subsequent experiments in this chapter were also resolved (**Figure 4.3**). There was little indication of protease activity in these samples.

4.3.2) The native purification of His-Tcj2 and His-Tcj2 H34Q

His-Tcj2 and His-Tcj2(H34Q) were produced and purified from *E. coli* BL21 (DE3) [pET28aTcj2] and *E. coli* BL21 (DE3) [pET28aTcj2(H34Q)] respectively. Both of these proteins expressed well using the pET28a expression vector and produced very similar profiles for samples taken from the soluble fraction of the cell lysate resolved on an SDS-PAGE gel (**Figure 4.4**). Likewise, the His-Tcj2 and His-Tcj2(H34Q) samples taken from the eluted protein solution and the dialysed protein solution (to remove imidazole from the eluted protein solution) produced a similar molecular weight profile of proteins. The proportional concentration of proteins of other molecular weight relative to His-Tcj2 and His-Tcj2(H34Q) is small. The majority of the contaminating bands are below the molecular weights of the purified proteins and are possibly the result of a certain amount of proteolytic cleavage of His-Tcj2 and His-Tcj2(H34Q), despite the inclusion of a protease inhibitor cocktail in the lysis buffer. A protein of higher molecular weight (~70kDa) than the purified Hsp40 proteins is more prominent in the His-Tcj2 purified protein solution, relative to the His-Tcj2(H34Q). This may be DnaK that was shown to be co-purified with His-Tcj2 when using the *E. coli* BL21(DE3) [pET28aTcj2] (**See section 4.3.5**)

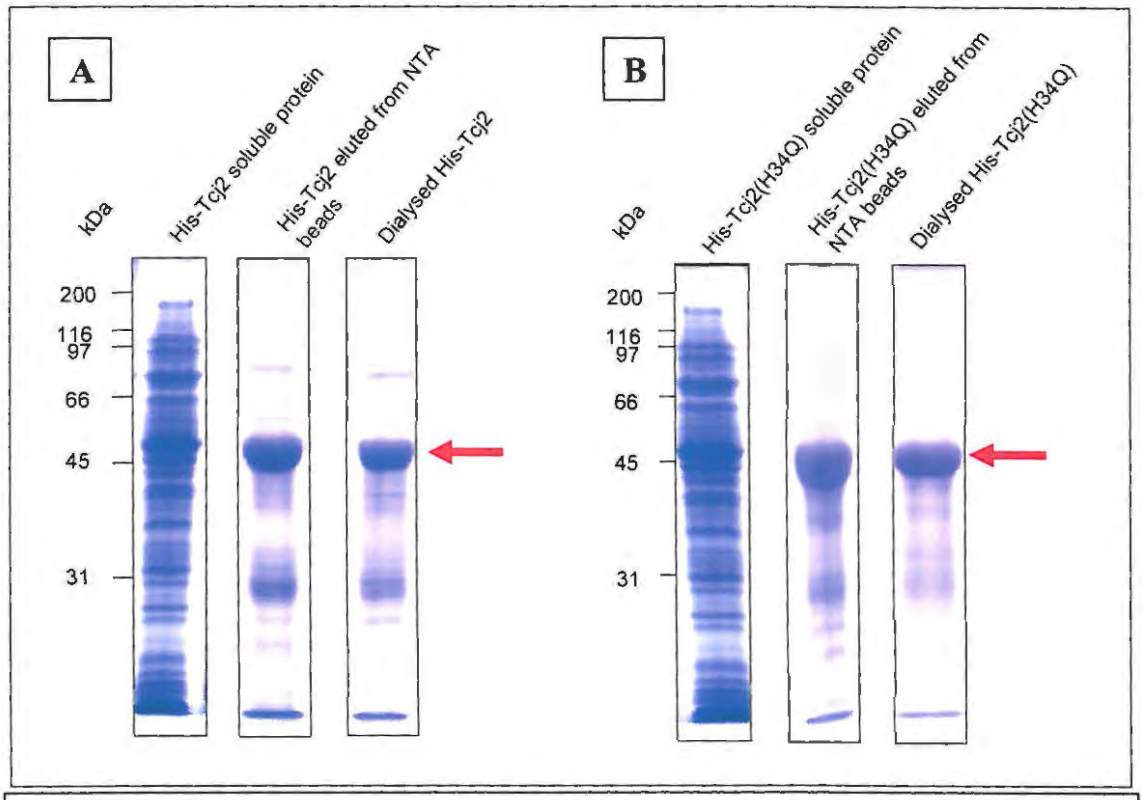


Figure 4.4: Purification of His-Tcj2 and its mutant protein His-Tcj2(H34Q).
Samples were resolved using SDS-PAGE (**Appendix C.1.3**). (A) The Purification of His-Tcj2. (B) The purification of His-Tcj2(H34Q). The soluble protein represents the fraction of the total cell lysate present in the supernatant after the clarification of this lysate through centrifugation at 16 000 x*g* for 25 min (**section 4.2.3**). The middle lanes show samples of the purified protein eluted from the nickel-doped NTA-bead matrix. The dialysed protein shows the purified proteins subsequent to the removal of imidazole using dialysis. The arrows show the bands of the purified proteins. The molecular weights to the left of the figures are shown in kilodaltons (kDa).

4.3.3) Assessment of the aggregated state of the various purified proteins

In order to confirm the native state of the purified proteins, the level of aggregation was detected by fluorescence spectrophotometry. Small aggregates were detected by transmitting light (350nm) through the solution and assessing the amount of light scattering that occurs by setting detection to scan from 300 nm to 600 nm wavelengths. Likewise, larger aggregates were detected by transmitting light (500nm) through the solution and again setting detection to scan from 300 to 500nm (Buchner *et al.*, 1998). A solution of the purified protein (10 µg/ml) in water was used in each case and bovine serum albumin (BSA) was used as a control solution (**Figure 4.5A, B, C and D**). A solution of BSA (10 µg/ml) was heated in boiling water bath for 10 minutes to denature

the protein to act as a control for the presence of aggregates in solution (**Figure 4.5 A and B**). Denaturation of the BSA produced a larger proportion of small aggregate complexes (**Figure 4.5A**) relative to large aggregates (**Figure 4.5B**). By contrast, an aliquot of the same solution of BSA that was not heat treated before measurement showed almost no detection of light scattering (**Figure 4.5C and D**). The assumption made here was that the non-heat treated BSA solution contained BSA in its completely native state, without any aggregates present. In comparison to the non-heat treated BSA the purified TcHsp70 (**Figure 4.5E and F**) used for the ATPase assay detection (**Section 4.3.1**), had similar peaks for detection of aggregates at 350 nm and 500 nm. This indicated that the majority of His-TcHsp70 was likely to be in a non-aggregated state.

Native and denatured samples (10 μg/ml) of His-Tcj2 and His-Tcj2(H34Q) were measured for the presence of aggregates (**Figure 4.6**). Denatured His-Tcj2 (**Figure 4.6A and B**) showed a similar proportion of both large and small aggregates relative to His-Tcj2(H34Q) (**Figure 4.6E and F**). However, both of these proteins had a reduced proportion of aggregation in their denatured samples (small aggregates: ~2200 units; large aggregates: ~700 units) relative to the denatured BSA (small aggregates: ~3500 units; large aggregates: ~1000 units) (**Figure 4.5A and B**).

Native samples of His-Tcj2 (**Figure 4.6 C and D**) and Tcj2(H34Q (**Figure 4.6 G and H**) showed higher levels of aggregation for both large and small aggregates relative to native BSA (**Figure 4.5 C and D**) However, comparison of the data for denatured and native samples in **Figure 4.6**, suggested that the majority of the protein population in both the native samples of His-Tcj2 and His-Tcj2(H34Q) was in a non-aggregated state.

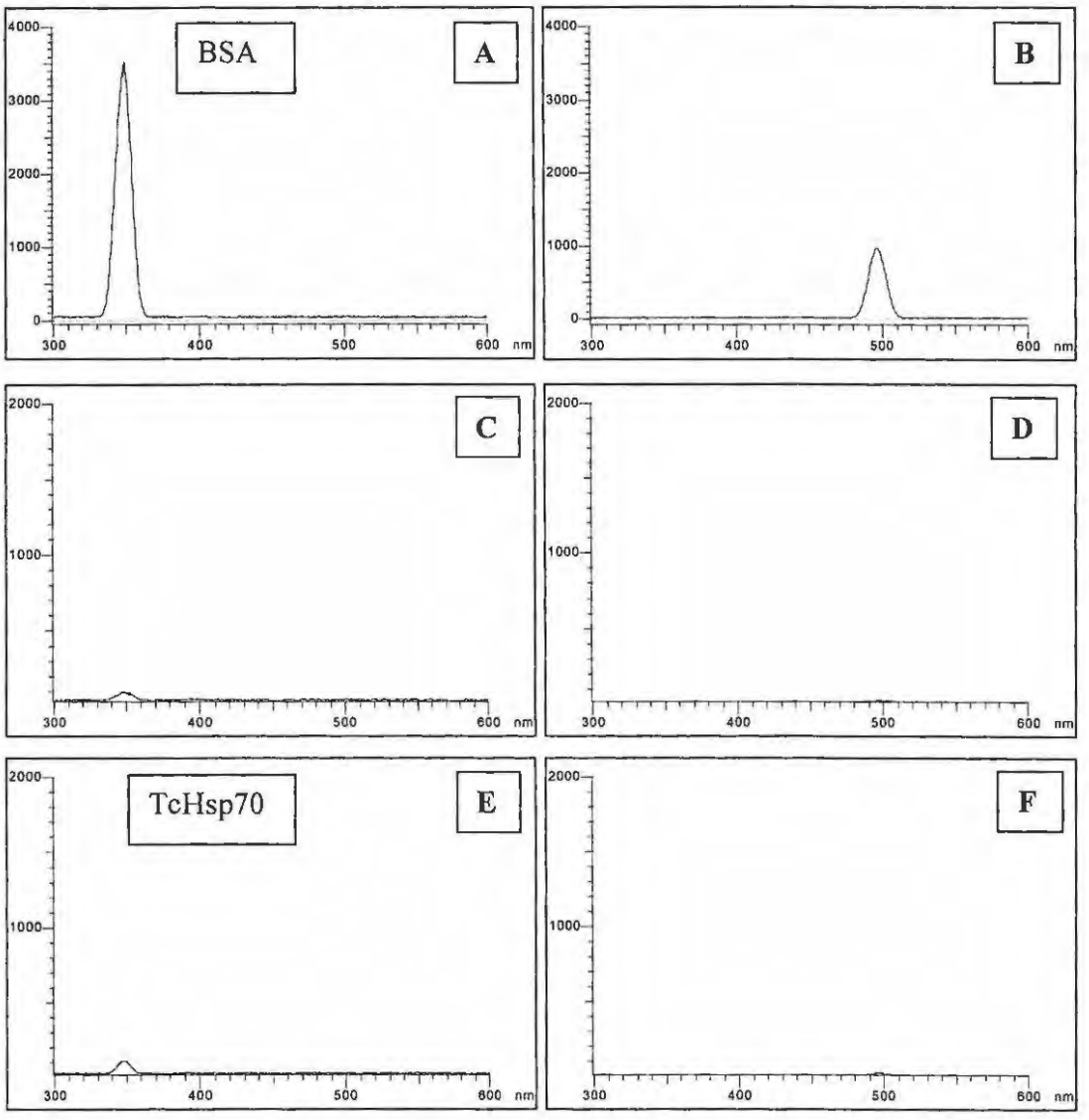


Figure 4.5: TcHsp70 is largely unaggregated.
Aggregates were detected through the measurement of light scattering. Denaturation of BSA (10 μg/ml concentration) in a boiling water bath for 10 minutes was used as a positive control for aggregates. Native protein samples of the same concentration were also tested for the scattering of light at 350nm (small aggregate detection) and 500nm (large aggregate detection). Light emission was monitored in the range of 300nm to 600nm (x-axis). The y-axis indicates the relative amount of light scattering observed. (A) Denatured BSA excited at 350 nm. (B) Denatured BSA excited at 500 nm. (C) Native BSA excited at 350nm. (D) Native BSA. (E) Native TcHsp70 (F) Native TcHsp70.

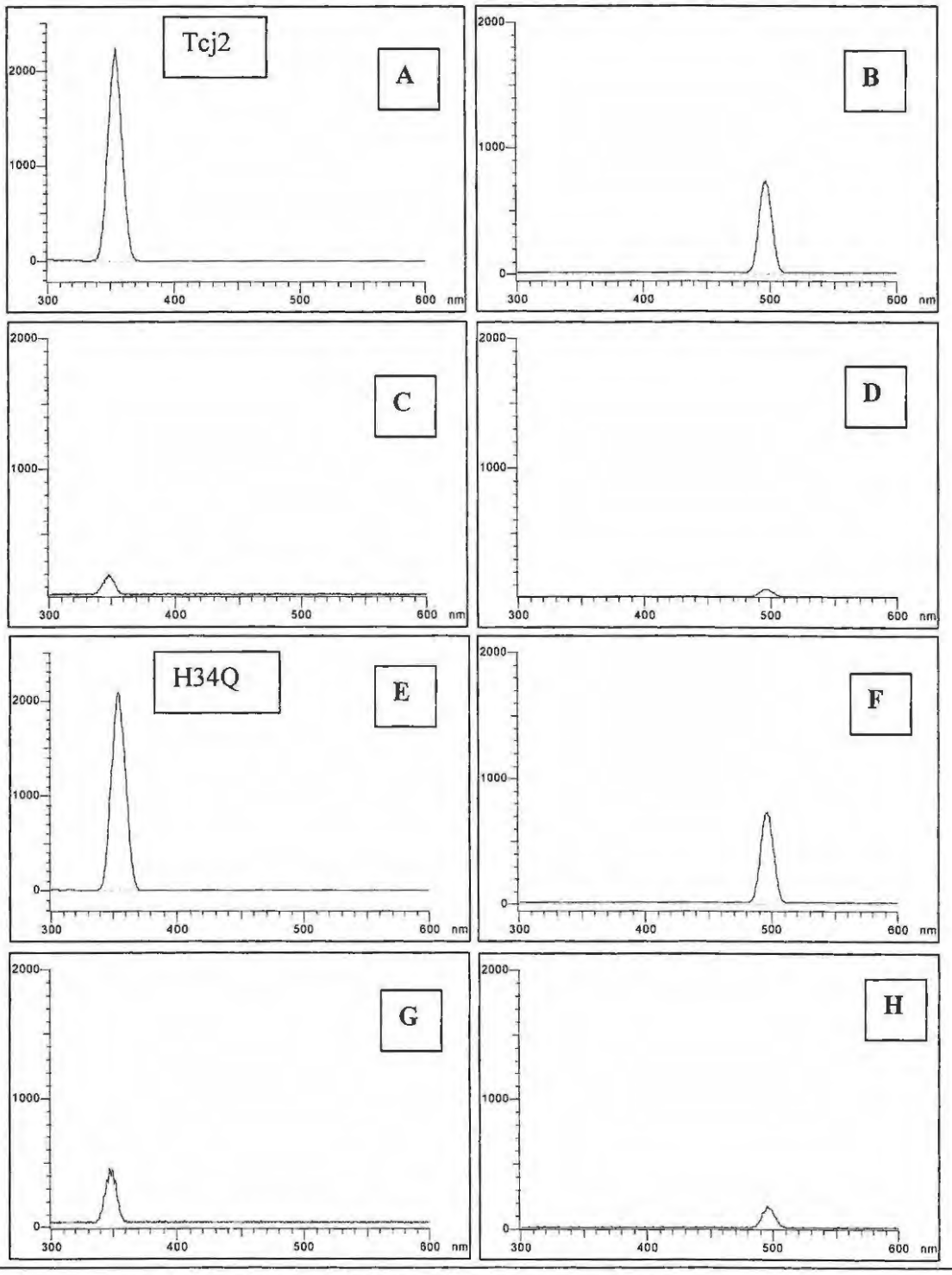


Figure 4.6: His-Tcj2 and His-Tcj2(H34Q) are largely unaggregated.
Protein solutions (10 μg/ml concentration) were denatured in a boiling water bath for 10 minutes. Native samples of the same concentration were also tested for the scattering of light at 350nm (small aggregate detection) and 500nm (large aggregate detection). Light emission was monitored in the range of 300nm to 600nm (x-axis). The y-axis indicates the relative amount of light scattering observed. (A) Denatured His-Tcj2 excited at 350 nm. (B) Denatured His-Tcj2 excited at 500 nm. (C) Native His-Tcj2 excited at 350nm. (D) Native His-Tcj2 excited at 500nm. (E) Denatured His-Tcj2 H34Q. (F) Denatured His-Tcj2(H34Q) excited at 500nm. (G) Native His-Tcj2(H34Q) excited at 350nm. (H) Native His-Tcj2(H34Q) excited at 500nm.

4.3.4) Tcj2 is able to stimulate the ATPase activity of various Hsp70s

ATPase assays were used to test the ability of His-Tcj2 to interact with various Hsp70 proteins in a functional manner *in vitro*. The basal ATPase activity (Hsp70 + ATP without substrate or Hsp40 co-chaperone) of untagged *Bos taurus* brain Hsc70 (0.72 nmoles of phosphate per minute per mg of protein [nmol/min/mg]) (A kind donation of by Dr Michael E. Cheetham, University College London Institute of Opthalmology, London, UK) and His-TcHsp70 (0.82 nmol/min/mg) were very similar (**Figure 4.7A**).

Addition of His-Tcj2 was able to stimulate the ATPase activity of *Bos taurus* brain Hsc70 9.5 fold, compared to the 1.4 Fold stimulation of His-Tchsp70. *M. sativa* Hsp70 (purchased from Alfa Biogene International, The Netherlands) has a high basal ATPase activity (28.9 nmols/min/mg) relative to *Bos taurus* Hsc70 and His-Tchsp70 (**Figure 4.7A and Figure 4.7B**). The ATPase activity of *M. sativa* Hsp70 was stimulated to a greater extent (in the presence of ATP) in a solution of 0.08 μM Hsp70 and 0.4 μM His-Tcj2 (196.6 nmol/min/mg) relative to a solution of 0.4 μM Hsp70 and 0.08 μM His-Tcj2 (32.6 nmol/min/mg). It would appear as if the extent of stimulation of the ATPase activity of *M. sativa* Hsp70 is greater when there is more Hsp40 relative to Hsp70 in an *in vitro* reaction system. This finding assumes that all the His-Tcj2 is native and is not being recognised as substrate by the Hsp70. The requirement for very much larger concentration of Hsp40 relative to Hsp70 (without unfolded protein substrate) for efficient stimulation of Hsp70 ATPase activity has been observed in *E. coli* DnaJ and DnaK studies (Laufen *et al.*, 1999).

The reason for the difference in ATPase activity stimulation of these different Hsp70s by His-Tcj2, particularly the stimulation of *B. taurus* Hsc70 relative to His-Tchsp70, is not immediately apparent. However, it does indicate the capacity for Tcj2 to interact with a diverse range of Hsp70s of divergent origins *in vitro*. The low stimulation of Tchsp70 by Tcj2, relative to Bovine Hsc70 and *M. sativa* Hsp70, may indicate appropriate levels of stimulation of the ATPase activity as opposed to over stimulation by an inappropriate partner Hsp40.

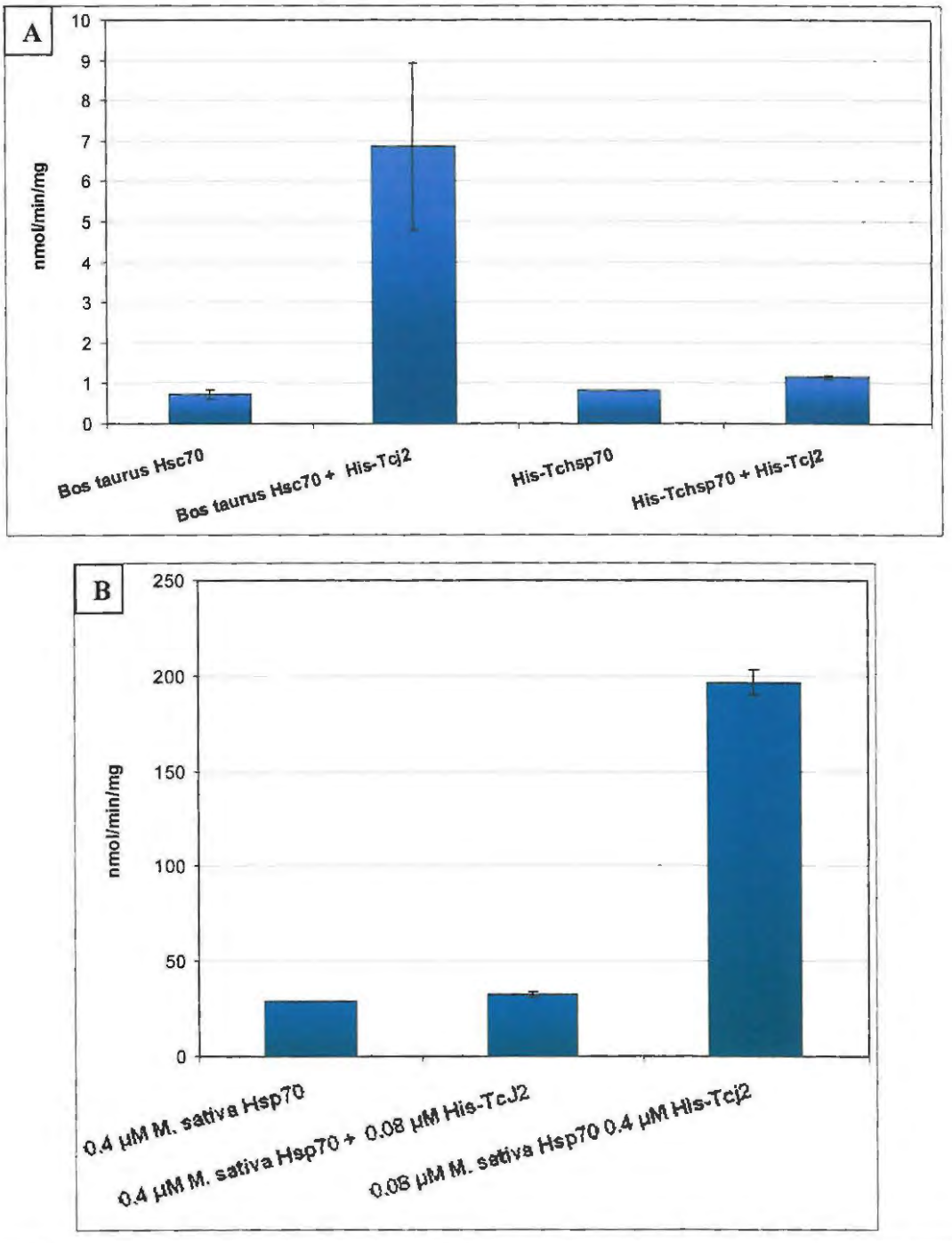


Figure 4.7: Tcj2 is able to stimulate the ATPase activity of Hsp70s from diverse sources.
(A) The basal ATPase activity (nmol of phosphate released per minute per mg of Hsp70) of *Bos taurus* brain Hsc70 (0.4 µM) and His-Tchsp70 (0.4 µM) was used to determine the proportion of stimulation of this activity generated by His-Tcj2 at the incubation temperature of 37°C. All reactions were carried out in the presence of 600 µM ATP (final concentration). **(B)** The basal ATPase activity of *M. sativa* Hsp70 relative to the stimulation of this activity by His-Tcj2 (0.4 µM and 0.08 µM). All reactions were carried out in the presence of 600 µM ATP (final concentration). The error bars indicate standard deviation recorded for at least three replicate ATPase assays.

4.3.5) DnaK contaminates Tcj2 purified from *Escherichia coli* BL21

During the course of the ATPase experiments, the capacity for the control containing 0.04 μM His-Tcj2 and ATP to produce an ATPase activity above the level of background ATP decay in the solution was surprising. Due to the absence of an identified ATPase domain in the Hsp40 domain structure, it was proposed that this ATPase activity may be due to the presence of *Escherichia coli* DnaK that copurified with His-Tcj2. This was tested using Western blot analysis and probing with a DnaK specific antibody and DnaK was detected in all samples taken to monitor the purification of His-Tcj2 including the final purified His-Tcj2 protein solution. It was attempted to remove contaminating DnaK from the purification using ATP (5 mM) in the wash buffers based on a protocol by Rial and Ceccarelli (2002) (**Figure 4.8**). The *Escherichia coli* BL21 cells expressing His-Tcj2 from pET28aTcj2 were harvested from a 1 litre culture by centrifugation according to the same method outlined in **section 4.2.3**. The cells were resuspended in 20 ml of lysis buffer, which was then divided into two 10 ml fractions. The His-Tcj2 was purified from both of these fractions according to the purification method in **section 4.2.9**, except that one fraction (Fraction A) was subjected to a wash with 5 mM ATP, while the other fraction (Fraction B) was subjected to a wash without ATP.

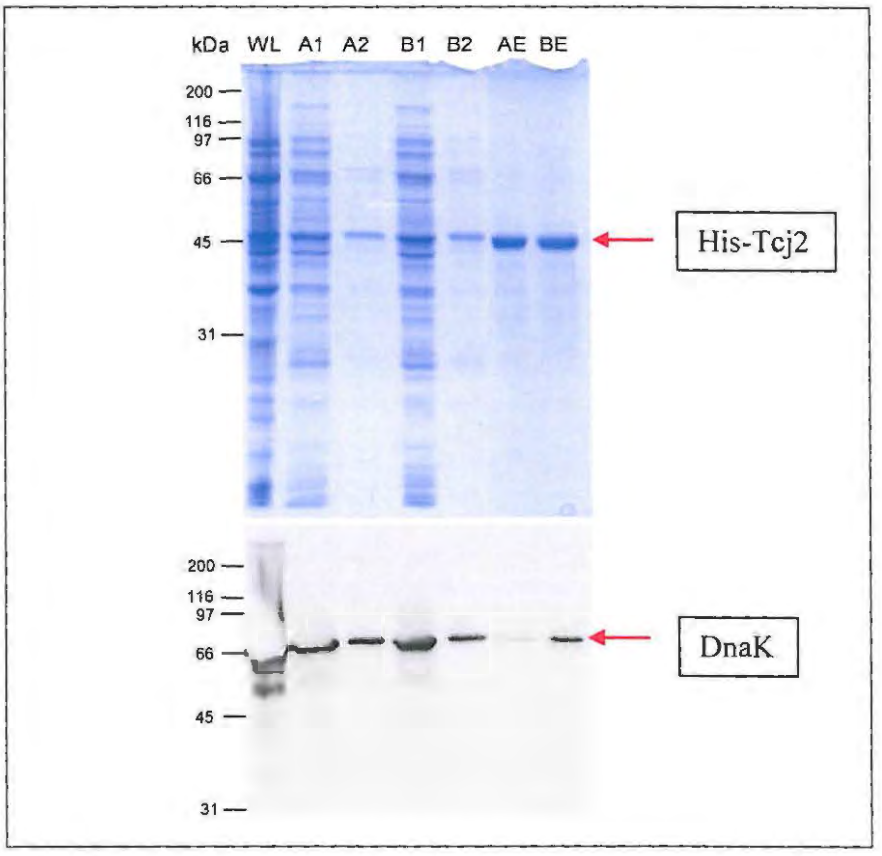


Figure 4.8: DnaK copurifies with His-Tcj2 when His-Tcj2 is produced in *Escherichia coli* BL21 (DE3).
Cells from 1L of cell culture were resuspended in 20 ml of native Lysis buffer (Constituents) to produce the whole cell lysate (WL). This was divided into two 10 ml fractions. Fraction A contained ATP (5mM) in the wash buffer for the washing steps (A1 and A2), while Fraction B was washed with wash buffer that did not contain ATP (B1 and B2). A1- ATP wash 1. A2 – ATP wash 2. B1 - Wash 1 without ATP in the wash buffer. B2 – Wash 2 without ATP in the wash buffer. AE- Eluted ATP washed protein. BE – Eluted protein that was not exposed to ATP while bound to the NTA sepharose column. These samples (20µl) were resolved on an SDS-PAGE gel (upper panel) and blotted onto a nitrocellulose membrane for Western analysis (lower panel) (primary antibody 1:5000 anti-DnaK; secondary antibody: 1:5000 ECL advance kit horse radish peroxidase labeled anti-rabbit IgG, GE Healthcare, Amersham). The upper red arrow on the right side of the panels indicates the bands of purified His-Tcj2. The lower red arrow shows the detection of DnaK by Western blot.

The whole cell lysate (**Figure 4.8WL**) showed a large amounts of DnaK contamination to the extent that the signal detection on the Western blot (using the Chemidoc gel and Western blot documentation system, Biorad) was saturated, producing a light smear between 66 kDa and 97 kDa. The first washes of the NTA-bead matrix for both Fraction A and B (**Figure 4.8 A1 and B1**) did not show very different levels of DnaK removal and neither did the second washes (**Figure 4.8 A2 and B2**). However, the elution fractions

showed significantly less DnaK contamination in the ATP washed fraction (**Figure 4.8AE**) relative to the fraction that did not have ATP washes (**Figure 4.8BE**).

4.3.6) Setup of the QCM-D lipid bilayer and immobilization of Tcj2 onto the SiO_2 covered quartz crystal

In order to investigate the binding of Tcj2 to an Hsp70 it is required that one of these partner proteins is immobilized onto the quartz crystal resonator surface. It was decided to immobilize Tcj2 as the protein produced from the pET28aTcj2 plasmid already contained a $(His)_6$-tag to purify this protein from *E. coli* BL21 lysates.

The introduction of lipid vesicle solution (**Section 4.2.10.3**) into the liquid surrounding the quartz crystal resonator led to the adsorption of these vesicles to the SiO_2 coating of the crystal. The binding of the vesicles caused a shift in the oscillation frequency of the crystal by approximately 60 Hz to 65 Hz (**Figure 4.9.1A and Figure 4.9.2A**). As the adsorbed vesicles ruptured to form a lipid bilayer on the crystal surface, the buffered solution trapped in the centre of the vesicles was decoupled from the reverse piezoelectric oscillations of the crystal. This produces a net loss of observed mass bound to the surface of the crystal and a frequency shift from between -65 and 60 Hz to -30 Hz and -26 Hz at which point the frequency shift stops and levels off. In **Figure 4.9.1**, this stable frequency is different from the figure quoted in literature of -26 Hz to -28 Hz shift from the start of measurement of a clean crystal, when using the same parameters (temperature, pH and ionic strength) as outlined for lipid immobilization in these experiments (Reimhult *et al.*, 2002; Seantier *et al.*, 2005; Reimhult *et al.*, 2003; Richter *et al.*, 2003; Keller and Kasemo, 1998). The lipid immobilization data (**Figure 4.9.2**) shows greater similarity to this frequency shift observed in this literature. This could be due to different lipids of different molecular mass being used in these experiments. It could also possibly be due to an older or damaged crystal used for the experiment shown in **Figure 4.9.1** relative to that in **Figure 4.9.2,** where the nanotopography of the crystal may have altered the binding of the lipid bilayer. Alternatively there may be some unruptured vesicles on the surface of the crystal. A subsequent wash with buffer not containing lipid vesicles (Stable frequency/dissipation measurements between **Figure 4.9A and B [Figure4.9 A/B**

interface]) produced no change in the frequency suggesting that all mass was bound to the surface in a stable manner.

The dissipation factor measurements for the lipid immobilization mirrors the shift in frequency. The adsorption of lipid vesicles to the surface of the crystal causes an increase in the dissipation factor suggesting that the vesicle layer has a high viscoelatsticity and is flexible and extends into the liquid medium (**Figure 4.9.1A and Figure 4.9.2A**). As the lipid vesicles rupture and form a supported lipid bilayer the dissipation factor decreases as the lipid layer is more compact and rigid on the crystal surface than the intact vesicles. As was observed for the frequency shift involving the lipid immobilization, the data shown in **Figure 4.9.2** more closely follows the results reported in literature than the data in **Figure 4.9.1** (Reimhult *et al.*, 2002; Seantier *et al.*, 2005; Reimhult *et al.*, 2003; Richter *et al.*, 2003; Keller and Kasemo, 1998). Gläsmaster and coworkers (2002) reported a very different frequency and dissipation shift relative to other literature with measurements using egg phosphatidylcholine (EggPC), which is the same lipid mixture used by many other studies using the third overtone of the resonator oscillations (Reimhult *et al.*, 2002; Seantier *et al.*, 2005; Reimhult *et al.*, 2003; Richter *et al.*, 2003; Keller and Kasemo, 1998).

The resultant supported phospholipid bilayer (SPB) was washed with buffer (100 mM Tris-HCl; 300 mM NaCl; pH8.0) (**Figure 4.9.1(A/B interface**) and **Figure 4.9.2(A/B interface**)) to remove any unbound lipid vesicles from the flow cell. The frequency and the dissipation factor both remained constant through this wash. The NTA groups of the DOGS-NTA lipids were doped with Ni^{2+} ions by passing a 40mM $NiSO_4$ solution over the lipid layer (**Figure 4.9.1B and Figure 4.9.2B**). This solution produced a significant decrease in the resonance frequency of the crystal. Most of this frequency decrease was reversed when the excess nickel solution was washed off with buffer. A slight decrease in resonance frequency (~5Hz) was maintained (**Figure 4.9.2**) relative to the SPB layer on its own, while **Figure 4.9.1** showed an additional loss of mass after the washing away the nickel solution was observed. This suggests that instead of Ni^{2+} ions binding to the NTA groups of the DOGS-NTA in **Figure 4.9.1**, part of the SPB was removed. However,

despite this, $(His)_6$-tagged Tcj2 was able to bind to both SPB layers (**Figure 4.9.1C and 4.9.2C**.). It is possible that some of the lipid vesicles did not rupture on contact with the SiO_2 surface **(Figure 4.9.1)**. This might explain why the lipid layer formed a stable frequency oscillation at -32 Hz as opposed to the ~-26 Hz of **Figure 4.9.2** and lipid immobilization data in the literature (Seantier *et al.*, 2005). The ionic strength of the $NiSO_4$ solution may have destabilized any adsorbed lipid vesicles and caused them to rupture releasing their water molecules and hence recording a loss of mass, despite the binding of Nickel. The importance of ionic strength (with Ca^{2+} and Na^{+}) of buffer solutions in the destabilization of lipid vesicles, causing them to rupture and form a lipid monolayer on a solid surface has previously been reported (Seantier *et al.*, 2005). The dissipation of the system increased with the addition of the nickel solution **(Figure 4.9.1B and Figure 4.9.2B)**. This could have been the result of an increase in the viscosity of the buffer medium containing $NiSO_4$. A higher baseline compared to the undoped SPB even after the $NiSO_4$ solution was washed from the flow cell was observed in **Figure 4.9.2**. This post wash change in the dissipation baseline is less evident in the data shown in **Figure 4.9.1**. The data displayed in **Figure 4.9** therefore show the two scenarios that were encountered during the construction and nickel doping of the SPB, with the data shown in **Figure 4.9.1** more closely replicating data reported in the literature.

The binding of His-Tcj2 produced a strong oscillation frequency decrease in both **Figure 4.9.1 (C)** and **4.9.2 (C)**, accompanied with a strong increase in the dissipation factor of the system. The increase in dissipation is to be expected as the $(His)_6$-tagged proteins are bound to the SPB through a flexible N-terminal peptide that contains the $(His)_6$-tag. A complicating factor in the immobilization of the Tcj2 is that no stable baseline was reached even after approximately two hours of exposure to the protein solution (~250 to 300 µg/ml) (**Figure 4.9.2C**), which is the equivalent of 6 ml of protein solution. It would have been anticipated that the frequency and dissipation shifts would have leveled to a stable baseline after all the NTA sites were occupied by a histidine tagged protein. It is unclear as to why this was not observed. It is possible that the lipid ratio was not 100% correct. The lyophilized lipid powders are extremely hygroscopic and become sticky

when hydrated, complicating accurate weight measurements of these compounds. More DOGS-NTA would mean more binding sites available for His-Tcj2.

When the flow cell was washed of Tcj2 solution with buffer (100 mM Tris-HCl; 300 mM NaCl; pH8.0) (**Figure 4.9.1D** and **Figure 4.9.2D**), a slow loss of mass coupled to the motion of the resonator was noticeable, particularly in **Figure 4.9.1**. This could have been caused by:

1) Detachment of nickel doped NTA bound His-Tcj2.
2) Detachment of Tcj2 interacting with the immobilized film in a non – His tag-NTA interaction. Tcj2 is predicted to dimerise (**Chapter 2 section 2.3.9**).
3) DnaK copurified from the *E. coli* BL21 [pET28aTcj2] system was slowly detaching from His-Tcj2 with which it was copurified (**Section 4.3.5**).
4) Part of the lipid scaffolding was detaching, either due to delayed vesicle rupture or lifting of a portion of the SPB.

The introduction of 10 mM ATP with 2 μM Human Hsp70 (His-tag deficient) (in 100 mM Tris-HCl; 300 mM NaCl; pH8.0) (**Figure 4.9.1E**) caused a decrease in resonance frequency and an increase in the dissipation. Washing of the Hsp70/ATP solution resulted in a substantial increase in frequency (mass loss) of ~12 Hz and more rapid stabilization of the frequency baseline, with a slower drift (**Figure 4.9.1F**). This suggests that the Hsp70 or the ATP had an effect on the baseline stabilization. As it is known that DnaK contaminates His-Tcj2 purified from *E. coli* BL21 [pET28aTcj2] and ATP is known to assist with the removal of this DnaK contamination (**Chapter 4 section 4.3.2**), the ATP may be causing DnaK to release from its interaction with His-Tcj2. The His-Tcj2 immobilised onto the lipid layer in **Figure 4.9.2** was stored at 4°C for a substantially longer time before use compared to the His-Tcj2 in **Figure 4.9.1**. This may suggest that the DnaK naturally detached from its interaction with His-Tcj2 in the older protein solution.

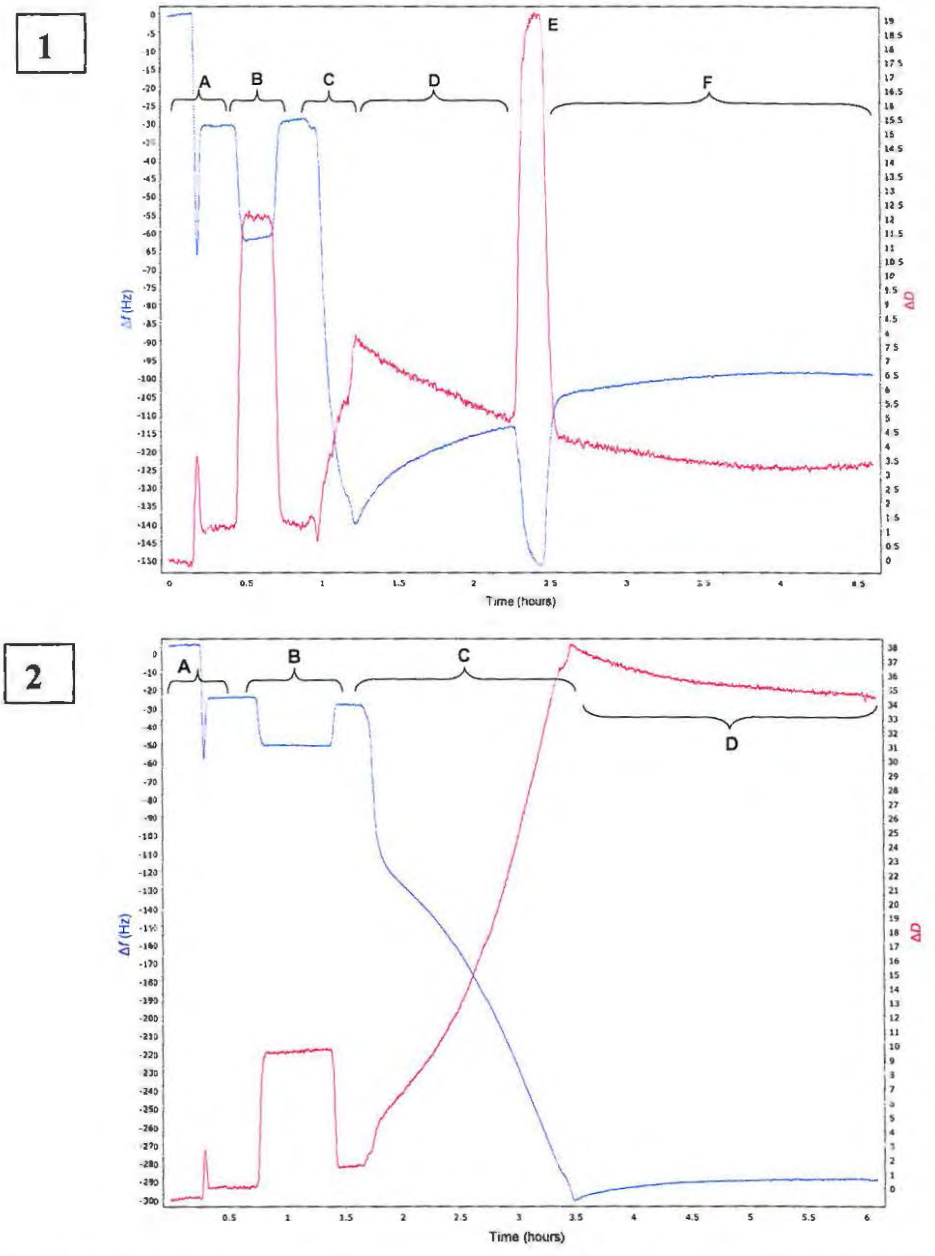


Figure 4.9: The immobilization of Tcj2 onto the quartz crystal sensor.
(1) and (2) show the immobilization of His-Tcj2 onto a SiO_2 coated quartz crystal resonator, representing the two scenarios observed during the construction of the nickel doped supported phospholipids bilayer. The data presented in (1) more closely represents data reported in the literature. The frequency change (Δf) is recorded in blue (left axis) and the change in the dissipation factor (ΔD) is recorded in red versus Time (hours) (x-axis). Measurements are shown for the third overtone of the crystal resonance. **(A)** The application of the lipid bilayer onto the SiO_2 coated crystal. **(B)** Ni^{2+} (40 mM $NiSO_4$) doping of the NTA groups of the DOGS-NTA lipids. **(C)** The application of Tcj2 protein [(1) 281µg/ml; (2) 310 µg/ml] to the system. As the Tcj2 proteins bind to the NTA groups by their His-tag, they add mass to the crystal sensor. Between each of these steps a wash with a buffered solution (100mM Tris-HCl (pH8.0); 300mM NaCl) was performed until a stable baseline was achieved. **(D)** wash step with the same buffer to remove any unbound Tcj2. Note the loss of mass as the wash step continues. **(E)** Introduction of 2µM Human Hsp70 + 10 mM ATP (in the buffered solution used for the washes. **(F)** wash with 100 mM Tris-HCl (pH 8.0); 300mM NaCl). Note the decreased loss of mass and stabilization of the baseline after the addition of ATP.

4.3.7) *M. sativa* (Alfalfa) Hsp70 interacts with Tcj2 and Tcj2 H34Q *in vitro*

In an attempt to compare the interaction Tcj2 and Tcj2(H34Q) with a partner Hsp70, His-Tcj2 and His-Tcj2(H34Q) were immobilized onto two separate SiO_2 coated quartz crystal resonators. Tcj2 and its mutant were chosen to be the immobilized proteins of this protein-protein interaction study as they contained a $(His)_6$-tag used in their purification. *M. sativa* Hsp70 was used in these experiments as it did not contain a histidine tag which could have produced false binding signals with the NTA lipid layer. In addition, this Hsp70 was more easily available commercially from its natural source than Human Hsp70 and is claimed to possess human Hsp70-like properties (Alfa Biogene International B.V.; http://www.alfabiogene.de/HSP70_page1.pdf). The buffer containing 100 mM Tris-HCl; 300 mM NaCl and a pH of 8.0 was used as the baseline liquid medium in these measurements, with various additives (Lipid vesicles, $NiSO_4$, ATP and proteins).

The immobilization of the supported phospholipid bilayer and nickel doping did not give the same measurements for each crystal. The SPB formation and Ni^{2+} doping on the crystal for immobilization of His-Tcj2 (**Figure 4.10.1A**) was similar to that shown in **Figure 4.9.1**, while the SPB formation on the His-Tcj2(H34Q) crystal (**Figure 4.10.2A**) was more similar to that shown in **Figure 4.9.2**. Despite these differences in the SPB measurements, both His-Tcj2 and His-Tcj2(H34Q) produced a strong decrease in the resonance frequency of their respective crystals suggesting that they bound to the NTA groups of the SPB. However, the frequency change produced by the immobilization of His-Tcj2 was approximately 2.8 times that for His-Tcj2(H34Q) over the same time interval, suggesting that more His-Tcj2 was bound to the lipid NTA sites relative to His-Tcj2(H34Q). As was the case in **Section 4.3.6**, saturation of binding does not appear to have been reached. The immobilization of His-Tcj2 generated approximately 2.5 times the shift in the dissipation factor relative to the immobilization of His-Tcj2(H34Q) (**Figure 4.11.1 A and Figure 4.11.2 A**).

Flushing of the flow cells with buffer (100 mM Tris-HCl; 300 mM NaCl; pH8.0) produced the same slow drifting frequency due to loss of mass observed in **Figure 4.9.1**.

Therefore, three washes with 60 mM ATP solution (in 100 mM Tris-HCl; 300 mM NaCl; pH8.0) were used (**Figure 4.10.1B** and **Figure 4.10.2B**), which allowed for the partial, but not complete stabilization of the frequency and dissipation measurements (**Figure 4.10B and Figure 4.11B**). The residual baseline frequency shift appears to be more pronounced in the His-Tcj2(H34Q) system (**Figure 4.10.2**).

Attempts at studying the interaction of His-Tcj2 with Human Hsp70, using QCM-D showed no detectable difference in the frequency shift produced by passing 0.1μM, 0.2 μM, 0.4 μM, 0.6 μM and 0.9 μM of Human Hsp70 (Histidine tag deficient) (contained in 100 mM Tris-HCl; 300 mM NaCl; 10 mM ATP; pH8.0) over immobilized His-Tcj2 (Data not shown). A dilution series of 2μM, 4 μM, 6 μM, 8 μM and 5 μM *M. sativa* Hsp70 (histidine tag deficient) (in 100 mM Tris-HCl; 300 mM NaCl; 10 mM ATP; pH8.0) was therefore chosen for this study (**Figure 4.10.1C** and **Figure 4.10.2C**). This dilution series produced a dose response in which a different frequency shift was generated proportional to the concentration of the *M. sativa* Hsp70. The addition of the same dilution series of bovine serum albumin (BSA) in the same buffer containing 10 mM ATP produced the same frequency shift response in each BSA concentration (**Figure 4.10.1D** and **Figure 4.10.2D**). A protein deficient control was also used (0 μM; **Figure 4.10.1D** and **Figure 4.10.2D**) in which the other solution constituents remained the same as for the dilution series of Hsp70 and BSA (100 mM Tris-HCl; 300 mM NaCl; 10 mM ATP; pH8.0), excluding the addition of any protein to the cell. This control suggested that the frequency and dissipation shifts observed for the addition of BSA was completely attributable to the 10 mM ATP in the solution, as it produced the same frequency shift as the BSA dilution series.. This further supports the interpretation that BSA did not bind to the quartz crystal or to its immobilized constituents. The apparent loss of mass in the 8 μM BSA measurement in **Figure 4.10.2D** was potentially due to a gas bubble becoming introduced into the flow cell. After the removal of the 8 μM BSA, the baseline settled approximately 2.5 Hz higher than it was previously. This change could be factored into subsequent data analysis. The *M. sativa* Hsp70 dilution series dose response in the presence of 10 mM ATP suggests that it interacted with both His-Tcj2

and His-Tcj2(H34Q), relative to the lack of interaction observed with BSA solutions. This apparent interaction between His-Tcj2/His-Tcj2(H34Q) and Hsp70 appears to be

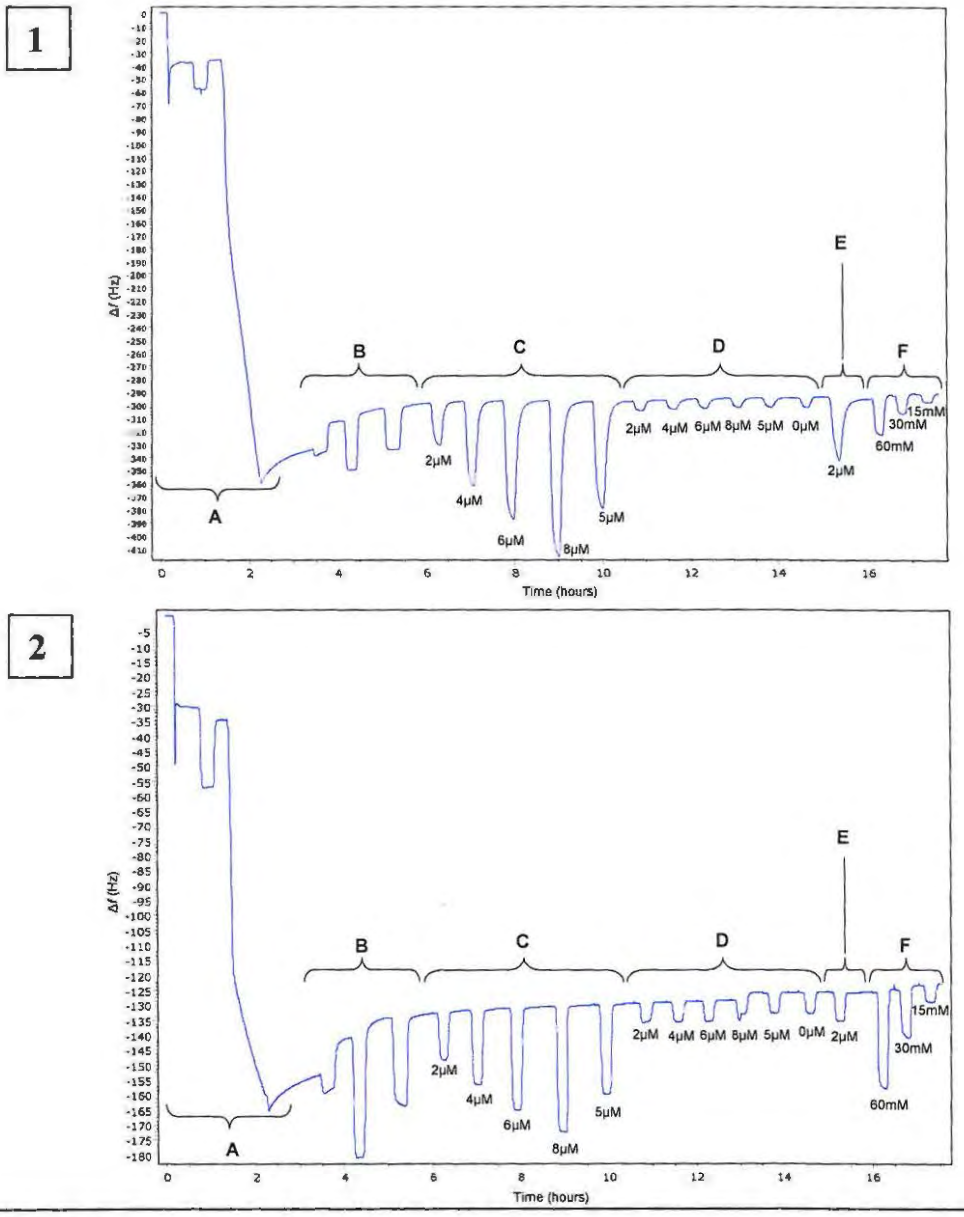


Figure 4.10: Change in resonance frequency of the Quartz Crystal Sensor in relation to time measured at the third overtone.
1) shows Tcj2 immobilised on the crystal and *M. sativa* Hsp70 passed over. 2) Tcj2 (H34Q) immobilized and *M. sativa* Hsp70 passed over. Change in frequency (Δf), in Hz, is shown on the Y-axis versus time on the X-axis. **(A)** indicates the lipid immobilization, Ni^{2+} doping of the NTA groups and the immobilization of Tcj2/Tcj2 (H34Q). **(B)** A series of 3 ATP washes (60mM ATP) to remove any co-purified DnaK. There is a substantial loss of mass after the ATP washes, indicating a loss of material (possibly DnaK). These washes stabilize the baseline and minimize the leakage of mass from the crystal sensor surface. **(C)** A number of concentrations of *M. sativa* Hsp70 (in 10 mM ATP, 100mM Tris-HCl (pH 8.0); 300mM NaCl). **(D)** The same concentrations of BSA were applied in the same solution constituents as for the Hsp70 as a no binding control. **(E)** 2 μM of *M. sativa* Hsp70 was applied without ATP in the solution (100mM Tris-HCl (pH 8.0); 300 mM NaCl). **(F)** Differing concentrations of ATP were applied to the system.

transient. The change in frequency and dissipation upon introduction of Hsp70 to the flow cell reverses during the wash step and returns to the original baseline frequency. The lack of permanent interaction of either Hsp70 or BSA with the resonator system provides evidence that the crystal surface is completely covered by the SPB, or at least that any gaps in the layer are equally inaccessible to both proteins. This is likely to be so as Hsp70 (~70 kDa) and BSA (~66kDa) are of a very similar molecular size in kilodaltons (**Section 4.3.1**).

Varying concentrations of ATP (in 100 mM Tris-HCl; 300 mM NaCl; pH8.0) were passed over the crystal and resulted in a dose response causing an increase in frequency shift with increase in ATP concentration (**Figure 4.10F and 4.11F**). It is therefore important to maintain the same concentration of ATP in all the protein solutions tested for interaction with the immobilized proteins in the presence of ATP. This was done for the data shown in **Figure 4.10** and **Figure 4.11**.

The addition of a 2 μM solution of *M. sativa* Hsp70 without ATP to both flow cell systems produced a substantially different shift in frequency relative to the 2 μM solution of Hsp70 with 10 mM ATP (**Figure 4.10.1E** and **Figure 4.10.2E**). The His-Tcj2 system showed a frequency shift of approximately 30 Hz for the Hsp70 and ATP solution relative to a shift of approximately 50 Hz for the Hsp70 solution lacking ATP. In contrast, the ATP containing Hsp70 solution (2 μM) produced an approximate 20 Hz shift from the baseline frequency in the His-Tcj2(H34Q) system, relative to an approximate 10 Hz shift in the ATP deficient Hsp70 solution (2 μM). However, as ATP appears to produce a frequency change, a better comparison may be to examine the 2 μM Hsp70 + 10 mM ATP response, subtracted from the 10 mM ATP alone response relative to the 2 μM Hsp70 excluding ATP response. This normalization produces a shift in frequency of 9.9 Hz (Hsp70 + 10 mM ATP) relative to 10 Hz (Hsp70 excluding ATP) for the His-Tcj2(H34Q) study and 24.2 Hz (Hsp70 + 10mM ATP) relative to 48.5 Hz (Hsp70 excluding ATP). This suggests that there is very little difference in the responses generated by the presence or absence of ATP for the Tcj2(H34Q) study. TheTcj2 study showed an almost 50% increase in the frequency shift response during exposure to Hsp70

without ATP relative to Hsp70 + ATP. The dissipation data shows a similar trend for this comparison.

Qualitative analysis of the change in dissipation data (**Figure 4.11**) did not supply any conclusive information regarding conformational changes of *M. sativa* Hsp70 or of His-Tcj2/His-Tcj2(H34Q) upon their interaction/exposure to one another. The tethered status of His-Tcj2/His-Tcj2(H34Q) generates a large dissipative effect in the system. The further increase in dissipation in a dose dependent manner during exposure to *M. sativa* Hsp70, relative to the lack of such a dose response during exposure to BSA appears to confirm the interpretation of the frequency shift data. The Hsp70 interacts with His-Tcj2/His-Tcj2(H34Q) and causes a further increase dissipation due to the increased coupled protein mass. However, the increase in dissipation associated with exposure of the BSA dilution series appears to be due to changes in the viscosity of the medium rather than due to the physical binding of BSA to Tcj2. Further viscoelastic modeling of this data would be required to determine any possible changes in the conformation of any of the proteins.

A difference in the shape of the frequency shift spikes resulting from the addition of *M. sativa* Hsp70 was observed for the His-Tcj2 (**Figure 4.10.1C**) relative to the His-Tcj2(H34Q) (**Figure 4.10.2C**) data. However, this appears to be due to differences in the crystal used in the experiment as the same difference was observed when differing concentrations of BSA were added to the respective systems. This same phenomenon was observed in the measurements of the dissipation of the two systems (**Figure 4.11.1** and **Figure 4.11.2**), but to a lesser extent. The application of ATP solutions (excluding protein) (**Figure 4.10.1B and F; Figure 4.10.2B and F**) did not show this difference in the frequency shift spike to the same extent as the addition of the protein containing solutions (*M. sativa* Hsp70 and BSA).

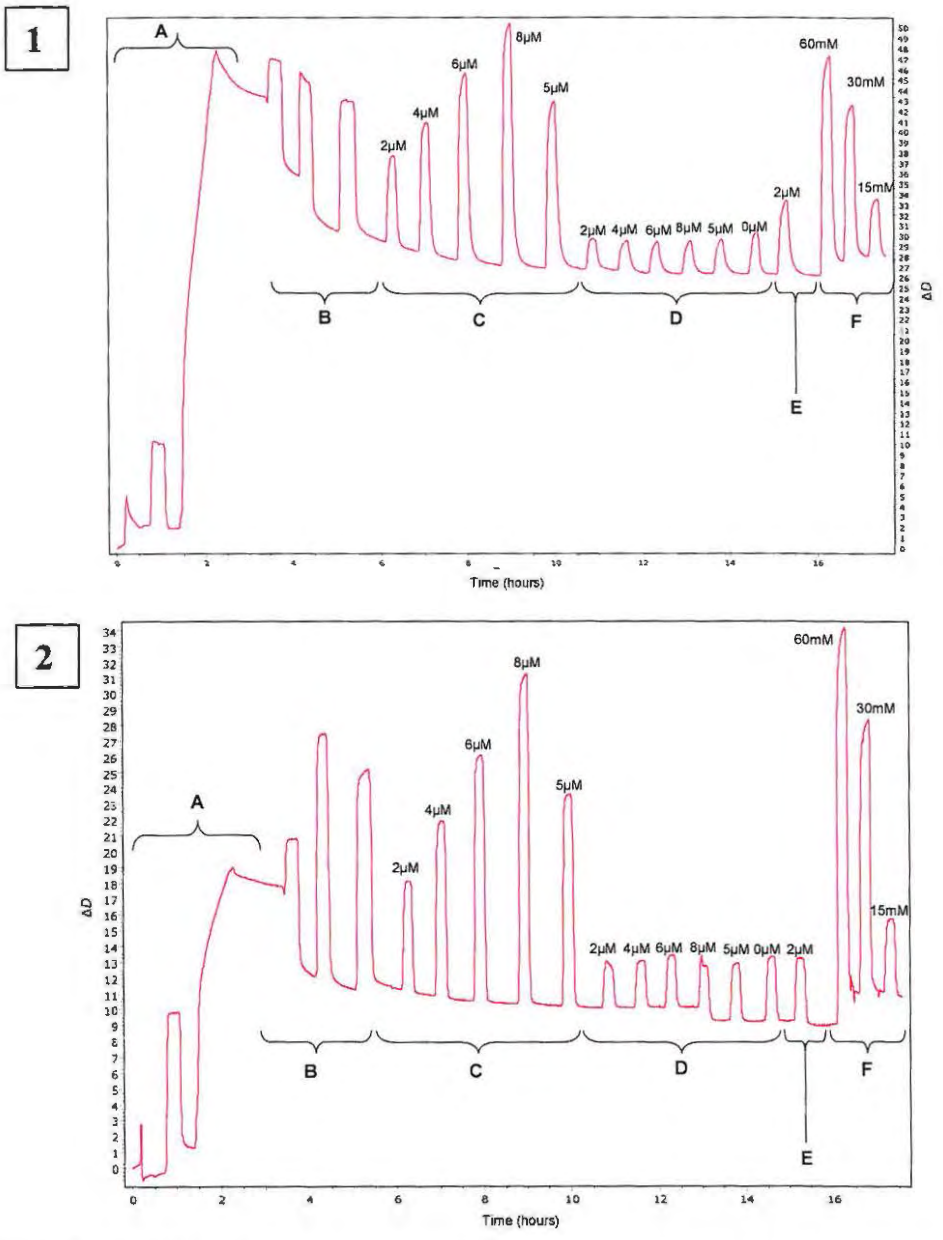


Figure 4.11: Change in the rate of resonance energy dissipation frequency of the Quartz Crystal Sensor in relation to time measured at the third overtone.
The change in dissipation measurements that were taken simultaneously with the change in resonance frequency data displayed in **Figure 4.10**. 1) shows the dissipation measurements for Tcj2 wild type immobilized onto the crystal sensor. 2) shows the measurements for Tcj2 H34Q immobilized onto the crystal sensor. Change in the dissipation factor (ΔD), in Hz, is shown on the Y-axis versus time on the X-axis. **(A)** indicates the lipid immobilization, Ni^{2+} doping of the NTA groups and the immobilization of Tcj2/Tcj2 (H34Q). **(B)** A series of 3 ATP washes (60mM ATP) to remove any co-purified DnaK. There is a substantial loss of mass after the ATP washes, indicating a loss of material (possibly DnaK). These washes stabilize the baseline and minimize the leakage of mass from the crystal sensor surface. **(C)** A number of concentrations of *M. sativa* Hsp70 (in 10 mM ATP, 100mM Tris-HCl (pH 8.0); 300mM NaCl). **(D)** The same concentrations of BSA were applied in the same solution constituents as for the Hsp70 as a no binding control. **(E)** 2 μM of *M. sativa* Hsp70 was applied without ATP in the solution (100mM Tris-HCl (pH 8.0); 300 mM NaCl). **(F)** Differing concentrations of ATP were applied to the system.

4.3.8) Comparison of the binding of *M. sativa* to His-Tcj2 and His-Tcj2(H34Q) in the presence of ATP

The different amounts of His-Tcj2 and His-Tcj2(H34Q) immobilized on their respective crystal resonators makes it difficult to compare the binding of Hsp70 to this protein and its mutant. An attempt was made to normalize the amount of Hsp40 bound in both systems using the following assumptions:

1) His-Tcj2 and His-Tcj2(H34Q) are of the same molecular size, structure and have the same hydration shell. In addition their binding to the nickel-doped NTA groups of the supported lipid bilayer is the same.
2) There are the same number of accessible NTA groups in both of the supported phospholipid bilayers on their respective crystals.

The total frequency shift from the start of His-Tcj2/His-Tcj2(H34Q) binding until the stabilization of the frequency after the ATP washes (**Figure 4.10.1B and Figure 4.10.2B**) was measured. This was approximately 265 Hz for His-Tcj2 (**Figure 4.10.1**) and approximately 95 Hz for His-Tcj2(H34Q) (**Figure 4.10.2**). The frequency shift was therefore 2.8 times greater for His-Tcj2 than for His-Tcj2(H34Q), implying that 2.8 times the amount of His-Tcj2 was bound (provided the above assumptions are valid). The frequency shift produced by each concentration of *M. sativa* Hsp70 and BSA in the presence of ATP (**Figure 4.10.1C and D; Figure 4.10.2C and D**) was measured. The frequency shifts measured for the binding of *M. sativa* Hsp70 to His-Tcj2 were then divided by 2.8 to account for the different amounts of His-Tcj2 and His-Tcj2(H34Q) that were bound. The resulting values were plotted as change in frequency (Hz) versus protein concentration (**Figure 4.12**). The plot of the unadjusted frequency shifts for *M. Sativa* Hsp70 binding to His-Tcj2 has a gradient more than double that of the plot for *M. Sativa* Hsp70 binding His-Tcj2(H34Q). However, the plot of the normalized frequency shift of *M. sativa* Hsp70 binding His-Tcj2 has a similar gradient to that of Hsp70 binding His-Tcj2(H34Q). This would suggest that both His-Tcj2 and His-Tcj2(H34Q) have a similar affinity for *M. sativa* Hsp70 *in vitro*. The plots for the different concentrations of BSA passed through both systems revealed gradients of almost zero, suggesting that no significant amount of BSA bound to the SPB or the immobilized Hsp40s. However, no

accurate deduction of affinities can be made without the elimination of the variables associated with using two different resonators and different lipid bilayers. Alternatively, a ratio of the frequency and dissipation measurements could be taken and plotted against the dilution series protein concentration. The resultant gradients for the Hsp70 dilution series in both cases were very similar to that shown in **Figure 4.12**.

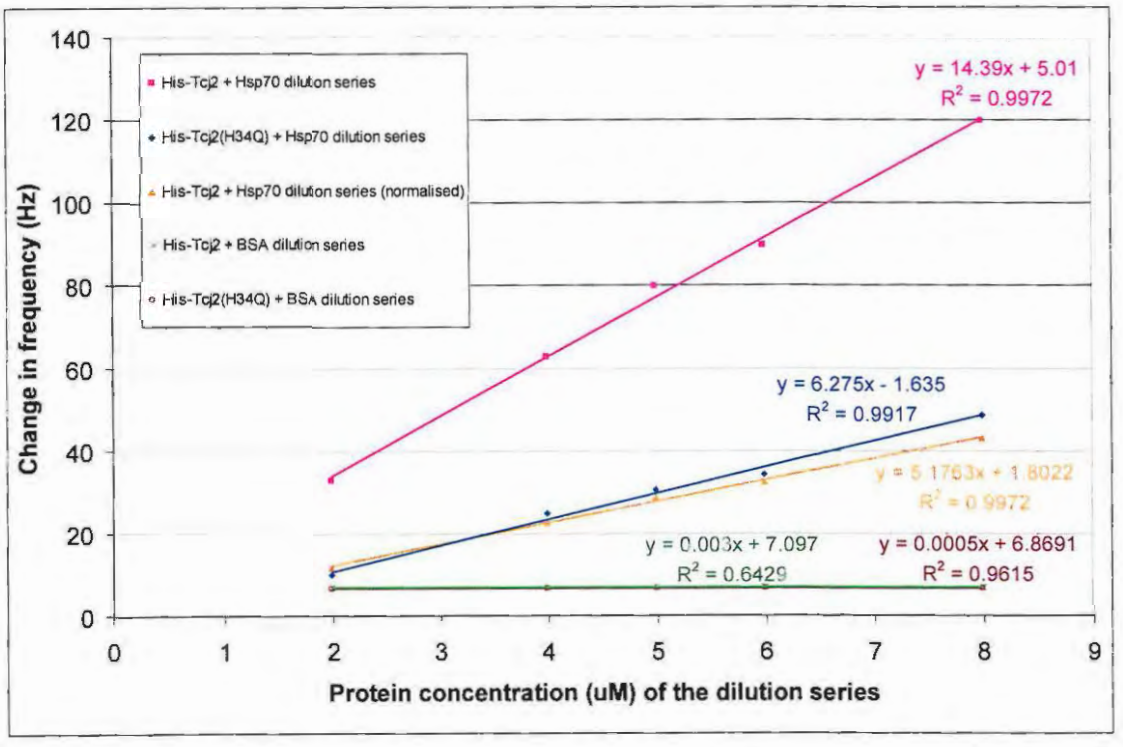


Figure 4.12: The comparison of the interaction of *M. sativa* Hsp70 and Bovine serum albumin (BSA) to surface immobilized His-Tcj2 and His-Tcj2(H34Q) protein films measured with the QCM-D.
The different amounts of His-Tcj2 and His-Tcj2(H34Q) that were immobilized onto their respective quartz crystal resonators was different. Therefore an attempt was made to normalize the shift in frequency caused by the M. sativa Hsp70 between the two systems (See text). His-Tcj2 + *M. sativa* Hsp70 diluition series. His-Tcj2(H34Q) + *M. sativa* Hsp70 dilution series. His-Tcj2 + *M. sativa* Hsp70 dilution series. His-Tcj2 + BSA dilution series. His-Tcj2(H34Q) + BSA dilution series. The lines represent a regression plot of each data series and the accompanying equation and R^2 values are displayed.

4.4) Discussion

4.4.1) *In vitro* ATPase assays

The basal ATPase activity for His-Tchsp70 that is reported in this chapter is vastly different from activities for this protein reported in the literature. GST-Tchsp70 had an ATPase activity of 520 nmol/min/mg (Olson *et al.*, 1994), while His-Tchsp70 was

reported to have a basal ATPase activity of 40 nmol/min/mg (Edkins *et al.*, 2004). This may suggest that the longer $(His)_6$-tag associated peptide of the His-Tchsp70 purified from *E. coli* BL21[pET14bTchsp70] relative to His-Tchsp70 purified from *E. coli* XL1blue[pQETchsp70] was able to interact with the substrate binding domain of His-Tchsp70 and stimulate an increased level of ATPase activity. The fold stimulation by His-Tcj2 (1.4 fold) is also lower than the 1.5 fold reported by Edkins and coworkers (2004). The basal ATPase activity of *Bos taurus* Hsc70 has reported as 0.2 nmol/min/mg (Minami *et al.*, 1996) to 1.08 nmol/min/mg (Chamberlain and Burgoyne, 1997). The basal ATPase activity of this protein was calculated as 0.72 nmol/min/mg, correlates closely with the data in the literature. The fold stimulation by His-Tcj2 (9.5 fold) also fits in the middle of the range of fold stimulation reported by a number of other Hsp40s (8-fold by DnaJ and GrpE, 6-fold by Hsp40 [Minami *et al.*, 1996] and 13 fold by Csp1 [Chamberlain and Burgoyne, 1997]). Although the basal ATPase activity of *M. sativa* Hsp70 was higher than the other Hsp70s in the study, these results were consistent with those obtained by others (38 nmol/min/mg; Louw *et al.* 2010). The ability of His-Tcj2 to stimulate the ATPase activity of all three of the Hsp70s tested, suggests that it is a functional protein *in vitro*.

4.4.2) The immobilization of His-Tcj2/His-Tcj2(H34Q) onto the surface of the quartz crystal resonator

Possible reasons for the differences in the lipid immobilization measurements (**Figure 4.9)** could be due to the age of the crystals. The data produced in **Figure 4.9.1** was measured on an older crystal than **Figure 4.9.2**. In addition, this older crystal could have been etched in the cleaning procedures it had been subjected to and nanotopography is known to affect layer immobilization (Pfeiffer *et al.*, 2008). This older crystal may also have been dirtier, and crystal cleanliness is known to affect the quality of the supported lipid bilayer that is formed (Gläsmaster *et al.*, 2002). This may have left some vesicles intact on the crystal surface.

The ATP washes only partially stabilized the baseline for the frequency and dissipation measurements, which could indicate that the number of washes should be increased to

five or six. As Tcj2 is thought to dimerise (**Chapter 2 section 2.3.9**), it is possible that the immobilized His-Tcj2 and His-Tcj2(H34Q) were dimers, where one or both monomers were tethered to the lipid bilayer. However, the H34Q mutation is unlikely to affect dimersiation, due to the active domains being in the C-terminal portion of the protein relative to the N-terminal J-domain containing H34/Q34.

4.4.3) The measurement of the interaction between His-Tcj2/His-Tcj2(H34Q) with *M. sativa* Hsp70

To the best of the authors knowledge, this chapter presents the first work showing an interaction between a full length Hsp70 and a full length Hsp40 *in vitro* using the QCM-D system and it should be highlighted that this work constitutes a preliminary study, that leaves many question unanswered.

Previous *in vitro* studies of the interaction of Hsp70 with Hsp40 were conducted using surface plasmon resonance spectroscopy, where prokaryotic DnaK mutant protein interaction with wild type DnaJ immobilized onto the chip surface was studied (Mayer *et al.*, 1999; Gässler *et al.*, 1998; Sugimoto *et al.*, 2007; Acebron *et al.*, 2008). Other studies involved the characterization of the interaction of wild type DnaK with immobilized mutant DnaJ proteins (Mayer *et al.*, 1999; Suh *et al.*, 1999).

Mayer and coworkers (1999) reported that a functional DnaJ protein is necessary for an interaction with Hsp70. They showed that there was no binding between DnaK and DnaJ H33Q. In contrast, data in this chapter tentatively shows the interaction of Alfalfa Hsp70 with His-Tcj2(H34Q) and His-Tcj2 in the presence of saturating levels of ATP. However, there are differences between the prokaryotic and eukaryotic Hsp40/Hsp70 systems, which may make them difficult to compare directly (e.g. the greater requirement for a nucleotide exchange factor in prokaryotic systems, Alberti *et al.*, 2003). The similar incremental dose response of both His-Tcj2 and its mutant for their interaction with *M. sativa* Hsp70 indicates that they may have similar affinities for *M. sativa* Hsp70. Therefore, the H34Q mutation may not affect the physical interaction of Hsp70 and Tcj2 at all, and that it is rather the positively charged residues of Helix II and Helix III of the J-

domain that are responsible for this physical interaction as was suggested by Gässler and coworkers (1998). Equivalent mutations in other Hsp40s have disrupted the ability of the Hsp40 to stimulate the ATPase activity of an Hsp70. Also, we have shown that the H34Q mutation disrupts the ability of Tcj2 to substitute for Ydj1 in vivo (Chapter 3). Taken together these data suggest that while mutations of the HPD motif (Such as H34Q in Tcj2/Tcj3) disrupt functional interaction of an Hsp40 with an Hsp70, they do not necessarily disrupt physical binding.

Unfortunately the experiments presented here can't rule out the possibility that His-Tcj2 and His-Tcj2H34Q are being recognised as an unfolded substrate by *M. sativa* Hsp70 as was suggested for the interaction between *E. coli* DnaJ and DnaK (Mayer *et al.*, 1999). There is therefore a need for confirmation that the proteins used were close a 100% native state. This could have been better achieved using circular dichroism which detects the secondary structure content of a protein solution and therefore is a more definitive measure of protein native state than light scattering.

The 2 μM *M. sativa* Hsp70 without ATP passed over the immobilized His-Tcj2 and His-Tcj2(H34Q) showed an interaction. It has been shown that DnaK does not interact with DnaJ in the presence of ADP (Mayer *et al.*, 1999; Suh *et al.*, 1999). ADP bound Hsp70 is reported to have an increased affinity for unfolded substrates, so it is surprising that the absence of nucleotide shows a higher affinity than DnaK/DnaJ in the presence of ADP, unless the DnaK is not recognizing DnaJ as substrate. For a more complete comparison of the effect of a lack of ATP or ADP in the system, a complete dilution series of *M. sativa* Hsp70 like the one performed in the presence of ATP would be required. The accompanying control dilution series with BSA would also be required.

The transience of the interaction between Hsp70 and Hsp40 make detection more complicated. This is especially the case when measurements are conducted with unfolded protein substrates or with ATP/ADP in the solution that alters the frequency/dissipation due to their effect on the solution viscosity.

The different amounts of His-Tcj2 and His-Tcj2(H34Q) bound to their respective quartz resonators makes it difficult to make a direct comparison regarding their affinity for *M. sativa* Hsp70. The attempt at normalization in section 4.3.9 is not ideal. Therefore voight viscoelastic modeling is required for estimation of the amount of His-Tcj2 and His-Tcj2(H34Q) bound to their respective crystals. This will allow more accurate comparison of the binding of the two proteins to Hsp70 and will also provide more information for the calculation of preliminary binding kinetics.

Chapter 5:
The *in vivo* Characterization of Tbj2 in the Proliferative Bloodstream Stage of *Trypanosoma brucei brucei*

CHAPTER 5
The *in vivo* characterisation of Tbj2 in the proliferative bloodstream stage *Trypanosoma brucei brucei*

5.1) Introduction

The characterization of a protein *in vivo* is preferable to *in vitro* characterization as this studies the protein in its natural cellular environment. However, it can often be difficult to study the workings of a protein in the complexity of the cellular environment and its incompatibility with *in vitro* assay detection systems. In addition, it is preferable to study a protein in the organism from which it originates (the homologous system) as opposed to a different organism (heterologous system). This reduces the chance of misinterpretation of signals such as sub-cellular localization signals or other differences in the cellular environment of the protein under study relative to its natural cellular environment.

Procyclic *Trypanosoma brucei brucei* has been cultivated *in vitro* in continuous culture in the laboratory since the early 1900's (Hirumi and Hirumi, 1989). However, bloodstream form cells were only successfully cultured continuously *in vitro* in 1977 (Hirumi *et al.*, 1977; Hirumi and Hirumi, 1994). This early *in vitro* culture of bloodstream form cells required the presence of mammalian feeder cells. The ability to culture bloodstream form cells without these feeder cells was first reported by Baltz *et al.* (1985) and Duszenko *et al.* (1985) and required only a 10% mammalian serum content in the culture medium. These protocols have since been modified (Hirumi and Hirumi, 1989; Hirumi and Hirumi, 1994). Culture of the proliferative bloodstream form cells in the laboratory is further complicated by the population growth arrest and differentiation of cells to the non-proliferative short stumpy forms (Vasella and Boshart, 1996). A density sensing mechanism utilizing a factor referred to as stumpy induction factor (SIF) and cyclic AMP (cAMP) has been implicated in this phenomenon of bloodstream form population control (Vasella *et al.*, 1997; Reuner *et al.*, 1997). Bloodstream form populations cultured *in vitro* therefore need to be maintained below a threshold cell density ($2x10^6$ cells/ml) in order to prevent the irreversible differentiation to the non-proliferative stumpy form.

RNA interference (RNAi) is a cellular reaction to the presence of double stranded RNA (dsRNA), in which mRNA with a sequence identical to one of the strands of a dsRNA molecule is degraded. The net result is cessation of translation of that particular gene transcript (Hannon, 2002). Discovery of this phenomemon was credited to Fire, Mello and coworkers (1998) for their observation of the phenomenon in *Caenorhabditis elegans* (Fire *et al.*, 1998). It has been proposed that this mechanism was developed in certain organisms for pathogen resistance (e.g. double stranded RNA viruses) or for the regulation of certain endogenous genes (Hannon, 2002). This natural phenomenon can be used to study gene function in a RNAi compliant-species, by introducing double stranded RNA sequences of which one strand is identical to the gene sequence intended to be silenced (Hannon, 2002). Such an RNAi system is present in *T. brucei* (Ullu *et al.*, 2004; Clayton *et al.*, 2005) and has been interpreted as a constitutive process to protect the genome against retroposon transcripts (Djikeng *et al.*, 2001; Patrick *et al.*, 2008). Methods that have been used to induce RNA interference in *T. brucei* include: 1) Electroporation of dsRNA or 2) the transcription of dsRNAs from a nucleotide construct integrated into the host genome (Tschudi *et al.*, 2003; Ullu *et al.*, 2004; Ngo *et al.*, 1998; LaCount and Donelson, 2001). The latter has been achieved by 1) transcription of dsRNA using opposing T7 polymerase promoters and 2) the production of a hairpin dsRNA structure by cloning the target sequence in tandem in the forward and reverse orientations (Ullu *et al.*, 2004; Tschudi *et al.*, 2003). Each of these has their own separate advantages. The introduction of synthetic dsRNA into *T. brucei* by electroporation produces only a transient RNAi response and does not require cloning manipulations (Ngo *et al.*, 1998). The *in situ* genetic construct producing dsRNA can be designed to be inducible (Alibu *et al.*, 2005) and hence more versatile than homologous recombination gene knockout which is complicated in a diploid organism. The construct producing the hairpin version of dsRNA is complicated to produce, involving insertion of the target sequence in forward and reverse orientation (LaCount and Donelson, 2001), while the opposing T7 promoter system is easier to clone and therefore more suited to high throughput analysis such as whole chromosome analysis. However, the opposing T7 promoter system is known to produce dsRNA even under non-induction conditions, which could be

problematic for essential genes (Wang *et al.*, 2000). This is due to the tet repressor not completely shutting down T7 polymerase production in this genetic system (Wirtz *et al.*, 1999). More recent RNAi genetic constructs for *T. brucei* make use of a ribosomal promoter thus eliminating the need to conduct experiments in a strain that expresses the T7 promoter and the problems of T7 polymerase toxicity at high expression levels (Alsford *et al.*, 2005; Alsford and Horn, 2008).

The sequencing of trypanosomatid genomes (El-Sayed *et al.*, 2005b; Berriman *et al.*, 2005; Ivens *et al.*, 2005; Peacock *et al.*, 2007) has produced a large amount of nucleotide and sequence information, without any corresponding functional information for many of these ORF's. The apparent absence of RNAi in *L. major* and *T. cruzi* (DaRocha *et al.*, 2004; Robinson and Beverly, 2003) and the overlap of a large proportion of the genes in these three organisms (El-Sayed *et al.*, 2005b) identifies *T. brucei* as the starting candidate organism for functional studies of trypanosomatid proteins, from which inferences about orthologues in the other trypanosomatids can be made. A large scale functional analysis has already been carried out on the ORF's of *T. bruce brucei* chromosome I (Subramaniam *et al.*, 2006) and there are countless other examples of *T. brucei brucei* protein characterization using RNA interference (Wang *et al.*, 2000; Alsford *et al.*, 2005; Motyka and Englund, 2004). *T. brucei* is therefore the best suited of *L. major*, *T. cruzi* and *T. brucei* to study trypanosomatid Hsp40 function by silencing their expression using RNAi.

In the proliferative life-cycle stages of *T. brucei brucei*, cell division differs significantly from many other eukaryotic cells. This is due to the differences in cell physiology relative to some eukaryotic cells. The *T. brucei* mitochondrion contain a concatenated network of circular chromosomes that are organised into a disk-like structure near the flagellar basal body called the kinetoplast (Robinson and Gull, 1991). The cells contain a subpellicular corset of microtubules that are thought to have a result in a mechanistically different process of cytokinesis (Hammarton, 2007). In addition to the nucleus, *T. brucei* also has a number of single copy organelles (mitochondrion, golgi apparatus and the

basal body/flagellar complex) all of which need to be duplicated correctly to produce viable progeny in the cell division process (Hammarton *et al.*, 2007).

The generalized description of the cell division cycle in eukaryotic cells is divided into four phases. The DNA replication phase (S-phase) separates two gap phases (G1 before and G2 after), which display a lack of apparent nuclear activity. Mitosis (the division of the replicated chromosomes into two separate nuclei) and cytokinesis (the division of the cytoplasm) are then initiated (Woodward and Gull, 1990; Alberts *et al.*, 1997). The start of the *T. brucei* cell cycle is defined as the maturation of the probasal body which allows the formation of a new flagellum (which continues to grow until cytokinesis occurs), followed by replication of the golgi apparatus (Hammarton, 2007). The need to replicate the extensive mitochondrial genome prior to cell division led to the description of two S-phases in *T. brucei*. The kinetoplast S-phase (S_k) and the nuclear S-Phase (S_n). S_k starts immediately prior to S_n, defined as the start of the S-phase in most cell division cycles (Woodward and Gull, 1990). S_k is completed before S_n and after its own G2 phase, the kinetoplast divides during the nuclear G2 phase. This produces a cell with two kinetoplasts and one nucleus (1N:2K). Most other eukaryotic cells studied to date have multiple mitochondria and show continuous mitochondrial genome replication at all stages of the cell division cycle (Guttes *et al.*, 1967; Braun and Evans, 1969; Meyer and Simpson, 1968; Schultz and Nass, 1969). Mitosis occurs at the end of the nuclear G2 phase and proceeds without chromosome condensation or nuclear envelope breakdown (Hammarton *et al.*, 2007; Galanti *et al.*, 1998; Alsford and Horn 2004). Cytokinesis in *T. brucei* occurs after the completion of nuclear DNA segregation and results in a 2N:2K state. The process is mechanistically very different to the actomyosin ring constriction mechanism observed in other eukaryotic cells (Hammarton, 2007). *T. brucei* cytokinesis is initiated with a cleavage furrow at the anterior end beginning at the site of attachment of the new flagellum to the cell body. Furrow ingression is unidirectional and is believed to include the modification of the microtubules in the subpellicular corset (Hammarton *et al.*, 2007). Abscission occurs along the flagellar axis of the dividing cell and appears to require an active flagellum for completion of the separation of the daughter cells (Ralston *et al.*, 2006; Broadhead *et al.*, 2006).

In addition, the duration of the *T. brucei brucei* cell division cycle differs from 8.5 hours in the procyclic cell forms to 6 hours in the bloodstream form (Hammarton *et al.*, 2007). The cell cycle is also not universal across trypanosomatid species. In *T. cruzi* and *Leishmania*, the timetable of cell cycle events do not follow all of the same events outlined for *T. brucei*. *Leishmania* is known to divide the nucleus before the kinetoplast (Cosgrove and Skeene, 1970; Hammarton *et al.*, 2007), while in *T. cruzi* flagellar elongation and flagellar pocket division occurs much later than in *T. brucei* (Elias *et al.*, 2007). The *T. brucei* cell cycle has also been found to have some novel regulatory checkpoints relative to the mammalian system, though many overlap (McKean, 2003; Hammarton *et al.*, 2003; Ploubidou *et al.*, 1999). There is also evidence that not all of these checkpoints are present in each *T. brucei* life-cycle stage (Hammarton *et al.*, 2003). The cell cycle of *T. cruzi*, *T. brucei* and *L. major* also display significant differences relative to one another (Ploubidou *et al.*, 1999; Hammarton *et al.*, 2007).

Cells of *T. brucei brucei* population, that are asynchronous in terms of their position in the cell cycle, can be categorized into various positions in the cell cycle by the use of monoclonal antibodies to specific organelles (e.g. golgi apparatus or mitochondrion) and the fluorescent staining of their nucleic acids (with e.g. DAPI (4,6-Diamidino-2-phenylindole)) (Woodward and Gull, 1990; Sherwin and Gull, 1989). The extensive kinetoplast genome and its replication during the cell cycle provides a novel and commonly used cytological marker for the identification of certain cell division stages (Woodward and Gull, 1990; Ploubidou *et al.*, 1999; Hammarton *et al.*, 2003 and Siegel *et al.*, 2008). Cells in G0 (non-dividing) or G1 (prior to nucleic acid replication) contain one nucleus and one kinetoplast (1N:1K). Cells in the G2 (gap phase after nucleic acid synthesis) have one nucleus and two kinetoplasts (1N:2K) (Woodward and Gull, 1990; Siegel *et al.*, 2008). Subsequent to mitosis and prior to cytokinesis cells have 2 nuclei and 2 kinetoplasts (2N:2K) (Sherwin and Gull, 1989; Woodward and Gull, 1990; Hammarton *et al.*, 2007). Differences in the ratio of 1N:1K, 1N:2K and 2N:2K between cell populations can indicate a cell cycle arrest in one of them. An example is, the knockdown of expression of the variable surface glycoprotein in bloodstream form parasites results in

an accumulation of 2N:2K cells in a population of cells, indicating a cell division arrest prior to cytokinesis (Sheader *et al.*, 2005).

The heat shock or stress response is a mechanism, found in most organisms, to deal with physiological and environmental stresses. This is achieved by an increase in expression of heat shock proteins (and others) and a decrease in activity of certain genes expressed during normal cellular activity (Lindquist and Craig, 1988; Morimoto, 1998; Morimoto, 2008). The transcriptional regulation of expression of heat shock proteins in most eukaryotic cells is controlled primarily by a proteins called Heat Shock transcription Factors (HSFs) (Morano and Thiele, 1999; Pirkkala *et al.*, 2001; Anckar and Sistonen, 2007). Under non-stress conditions the HSF is loosely bound to heat shock proteins, inhibiting their transcription stimulatory role. Under stress conditions, the available heat shock protein population is increasingly scavenged to bind the growing number of misfolded/unfolded proteins. The HSF's are released from the heat shock protein interaction and allowed to produce homo-oligomers, which are imported into the nucleus. In the nucleus the heat shock factors bind to the Heat Shock Elements (HSEs) (consisting of multiple contiguous inverted repeats of the sequence nGAAn) in the promoter regions of the stress-inducible heat shock protein genes and increase the transcription levels (Westerheide and Morimoto, 2005). As more heat shock proteins unbound by unfolded substrate proteins become available, HSFs become reassociated with these heat shock proteins and are unable to enter the nucleus to stimulate transcription of more heat shock proteins (Prahlad and Morimoto 2008; Westerheide and Morimoto, 2005). Small molecules are also known to have an effect on the regulation of expression of heat shock proteins (Westerheide and Morimoto, 2005). Different organs/systems in multicellular organisms have been known to react differently to different stresses (Prahlad and morimoto, 2008).

Some trypanosomated molecular chaperone proteins and cochaperones are known to increase expression during stress conditions. Hsp70 mRNA transcript (Engman *et al.*, 1995) and protein (Requena *et al.*, 1992) populations are known to increase in *T. cruzi* during heat stress. In *T. brucei brucei*, Hsp70s are also known to be stress inducible (Lee

et al., 1990) and putative heat shock elements have been detected upstream of some Hsp70 genes (Lee and Van der Ploeg, 1990). Of the *T. cruzi* Hsp40s examined experimentally in epimastigotes (Tcj1, Tcj2, Tcj3, Tcj4 and Tcj6), only Tcj2 (Type I Hsp40) and Tcj6 (Type II Hsp40) have displayed a significant increase in their mRNA transcript population, while information about the protein population of Tcj2 remains unknown (Tibbetts *et al.*, 1998; Salmon *et al.*, 2001). The mRNA transcripts of Tcj2, and of most other heat-inducible mRNAs in *T. cruzi*, contain direct repeats of the trinucleotide UUA in their 3' untranslated region (UTR) (Tibbetts *et al.*, 1998; Sullivan *et al.*, 1994). Total protein synthesis is known to decline in *T. cruzi* under conditions of heat stress (Alcina *et al.*, 1988), while the spliced-leader RNA silencing in *T. brucei* during stress conditions indicates that it has a similar characteristic (Lustig *et al.*, 2007).

The objectives of the experiments in this chapter were to characterize the function of Tbj2 *in vivo* in *T. b. brucei* by:

1) Determination of the ability of Tbj2 to be heat stress inducible.
2) Determination of the phenotypic effects of Tbj2 deficiency.
3) Assessment of the subcellular localization of Tbj2.

5.2) Materials and Methods

5.2.1) RNA interference target sequence design

The double stranded RNA interference target sequences to Tbj2 were designed using the RNAit computer software programme (Redmond *et al.*, 2003; http://trypanofan.path.cam.ac.uk/software/RNAit.html) to analyse the nucleotide sequence of Tbj2 posted on the GeneDB database (www.genedb.org; Hertz-Fowler *et al.*, 1999). RNAit enables the rapid generation of PCR primers for potential RNA interference targets in a segment of the *T. brucei brucei* genome sequence. RNAit screens these targets for their suitability based on their ability to cause RNAi knockdown in genes other than the target coding sequence. Two sets of primers were chosen for 2 individual RNAi targets (**Appendix 5.2**), two in the coding sequence of Tbj2 and one in the downstream untranslated region (UTR) of Tbj2.

5.2.2) Cloning of RNA interference constructs

The Tbj2 nucleotide coding sequence was amplified from *T. brucei brucei* Lister 427 genomic DNA (Kindly donated by Dr David Horn, London School of Hygiene and Tropical Medicine, UK), using PCR (**Appendix C.1.15**) primers Tbj2*Eco*RI,*Bam*HI, *Nde*IFor3 and Tbj2*Hind*III,*Xho*I,*Sal*IREV (**Appendix A2 Table A2**) designed from the Tbj2 nucleotide sequence deposited on the *T. brucei brucei* genome sequence database on GeneDB (www.genedb.org; Hertz-Fowler *et al.*, 1999) The PCR product was purified from a band resolved on an agarose gel (**Appendix C.1.9**) and cloned into pGEMT-Easy (**Appendix C.1.14**). The constructs were confirmed by sequencing. The internal RNAi targets were amplified from pGEM-T EasyTbj2 by PCR (**Appendix C.1.15**), using the primers (Tbj2RNAi2FOR, Tbj2RNAi2REV; **Appendix A2 Table A2**) designed with RNAit (Redmond *et al.*, 2003). The UTR target was amplified from *T. brucei brucei* Lister 427 genomic DNA, using Tbj2UTRdFOR and Tbj2UTRdREV (**Appendix A2 Table A2**) designed from RNAit (Redmond *et al.*, 2003) (**Section 5.2.1**). The PCR products were purified after resolution on an agarose gel **(Appendix C.1.10)**, which revealed them to have the anticipated size in bp. p2T7^{Tablue} (Alibu *et al.*, 2005; Kindly donated by Dr David Horn, London School of Tropical Hygiene and Medicine, London, UK) was digested with *Eam*1105I restriction endonuclease. This released a LacZ stuffer DNA sequence located between 2 opposing T7 promoters. This digest produces A/T overhangs in p2T7^{Tablue} that are similar to those found in pGEM-T Easy. The PCR products of the RNA interference targets were ligated (**Appendix C.1.13**) into p2T7^{TaBlue} and clones were screened by *Xho*I and *Bam*HI digest and nucleotide sequencing. Plasmid maps of the resultant plasmids are displayed in **Figure 5.1**.

5.2.3) *Trypanosoma brucei brucei* laboratory strains and culture media

Other than the Heat shock/stress inducibility experiments (**section 5.3.1**) and the subcellular localization experiments (**section 5.3.4**), the *T. brucei brucei* Single Marker Bloodstream form (SMB, *T7RNAP::TETR::NEO*) (Wirtz *et al.*, 1999) cell line was used. These cells express the phage T7 RNA polymerase and the tetracycline-repressor, using neomycin phosphotransferase as a single selectable marker for both of these genetic modifications. This cell line is a genetic modification of *T. brucei brucei* Lister 427

bloodstream form cells (Wirtz *et al.*, 1999). Wild type Lister 427 *T. brucei brucei* bloodstream form cells were used for the heat shock/stress inducibility experiments.

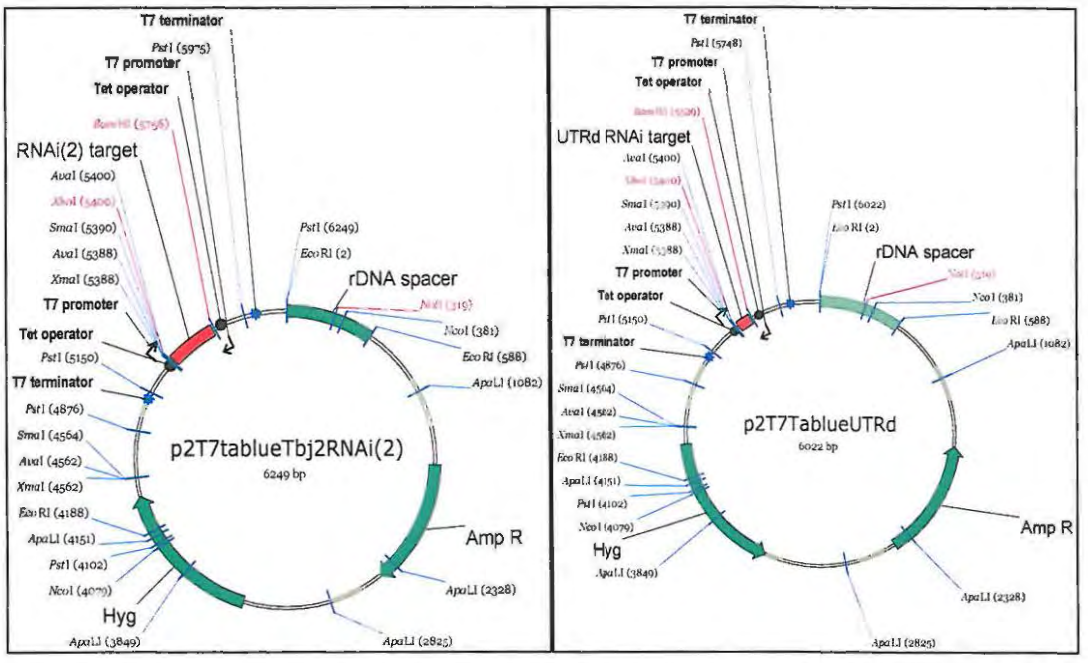


Figure 5.1: Plasmid map of p2T7tablueTbj2RNAi(2) and p2T7^{Tablue}UTRd. The plasmids are derived from the insertion of selected target sequences from Tbj2 (Tb927.2.5160) into the p2T7^{TAblue} vector (Alibu *et al.*, 2005; Subramaniam *et al.*, 2006). The RNAi target (shown in red) was inserted between the two opposing T7 promoters that would lead to the generation of forward and reverse RNA transcripts that anneal subsequent to transcription. The Tet operator downstream of each T7 promoter makes the RNAi system inducible with tetracycline. The T7 terminators prevent transcription into the remainder of the plasmid. An ampicillin resistance gene (Amp R) acts as a selectable marker for manipulations in *E. coli*. Hygromycin resistance (Hyg) acts as a selectable marker in *T. brucei brucei*. Linearisation within the rDNA spacer provides a location for integration into the host genome by homologous recombination. p2T7tablueTbj2RNAi(2) produced double stranded RNA homologous to a sequence inside the Tbj2 coding sequence, while p2T7^{Tablue}UTRd produced double stranded RNA homologous to a portion of the downstream untranslated region of the Tbj2 gene.

5.2.4) *Trypanosoma brucei brucei* laboratory culture

The cells were cultured in HMI-9 medium **(Appendix 5.1)** (Hirumi and Hirumi 1989; Hirumi and Hirumi, 1994) containing 10% Foetal Bovine Serum (heat inactivated at 56 °C). The medium was prewarmed to 37 °C prior to the addition of cells and incubated at 37 °C with a 5% CO_2 atmosphere with a starting density of $1x10^5$ cells/ml for 24 hours. Cell densities were assessed using a haemocytometer. *T. brucei brucei* bloodstream form cultures were maintained below $2x10^6$ cells/ml by transferring a small volume of the

population into fresh medium. This is necessary as the bloodstream form cells secrete a factor which inhibits cell division at high population densities (Reuner *et al.*, 1997; Vasella *et al.*, 1997). All cell manipulations involving open containers were performed using a sterile flow hood system.

5.2.5) Testing the induction of Tbj2/Tbj3 expression by heat stress

Lister *T. brucei brucei* 427 proliferative bloodstream stage cells were incubated at 37 °C, 5% CO_2 for 22 hours and then heat shocked for 1 hour at 42 °C. The control was maintained at 37 °C for this time. The lack of a second CO_2 incubator led to the sealing of the flasks to prevent differences in CO_2 exposure from affecting the experiment.

5.2.5.1) Preparation of total Protein and total RNA extracts

A 75 ml culture was prepared for each individual clone and incubated at 37 °C (5% CO_2) and grown to a density of ~$1x10^6$ cells/ml. 25 ml of each culture were harvested by centrifugation at 1000 x*g* for 10 minutes. The supernatant was removed and the cell pellets were resuspended in 5 ml PBS (10 mM Na_2HPO_4; 2 mM KH_2PO_4; 137 mM NaCl; 2.7 mM KCl). The resuspended cells were collected by repeating the centrifugation with the same force and duration. The supernatant was removed completely and the cells were resuspended in SDS-PAGE sample preparation buffer (0.0625 M Tris-HCl pH 6.8, 10% glycerol, 2% SDS, 5 % β-mercaptoethanol, 0.05% Bromophenol blue) to yield 500 000 cells/µl. Samples were stored at -20 °C until gel analysis.

5.2.5.2) Analysis of Tbj2 protein levels in different *T. brucei brucei* clones

Equal numbers of total cell extracts ($5x10^6$ cells per sample) in SDS-PAGE loading buffer were resolved on a discontinuous SDS-PAGE gel (**Appendix C.1.3**) and blotted on to a membrane for Western Blot analysis (**Appendix C.1.4**). Anti –Tbj2 (1:4000 dilution; design described in **chapter 2 section 2.3.11**) was used as primary antibody and goat anti-rabbit IgG conjugated to horse raddish peroxidase (1:5000 dilution) was used as the secondary antibody.

5.2.5.3) Preparation of total RNA extracts

The remaining 50 ml of each culture were collected with centrifugation at 1000 xg for 10 minutes. The cells were washed with 10 ml of PBS (10 mM Na_2HPO_4; 2 mM KH_2PO_4; 137 mM NaCl; 2.7 mM KCl) and transferred to a 15 ml falcon tube and the cells collected with the same centrifugation parameters. The supernatant was completely removed and the cells resuspended in 1 ml of PBS (10 mM Na_2HPO_4; 2 mM KH_2PO_4; 137 mM NaCl; 2.7 mM KCl) and transferred to 1.5 ml microcentrifuge tubes. The cells were collected with 10 000 RPM centrifugation for 5 minutes in a microcentrifuge, after which the supernatant was removed and the pellets were stored at -80 °C until the total RNA was extracted for Northern blotting. Total RNA was extracted from the cell pellets using a total RNA extraction kit (for extraction from mammalian cells) (Qiagen, USA). RNA concentrations were quantified using the Genequant (Pharmacia Biotech, GE Healthcare, USA) spectrophotometer and volumes for 1 µg of RNA per sample were calculated. These concentrations were confirmed using a conventional agarose gel, to ensure equal amounts of RNA per sample that was compared.

5.2.5.4) Northern Blot to assess Tbj2 and Tbj3 mRNA levels

5.2.5.4.1) Denaturing RNA agarose gel electrophoresis

Equal quantities of total RNA extracts (6 µg) from the individual clonal cell lines of *T. brucei brucei* were resolved using a 1.5% denaturing agarose gel (in 20 mM MOPS, 2 mM $NaCH_3COOH$; 1 mM EDTA; 1.2% formaldehyde). 11 µl of total RNA extract (0.55 µg/ µl) was added to 3 µl formaldehyde and 10 µl formamide and incubated at 65 °C for 10 minutes. After an incubation on ice for 10 minutes, the condensate was collected by a short centrifugation in the microfuge (2 to 3 seconds) and 2 µl of 10x Formaldehyde gel loading buffer (50% glycerol; 10 mM EDTA; 0.25% Bromophenol blue; 5 µg ethidium bromide) was added to each. The gel was prerun for 5 minutes (5V/cm) submerged in 1x Mops-Acetate-EDTA buffer (20 mM MOPS, 2 mM $NaCH_3COOH$; 1 mM EDTA). Samples were loaded and resolved at 5 V/cm until the bromophenol blue had migrated approximately 8 cm.

5.2.5.4.2) Northern Blot onto membrane

The resolved RNA was transferred from the formaldehyde-containing agarose gel to a nylon membrane using the capillary transfer technique. The gel was rinsed in Diethylpyrocarbonate (DEPC) treated water and soaked for 20 minutes in at least 5 gel volumes of 0.05 N NaOH. The gel was transferred to 10 gel volumes of 20x SSC (3.0 M NaCl; 0.3 M $NaCH_3COOH$, pH7.4) for 40 minutes. Excess gel area was trimmed away and placed on a piece of thick blotting paper (Soaked in 20x SSC). This was then placed on an elevated support in a dish, which was filled with transfer buffer (3.0 M NaCl; 0.3 M $NaCH_3COOH$, pH7.4) until just below the top of the support, ensuring that the ends of the blotting paper were resting in the transfer buffer reservoir. A nylon membrane, 1 mm larger than the gel (in both dimensions) was wet by laying on the surface of deionised H_2O and soaked in 10x SSC for 5 minutes. The membrane was carefully positioned on top of the gel and all bubbles are removed, after which 2 pieces of blotting buffer (Soaked in 10x SSC) were layered over the membrane. A stack of paper towels (5-8 cm high) was placed over the blotting paper and weighted down with a ~400g weight. After upward transfer was completed, the capillary transfer system was dismantled the membrane was removed and excess transfer buffer was allowed to drain away. The RNA was fixed to the membrane by irradiation at 254 nm for 1 minute 45 seconds at 1.4 J/cm^2. The gel was stained in 0.5 µg/ml ethidium bromide; 0.1M ammonium acetate for 45 minutes and photographed under ultraviolet light to confirm the RNA transfer.

5.2.5.4.3) Probe hybridisation and detection

The membrane was incubated for two hours at 68 °C in ~15 ml of prehybridisation solution (6x SSC; 2x Denhardts solution; 0.1% SDS; 100 µg/ml single stranded DNA). Denhardts solution: 1 % w/v Ficoll; 1 % W/V polyvinylpyrrolidone; 1% w/v Bovine Serum Albumin. The double-stranded ^{32}P-labelled DNA probes were denatured by heating at 100 °C for 5 minutes, followed by rapid chilling on ice. The probe was added directly to the prehybridisation solution and incubated at 65 °C for 12 to 16 hours. The membrane was washed in 0.2x SSC, 0.25 % SDS for 45 minutes, followed by a second wash of 30 minutes all at 65 °C. The membrane was dried on blotting paper and detection was produced using a phosphorimager/autoradiograph. A full length PCR derived

sequence (using primers J3NGFPF and J3NGFPR) of Tbj3 and a partial sequence of Tbj2 (digestion of the PCR derived full length Tbj2 (primers J2NGFPF and J2NGFPR) with *Cla*I restriction endonuclease) were used as the radiolabelled probes for the detection of the respective mRNAs. The Tbj2 sequence (891 bp to 1215 bp) was used to prevent hybridization to the dsRNA derived section of Tbj2. This may have significantly increased background signal on the Northern blot.

5.2.6) RNA interference in *Trypanosoma brucei brucei*

5.2.6.1) Preparation and Transfection of plasmid vectors into *T. brucei brucei* cells

Approximately $2.5x10^7$ to $4x10^7$ cells were required per transfection. Cells were harvested from an overnight culture (~$1x10^6$ cells/ml) by centrifugation at 1000 x*g* for 10 minutes in sterile falcon tubes. The cells were resuspended in a total of 15 ml of PBS (10 mM Na_2HPO_4; 2 mM KH_2PO_4; 137 mM NaCl; 2.7 mM KCl), transferred to a 15 ml falcon tube and collected by centrifugation at 1000 x*g* for 10 minutes. The supernatant was removed until approximately 1 ml remained. The cells were resuspended and transferred to a 1.5 ml microcentrifuge tube and the cells were collected by a centrifugation at 7000 RPM for 20 seconds. All of the supernatant was removed and the cells were resuspended in Amaxa transfection solution (100 µl per transfection) (Human T cell nucleofection kit, Amaxa Biosystems, USA). An aliqout of cells (100 µl) were mixed with 5 µg linearised plasmid DNA and transferred to a 2mm gap electroporation cuvette, ensuring that the solution was at the bottom of the cuvette. One pulse (1.4 kV, 25 µF) was administered, after which 1 ml of HMI-9 medium was mixed with the cells in order to facilitate their extraction from the cuvette. The contents of the cuvette were transferred to 36 ml of HMI-9 medium and incubated at 37 °C (5% CO_2) for 6 hours. A mixture of the relevant antibiotics (Hygromycin B, 2.5 µg/ml and G418 (Neomycin 2 µg/ml) was then added and subsequent to mixing, 1.5 ml of the culture were transferred to separate wells in sterile 24-well culture plates. These were incubated at 37 °C (5% CO_2) for 6 days and positive wells were visible as having dense populations of swimming (living) cells. At least 2 clones for each plasmid construct were picked for further analysis.

5.2.7) Assessment of knockdown of expression

Knockdown was assessed at the level of protein and mRNA, by using Western blotting and Northern blotting respectively. Total protein extracts and total RNA extracts were prepared as described in **sections 5.2.5.1**. *T. brucei brucei* SMB (Wirtz *et al.*, 1999) cell lines were prepared at $1x10^5$ cells/ml with tetracycline (1μg/ml) added to induce RNAi and duplicate cultures in which RNAi was not induced to serve as a negative control.

5.2.8) Testing of the phenotype of Tbj2 knockdown

5.2.8.1) Growth Curve assay

To determine if the knock down of Tbj2 has an effect on the growth rate of *T. brucei brucei*, the growth of the cell cultures were monitored for 72 hours (3 days). SMB wild type cells and two separate clones containing p2T7RNAi(2) were used. As in other experiments, neomycin and hygromycin B was used to maintain selective pressure on the SMB cell strain and p2T7 clones respectively. The two separate RNAi construct containing clones were split into two cultures in which one had RNA interference induced by the addition of tetracycline (1μg/ml final concentration) and one without RNAi induced. Each culture was produced in triplicate to ensure the reproducibility of results. Cells from starter cultures were diluted to $1x10^5$ cells/ml in 10 ml volumes and incubated at 37°C for 24 hours. The cell densities of each culture were assessed using a haemocytometer, after which a portion of these cultures was used to produce fresh $1x10^5$ cells/ml cultures that were incubated for a further 24 hours. Cell density was again assessed and the process was repeated for the following 24 hours. The dilution of the cultures after 24 hours is necessary as the bloodstream form cells produce a factor that induces cell cycle arrest at high cell densities (above $2x10^6$ cells/ml) (Reuner *et al.*, 1997; Vasella *et al.*, 1997).

5.2.8.2) Determination of the effect of Tbj2 knockdown on the cell cycle of *T. brucei brucei* proliferative bloodstream form cells

Approximately $1x10^6$ cells were collected by centrifugation at 1000 xg for 10 minutes and the supernatant removed until 100 μl remained, in which the cells were resuspended and transferred to a 1.5 ml microcentrifuge tube and collected using the same

centrifugation parameters. The supernatant was poured off until approximately one drop remained, in which the cells were resuspended. Formaldehyde solution (1ml) (1% in PBS) was added to each pellet and mixed by inverting the tubes four times and incubated at 4 °C for no more than 7 days. The cells were collected by microcentrifugation (10 000 RPM for 30 seconds) and the supernatant removed. The cells were washed by resuspending in two consecutive volumes of 1 ml chilled PBS and 1 wash with 0.5 ml of chilled BSA (1% in H_2O). Each wash was followed by a recollection of the cells with microcentrifugation and careful removal of the supernatant. The cells were resuspended in 30 µl of BSA (1% in H_2O) cell suspension and 5 µl of each cell line were applied to different wells in a twelve well microscope slide and allowed to dry at 37 °C up to 3 hours or until fixed. A mixture of alkaline glycerol and DAPI (4,6-Diamidino-2-phenylindole) stain with Vectashield mounting medium (Vector Laboratories Inc., USA, #H-1200) (5µl) was added to each well and a coverslip was applied and sealed with nail varnish. 200 cells per well were observed under a fluorescence microscope and assessed for their possession of 1 nucleus:1 kinetoplast (1N:1K), 1N:2K or 2N:2K to indicate their position in the cell division cycle.

5.2.9) *In vivo* localization of Tbj2 using Green Fluorescent Protein Fusions

5.2.9.1) The construction of pRPaGFPTbj2

To determine the sub-cellular localization of Tbj2 in proliferative bloodstream form cells, a fusion protein with GFP at the N-terminus of Tbj2 was constructed. The Tbj2 CDS was PCR amplified (**Appendix C.1.15**) from pGEM-T EasyTbj2 (1) (**section 5.2.2**), using primers j2NGFPF and j2NGFPR (**Appendix A2 Table A2**) with an annealing temperature of 55 °C. The resultant PCR product was purified from an agarose gel (**Appendix C.1.10**) and cloned into pGEM-T Easy. The resultant pGEM-T EasyTbj2(2) clone was digested with *Xba*I and *BamH*I (New England Biolabs, USA) and the resulting ~1200 bp fragment was purified from an agarose gel and ligated (**Appendix C.1.13**) into pRPaGFP (A kind donation of Dr David Horn, London School of Hygiene and Tropical Medicine, London, United Kingdom) (Alsford *et al.*, 2005; Alsford and Horn, 2008) which had been digested with the same two enzymes and purified from an agarose gel. The resulting plasmid construct, pRPaGFPTbj2 (**Figure 5.2**) was verified by restriction

digest analysis with *BamHI*/*Xba*I, *Bste*II, *Ear*I and *Sph*I and sequencing to ensure that the GFP-Tbj2 fusion is in the same reading frame.

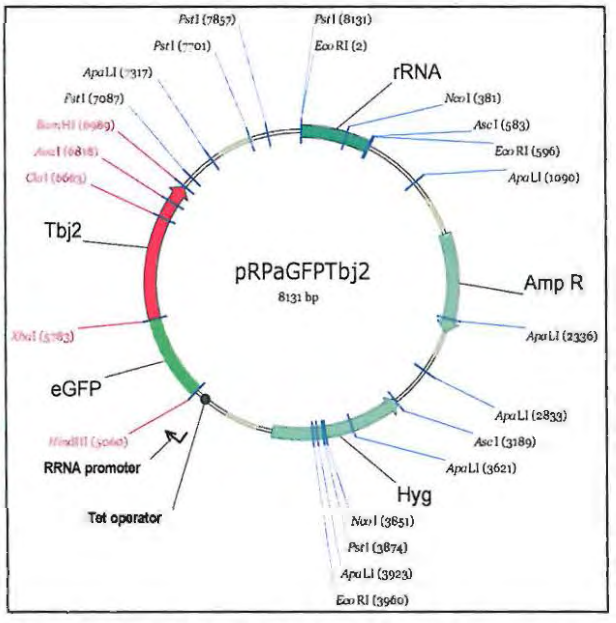


Figure 5.2: Plasmid map of pRPaGFPTbj2 fusion protein expression vector. GFP is fused to the N-terminus of Tbj2 and is controlled by a rRNA promomter. The plasmids are linearised with *Not*I prior to transfection and integrate into a non-transcribed rDNA spacer of the trypanosomal genome through homologous recombination (Alibu *et al.*, 2005). Successful integration of this plasmid into the host genome is selected through the Hygromycin selectable marker. The Ampicillin resistance selectable marker was used for plasmid construct manipulations in *Escherichia coli* laboratory strains.

5.2.9.2) Determination of the sub-cellular localization of Tbj2

T. brucei brucei Lister 427 proliferative bloodstream form cells were transfected with pRPaGFPTbj2 as per **section 5.2.6.1**. The cells were immobilized onto a glass microscope slide and stained with DAPI stain as per **section 5.2.8.2**. These cells were then observed under a fluorescence microscope with excitation at different wavelengths of light to view GFP or DAPI fluorescence.

5.3) Results

5.3.1) Tbj2 is heat inducible

The induction of Tbj2 expression under heat stress was assessed by incubating wild type *T .brucei brucei* SMB cells overnight at 37 °C for 22 hours and then exposing them to heat shock for 1.5 hours at 42 °C. Both protein (**Figure 5.3A**) and mRNA (**Figure 5.3B**) levels were assessed for Tbj2. As Tbj3 is possibly a paralogue of Tbj2 and is also predicted to localize to the cytoplasm (**See Chapter 2**), it was considered interesting to monitor the expression of Tbj3 in parallel with Tbj2. However, due to the absence of a reliable and specific antibody against Tbj3, only the Tbj3 mRNA levels were studied (**Figure 5.3B**).

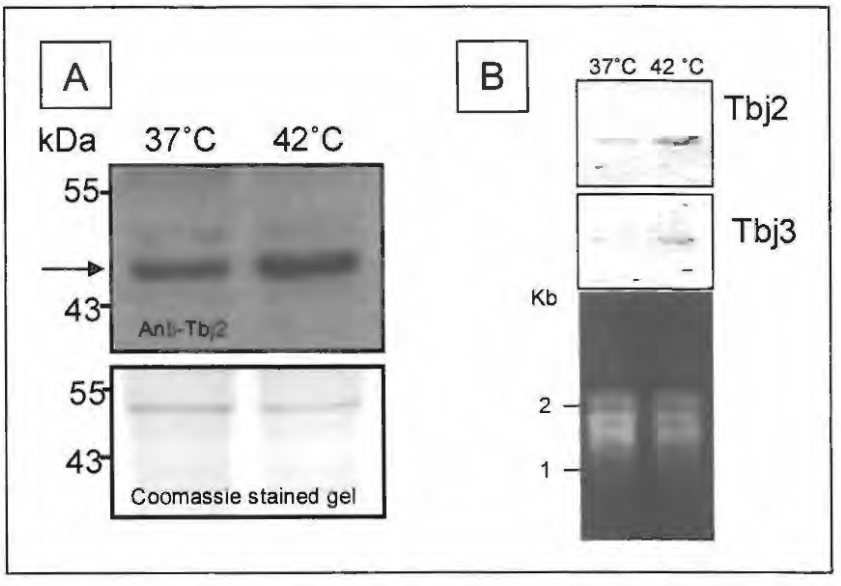


Figure 5.3: Tbj2 is heat inducible. (A) Total cell lysates of *T. brucei* SMB cells (5 million cells/lane) heat shocked at 42 °C or 37 °C for 1.5 hours after incubation at 37 °C for 22 hours. Tbj2 was probed with anti-Tbj2 peptide polyclonal antibodies (Upper panel). A duplicated SDS-PAGE gel (lower panel) was used to ensure that equal amounts of material were loaded in each lane. **(B)** Northern blot analysis to assess mRNA levels in the same total cell lysates from heat shocked (42 °C) and non- heat shocked (37 °C) cell cultures as shown in (A). Tbj2 – indicates the Northern blot probed with the Tbj2 probe and Tbj3 – indicates the Northern blot probed with the Tbj3 probe. An ethidium bromide stained agarose gel of the RNA samples was run to ensure equal amounts of RNA were from each sample were used for the Northern blots (data not shown).

The amount of Tbj2 protein in the total cell lysates increased significantly after heat shock (**Figure 5.3A**) indicating that Tbj2 is heat stress inducible. The control incubated at 37 °C also shows that Tbj2 is expressed in the cell under normal culture conditions, indicating constitutive expression in bloodstream form cells. Tbj2 therefore appears to be required for monitoring protein quality control of cells under non-stress conditions. The Northern blot detection of Tbj2 mRNA (**Figure 5.3B upper panel**) showed that the induction of Tbj2 expression by heat shock at 42 °C also results in a greater population of Tbj2 mRNA transcripts. The Northern blot detection of Tbj3 mRNA (**Figure 5.3B** middle panel) also indicates an increase in the Tbj3 mRNA transcript population during heat stress. However an increase in the mRNA transcript level does not necessarily equate to a greater amount of protein being produced, which is dependent on the half life of this particular protein under given cellular conditions. Therefore Western blot analysis of Tbj3 would be required to determine if the protein population is increased during stress.

Two putative nGAAn heat shock elements were observed upstream of the Tbj2 coding sequence (145bp and 180bp upstream of Tbj2 ORF start) derived from the *T. brucei brucei* genome database (Hertz-Fowler *et al.*, 1999). However, these do not seem to be as extensive as the ones upstream of some of the Hsp70 sequences in *T. brucei* (Lee and Van der Ploeg, 1990). The presence of these putative heat shock elements complements nicely with the heat stress inducible expression of Tbj2 shown in **Figure 5.3.** In addition to these putative HSE sequences, 10 TTA (UUA repeats on RNA) were detected between 251 bp and 286 bp downstream of Tbj2. This could possibly correspond to the UUA repeats found on mRNA transcripts downstream of many stress inducible heat shock protein ORF's in *T. cruzi* (Tibbetts *et al.*, 1998; Sullivan *et al.*, 1994). Until recently GeneDB had annotated a hypothetical protein (Tb927.2.5170) 122 bp downstream of Tbj2 ORF stop, preventing the identification of these putative features until now (**Appendix 5.2**). The nGAAn or UAA repeats were not clearly identified upstream or downstream of Tbj3. *T. brucei* Type I Hsp40s do not appear to be found in a gene cluster (GeneDB database, Hertz-Fowler *et al.*, 1999) as has been found for many Hsp70s (Louw, 2009).

5.3.2) Assessment of Tbj2 knockdown by RNA interference

Two RNAi target sequences were chosen and produced for the determination of the phenotype of Tbj2 knockdown (**Figure 5.4**). One of these, RNAi target 2 (RNAi (2)) (position **271bp** to **598bp**) were within the coding sequence of Tbj2. The second was targeted to the downstream untranslated region (UTR) of Tbj2. The downstream UTR was also targeted to explore the feasibility of using *T. brucei brucei* as an *in vivo* complementation system for studying *T. brucei brucei* Hsp40/Hsp70 protein-protein interactions, similar to the systems in *S. cerevisiae* and *E. coli* used in chapter 3.

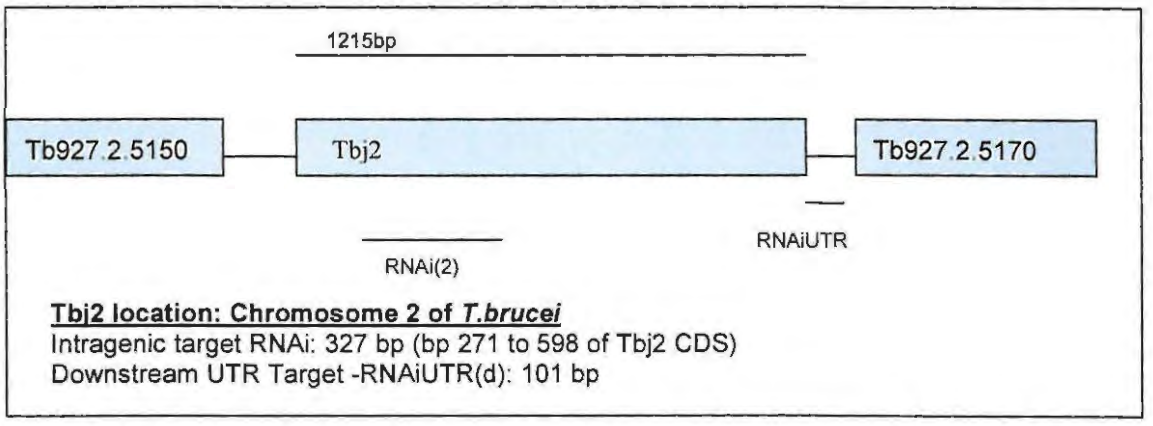


Figure 5.4: Scaled schematic diagram showing Tbj2 (GeneDB accession number Tb927.2.5160), the intergenic regions and portions of the flanking genes (GeneDB accession numbers Tb927.2.5150 and Tb927.2.5170). The location of the dsRNA targets and their sizes are also indicated.

Initial analysis of RNAi knockdown by detection of Tbj2 protein levels in two independent clones for each plasmid revealed that RNAi (2) produced the strongest reduction in expression of Tbj2 (**Figure 5.5**). The UTR RNAi target produced a less complete knockdown of Tbj2 (Data not shown). This low signal could possibly be due to the short length of this target sequence due to the short downstream UTR that was annotated for Tbj2 at the time. An RNAi target of 400bp to 600bp is considered optimal for RNAi in *T. brucei brucei.*

RNAi(2) was chosen as the RNAi target to use for further characterization of the knockdown of Tbj2, due to its very strong knockdown effect in all clones tested. **Figure 5.5A** shows a Western Blot detection of Tbj2 for both RNAi(2) clones in both the RNAi

uninduced and RNAi induced states. In comparison to the SMB wild type cells, the cell populations containing the RNAi(2) construct in which RNAi was not induced appear to cause a certain level of Tbj2 knockdown. This is likely due to the tet repressor not completely preventing transcription of the T7 polymerase and the RNAi(2) dsRNA (Wirtz *et al.*, 1999). The SDS-PAGE gel depicted below the Western Blot image (**Figure 5.5A**) shows that the loading of the total protein extracts was approximately equal. Although the first RNAi(2) cell population that had RNAi induced (I1) appears to show a slightly lower amount of protein, this is not great enough to correspond to the reduction in Tbj2 expression.

As Tbj3 is thought to be a paralogue of Tbj2 (**See Chapter 2**), it was considered interesting to monitor the effect of Tbj2 knockdown on the expression of Tbj3. However, problems with the anti-Tbj3 antibody prevented analysis of Tbj3 expression at the protein level. Northern blot detection of the Tbj2 mRNA levels (**Figure 5.5B**) showed almost complete removal of Tbj2 transcripts from the cell populations in which RNAi(2) dsRNA transcription was induced. Corresponding with the Tbj2 protein detection in **Figure 5.5A**, Tbj2 mRNA populations (**Figure 5.5B**) showed a reduction in Tbj2 mRNA in the RNAi(2) containing cell populations in which dsRNA production was not induced in comparison to the *T. brucei brucei* SMB wild type cell population. The Tbj3 mRNA population appears to increase in proportion to the decrease in the Tbj2 mRNA population, with the highest levels in the cell populations in which RNAi(2) dsRNA production has been induced. This could indicate that Tbj3 expression was increased in an attempt to compensate for the scarcity of Tbj2 in the cell populations in which Tbj2 expression was knocked down. Alternatively, Tbj2 knockdown of expression could produce a state of stress in the cells and result in a generalized increase in the production of all of the *T. brucei brucei* stress inducible molecular chaperone family. A denaturing agarose gel (**Figure 5.5B lowest panel**) is also shown to indicate that the amount of total RNA loaded for each lane is equal.

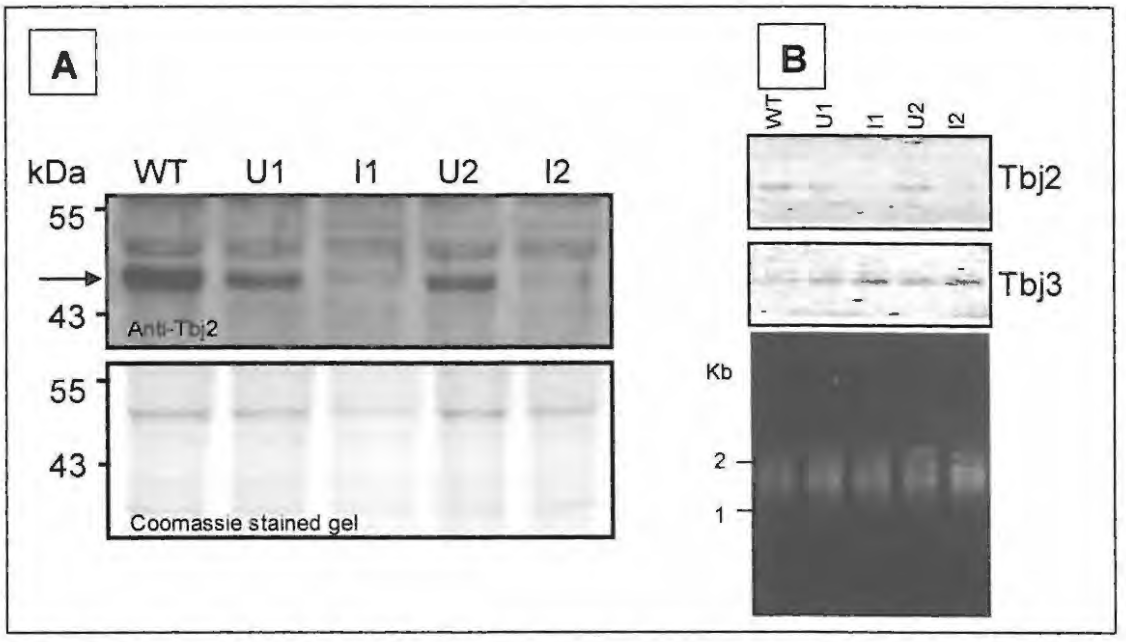


Figure 5.5: RNAi knockdown of Tbj2 was assessed at the protein level by means of a Western Blot probed with Anti Tbj2 peptide polyclonal Antibody (panel A) and a SDS-PAGE gel is shown below it to show that similar levels of protein were loaded into each lane. RNAi induced knockdown of mRNA levels was assessed through Northern Analysis (panel B) Tbj2- Northern blot probed with Tbj2 DNA probe and Tbj3- Northern blot probed with Tbj3 DNA probe. An ethidium bromide stained Agarose gel of the RNA samples was run to ensure equal amounts of RNA were from each sample were used for the northern blots (data not shown). WT- Wild Type SMB *T. brucei* cells; U1- uninduced RNAi knockdown of Tbj2 replicate 1; I1- Induced RNAi knockdown of Tbj2 replicate 1; U2 – uninduced knockdown of Tbj2 replicate 2; Induced RNAi knockdown of Tbj2 in replicate 2).

5.3.3) The phenotypic effects of Tbj2 knockdown

5.3.3.1) Effect of Tbj2 on cell population growth

Preparation of total protein and total RNA extracts to confirm knockdown of Tbj2 expression, revealed that induction of dsRNA production resulted in a very much reduced population growth. This finding was further characterized by growth curve studies (**Figure 5.6**). The growth of Tbj2 deficient and Tbj2 containing *T. brucei brucei* SMB cell lines was monitored for 3 days. As *T. brucei* stops dividing at a cell density of ~$2x10^6$ cells/ml, the cell populations required dilution every 24 hours. This accounts for the drop in cell population at 24 hour intervals. Induction of RNAi mediated knockdown of Tbj2 produced a severe growth defect in *T. brucei brucei*, resulting in slower population growth within the first 24 hours subsequent to RNAi induction and an almost

complete stop to cell division in the subsequent 2 days. One of the cell populations in which RNAi was induced (the second p2T7^{Tablue}RNAi(2) clonal population (clone 2 in **Figure 5.6**)) appeared to recover some of its capacity for growth on the third day. It is unclear as to whether this is selection for cells in the population that are able to shut down their RNA interference machinery or whether some other form of compensation for Tbj2 deficiency has been activated. This phenomenon has been observed in the knockdown of expression of other proteins in *T. brucei brucei* (Alsford *et al.*, 2005 GeneDB: [Tb10.6k15.3240]; Subramaniam *et al.*, 2006, http://trypanofan.path.cam.ac.uk/trypanofan/allCompleted/) and therefore possibly indicates partial downregulation of the RNAi machinery in these cell populations.

The two clonal populations of *T. brucei brucei* containing p2T7^{TAblue}RNAi(2) (population 1 and 2) in which dsRNA transcription was not induced showed a slightly slower growth rate than the wild type SMB cell line, most notable in the first 24 hours. It is possible that this is due to the slightly "leaky" nature of the T7 promoters in the p2T7^{Tablue} construct, which would result in the production of a small amount of dsRNA target (in the absence of tetracycline inducer) that would result in the partial silencing of Tbj2 in this cell line. This recovery that is observed in the growth of these p2T7^{Tablue}RNAi(2) containing cell populations could be due to an increase in the production of other chaperones to compensate for the reduced levels of Tbj2. Alternatively, the half life of Tbj2 protein in the cell or translation from the remaining mRNA could possibly be adjusted to compensate for the smaller Tbj2 mRNA population. This indicates that the *T. brucei brucei* cells are able to compensate for an approximate 50% reduction in the amount of Tbj2 (**see Figure 5.5**), but are unable to continue dividing if there is almost no Tbj2 present. Controls involving the addition of tetracycline to untransfected *T. brucei brucei* SMB cells and transfection of a vector excluding a dsRNA insert were not included as it is well accepted that the vector and the tetracycline do not affect *T. brucei brucei* growth (Horn, Pers. Comm., July 2007).

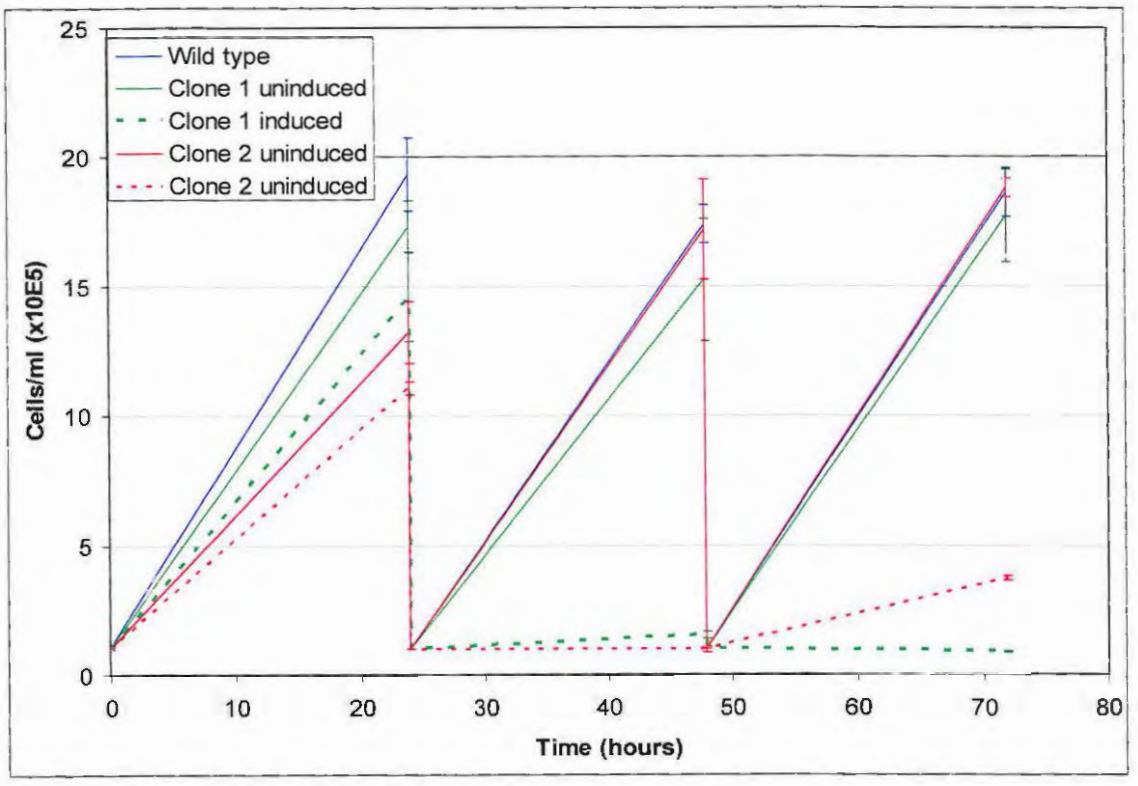


Figure 5.6: Growth curve of *T. brucei* wt SMB cells (Wirtz *et al.*, 1999) and two independent RNAi clones to the intragenic target (RNAi(2)) (induced with tetracycline and uninduced). The Y-axis units are x 10^5 cells/ml. Note the almost complete halt in cell division, after 24 hours, of the cells in which RNAi knockdown of Tbj2 was induced. Standard deviation bars indicating the variation between replicates are included for each data point.

It is important to consider that the knockdown of Tbj2 within these cells could result in the normal culture temperature of 37 °C being a stress temperature, and hence inhibit the growth of the cells at this temperature. This phenomenon has been observed in *S. cerevisiae* cells in which Ydj1 (the homologue of Tbj2) has been knocked out by homologous recombination (Caplan and Douglas, 1991; Atencio and Yaffe, 1992; Meacham *et al.*, 1999; Johnson and Craig, 2000). Ydj1 knockouts of *S. cerevisiae* are unable to grow at the normal culture temperature of 30 °C but grow at 23 °C to 24 °C (Caplan and Douglas, 1991; **Chapter 3**). It is therefore possible that Tbj2 deficient cells are unable to grow normally at 37 °C, but would be able to grow easily at temperatures lower than approximately 30 °C.

In an attempt to characterise this temperature sensitive phenotype, cell cultures were subjected to various temperatures (37 °C, 42 °C, 27 °C). However, due to insufficient CO_2 incubators essential for culturing of bloodstream form *T. brucei*, exposure to the various temperatures was limited to a maximum of 4 hours before replacement at 37 °C. RNAi was not induced in these experiments due to the severe growth defect observed in samples where Tbj2 was completely knocked down. The leaky expression of the dsRNA fragments against Tbj2 were thought to be significant enough to produce a marked difference in growth compared to wild type cells at these various temperatures. Results from these experiments proved to be inconclusive and possibly require prolonged exposure the various temperatures to see a conclusive effect, and perhaps in conjunction with the complete knockdown of Tbj2 with induction of RNAi.

5.3.3.2) The effect of Tbj2 Knockdown on the cell cycle

The number of kinetoplasts and nuclei present in a *T. brucei brucei* cell can be used as markers to determine the stage of the cell cycle. Differences in the proportions of cells, among different populations, containing 1N1K, 1N2K or 2N2K could indicate problems in the cell division cycle of one of the cell populations. Comparison of the proportion of cells with these different numbers of nuclei and kinetoplasts (**Figure 5.7**) did not reveal significant differences between SMB wt cells and heat shocked *T. brucei brucei* Lister 427 cell populations. Tbj2 knockdown showed a slight increase in the proportion of cells in the 1N:1K state and a slightly lower number of cells in the 1N:2K state. This could possibly indicate slightly fewer cells have entered the cell division cycle in Tbj2 deficient cells, possibly due to misfolded protein and protein aggregation stress. However, this difference could be classed as insignificant as the standard deviation indicator bars overlap. The RNAi knockdown samples were taken 24 hours post dsRNA induction of transcription. It would be interesting to see if there was any change to the proportion of cells in the different stages of cell cycle division if further samples had been taken at 48 hours post RNAi induction, as this was where a complete lack of growth was observed in the growth curve data (**Figure 5.7**).

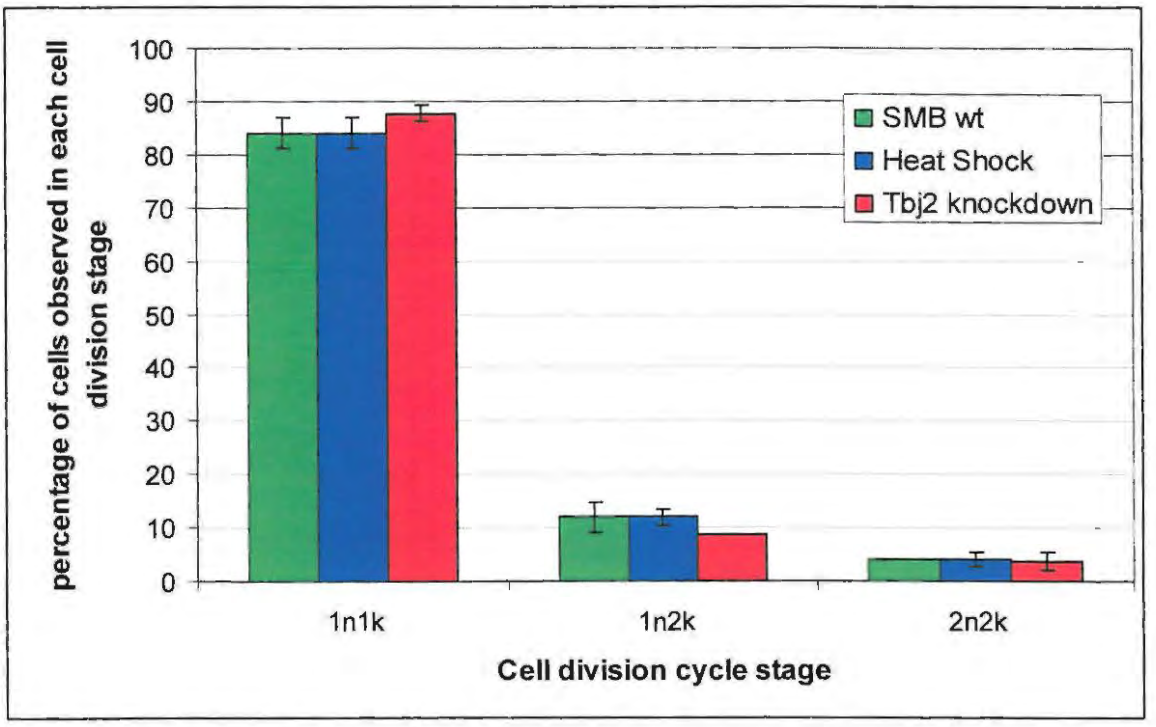


Figure 5.7: The effect of Tbj2 knockdown on the cell cycle of proliferative bloodstream form *T. brucei brucei*. The percentage of cells counted that contain 1N:1K, 1N:2K and 2N:2K were compared between *T. brucei brucei* SMB (Wirtz *et al.*, 1999), heat shocked *T. brucei brucei Lister* 427 cells and a Tbj2 knockdown cell line of SMB cells. The standard deviation between the 3 replicates for each value is shown.

The lack of difference between the proportions of cells in the different stages of division between the *T. b. brucei* SMB wild type cells and the heat shock treated cells is surprising. If general protein synthesis is likely to decline due to silencing of the mRNA spliced-leader sequence (Lustig *et al.*, 2007), then this should decrease the ability of cells to increase their size and divide. The cells were only exposed to heat stress of 42 °C for 1.5 hours, which may not be long enough to register the effect on the cell division within the population. The heat stress may have caused the dividing cells (1N2K and 2N2K) to arrest their division, but this would only be detectable if the stress was maintained for a longer period of time. It would be interesting to see the effect on the cell division cycle within the population if the cells were exposed to a higher stress temperature for the same length of time, or incubated at the same stress temperature for a longer period of time.

5.3.4) The *in vivo* subcellular localization of Tbj2 in proliferative bloodstream form *T. brucei brucei*

An important determinant in the cellular function of a protein is its subcellular localization. This determines in which cellular compartment a protein will be present and therefore which protein partners it will have to perform its function. Initial studies of the Tbj2 subcellular localization using immunofluorescent antibody staining resulted in the antibodies nonspecifically binding to the flagellar complex (Data not shown), which did not correspond to the *in silico* predictions of Chapter 2 or to the predicted orthologues such as Ydj1 (Caplan and Douglas, 1991; Caplan *et al.*, 1992a) (**See Chapter 2**). As it is known that many antibodies bind non-specifically to components of the flagellar assembly (David Horn, Pers. Comm; 2007), a green fluorescent protein (GFP)-Tbj2 fusion protein was therefore created to confirm Tbj2 cellular localization. An N-terminal fusion was chosen as the C-terminus of Tbj2 was known to have a CAAX prenylation motif (www.geneDB.org; Hertz-Fowler *et al.*, 1999). There was no known localization signal at the N-terminus of Tbj2.

The micrographs in **Figure 5.8** show that GFP-Tbj2 had a generalized cytoplasmic subcellular localization in the *T. brucei brucei* Lister 427 cells transfected with the pRPaTbj2GFP nucleotide construct. This localization does not appear to extend to the nucleus as there is a darker patch at the position where the nucleus is mapped to be through correlation with the DAPI stained image. This is especially visible in **micrograph set B** in **Figure 5.8.** The composite of the GFP-Tbj2 and DAPI images **(Figure 5.8 A4, B4** and **C4)** show the nucleus as a shade of orange, which could be misinterpreted as the presence of some GFP-Tbj2 in the nucleus. However, these images were produced using a fluorescence microscope without confocal abilities. Therefore the image most probably shows a section of cytoplasm over the nucleus which is on top of another section of cytoplasm.

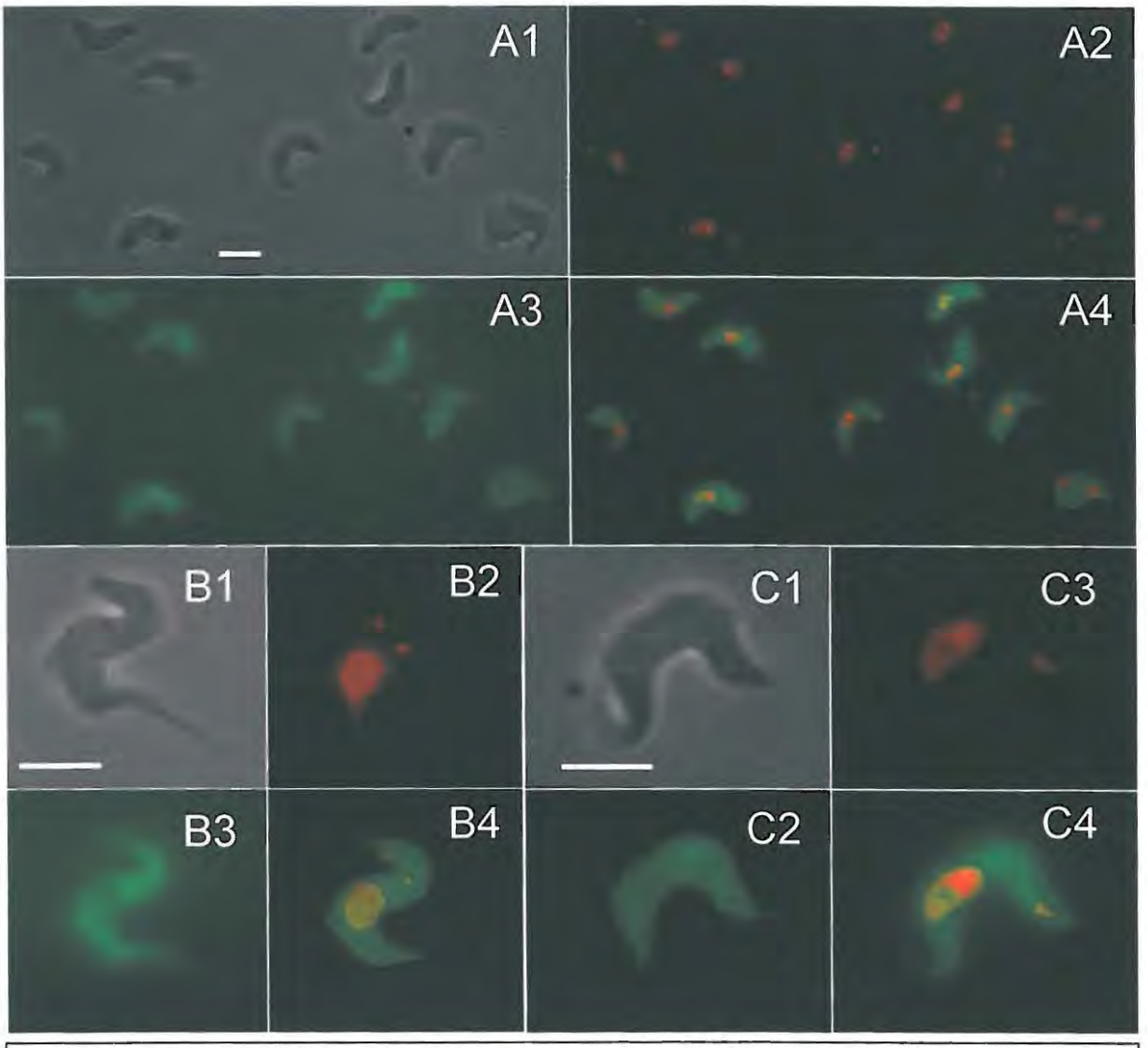


Figure 5.8: Fluorescence light micrographs showing the localization of GFP-Tbj2 in *T. brucei brucei*. A1,B1 and C1 are phase contrast images. A2, B2 and C2 are DAPI images showing the position of the nucleus and kinetoplast. A3, B3, and C3 show the localization of GFP-Tbj2. A4, B4 and C4 show a composite of the corresponding 2 and 3 images of each set. Scale bars indicate 5 μm.

There does not appear to be any difference in the localization of GFP-Tbj2 at particular times of the *T. brucei brucei* proliferative bloodstream form cell cycle. **Panel set B** shows a cell in the 1N:2K position, while **panel set C** shows a cell in 1N:1K position in the cell division cycle. The cell in the extreme bottom right hand side of the photos in **panel set A** may be in the 2N:2K state, but this is not clear enough to confirm. It may just be two 1N:1K cells in close proximity.

5.4) Discussion

Tbj2 is known to be expressed in both the procyclic and the bloodstream life cycle stages under normal culture conditions for these different stages (Vertommen *et al.*, 2008; Jones *et al.*, 2006). This indicates an important cellular function for Tbj2 even under non-stress conditions. Previous work on the heat stress induction of trypanosomatid heat shock proteins focused on mRNA and not protein (Tibbetts *et al.*, 1998; Engman *et al.*, 1995; Lee *et al.*, 1990). This study is therefore one of the first to observe the increase in the cellular protein population of a *T. brucei* Hsp40 after exposure to heat stress. Tcj6, a Type II Hsp40 in *T. cruzi*, has been shown to increase its protein population by two-fold subsequent to heat stress (Salmon *et al.*, 2001). The Tbj2 orthologue in *T. cruzi*, Tcj2, **(see Chapter 2, section 2.3.2)** has been shown to have increased mRNA transcript populations subsequent to heat stress, and the putative Tbj2 orthologue in *S. cerevisiae*, Ydj1, is known to be heat stress inducible at the protein level (Caplan and Douglas, 1992). This adds evidence to a similar function for these proteins. It is interesting that another putative Tbj2/Tbj3 orthologue in *H. sapiens*, Hdj2 also known as DnaJA1 (Kampinga *et al.*, 2009), was not stress inducible with regard to its mRNA transcript or protein populations (Davis *et al.*, 1998).

The increase in the mRNA population of Tbj3 could be caused by a number of possibilities: 1) Tbj3 mRNA increases to produce more Tbj3 in an attempt to take over the function of Tbj2. 2) Alternatively, the increase in Tbj3 mRNA could be a manifestation of the generalized increase in stability of the mRNA transcripts of heat shock proteins that has been proposed in trypanosomatids under stress conditions (Salmon *et al.*, 2001; Argaman *et al.*, 1994). This increase in stress could be due to the lack of Tbj2 in the cell and hence possibly more misfolded polypeptides. The inability of Tbj3 to rescue the severe growth defect produced by Tbj2 knockdown indicates that there is a large gap in the functional redundancy of these two proteins despite their predicted similar subcellular localization **(See Chapter 2, section 2.3.4)**, their high percentage amino acid identity (44%) and their similar domain structure. It is therefore likely that Tbj3 mRNA is being increased due to a generalized stress response. In the future, the

possibility of partial functional redundancy of Tbj2/Tbj3 could be further tested by knockdown of Tbj2 and Tbj3 expression in the same clonal population.

RNA interference has already been conducted on other Hsp40s (Subramaniam *et al.*, 2006; http://trypanofan.path.cam.ac.uk/trypanofan/main/), and knockdown of expression of a Type IV/I Hsp40 Tbj47 (Tb927.1.1230) was shown to cause reduced growth rate of *T. b. brucei*. Nevertheless, this is the first study to report the knockdown of a *T. brucei* Type I Hsp40, and to show that this Hsp40, Tbj2, was essential for the growth of the parasite.

The use of a UTR as a target for RNAi knockdown has been reported previously (Rusconi *et al.*, 2005) and could possibly have been useful to develop a strain of *T. brucei* for the *in vivo* complementation of Tbj2 in its homologous *in vivo* system. This would have been beneficial as **Chapter 3** highlights the disadvantages of using a heterologous *in vivo* complementation system. The reduced effectiveness of the UTR dsRNA target could have been due to its short length (101 bp) that was chosen due to the previous annotation of an ORF (Tb927.2.5170) 122 bp downstream of Tbj2 on GeneDB (Hertz-Fowler *et al.*, 1999). It would be interesting to see the effectiveness of a longer UTR based dsRNA target.

The growth curve data indicated that Tbj2 was essential for cell population growth of bloodstream form *T. b. brucei* at normal culture temperature of 37 °C. However, this could have been due to direct involvement of Tbj2 in the mechanism of cell cycle division, or because the shut down of cell division caused protein misfolding/aggregation stress. The comparison of the proportion of cells in various cell division cycle stages between *T. b. brucei* SMB wild type and Tbj2 deficient *T. b. brucei* SMB cells revealed no difference after a period of 24 hours (**Figure 5.7**). However, the real growth deficiency was detected after 24 hours. Deficiency of Ydj1 (a putative *S. cerevisae* orthologue of Tbj2) results in a temperature sensitive phenotype in *S. cerevisiae*, which is still able to grow at temperatures lower than the normal culture temperature of 30 °C at a reduced growth rate (Caplan *et al.*, 1992a; Johnson and Craig, 2000; Caplan and Douglas,

1991). This implies that Ydj1 is not directly involved in the cell division mechanism of *S. cerevisiae*. Extrapolation of this on to Tbj2 would imply that Tbj2 is involved in protein structural quality control in the cell and that lower culture temperatures may decrease the rate of protein misfolding/aggregation, thus allowing other chaperones to compensate well enough for the absence of Tbj2. This compensation does not appear to be possible at normal growth temperatures. Tbj2 deficient *T. b. brucei* may therefore be able to divide properly at lower incubation temperatures. A limited attempt was made to test this over a four hour period at 37 °C, 42 °C and 27 °C, but no conclusive result was observed. An obvious future experiment would be to incubate Tbj2 deficient cells at various temperatures for 24 hours or longer to determine if *T. b. brucei* is able to grow at temperatures below its conventional culture temperature even if it is at a reduced growth rate.

Although not all the available cell cycle markers (Robinson and Gull, 1991; Woodward and Gull, 1990) were utilized to detect a possible cell cycle arrest, it is possible that a change in the proportion of cells in the 1N1K:1N2K and 2N2K stages of cell division would indicate a cell division problem. It is likely that cell division arrest was not observed due to the utilization of cells that were exposed to heat stress for a short length of time and the populations in which RNAi was induced, had only had dsRNA induced for 24 hours. Lack of growth was only properly observed after 48 hours.

Cytoplasmic localization of GFP-Tbj2 is in agreement with the *in silico* predicted localization for Tbj2 generated in **Chapter 2 section 2.3.4.** The predicted yeast orthologue of Tbj2 (Ydj1) is known to have a generalized cytoplasmic localization under normal culture conditions (Caplan and Douglas, 1991; Caplan *et al.*, 1992a). The *H. sapiens* Hdj2 (also known as DnaJA1 [Kampinga *et al.*, 2009]), another putative orthologue of Tbj2, also has a general cytoplasmic localization (Davis *et al.*, 1998). An interesting feature of both Ydj1 and DnaJA1 is that their subcellular localization changes during heat stress (Caplan *et al.*, 1992a; Davis *et al.*, 1998). Ydj1 moves from a generalized cytoplasmic localization under normal conditions to association with the cytoplasmic face of the nuclear envelope and associated endoplasmic reticulum

membranes (Caplan *et al.*, 1992a), while DnaJA1 becomes associated with the golgi complex, perinuclear envelope and the nucleolus upon heat stress (Davis *et al.*, 1998). There has been debate as to whether this is due to membrane association through the farnesyl moiety attached to the CaaX box signal (Davis *et al.*, 1998; Caplan *et al.*, 1992a). It would be interesting to determine if Tbj2 or other Type I Hsp40s of *T. b. brucei* alter their subcellular localization after cellular exposure to heat stress.

It would have been preferable to use antibodies for the determination of the subcellular localization of Tbj2 as a GFP fusion operates on the assumption that the fusion does not mask any localization signal on the Tbj2 sequence. Tbj2 was annotated in GeneDB as having a tetrapeptide (MQAR) preceding the highly conserved sequence that it has in common with its orthologues (MVKET.....) (**See Chapter 2 section 2.3.9 Alignment Figure 2.5**). This did not correspond to any known localization signal in trypanosomes or other eukaryotes. Utilisation of a C-terminal Tbj2-GFP fusion in tandem to the GFP-Tbj2 N-terminal fusion would help to confirm this. It is possible that this MQAR tetrapeptide is a misannotation of the Tbj2 ORF.

Tbj2 is known to be expressed in both the procyclic (insect stages) and the bloodstream stages of *T. b. brucei* (Vertommen *et al.*, 2008; Jones *et al.*, 2006). However, given the differences in cell physiology and environment between the relative stages, it may be difficult to predict whether the protein would behave in exactly the same way in the insect stages compared to the bloodstream form. Investigation of the Tbj2 deficient phenotype, the reaction of Tbj2 to heat stress and its subcellular localization is required for insect stage cells. In addition, it would be interesting to similarly characterize all of the Type I Hsp40 proteins in *T. b. brucei* and this is likely to clarify the relative function of Tbj2 and Tbj3.

Chapter 6:
Final Conclusion and Future Work

CHAPTER 6
Final Conclusions and Future Work

Chapter 2: Bioinformatic comparison of the Tryanosomal Type I Hsp40s

Comparison of the Trypanosomatid species Hsp40 families with those of other organisms showed that the number of these proteins per species can differ greatly. The eight trypanosomatid species examined had approximately the same number of Hsp40s encoded on their nuclear genome, with the exception of *T. cruzi*, whose genome contains a number of gene duplications. Many of the findings and predictions made from the *in silico* analysis of the trypanosomal Type I Hsp40s need to be confirmed experimentally *in vivo* and *in vitro*. *In silico* analysis is useful in processing large amounts of genome/protein sequence information, but there are limitations to this system. An example is the prediction of the subcellular localization of the trypanosomal Type I Hsp40s. The use of an *in silico* subcellular localization prediction tool not specific for trypanosomal systems could mean that localization signals of the trypanosome system are not recognised by the programme and false predictions are made. The *in silico* characterization of genome information allows the researcher to more effectively prioritise which proteins or genes should be characterised experimentally. A future study involving this approach would be to identify the trypanosomal Hsp40s without a homologue in the human genome. These could possibly have a novel function in the trypanosome life cycle of this parasite that may have medical significance. In conjunction with this, the complete categorization of the trypanosomal Hsp40 families into Type II, Type III and Type IV Hsp40s still needs to be completed.

The hypothesis that the J66 protein family (and TbgambJ2/TbgambJ3) is the C-terminal section of an Hsp40 that interacts with a separately transcribed and translated J-domain makes this group of Hsp40s of fundamental interest to molecular chaperone research. Examples of these fractured Hsp40s have been identified (Bursać and Lithgow, 2007) Many pathogens are being implicated in the remodeling of the behaviour and/or

physiology of their host. *Toxoplasma gondii* is reported to change the behaviour of infected rhodents so that they are more easily eaten by cats (Berdoy *et al.*, 2000). *Plasmodium falciparum* is known to alter the physiology of the erythrocytes that it infects using exported molecular chaperones (Maier *et al.*, 2009). It is tempting to hypothesise that some trypanosomes might be shown to do the same thing in subsequent years. The J4 Type I Hsp40 group was found to have orthologues in those trypanosomatids that had an intracellular life cycle stage. Characterisation of this particular Type I Hsp40 family may reveal important details of the intracellular stage for these parasites.

Chapter 3: Analysis of the amino acids of the Tcj2 and Tcj3 J-domains important for the interaction with Hsp70

The *in vivo* complementation data in chapter 3 supports the hypothesis that Hsp40s interact with their partner Hsp70 through the J-domain to produce a functional interaction. Despite the use of different J-domains, certain mutations (R27A, H34Q and K37A) consistently abrogated the ability of the Hsp40 to interact functionally with Hsp70. This may suggest that certain aspects of the interaction between the two proteins is evolutionarily conserved from prokaryotic systems to eukaryotic systems. These include the need for the histidine of the HPD motif and the positively charged residues in Helix II and Helix III for functional interaction with Hsp70. It should be highlighted that all the Hsp40s in these experiments were Type I Hsp40s (both those tested, Tcj2 and Tcj3, and those substituted, *A. tumefaciens* DnaJ and Ydj1). Differences in the ability of the Helix IV mutation, R63A, to functionally interact with DnaK/Ssa1, supports the hypothesis proposed by Hennessy and coworkers (2005b) that Helix IV is involved in defining with which Hsp70 a given Hsp40 can interact. These differences also highlight the potential problems associated with conducting *in vivo* functional substitution assays with a protein in a non-homologous system. Additionally, such assays take place in the complex mixture of the cell, where there could be factors (e.g. other chaperone proteins) that mask the effect of a mutation in a particular protein and its ability to functionally interact with other proteins. These assays do not give conclusive information about the binding kinetics or affinity of interaction of these proteins. Farnesylation appears to be important for the *in vivo* function of Tcj2 and implies that membrane association is

integral in its function. Farnesylation has been shown to be important for the function for a number of proteins and inhibitors of this process have been investigated as possible means to treat cancers, malaria and trypanosomiasis (Kraus *et al.*, 2009; Eastman *et al.*, 2006).

As a future experiment, it would be interesting to determine the ability of Tcj3-Agt or full-length *A. tumefaciens* DnaJ to functionally substitute for Ydj1 in *S. cerevisiae*. *In vitro* binding studies using QCM-D and SPR could be used to determine how many of the positively charged residues of Helix II and Helix III of the J-domain are necessary to maintain a physical interaction with a partner Hsp70 and how the kinetics of this physical interaction are affected.

Chapter 4: The *in vitro* characterization of the interaction between Tcj2 and Hsp70

The interaction between Hsp70 and Hsp40 is a complex process, involving not only the proteins themselves, but also the presence of nucleotides (ATP/ADP) and unfolded peptide substrate. In addition it has been speculated that there are multiple contact points between the two proteins (Davis *et al.*, 1999; Mayer *et al.*, 1999; Suh *et al.*, 1999). The transient nature of the interaction between Hsp40 and Hsp70 adds a further complication to this study. Therefore to elucidate the process of the binding kinetics of these two proteins, multiple aspects need to be examined. Although the data in Chapter 4 suggested that His-Tcj2/His-Tcj2(H34Q) interact with *M. sativa* Hsp70 in the presence and absence of 10 mM ATP, there is still speculation as to whether this Hsp40 and its mutant are recognised as substrate by *M. sativa* Hsp70. This could possibly be determined (in addition to circular dichroism studies) by conducting competitive binding studies between His-Tcj2 and a peptide (e.g. σ^{32} [Mayer *et al.*, 1999]) for the interaction with *M. sativa* Hsp70. Competition between these two molecules for interaction with Hsp70 could indicate that His-Tcj2 is being recognised as substrate, as was concluded for *E. coli* DnaJ by Mayer and coworkers (1999). However, the recognition of part of the Hsp40 as substrate by Hsp70 has been postulated as an integral function for this protein (Hennessy *et al.*, 2005b; Suh *et al.*, 1999). It has also been proposed that Hsp40s containing a

peptide binding cleft in their C-terminal domain I recognize part of Hsp70 as substrate. This is thought to enable the transferal of unfolded proteins from Hsp40 to Hsp70 (Li *et al.*, 2009).

In addition to studying the interaction between His-Tcj2/His-Tcj2(H34Q) with Hsp70 in the presence of ATP, studies should be done in the absence of ATP, presence of ADP and in the presence of unfolded protein substrates to deliver a more complete picture of the binding kinetics of His-Tcj2/His-Tcj2(H34Q) with *M. Sativa* Hsp70. Each of these should be done as a dilution series of varying concentrations of Hsp70 and accompanied by the corresponding dilution series of BSA. The use of non-hydrolysable ATP (ATPγS) could compensate for the effect that ATP has on the viscosity in the no ATP studies. In addition, it should be ascertained as to whether His-Tcj2(H34Q) is able to functionally interact with *M. sativa* Hsp70 to stimulate its ATPase activity. It is likely that it would not be able to, as functional interaction between Tcj2(H34Q) did not occur during *in vivo* complementation studies in *S. cerevisiae* (*Chapter 3*).

The QCM-D binding assay for Hsp70/Hsp40 interaction studies has a number of problems that need to be addressed in future work. It is required that as many variables as possible are removed from the comparison of the interaction of His-Tcj2 and its mutants with Hsp70. This could possibly be done by immobilizing the Hsp70 onto the lipid bilayer and passing the Hsp40s over. However, this would require the use of a His-tagged Hsp70 and His-tag deficient Hsp40s. This would also confirm that the results obtained from the other system are not an artifact of the system. Perhaps an alternative method of immobilization of Hsp40/Hsp70 to confirm results could be used. For example, Hsp40 specific antibodies could be immobilized onto the crystal, BSA used to adsorb and thus block any spaces between the antibodies on the crystal surface before binding Hsp40. Hsp70 could then be passed over and binding to Hsp40 assessed. However, this is assuming that all proteins in the system are native, and that antibody binding does not compromise the ability of a target protein to function.

The use of Voight viscoelastic modeling is necessary to estimate the amount of protein immobilized onto the lipid-NTA sites and then the amount of protein bound to these proteins. QCM-D shows great potential for protein-protein interaction studies. However, it is becoming apparent that a multi-technique approach to these studies may be preferable. QCM-D and Surface Plasmon Resonance spectroscopy are viewed as complementary techniques giving slightly different information about the molecules involved in the binding. Both measurements should perhaps be used to determine the true kinetics of the binding events (Reimhult *et al.*, 2004; Su *et al.*, 2005). The supported lipid bilayer formation can be performed using SPR (Ohlsson *et al.*, 1995) and therefore the QCM-D studies can be replicated on the SPR to provide slightly different information. These binding studies should be done with a dilution series of the protein being passed over the immobilized protein in the presence and absence of ATP, and in the presence of ADP, to examine the interaction between these two proteins through the course of the Hsp70 chaperone cycle.

Chapter 5: The in vivo characterization of Tbj2 in the proliferative bloodstream stage of *Trypanosoma brucei brucei*

The *in vivo* characterization of Tbj2 indicated that this protein was expressed under homeostatic conditions in *T. b. brucei*, but the expression of Tbj2 was induced after a short exposure to heat stress. Tbj2 deficiency leads to a severe growth defect of bloodstream form cells at 37 °C. Tbj2 shows a generalized cytoplasmic localization in *T. b. brucei* bloodstream stage cells at normal incubation temperatures of 37 °C. These findings, along with the knowledge that Tbj2 is expressed in both the insect and the bloodstream stages of *T. b. brucei* (Vertommen *et al.*, 2008; Jones *et al.*, 2006), indicates that Tbj2 has an important function for the parasite. However, assessment of the effect of knockdown of expression of Tbj2 in the insect stage of the parasite have the same effects as for the bloodstream stages. Much of the data collected in this chapter adds supporting evidence for the classification of Tbj2 as an orthologue of Ydj1, as the this yeast Type I Hsp40 has many of the same functional characteristics *in vivo* in the *S. cerevisiae* system.

Many of the possible future experiments for the characterization of Tbj2 *in vivo* have been proposed in the discussion of Chapter 5. However, an important future experiment in the *in vivo* characterization of Tbj2 would be to test the thermosensitivity of *T. b. brucei* populations in which Tbj2 expression has been knocked down. Using a broader perspective, RNA interference could be used to test the function of all the Hsp40s in *T. b. brucei* in both the bloodstream and procyclic stages. Tentative extrapolations of function could then be made about Hsp40s in other trypanosomatid species that have been identified as orthologues of *T. b. brucei* proteins. This is especially pertinent to those trypanosomatid species reported to be non-RNAi compatible. Hsp40s are a large protein family in the trypanosomatid species and it is therefore necessary to prioritise which proteins should be studied. Those Hsp40s that do not have a clear orthologue in the human host or in other organisms may have a trypanosome specific function that may be involved in the development of parasitaemia in the host. Alternatively, a study of the function of Tbj66 may be interesting due to the lack of a J-domain in its sequence. In addition to the functional characterization by phenotypic screening of Hsp40 knockdown in *T. b. brucei*, subcellular localization studies could be performed to confirm the tentative predictions made in chapter 2. It would also be pertinent to investigate the *in vivo* function of the Hsp70 family in *T. b. brucei*, thereby enabling the elucidation of their partnerships with the Hsp40 proteins.

Appendices

APPENDICES:

Appendix A1: Laboratory organisms and strains used in this thesis

Table A1: Genotypes of the various organisms and strains used in this thesis.

Organism	Strain	Genotype
E. coli	OD259	MC4100 *araD139* Δ*ara714* Δ*cbpA::kan dnaJ*::Tn10-42
E. coli	BL21	*hsds gal* (λcI*ts*857 *ind*1 Sam7 *nin*5 *lac*UV5-T7 gene *1*)
E. coli	DH5α	*supE*44 Δ*lac*U169 (Φ80 *LacZ* ΔM15) *hsdR*17 *recA*1 *endA*1 *gyrA*96 *thi*-1 *relA*1
E. coli	JM109	*recA*1 *supE*44 *endA*1*hsdR*17*gyrA*96 *relA*1*thi* Δ(*lac-proAB*) F'[*TraD*36 *proAB*$^{+}$ *lacI*^{q} *lacZ* ΔM15]
E. coli	XL1blue	*supE*44 *hsdR*17 *recA*1 *endA*1 *gyrA*46 *thi relA*1 *lac*$^{-}$ F' [*pro*AB^{+} *lacI*^{q} *lacZ* ΔM15 *Tn10* (*tet*^{r})]
S. cerevisiae	JJ160	***mat α*** *trp1-1 ura3-1 leu2-3,112 his3-11,15 ade2-1 can1-100 GAL2+ met2-Δ1 lys2-Δ2* ***ydj1::HIS3***
T. brucei brucei	SMB	(SMB, *T7RNAP::TETR::NEO*)
T. brucei brucei	Lister 427	Wild Type Laboratory strain

Appendix A2: Primers

Table A2:PCR primers

Name	Sequence (5'-3')	Use	Tm (°C)
Tcj2F (EcoRI)	GAATTCAAGATGGTTAAGGAGACTAAGTTT	**PCR product cloning of Tcj2**	**63.3**
Tcj2R (XhoI)	CTCGAGCTACTGTTGCGTACAAGT	**PCR product cloning of Tcj2**	**60.6**
Tcj3F (BamHI)	GGATCCGAGAAGATGGTAAAGGAAACAGAG	**PCR product cloning of Tcj2**	**70.6**
Tcj3R (EcoRI)	GAATCCAAGCTTTTATCTCCGCCGGGCCGC	**PCR product cloning of Tcj2**	**82.6**
Tbj2PCREcoRI,BamHI,NdeIFor3	GAATTCGGATCCCATATGGTGAAAGAAACAAAATAC	**PCR product cloning of Tcj2**	**59.3**
Tbj2HindIIIXhoISalIREV	AAGCTTCTCGAGGTCGACCTATTGCTGCGTACACG	**PCR product cloning of Tbj2**	**67.0**
Tbj3REFOR(EcoRIBamHINdeI)	GAATTCGGATCCCATATGGTCAAGGAGACAGAG	**PCR Product cloning of Tbj3**	**62.2**
Tbj3REREV(HindIIIXhoISalI)	AAGCTTCTCGAGGTCGACCTACTGTTGAGTACACTG	**PCR Product cloning of Tbj3**	**64.8**
Tbj2RNAiFOR	TGTGGAGAAAAGGAGACGC	**Cloning RNAi fragment 1 into $p2T7^{TABlue}$**	**54.8**
Tbj2RNAiREV	GCAACTCTGGCACTTATGC	**Cloning RNAi fragment 1 into $p2T7^{TABlue}$**	**54.3**
Tbj2RNAi2FOR	TTTCCAGTTTCTTTGGTGGG	**Cloning RNAi fragment 2 into $p2T7^{TABlue}$**	**53.2**
Tbj2RNAi2REV	TGCAACTCTGGCACTTATGC	**Cloning RNAifragment 2 into $p2T7^{TABlue}$**	**55.8**
Tbj2UTRf	TAATTCCAAGCATGAAAAGG	**Cloning UTR d RNAi fragment into $p2T7^{TABlue}$**	**52.5**
Tbj2UTR	GAAGTAAAAGGTTCGCATTG	**Cloning UTRd RNAi fragment into $p2T7^{TABlue}$**	**52.2**
J2NGFPF	GATCTCTAGAGTGAAAGAAACAAAATACTACG	**PCR cloning of Tbj2 into pRPaGFP**	**53**
J2NGFPR	GATCGGATCCCTATTGCTGCGTACACGTTG	**PCR cloning of Tbj2 into pRPaGFP**	**59**
Tc70F	GGATCCACGATGACGTACGAGGGAGC	**Cloning pQETchsp70**	**65.6**
Tc70R	GTGGAGGAAGTGGACTGAGAATTCAAGCTT	**Cloning pQETchsp70**	**63.9**

Names of primers that contain restriction enzyme names had those restriction enzyme recognition sequences included for the possibility of sub-cloning into a number of vectors from pGEM-T Easy

Table A3: Mutagenesis primers

Name of Primer	Sequence (5'-3')	Mutation detection (RE enzyme)	Tm °C
Tcj2Y8AFor	GTTAAGGAGACTAAATTTCGGGACTCTCTGGGTG	*Apo*I	62.0
Tcj2Y8ARev	CACCCAGAGAGTCCCGAAATTTAGTCTCCTTAAC	*Apo*I	62.0
Tcj2L11AFor	GTTTTACGACTCTGCCGGTGTTTCCCCAGATG	*Bsr*FI	64.8
Tcj2L11ARev	CATCTGGGGAAACACCGGCAGAGTCGTAAAAC	*Bsr*FI	64.8
Tcj2R27AF	GAAATCAAAAGGGCGTACGCGAGGCTTGCTTTGAAG	*Bsi*W I	69.0
Tcj2R27AR	CTTCAAAGCAAGCCTCGCGTACGCCCTTTTGATTTC	*Bsi*W I	69.0
Tcj2H34QF	GCTTGCTTTGAAGTACCAACCGGATAAGAATAAGGATC	*Bsa*W I	65.0
Tcj2H34QR	GATCCTTATTCTTATCCGGTTGGTACTTCAAAGCAAGC	*Bsa*W I	65.0
Tcj2K37AFor	CTTTGAAGTACCACCCTGATGCGAATAAGGATCCCG	*Msl* I	64.7
Tcj2K37ARev	CGGGATCCTTATTCGCATCAGGGTGGTACTTCAAAG	*Msl* I	64.7
Tcj2F47AFor	GGTTCACAGGAGAAGGCCAAGGAGGTTTCCG	*Eco*T14I	66.4
Tcj2F47ARev	CGGAAACCTCCTTGGCCTTCTCCTGTGAACC	*Eco*T14I	66.4
Tcj2L57AFor	GCCTATGAATGCGCATCAGACCCTGAAAAGC	*Fsp*I	64.2
Tcj2L57ARev	GCTTTTCAGGGTCTGATGCGCATTCATAGGC	*Fsp*I	64.2
Tcj2D59AFor	CTTATGAATGCCTTTCAGCTCCGGAAAAGCGCACG	*Bse* AI	67.2
Tcj2D59ARev	CGTGCGCTTTTCCGGAGCTGAAAGGCATTCATAAG	*Bse* AI	67.2
Tcj2R63AFor	CTTTCAGATCCGGAGAAGGCGACGCGTTACGACCAATTTG	*Bsi*YI	67.5
Tcj2R63ARev	CAAATTGGTCGTAACGCGTCGCCTTCTCCGGATCTGAAAG	*Bsi*YI	67.5
Tcj2(C396S)For	GGGCACACAGGAGCAACTAGTACGCAACAGTAG		65.4
Tcj2(C396S)Rev	CTACTGTTGCGTACTAGTTGCTCCTGTGTGCCC		65.4
Tcj2(Q199G)For	GACAGTTGCCGTGGCCAGGGGATCAAGAAGG	*Bal* I	68.2
Tcj2(Q199G)Rev	CCTTCTTGATCCCCTGGCCACGGCAACTGTC	*Bal* I	68.2
TCJ2S64TFor	CCCTGAAAAGCGCACGCGGTACGACC		67.3
TCJ2S64TRev	GGTCGTACCGCGTGCGCTTTTCAGGG		67.3
Tcj3Y8AF	GGAAACAGAGTATGCAGAGATCTTAGGCCTAGAAGC	*Bgl*II	62.9
Tcj3Y8AR	CCTTTGTCTCATACGTCTCTAGAATTCGGATCTTCG	*Bgl*II	62.9
Tcj3L11AF	GTATTACGAGATTGCCGGCCTAGAAGCG	*Nae*I	62.2
Tcj3L11AR	CATAATGCTCTAACGGCCGGATCTTCGC	*Nae*I	62.2
Tcj3R27AF	GACATCAAGAGGGCATATGCGCGCTTGGGCCTGAAGTATC	*Bss*HII	68.8
Tcj3R27AR	CTGTAGTTCTCCCGTATACGCGCGAACCCGGACTTCATAG	*Bss*HII	68.8
Tcj3H34QF	CTTGGCCTGAAGTATCAACCGGACAAGAACCCC	*Bsa*WI	65.8
Tcj3H34QR	GAACCGGACTTCATAGTTGGCCTGTTCTTGGGG	*Bsa*WI	65.8
Tcj3K37AF	CTGAAGTATCATCCAGACGCGAACCCCGGAGACCAGGAGGC	*Acc*III	70.6
Tcj3K37AR	GACTTCATAGTAGGTCTGCGCTTGGGGCCTCTGGTCCTCCG	*Acc*III	70.6
Tcj3F48AF	GCGGAAATGGCCAAGCGGATCGGTCATG	*Bal*I	66.9
Tcj3F48AR	CGCCTTTACCGGTTCGCCTAGCCAGTAC	*Bal*I	66.9
Tcj3L58AF	CATGCATACGAGATAGCTAGCGACGAAGAGAAG	*Nhe* 1	62.1
Tcj3L58AR	GTACGTATGCTCTATCGATCGCTGCTTCTCTTC	*Nhe* 1	62.1
Tcj3D60AF	GCATACGAGATTCTGAGCGCTGAAGAGAAGC	*Aor*51H1	63.2
Tcj3D60AR	CGTATGCTCTAAGACTCGCGACTTCTCTTCG	*Aor*51H1	63.2
Tcj3R64AF	GTGACGAAGAGAAGGCGCGCATTTACGACCAGC	*Bss*HII	67.5
Tcj3R64AR	CACTGCTTCTCTTCCGCGCGTAAATGCTGGTCG	*Bss*HII	67.5

Mutagenesis primers were labeled according to the target protein (e.g. Tcj3), followed by the original amino acid at position number, which was changed to (e.g. Y8A). The F or R designates 'Forward' or 'Reverse' primers.

Table A4: DNA Sequencing Primers

Name	Sequence	Use
pQE30For	CCC GAA AAG TGC CAC CTG	Sequencing pQE30 vectors
pQE30Rev	GTT CTG AGG TCA TTA CTG G	Sequencing pQE30 vectors
pUCF	CGC CAG GGT TTT CCC AGT CAC GAC	Sequencing pGEM vectors
pUCR	TCA CAC AGG AAA CAG CTA TGA C	Sequencing pGEM vectors
Tcj2F (EcoRI)	GAATTCAAGATGGTTAAGGAGACTAAGTTT	Tcj2 J-domain mutagenesis sequencing confirmation
Tcj3F (BamHI)	GGATCCGAGAAGATGGTAAAGGAAACAGAG	Tcj3 J-domain mutagenesis sequencing confirmation

Appendix B: Recipes

Yeast-Tryptone (YT) Broth growth medium

Tryptone	16g/L
Yeast Extract	10g/L
NaCl	5g/L

Make up to 1L with water and autoclave (121°C and 119 kPa for 20 minutes)

Yeast-Tryptone (YT) Agar

Same recipe as for YT broth, except add 15g bacteriological agar per litre of broth. Autclave the solution (121°C and 119 kPa for 20 minutes)

RF1 250 ml(pH 5.8) (Store at 4 °C)

(100 mM KCl, 50 mM $MnCl_2$, 30mM CH_3COOK, 10mM $CaCl_2$, 15% glycerol)

- 2.45 g or 7.5 ml of 1M CH_3COOK and pH to 5.8 with HCl or acetic acid
- Add 37.5ml glycerol and make up to a final volume of 202.5 ml and Autoclave (121 °C, 119 kPa for 15 to 20 minutes).

Subsequent to autoclaving, add the following autoclaved stocks:

- 25 ml 1M KCl
- 12.5 ml 1M $MnCl_2$
- 2.5 ml 1M $CaCl_2$

RF2 150 ml (pH 6.8) (Store at 4 °C)

(10 mM MOPS buffer pH 6.8, 10 mM KCl, 75 mM $CaCl_2$, 15% glycerol)

- 0.313 g or 1 ml 1M MOPS (pH to 6.8 with KOH)
- Add 22.5 ml glycerol
- Make up to 150 ml with distilled water and autoclave (121 °C, 119 kPa for 15 to 20 minutes.

Subsequent to autoclaving, add the following autoclaved stocks:

- 1.5 ml 1M KCl
- 11.25 ml 1M CaCl2

Appendix C: Basic Protocols

C.1) Standard procedures.

C.1.1)Manufacture of Transformation competent cells through Chemical treatment:

A fresh colony of a pure culture of *Escherichia coli* (XL1blue, JM109, BL21 or OD259) was inoculated into 5 mL of YT broth and incubated at the relevant temperature for the strain (37 °C except for *E. coli* OD259 which requires a growth temperature of 30 °C, due to its temperature sensitivity) with shaking for 12 to 16 hours. Four flasks each of 100 mL YT broth were inoculated with 1.5 mL, 1 mL, 0.7 mL and 0.3 mL of the overnight culture and incubate with shaking until the 1.5 mL culture reaches an A_{600} of 0.6 to 0.8 absorbance units. The flasks should be equilibrated to the incubation temperature prior to inoculation. The contents of the four flasks was pooled and the cells were harvested using sterile 250 mL (Beckman JA14 rotor compatible) centrifuge tubes by centrifuging for 10 minutes at 5000 RPM at 4 °C. The pellets were resuspended in RF1 (pH 5.8, 100 mM KCl, 50 mM $MnCl_2$, 30mM CH_3COOK, 10mM $CaCl_2$, 15% glycerol (**Appendix B**) solution and incubated them on ice for 20 minutes. The cells were separated from the solution by centrifugation for 10 minutes at 5000 RPM at 4 °C in a Beckman JA14 rotor. Resuspend the pellets in a total of 4 mL of RF2 (10 mM MOPS buffer pH 6.8, 10 mM KCl, 75 mM $CaCl_2$, 15% glycerol. **Appendix B**) solution. The cell suspension was aliquoted and stored at -80 °C.

C.1.2) Transformation of competent cells with plasmid DNA

Competent cells (80 to 100 µl) were mixed with 450 to 600 ng of Plasmid DNA and incubated on ice for 30 minutes. The reactions were then heat shocked at 42 °C for 45 seconds and cooled on ice for 4 minutes. 1 ml of YT-broth was added to each reaction and incubated at 37 °C for 1 hour. Variable amounts of the cells (100 µl of the solution or all of the cells resuspended in 100 µl of YT-broth after centrifugation) were plated onto YT-Agar plates containing the appropriate antibiotic. Plates were incubated at 37 °C for 12 to 16 hours. *E. coli* OD259 was heat shocked at 37 °C due to its temperature sensitivity and agar plates from the transformation were incubated at 30 °C.

C.1.3) Discontinuous sodium dodecyl sulphate (SDS) polyacrylamide gel electrophoresis for resolution of proteins:

The Biorad Mini-PROTEAN 3 system was used to run a modification of the discontinuous SDS-PAGE method used by Laemmli (1970). Gels were cast using a 12 % resolving gel (0.375 M Tris, pH 8.8, 0.1 % SDS, 12 % Acrylamide/0.32 % Bis-Acrylamide, 0.1 % APS, 0.2 % TEMED) and 4 % stacking gels (0.125 M Tris, pH 6.8, 0.1 % SDS, 4 % acrylamide/0.1 % Bis-Acrylamide, 0.2% APS, 0.4 % TEMED). The ammonium persulphate (APS) solution must be made fresh and the N,N,N'N'-tetramethyleethylenediamine (TEMED) initiates polymerization of the mixtures into gel. Protein samples for SDS-PAGE were mixed in a ratio of 1 part sample buffer (0.0625 M Tris-HCl pH 6.8, 10% glycerol, 2% SDS, 5 % β-mercaptoethanol, 0.05% Bromophenol blue) to 4 parts protein solution and boiled at 95 °C for 10 minutes. The samples were resolved with SDS-PAGE running buffer (25 mM Tris-HCl, pH8.3, 192 mM glycine and 0.1 % SDS) for 1 to 2 hours at 150 V Gels were stained in Coomassie stain (40 %

Methanol, 2 % acetic acid, 0.25 % Coomassie blue R) and destained in Coomassie destain solution (40% Methanol, 2 % acetic acid).

C.1.4) Western Blotting of Proteins

Proteins resolved on a Discontinuous SDS-PAGE gel were transferred onto nitrocellulose as described by Towbin *et al.* (1979). A Biorad wet transfer apparatus was used with a freshly prepared methanol based transfer buffer (25 mM Tris, 192 mM glycine, 20 % methanol) stored at 4 °C. The membrane and nitrocellulose sandwich were arranged between 2 pieces of Whatman 3MM filter paper on either side, so that the gel is on the negative side of the circuit and the membrane on the positive side. The transfer conditions were 100 V for 1 hour at 4 °C. Transfer of the protein to the membrane was confirmed by Ponceau S stain (0.5 % Ponceau S, 1 % glacial acetic acid) for 5 minutes and destaining was performed with distilled water until the protein bands were visible.

C.1.5) Chemiluminescent-based immunodetection of proteins

After Ponceau S destaining, the membrane was washed in Tris buffered saline (TBS; 20 mM Tris, 140 mM NaCl; pH 7.4). The membrane was subsequently incubated in blocking solution (2 % block powder, in Amersham-ECL western blot detection kit, in TBS-tween) overnight at 4 °C. TBS-tween consists of TBS with 0.1% (v/v) Tween 20 added. The membrane was washed briefly in TBS-Tween and then incubated in blocking solution containing the primary antibody (antibody dilution was determined for each individual antibody, normally 1:5000) for 1 to two hours with agitation at room temperature. The membrane was washed in two brief washes of TBS-tween and again in TBS-Tween for 15 minutes (4 ml/cm2 of membrane). After a further three washes in TBS-Tween, each of 5 minutes long, the membrane was incubated with the secondary antibody (anti-rabbit or anti-mouse, depending on the source of the primary antibody, 1:5000 dilution) in blocking solution for 1 to 2 hours with agitation at room temperature. The wash steps were then repeated as after the primary antibody incubation. The Amersham-ECL western blot chemiluminescent detection kit was used in conjunction with the Chemidoc gel documentation system (Biorad). Ten exposures were taken over the space of 20 minutes, starting at 30 s.

C.1.6) DNA Sequencing to confirm site-directed mutagenesis and plasmid clone identity:

DNA for sequencing was prepared using either the Qiagen Miniprep Kit (Qiagen, USA) or the High Pure Plasmid Isolation kit (Roche) (**Appendix C.1.8**). Didedeoxy terminator DNA sequencing was performed using the ABI BigDye V3.1 Terminator cycle sequencing kit (Applied Biosystems, USA). Oligonucleotide primers used for cycle sequencing was dependent on the plasmid being sequenced (please refer **Appendix A2** for a list of primers).

Plasmid DNA (300 ng to 500 ng) was mixed with 3.2 pmol of the appropriate sequencing primer and 8 µl of BigDye. The reaction was made to a total of 20 µl with distilled water. The thermal cycling protocol used 25 cycles of: 96 °C for 10 seconds, X °C (dependent on individual primer T_m) for 5 seconds and 60 °C for 4 minutes. The polymerized products were purified using Zymogen DNA Cleanup and concentrator kit (Zymo

Research, USA) (**Appendix C.1.7**) to remove unincorporated dye terminators and proteinaceous components of the cycle sequencing reaction. The DNA was dried in a vacuum centrifuge and the different length strands resolved using the 3100 ABI Genetic Analyser. Sequence information was analysed using Vector NTI (Invitrogen, USA).

C.1.7) Zymogen DNA Cleanup and Concentrator Kit (Zymo, USA)
This was done according to the instruction manual within the kit and was used to purify the polymerized DNA strands produced during cycle sequencing. 40 µl of DNA binding buffer were added to the 20 µl cycle sequencing reaction This was loaded onto a Zymo-Spin column and centrifuged for 30 s at 16 000 x g in a microcentrifuge. The flow through was discarded and the column was washed with two 200 µl volumes of wash buffer. Each of these wash volumes was passed through the spin column by centrifugation at 16 000 x g for 30 s. A further centrifugation at 16 000 x g for 30 s was used to remove any remaining wash buffer. The DNA was eluted with 10 µl of distilled water and collected into a clean microcentrifuge tube by centrifugation at 16 000 x *g* for 30 s.

C.1.8) Preparation of plasmid DNA
Plasmid DNA was isolated from *Escherichia coli* cells using either the Qiagen Miniprep Kit (Qiagen, USA) or the High Pure Plasmid Isolation Kit (Roche). Both of these kits are modifications of the alkaline lysis method described by Birnboim and Doly, 1979). Plasmid DNA was isolated from *E. coli* XL1blue, JM109 and DH5α using these protocols. Cells containing the plasmid of interest were propagated in a 5 ml solution of YT-broth (containing the relevant selective antibiotic) for 16 to 18 hours. The harvested cells were resuspended in 250 µl buffer P1 (Qiagen, USA) or Suspension Buffer (Roche; 50 mM Tris-HCl, 10 mM EDTA, 0.1 mg/ml; pH 8.0). 250 µl of Buffer P2 (Qiagen) or lysis buffer (Roche; 0.2 M NaOH and 1% SDS) was mixed with the suspended cells and allowed to stand for a minute. 350 µl Buffer P3 (Qiagen) or Binding Buffer (Roche; 4 M guanidine hydrocholoride; 0.5 M potassium acetate; pH 4.2) was mixed into the solution by inversion and incubated on ice for 5 minutes. Solutions were centrifuged for 10 minutes at 16 000 x g and the supernatant applied to a Qiaprep spin column (Qiagen) or to High Pure Filter Spin Column assembly and centrifuged for 1 minute at 16 000 x g. the flow through was discarded and washed with 500 µl buffer PB (Qiagen) or Wash Buffer 1 (Roche; 5 M guanidine hydrochloride, 20 mM Tris-HCl (pH 6.6), 33% EtOH) that was washed through the spin columns by centrifuging at 16 000 x g for 1 minute. 750 µl buffer PE (Qiagen) or Wash Buffer 2 (Roche; 20 mM NaCl, 2 mM Tris-HCl pH 7.5, 20 % EtOH) and centrifuged as for the first wash step. The columns were then centrifuged for an additional minute at 16 000 x g to remove any traces of the wash buffers. The Filter sections of the spin columns were then inserted into clean microcentrifuge tubes and eluted by application 40 µl Tris-EDTA (TE) buffer (100 mM Tris-HCl pH 8.0, 10 mM EDTA) or distilled water to the filter for one minute and collecting the DNA solution by centrifugation at 16 000 x *g* for one minute.

C.1.9) Resolving of DNA using Agarose gel electrophoresis
Agarose (0.8 % to 1 % (w/v) Molecular Grade) was dissolved in Tris-Acetate-EDTA (TAE) buffer (pH 8.0; 40 mM Tris-Acetate, 1 mM EDTA) by heating. The solution was

allowed to cool and then ethidium bromide (0.3 µg/ml final concentration) was added and the solution was allowed to set in a casting mould. DNA loading buffer (6x loading buffer; 0.25 % bromophenol blue, 30 % glycerol), was added to DNA solutions in a ratio of 1:10. Lambda (λ) Phage genomic DNA restricted with *Pst* I restriction enzyme was used as a molecular weight marker. Gels were immersed in I x TAE buffer and, after the samples were loaded, they were run at 60 V (25 mAmps) until the tracking dye indicated that the DNA bands should be sufficiently resolved. DNA in the gel was visualized using ultra-violet light in the Chemidoc Gel documentation system (Biorad).

C. 1.10) Purification of DNA fragments from Agarose gels

Restricted fragments were separated on an agarose gel excluding ethidium bromide in duplicate adjacent lanes. The gel was cut along the axis between these two gels and one half was stained in an ethidium bromide solution (4 µg/ml ethidium bromide in water) for 30 minutes. The gel was washed in distilled water for a minute with agitation. The band of interest was located on an ultraviolet transilluminator and a clean scalpel used to cut above and below the fragment. The transilluminator was switched off and the unstained gel segment was aligned next to the stained segment. The incisions around the gel fragment were extrapolated onto the unstained segment to cut out the DNA fragment of interest in the unstained gel, thus avoiding damaging ultraviolet light exposure. GFX™ PCR DNA and Gel Purification kit, GE healthcare.The eluted DNA was precipitated by the addition sodium acetate (3 M) and 100 % ice cold ethanol in a ratio of 1 volume DNA: 2 volumes 100 % ethanol: 0.2 volumes sodium acetate. This was mixed by inverting the containing tubes and incubated at -20 °C for at least 2 hours. The DNA precipitate was collected by centrifugation at 16 000 x g for 40 minutes at 4 °C. The pellet was washed in 700 µl ice cold 70 % ethanol and centrifuged for a further 8 minutes at 16 000 x g before discarding the 70 % ethanol. The pellet was air dried for 30 minutes and resuspended in TE buffer (100 mM Tris-HCl pH 8.0, 10 mM EDTA) overnight before use.

C.1.11) Restriction endonuclease digestion of plasmid DNA

Plasmid DNA prepared as described in **Appendix C.1.8.** 300 ng was mixed with 2 µl of the appropriate 10 x restriction endonuclease buffer and 2 Units of the relevant enzyme. The reaction volume was made up to 20 µl with distilled water. After incubation for four hours at the relevant temperature, the resultant DNA fragments were resolved on an agarose gel (Appendix 1.9).

C.1.12) Alkaline Phosphatase treatement of restricted DNA fragments

The plasmid vector to undergo alkaline phosphatase treatment was cut with the appropriate restriction enzymes (**Appendix C.1.11**). Shrimp Alkaline Phosphatase (Roche, Germany) was used to prevent recircularisation of the vector during ligation reactions. The restriction enzyme was heat inactivated at 65 °C for 15 minutes. 6 µl of dephosphorylation buffer (10 x; 0.5 M Tris-HCl, 50 mM $MgCl_2$, pH 8.5), 6 µl of distilled water, 2 Units of enzyme and 60 µl of DNA solution (600 ng) were mixed and incubated at 37 °C for 15 minutes. The enzyme was subsequently heat inactivated at 65 °C for 15 minutes. The solution was cooled to 37 °C and a further 2 U of shrimp alkaline

phosphatase was added and incubated for 15 minutes at 37 °C before heat inactivation at 65 °C for 15 minutes. This DNA was used for ligation with other DNA fragments.

C.1.13) Ligation of DNA fragments during plasmid modification

Ligation reactions were set up in a molar ratio of 3 insert fragments to 1 vector fragment. This varied depending on the relative sizes of the fragments. Vector DNA (150 ng) was mixed with an appropriate amount of insert DNA, 1 µl of ligation buffer (10x; 300 mM Tris-HCl pH 7.8; 100 mM $MgCl_2$, 100 mM DTT, 10 mM ATP) and 1 Unit of ligase enzyme (Promega, USA) was made to a final volume of 10 µl. The reaction was mixed and incubated at 4 °C overnight. A portion of the reaction (15 %) was used to transform *E. coli* DH5α and the resultant colonies were screened for the correctly assembled plasmid.

C.1.14) Ligation of PCR products into pGEM-T Easy

PCR products were gel purified to eliminate any trace non-specific products of incorrect length (**Appendix C.1.10**). pGEM-T Easy (25 ng) was mixed with 50 ng PCR product, 1 x final concentration ligation buffer (2x; 60 mM Tris-HCl pH 7.8, 20 mM MgCl2, 20 mM DTT, 2 mM ATP, 10 % polyethylene glycol) and 1.5 U of T4 DNA ligase were mixed to a final volume of 5 µl. Reactions were incubated at 4 °C overnight before being transformed into *E. coli* DH5α or *E. coli* JM109. Selection of colonies was based on blue/white screening with X-gal and IPTG in the YT-Agar plates along with antibiotics for selection of plasmid retention. Plasmids were then isolated from selected colonies (**Appendix C.1.8**) and constructs checked by restriction digestion (**Appendix C.1.11**) and sequencing (**Appendix C.1.6**).

C.1.15) PCR Amplification of DNA

In a 50 µl reaction:
200 µM of each dNTP
5 µl 10 X reaction buffer (containing 15 mM $MgCl_2$)(Roche)
1 to 100 ng template DNA
300 nMol of each primer
1 U of Expand High Fidelity PCR enzyme mix (Roche)

Thermocycler reaction conditions:

Stage of cycle:	Temperature	Number of cycles
Stage 1	95 °C for 2 minutes	1 cycle
Stage 2 step 1	95 °C for 30seconds	25 cycles
Stage 2 step 2	*X* °C for 40 seconds	
Stage 2 step 2	72 °C for 45 seconds	
Stage 3	72 °C for 5 minutes	1 cycle

X is the average of the primer Tm's.

C.1.16) Site-Directed mutagenesis

Site directed mutagenesis was performed using a 50 µl reaction in a 0.2 ml PCR tube. Reaction constituents were 100 ng of template DNA, 125 ng forward mutagenesis primer, 125ng reverse mutagenesis primer, 200 µM final concentration for each dNTP, 5 mM

final concentration of extra $MgCl_2$, 5% final concentration of DMSO, 5 µl Pfu DNA polymerase buffer (Promega, USA) and 3 units of *Pfu* DNA Polymerase enzyme.

Reaction conditions for thermocycler:

Step	**Number of Cycles**	**Temperature**	**Time**
Denaturation	1	95 °C	30 seconds
Denaturation	18	95 °C	30 seconds
Annealing		x °C *	60 seconds
Extension		68 °C	5 minutes
Final Extension**	1	68 °C	7 min

* this is dependent on the annealing temperature of the mutagenesis primers, however approximately 55 °C was found to be best for these reactions.
** Final extension permits the completion of copies that were left unfinished by previous extension periods.

The reactions were then treated with 10 U *Dpn*I restriction endonuclease for 1 hour at 37 °C to shred the methylated (template) DNA. Prior to the addition of *Dpn*I, 10 µl of the reaction was removed for comparison with the DpnI treated reaction. Resolution of 10 µl of the before and after *Dpn*I treatment on a 0.8% agarose gel indicates if the reaction has worked. 1.5 µl of the *Dpn*I treated reaction was then transformed into *E. coli* JM109/DH5α competent cells (**Appendix C.1.1**). A sample of the resultant colonies on YT-agar was randomly selected and the plasmids isolated according to **Appendix C.1.8**. These plasmids were screened for a restriction enzyme recognition site that the successful mutagenesis introduces and further confirmation was produced by DNA sequencing (**Appendix C.1.6**) of the mutated area.

Appendices: Chapter 2:

Appendix 2A: Type I Hsp40s of the different Trypanosomatid species

The J-domains of each sequence is highlighted in yellow and the Zn finger-like region is highlighted in green. The CaaX box is highlighted in red. The HPD motif is underlined in every sequence if one was located. Type I/IV sequences have a KDPQ motif instead of the HPD motif.

Trypanosoma brucei Type I Hsp40s:

>Tbj2 (Tb927.2.5160) ||30J2.30|chaperone protein DnaJ, putative|Trypanosoma brucei|chr 2|||Manual
MQARMVKETK YYDALGVPPN ASEDDIKRAY RKLALKYHPD KNKEPGANEK FKEVSVAYEC
LSDVEKRRRY DQFGEKGVES EGVGIDPSDI FSSFFGGRRA RGEAKPKDIV HQQPVPLETF
YNGFTISLAI IRDRLCDSCN GSGSKDPKVS SRCVECDGRG VKIITRSIGP GFVQQMQVAC
PRCQGEGTDI KEEHKCQSCR GQQIVKDKKV FDVVVEKGMQ HGDSVTFQGE GDQIPGVRLS
GDIIIILDEK PHPVFTRKGD HLLIHHKISL AEALTGFTMN IKHLDERAIS IRSTNVIDPQ
KLWSVSREGM PIPGTGGTER GDLVIKFDVV YPSAQSLSGD GIEPLRRILG YPKQEEPAPE
ATEHTLAVTY VDLDREARRR RTAANDDDDD AGQHVHTGAT CTQI

>Tbj3 (Tb10.70.5440) |||chaperone protein DNAJ, putative|Trypanosoma brucei|chr 10|||Manual
MVKETEYYEL LGVAVDATEN DIKRAYRRLA LRYHPDKNPD NAEAAEMFKQ ISHAYEVLSD
EDKRKLYDQH GKDGLSGGGD EGEFDASDIF SMFFGGGRRQ RGERKPRDLV HELAVSLEDM
YNGRVKRVAV TRDRLCSQCD GSGVRPGACQ QMCEACNGQG IQVLVQHIIP GVRQQVQLTC
QNCGGCGKYV RESDVCPRCH GKQMVEDEKV LEVPIERGMK ADDAIRFEGE GDEVLGVRLK
GDVLIILAEK PHDVFRRVGD HLIMNYRITL QEALCGFELP VQHLDKRMLL IKIPAGQVID
PEAGWVVHRE GMPLPNTGGI ERGNLIIHFE VEYPTKLSSR QIDLIADAFH VSEGFPHVGG
QKVVLRDETA RRQRRNTASA RQAQQRRSRD TRGFDNPDVF SMGFGGGQTA HQETGC

>J46 (Tb927.3.1430) |||chaperone protein DnaJ, putative|Trypanosoma brucei|chr 3|||Manual
MRCTMSSPLF RLTRLSVVIT IAFFRNVHAE EEDDYYRVLG LDAEHEDVSE RDIKSAWRKL
SKKHHPDVAG ESQRVMYQRI QRAYEVLGDR RRRKIYDILG TEGLKKYEKP QEQGRHQSIF
DTFFSFIGGE SGGNADRGSD EELMLLVTLE DMYKGAAHTA KLPRIKICRK CRGTGARSKD
DYVKCPHCGG GGRVVRRVQI APGFIQQIEQ VCGQCGGGGR VVRRKCPVCG GHRLVKGESS
VSIDIEQGTP NGYKMTYEME ADQQPNKMPG DLIFTIVTIP HPEFARMSSG KEGVPDDLST
AVELTLKEAL LGFNKTLKHL DGRVLSLVET GVTKFGQIRK YAGEGMPRHH VPSERGNLLV
LYTVELPKIL TEEQRKAIER ALD

>J50 (Tb09.211.3680) |||chaperone protein DNAJ, putative|Trypanosoma brucei|chr 9|||Manual
MLRFTSVSSI WRRLAAAPPT AATAAFANVS KRQSSSNKDY YKILGVSQSA SQSDIKKAYR
KRALETHPDQ GGNKEDFAEV AEAYECLSNE EKRRIYDQYG SEAAANMNAG GGMGGFGGRS
AEDIFAEFFK GGMGGFGGNR SAGPPQVPPL EVTLRMTLEE VYKGASKSFR VNRPVVCSDC
RGFGTKSQKK KPKCSECDGS GHVVQHHRFG PGMVQQTVSQ CPRCGGAGTV AKPDDKCPKC
KGMGYFHLVQ SVSIDIPAGV PPDVTLVVRG EGGTMPEAEP GDLHVHVEVE EHNVFKRRGN
DLVVERDVTL SEALLEFDLS LKTLDGRSIT VKSPKSSVLQ PNSVLRVAGE GMPNSSGGNG
DLYIVTKLKL PRTLTDQQKE AVKKAFGEPK KKDPDSGSDK TVTASVLRGG VREMEEALHS
NWDSERGGGS QQGNGRRRSG RKQHTAECHE Q

>J45 (Tb11.01.8480) |||chaperone protein DNAJ, putative|Trypanosoma brucei|chr 11|||Manual
[illegible]CLGFAPV VHSHDDSGDG ADEIDYYAVL GLTPEATDRE
VRQRFRELSR KYHPDVSSGG DAREMFSRIT RANEVLSDKK KRRMYDMRGE EGLRQLKRLE
AESGGEQFGS ISQLFRHHAA RRLRGKNTEA TLHLPLDVVY T[illegible]
[illegible]
[illegible]VELTV DIDAGMPEGH VLSFEMEADE SPDTIPGDLL LSVQTKKHPR FSRRANDVDL
DMTLVVTLKE ALLGFQRRVE HLDGSEFFVN ETGITQYGSV LKVPGKGMPR HNVPSEFGDL
YVKVLFEMPD MLTKEQREEL AEHL

Leishmania major Type I Hsp40s

>LmJ2 (LmjF27.2400) |||heat shock protein DnaJ, putative|Leishmania major|chr 27|||Manual
MVKETGYYNA LGLSPDASED EIKRAYRKLA LKYHPDKNTE PGAQEKFKEV SVAYECLSDP
DKRKRYDQFG KDAVEMQGGG VDPSDIFASF FGGGSRPRGE PKPKDIVHEL PVPLEAFYCG
[illegible]
[illegible]KVFEV MVEKGMHRGD SVTFSGEGDQ IPGVKLSGDI
IIILDQKPHQ TFIRKGDHLF LEQTISLAEA LTGFSLNITQ LDGRELAISS TAGTIIDPAN
MYSVSREGMP VAHTGGMERG DLIIRFQVVF PKTLRQGCVP ELRKMLGYPQ QPPAKDGAEQ
YTLQESHIDL EKEARRNAYD DDGDQPRVQT AG[illegible]

>LmJ3 (LmjF21.0490) |||DNAJ protein, putative|Leishmania major|chr 21|||Manual
MVRETELYEV LNVSVEANEH EIKRSYRRLA LKYHPDKNTG DEAAADMFKK VSNAYEVLSD
PEKRKVYDKY GKEGLERGTG EGGGFHDATD IFSMFFGGGA RERGEPKPKD IVHELEVKLD
DLYN[illegible]
[illegible] SVLEVHIGRG ASKSDHFTFI GEGNQEPGIR
LSGDVLIFLR VRPHPVFHRI NDHLMMRCSI TLQEALCGFE VPIEHLDGRQ LVIKASPGQV
VHSDSAWSVY NEGMPVKGTG GLQKGKLFIY FDVEWPETLP REQIDKIVTA FNVPEKPGKL
GGHVVELREY KAAGKFKSGK NEKRGGTAGS RSAGAGRGRE AARGRQAHAE EDEFEDVTDD
DDDEQQQYFR AGPQGFNGNT QTVE[illegible]

>LmJ46 (LmjF25.1100) |||DNAJ-like protein|Leishmania major|chr 25|||Manual
MRGRRTTRVV AALLVLVWVA ALVAEVPVHM AGAADPRDED AKAVNAVLRL PEDDFYAVLG
LGEAREDATE RDIKNAFRRL SKKYHPDVAT GDQDSYRLVY QRVQRAYEVL GDRRKRKIYD
ILGIDGVTRL EKPQQQQQMN PFFAFFGVGQ QADAERGKDM VLLMVVPLED IYR[illegible]
[illegible]
[illegible]AVL SVDIEEGLPE GHVLTYELEA DQAPGQVPGD VLLTVISAPH
PVFHRSGNDL YANVSITLKE ALLGFKKTLA HLDGHNVELH WDGVMQNTQQ VRIAGEGMPR
HHVPSERGDL YITYNVLLPE ALTAEQRALF QEHFA

>LmJ45 (LmjF32.3300) |||chaperone protein DNAJ, putative|Leishmania major|chr 32|||Manual
MTIARSRRAT VRRTCVLVLL LLCVALSSRA FFDFGGGHRA DAPSAEVTHA QEVDYYMVLQ
LEDKREEATE KDIRQQFRRL SRLYHPDVAK TEEDKAKYSQ VNRAYEVLSD KRKRKVYDMR
GERGLEQLER IDRSKDTPGG GVNPLSRLFG MRVDDGLRGS DMKLEAKVDL AKLFT[illegible]
[illegible]
[illegible]S TVTLELEPGM EEGHVLKFEM EAEESPDRLP GDLLVHVHTL
PHPVFSRRRN QIDLDTSLTL TLQEALVGFD RNITHLDGVE QVRVQRLDTV SPYGTVLRLP
GKGMPKMNVA SERGDLYVRL QYNMPAQLTE EQRKLVDMLL

>LmJ50 (LmjF35.2980) |||chaperone protein DNAJ, putative|Leishmania major|chr 35|||Manual
MFRYLSASTA RRAGAWGAGA SLSAAPGASS TFCSSTNPAR LYSSGNKDYY KMLGVDRNSD
LKDIKKAYRK RALETHPDQG GNKEEFAEVA EAYEVLSNPE KRKVYDQYGS EAATNPNMGG
AGMGGFGTGG RSAEDIFAEF FRGGMGGMGG FGDMFGGGPG GRQGTPTLQP LEVRTRLTLE
DVYKGVTKTM RVNRPQVCSD CTGPGTKSKT EKPKCTQCNG SGHVVQQHRM GPGMVQQTIS
ECPRCRGTGT VAKLSDQCHK CHGKGYRTVS QDVTIEIPAG VPSNVTLVVR GEGGTIPGAP
PADMHLHIEL TPHRVFQRRG NDLIVNKDVT LQEALLGLHM PLRLLDGRTV NVETSADQIL
KPEGVIKVTG EGLPGTSGER GDVYIFTHLK MPNKLSDEQK AHIKKAFGTP GKDADASPGN
TVRARVLRES REQLEEQKRS LWASQESGGH SGGGGGSSRR SAGSTHGSGA QQVEGATQ

>Lmj27 (LmjF04.0940) |||chaperone protein DNAJ, putative|Leishmania major|chr 4|||Manual
MKAASCPPAA ARLYRVYGPV HLLPRTTAMR RLAAASAMAV TAGSLGETLL SSAAVSAGGA
AHSAMGGSTA LLQHRRWQSG GGSKKDLYSV LGVARNSTPE QIKSAYKKRA KALHPDVNPS
PTAAEDFAEA KQAYETLSDP QKRSMYDMTG NASAAGGFGG PGGTGGGFNP FGAGGNPFAA
GGNPFANMGG ANQPGGGQGG FSFNDFEEIF QKMSNSGKDK TRKPQGPEPG ADIHYKLVLS
FLEAVKGCQK EISYNTMRRC GACTGSGCQD TGSRTKCPHC GGRGKKVMST GFFHMQQDCT
HCGGTGELGR TTCTQCSGKG IVKDRSVQTL PVPKGVDNKE RLKVTGKGEA GVRNGPPGNL
YIEISVEDDP VFHREGSDIH VITPITLSCA VLGGTVRVPT LTGEVETRVP VGTQQGDKLV
LRGRGVHRPN HNKTGDFYIH FAVMLPKELT EEQKKAIADF AKDERPLHLN DAQLQELKGR
YRSWFAT

>LmJ4 (LmjF15.1220) |||hypothetical protein, unknown function|Leishmania major|chr 15|||Manual
MVKETGLYDE LGISPDATEP QIRSAYRRKA LQYHPDKNSG DPAAAEKFKK VAEAYEILSD
AERRKQYDTF GRNGPGGAAG GSGGVPGGGF GSSFGPGTDP MDIFSSFFGF ASGGGPGASR
QQRPAKPSFI LVVLQCSLEE LYCGTVKVLQ VRRRRLCPSC HGHGTSSGRP SPVCPQCKGN
RTVTKALNPG GLTAYQQCRC DHCEGLGKLP IPLPCARCGP VYDAHQRAPL SGTRKPDPLP
RGVVVDEKMI KVSIEPGTED SDALNYVGQG DELPGFGAVG DVLVVIEQLP HPHYHRLNST
DLLLSNCRVP LGCLFRDAFS IPIELLDGRV LRLATPLREG NVPHFLFDSQ HVFVVDGEGM
PLKGRKGGGA AADANHDDSG AADDVAYRKK GKLYLCVNVV FPSSLTPAQV DIIAKALGGD
HERDPAATGE DAGRVVTLQP HHSPAPSWYT VDTEGNAWYQ PNGPATAKKK RKATSTTA

Trypanosoma cruzi Type I Hsp40s

>Tcj4 (Tc00.1047053506445.121)
MVVDTSLYDE LGILPSAATD EIRTAYRRLA LKYHPDKNGG DARAAEKFKK VAEAYEILSD
PTKRRHYDQL GRASAVGQNG PSAANFPFGN VDAEELFRRF FGVSTGGGSA GGPPSARKPP
DIVIELQLSL EELYCGTRKR VAVRRVRRCP HCKGHGTSDQ IPLASCQLCG GRGERVHGMR
VGGLILQQMQ VCSSCNGTGK TALKHPCKHC FMSQHSDNRV AGTVECVKEL LLEVDPGTDN
EARFRFHGEG DEMPPPYQEP GDIIITTKAL PHPHYRRISK NDLLLLNCVV PLESVFQKDF
FIPIEHLDGR ILKIFPAEGT CAKMNILEPL FPHCLYSVAN KGMPIRGDPQ GRQGKLFVRI
HIVYPRALNV SQLTLLEEAF RYRLPDMTEP PQGKFVCLNY YSGNSAPSQK ASSKWGKEEQ
QGAKRNSNSR VSTSSSTTQL Q

>Tcj27 (Tc00.1047053510575.200) |||chaperone DnaJ protein, putative|Trypanosoma cruzi|chr unknown84|TIGR||Auto
MRQHRARTQR WVRRAVYPAS LHGGVSCGNL RFLVRSDLTA FTPCIAAAAD LTGLGVSRRF
SSTNAKKDLY SVLGVARNAS QEDIKSAYKK KAKQLHPDVN PNPRAAEDFA DVKQAFDVLS
DPQKRSMYDM TGNSGAADRF GSGPGFNPFA QAGGENPFAG ANPFGGANPF SSSAGGQGFS
FNDFEEIFQK MSGAGKDKTR KPQGPEPGAD IHFKLRLSFL EAVNGCSKEI SYNTFRRCGS
CTGSGPQDSG SKAKCVHCGG RGKKVMSTGF FHMQQDCTHC GGTGETGRSS CTHCSGKGIV
KDESVQTLPV PKGVDNKERL KVSGKGEAGV RNGPPGNLYV EITVDEHPVF HREGCDIHVI
TPISLATAVL GGTVRVPTLT GEVETKVPVG TQQGDKLVLR GRGVHRPNQN KTGDFFIHFA
VLLPKSLTEK QKSLIAEFSK EEKPLELTDP QMQELKGRYK SWFSS

PS51257 PROKAR_LIPOPROTEIN Prokaryotic membrane lipoprotein lipid attachment

>Tcj45 (Tc00.1047053511025.100) |||chaperone DnaJ protein, putative|Trypanosoma cruzi|chr unknown85|TIGR||Auto
MMITGEFRFA CPTLLIFFLL YVVYFAVPTA AFGGRGRAGA RAPAEEEEEV DYYAVLGLTE
NATEKEVRQK FRELSRKYHP DVAKTEEAKA MYGKINRANE VLTDKKKRRM YDMRGEEGLR
QLERALAQNE QGHSMDPFAR LFGMGSGGNL RGSDSQSTLH VELEDVYKGT QRSVVLEKQK
VCTKCKGTGA SRGSGVTTCS HCRGHGVVIQ RLQLGPGMYQ DIQQACPHCQ GQGRIAKHRC
PACNGKKVVR GEVTLTIDIE QGIPEGHKVT FEMESDESPD LVPGDLIMAV LTKPHPRFSR
RPNGLDLDMS LTVTLKEALL GFERRVEHLD GTEFLVEATG VTPYGEVLKV RGKGMPRHHM
PSEKGDLYVK VMFELPSFLT EAQRKEIEEH F

>Tcj46 (Tc00.1047053509233.80) |||heat shock protein DnaJ, putative|Trypanosoma cruzi|chr unknown80|TIGR||Auto
MSCHGGSHTW LFFFLLRGVL LLLLLLPLDA GEEVVVVAAA AKSAPKVEEE DFYEVLGLGK
ERDDASERDI KSAWRKLSKK HHPDLAGESQ REVYQRIQRA YEVLGDRKKR KVYDILGLDG
VKKIEQPQEQ QQQQQHMHSF FSFFGGGHQQ QQVDRGKNED LVLLVPLEDV YSGAAHTVKL
SKTKICRNCR GTGARSKDHL VRCPHCNGEG RVLRRVQLAP GFIQQMEQPC AHCNGQGVFI
TEKCLMCKGK KTVRSTSSIS IDIEQGIPDG HVLTYELEAD QQPNQVPGDV LFTVVTASHP
RFTRSDNDLT VTVVLTLKEA LLGFSKSLTH LDGHVVELEQ SGVTQHGERR KIAGEGMPKH
HVPSERGDLH VIFEVEVPSL LTKAQKEALE RAFG

>Tcj46 (Tc00.1047053509437.40) |||heat shock protein DnaJ, putative|Trypanosoma cruzi|chr unknown80|TIGR||Auto
MSWHGGSHKW LFFFLLRGVL LLLLLLPLEA GEVVVVAAAK SAPKVEEEDF YEVLGLGKER
DDASERDIKS AWRKLSKKHH PDLAGESQRE VYQRIQRAYE VLGDRKKRKV YDILGIDGVK
KIEQPQEQQQ QQHMHSFFSF FGGGHQQQQV DRGKNEDLVL LVPLEDVYSG AAHTVKLSKT
KICRTCRGTG ARSKDHLVRC PHCNGEGRVL RRVQLAPGFI QQMEQPCAHC NGQGVFISEK
CLTCKGKKTV RSTSSISIDI EQGIPDGHVL TYELEADQQP NQVPGDVLFT VVTASHPRFT
RSDNDLTVTV VLTLKEALLG FSKSLTHLDG HLVELEQSGV TQHGERRKIA GEGMPKHHVP
SERGDLHIIF EVEVPSLLTK AQKEALERAF G

>Tcj50 (Tc00.1047053510743.100) |||chaperone DnaJ protein, putative|Trypanosoma cruzi|chr unknown84|TIGR||Auto
MLRFTPPFML WRRWAPAAVA AALPGLSLRL SSSEKDYYKI LGVSRTASVS DIKKAYRKRA
LETHPDQGGK KEEFAEVAEA YECLSNEEKR RVYDQYGSEA AANMNAANGM GGFGAHSAND
IFAEFFKSRM GGFGDDMRRG PAQVQPIEVK LRMTLEEIYK GVTKKPRVNR PVKCADCRGF
GTKSQTKKPK CAHCDGSGHV VHQHRMGPGM VQQTVTQCPR CGGSGTMAKP DDQCPKCHGM
GYRHLSQEVN IDIPPGVPSN VTLVVRGEGG TMPDAEPGDL HVHVEVAPHK IFTRRGDDLL
MKKEISLSEA LLGTQFSVKM LDGRHVTVKV PHENVLRPDS VLKVSGEGMP SADGGRGDLY
VITHLKMPAK LTAQQREAII QAFGKPNEEK HTSADKTTTA RVMREGAQEL EDAMRDQWDA
QEGGGFRSGA GGHGRGRQSS QQAECVHQ

>Tcj50 (Tc00.1047053510659.210) |||chaperone DnaJ protein, putative|Trypanosoma cruzi|chr unknown84|TIGR||Auto
MLRFTPPFML WRRWAPAAVA AALPGLSLRL SSSEKDYYKI LGVSRTASVS DIKKAYRKRA
LETHPDQGGK KEEFAEVAEA YECLSNEEKR RVYDQYGSEA AANMNAANGM GGFGAHSAND
IFAEFFKSRM GGFGDDMRRG PAQVQPIEVK LRMTLEEIYK GVTKKPRVNR PVKCADCRGF
GTKSQTKKPK CAHCDGSGHV VHQHRMGPGM VQQTVTQCPR CGGSGTMAKS DRQCPKCHGM
GYRHLSQEVN IDIPPGVPSN VTLVVRGEGG TLPDAEPGDL HVHVEVAPHK IFTRRGDDLL
MKKEISLSEA LLGTQFSVKM LDGRHVTVKV PHENVLRPDS VLKVSGEGMP SADGGRGDLY
VITHLKMPAK LTAQQREAII QAFGKPNEEK HTSADKTTTA RVMRESAQEL EDAMRDQWDA
QEGGGFRSGA GGQGRGRQSS QQAECVHQ

>Tcj2 (Tc00.1047053507801.130)|||heat shock protein DnaJ (pseudogene), putative|Trypanosoma cruzi|chr unknown74|TIGR||Auto
MVKETKFYDS LGVSPDASVD EIKRAYRRLA LKYHPDKNKE PGSQEKFKEV SVAYECLSDP
EKRSRYDQFG EKGVEMESGG IDPTDIFASF FGGSRARGEP KPKDIVHELP VSLEAFYTGK
TIKLAIIRDR LCPACNGSGS KVPNASVTCK ECDGRGVKLI TRSIGPGFIQ QMPVACPQCR
EKGTDMREER QCDSCRGQQI KKDEKIS*IF RGKRYASWRQ CDISGRR*PN PRSPTLR*YH
YHL*TKTTSC FYA*RGSSCH GTHHLFGGGA HGVHIEHKAS GRP*CFNYIY WRC*SLQTVV
CEQGGYAYSQ HWWRRTRRPC C*VSRRLPKC PESAV**NFR SA*DSSLSST AVSASFCYAV
PFVRNEH*S* EGS*ASSTDR WR**RRCTTR AHWRDLYATV V
This CDS is annotated as pseudo. This is because, either:
The CDS translation contains stop codons
The CDS is frameshifted and the translation presented has been obtained by the inclusion of a join(s)
Protein doesn't end with a stop codon
Protein contains stop codons inside the peptide sequence Send to GeneDB omniBLAST Send to GeneDB BLAST Send to BLAST at NCBI

>Tcj27 (Tc00.1047053510243.30) |||chaperone DnaJ protein, putative|Trypanosoma cruzi|chr unknown83|TIGR||Auto
MRQHRARMQR WVRRAVYPAS LHGGVSCGNL RFLVRSDLTA FTPCIAAAAG LTGFGVSRRF
SSTNVKKDLY SVLGVARNAS QEDIKSAYKK KAKQLHPDVN PNPRAAEDFA DVKQAFDVLS
DPQKRSMYDM TGNSGAADKF GSGPGFNPFA QAGGENPFAG ANPFGGANPF SSAAGGHGFS
FNDFEEIFQK MSGAGKDKTR KPQGPEPGAD IHFKLRLSFL EAVNGCSKEI SYNTPRRCGS
CTGSGFQDSG SKAKCVHCGG RGKKVMSTGF FHMQQDCIHC GGTGETGRSS CTHCSGKGLV
KDRSVQTLPV PKGVDNKERL KVSGKGEAGV RNGPPGNLYV EITVDEHPVF HREGCDIHVI
TPISLATAVL GGTVRVPTLT GEVETKVPVG TQQGDKLVLR GRGVHRPNQN KTGDFFIHFA
VLLPKSLTEK QKSLIAEFSK EEKPLELTDP QMQELKGRYK SWFSS

PS51257 PROKAR_LIPOPROTEIN Prokaryotic membrane lipoprotein lipid attachment site profile :

>Tcj3 (Tc00.1047053511367.138)|||chaperone DnaJ protein, putative|Trypanosoma cruzi|chr unknown86|TIGR||Auto
MVKETEYYEI LGLEAEATEH DIKRAYRRLA LKYHPDKNPG DQEAAEMFKR IGHAYEILSD
EEKRRIYDQH GKAGLEGGSM DEGGLDAADI FSMFFGGGRR PRGERKPRDL VHEMRVSLED
MYNGKTKKIS VTRDRICGAC EGGGIKPGAE RRTCVACRGQ GVQTFVQELF IGMHQRMQQT
CQSCGGEGTT VREVDICGRC RGSGIVKDQK ILEVHIEKGM KHQDVVRFDG EGNEVVGVRL
KGDVLIILAQ KPHDVFRRVG NHLIMNYTIN LQEALCGFEL PVQHLDKRLR LITIPCGQVI
DPGAAWVVRG EGMPLPNTGG LDRGNLVIHF EVEYPTRLSA QQLKSIAKAL GVTESFPRVT
GQKLTLSDVS QRQSRRRSGS QRANAAARRR QMQMAGGMDE EGFTAFHGGH SGAQTVQQAH
E

>Tcj2 (Tc00.1047053511627.110) |||heat shock protein DnaJ, putative|Trypanosoma cruzi|chr unknown87|TIGR||Auto
MVKETKFYDS LGVSPDASVD EIKRAYRRLA LKYHPDKNKD PGSQEKFKEV SVAYECLSDP
EKRSRYDQFG EKGVEMESGG IDPTDIFASF FGGSRARGEP KPKDIVHELP VSLEAFYTGK
TIKLAITRDR LCPACNGSGS KVPNASVTCK ECDGRGVKLI TRSIGPGFIQ QMQVACPKCR
GKGTDMREED KCDSCRGQQI KKDKKIFEIF VEKGMHRGDN ATFRGEGDQI PGVRLSGDII
IIFEQKPHPV FTRKGDHLVM ERTISLAEAL TGFTLNIKHL DDRDVSITST GVVDPSKLWC
VSREGMPIPN TGGVERGDLV VKFHVVYPSA QSLQSNEISD LRKILHYPPQ QSPPPSAMLC
HLSETNIDLE KEAKRRRQTG GDDDDDAPQG HTGATCTQQ

Trypanosoma brucei gambiense Type I Hsp40s

>TbgambJ45 (Tbgamb.42991) |||chaperone protein DNAJ, putative|Trypanosoma brucei gambiense|chr 11|||Manual
MRGLAAAHTS NTKLVSFLFV SPACLGFAPV VHSHDDSGDG ADEIDYYAVL GLTPEATDRE
VRQRFRELSR KYHPDVSSGG DAREMFSRIT RANEVLSDKK KRRMYDMRGE EGLRQLKRLE
AESGGGQFGS ISQLFRHHAA RRLRGKNTEA TLHLPLDVVY TGGRRTVTIN KQKVCTKCRG
TGAKRPSGLK TCEHCGGHGI LRQRLQLAPG MFQSIRQTCP YCGGRGSIMK ERCGVCGGNG
VHRADVELTV DIDAGMPEGH VLSFEMEADE SPDTIPGDLL LSVQTKKHPR FSRRANDVDL
DMTLVVTLKE ALLGFQRRVE HLDGSEFFVN ETGITQYGAV LKVPGKGMPR HNVPSEFGDL
YVKVLFEMPD MLTKEQREEL AEHL

PS51257 PROKAR_LIPOPROTEIN Prokaryotic membrane lipoprotein lipid attachment site profile :

>TbgambJ27 (Tbgamb.23116) |||chaperone protein DNAJ, putative|Trypanosoma brucei gambiense|chr 9|||Manual
MRQRSPLFQQ YVRRAIPSTF SLVPSRHFTA PGVRPNTVCA SGSIICTAVV SAGLGVSRRF
SATNAKKDLY SVLGVARNAT QEEIKSAYKK KAKQLHPDVN PSPRAAEDFA DVKQAFDVLS
DPQKRSMYDM TGNSGAADKF GAGPGFNPFA QAGGENPFGA GANPFAGAAG GQGFSFNDFE
EIFQKMSGAG KDKTRKPQGP EPGADIHFKL RLSFLEAVNG CSKEISFNTP RRCGSCTGSG
FQDSGSKAKC VHCGGRGKKV MSTGFFHMQQ DCTHCGGTGE TGRSSCTHCS GKGIVKDRSV
QTLPVPKGVD NKERLKVSGK GEAGVRNGPP GNLYVEISVD EHPVFHREGC DIHVITPITL
ATAVLGGAVR VPTLTGEVET KVPVGTQQGD KLVLRGRGVH RPNQNKTGDF FIHFAVLLPK
SLTEKQKNLI ADFAKDEKPL ELTDPQLQEL KGRYKSWFAA

>TbgambJ46 (Tbgamb.2981) |||chaperone protein DnaJ, putative|Trypanosoma brucei gambiense|chr 3|||Manual
MRCTMSSPPF RLTRLSVVIT IAFFRNVHAE EEDDYYRVLG LDAEHEDVSE RDIKSAWRKL
SKKHHPDVAG ESQRVMYQRI QRAYEVLGDR RRRKIYDILG TEGLKKYEKP QEQGRHQSIF
DTFFSFIGGE SGGNADRGSD EELMLLVTLE DMYKGAAHTA KLPRIKICRK CRGTGARSKD
DYVKCPHCGG GGRVVRRVQI APGFIQQIEQ VCGQCGGGGR VVRRKCPVCG GHRLVRGSSS
VSIDIEQGTP NGYKMTYEME ADQQPNKMPG DLIFTIVTIP HPEFARMSSG KEGVPDDLST
AVELTLKEAL LGFNKTLKHL DGRVLSLVET GVTKFGQIRK YAGEGMPRHH VPSERGNLLV
LYTVELPKIL TEEQRKAIER ALD

>TbgambJ50 (Tbgamb.24631) |||chaperone protein DNAJ, putative|Trypanosoma brucei gambiense|chr 9|||Manual
MFRFTSVSSI WRRLAAAPPT AATAAFANVS KRLSSSNKDY YKILGVSQSA SQSDIKKAYR
KRALETHPDQ GGNKEDFAEV AEAYECLSNE EKRRIYDQYG SEAAANMNAG GGMGGFGGRS
AEDIFAEFFK GGMGGFGGNR SAGPPQVPPL EVTLRMTLEE VYKGASKSPR VNRPVVCSDC
RGFGTKSQKK KPKCSECDGS GHVVQHHRFG PGMVQQTVSQ CPRCGGAGTV AKPDDKCPKC
KGMGYRHLVQ SVSIDIPAGV PPDVTLVVRG EGGTMPEAEP GDLHVHVEVE EHNVFKRRGN
DLVVERDVTL SEALLEFDLS LKTLDGRSIT VKSPKSSVLQ PNSVLRVAGE GMPNSSGGNG
DLYIVTKLKL PRTLTDQQKE AVKKAFGEPK KKDSDSGSDK TVTASVLRGG VREMEEALHS
NWDSERGGGS QQGNGRRRSG RKQHTAECVH Q

Trypanosoma vivax Type I Hsp40s

>TvivJ27 (Tviv1192b08.p1k_5) |||chaperone protein DNAJ, putative|Trypanosoma vivax|chr unknown|||Manual
MRRRGVLPFH GFARAVTPPF IHRVCEPFRR FGHATSPALS CGYTAATAHA PLRCPKRFSA
TNSKKDLYST LGVSRNATQE EIKSAYKKKA KQLHPDVNPN PRAAEDFADV KQAFDVLSDP
QKRSMYDMTG NSGAADKFGA GPGFNPFAQA GGENPFGAGG NPFANAAGGQ GFSFNDFEEI
FQKMSGAGKD KTRKPQGPEP GADIHFKLRL SFLEAVNGCS KEISFNTFRR CGSCTGSGFQ
DSGSKAKCVH CGGRGKKVMS TGFFHMQQDC THCGGTGETG RTSCTHCSGK GIVKDRSVQT
LPVPKGVDNK ERLKVSGKGE AGVRNGPPGN LYVEISVDEH SVFHREGCDI HVITPVTLAT
AVLGGTVRVP TLTGEVETKV PVGTQQGDKL VLRGRGVHRP NQNKTGDFFI HFAILLPKSL
TEQQKSLIAE FAKDEKPLEL TDPQLQELKN RYKSWFSV

>TvivJ3 (tviv1100a12.q1k_0) |||DNAJ protein, putative|Trypanosoma vivax|chr 10|||Manual
MVKETEYYDL LGVPPDASEN DIKRAYRRLA LRYHPDKNPG DENAADMFKK IGQAYEILSD
EEKRRIYDQS GKDGLSGGGY EGEFDPSDIF AAFFGGSRRP RGERKPKDLV HELRVSLEDM
YNGRVKRVSV VRDRLCGSCE GTGVPPGAQL QPCAACQGCG VQVLVQQLFP GVQQRVQVAC
QTCGGEGRCV RSIDVCTECR GNREVKMEKV LEVHIERGAK HEDVLRFEGE GDEIPGMRLK
GDVLIILDEK PHDVFRRAGN HLIMNYRITL QEALCGFELP VQQLDKRMLL VKVPSGQVVD
PEVAWVLHHE GMPLANTGGC EKGNLIIHFE VDFPSKLSER MINQIAEAFN LPSKFPPVAG
QKVKLQDPNS RRKRRVDVQR APESQQHSQT PFGFHGTEFM SFDSHGSGGQ TARSTHG

>TvivJ45 (tviv1163f03.q1k_15) |||chaperone protein DNAJ, putative|Trypanosoma vivax|chr 11|||Manual
MTHVLTSEKS FTMHFISSFH AFPVPIGWFL FLCSAMLSTL VAGDDQRQEE MEEIDYYAVL
GLNEDATAKD IRQKFRELSR KYHPDVARTA EAREMFTKIS RANEVLSDKK KRRMYDMRGE
EGLRQLQRAE ASGNSGQSNS IFSQLFSMRN QQFKGQNSEA TFRVPLETVY TGGRQVLSLN
KQKVCTQCKG TGAEKNSGTV TCPRCSGHGV LIQRMQLGPG MYQSMRTCPT SCGGKGHVVK
KQCSACHGRR VVRADVELVL DVEAGIPEGH TVTFEMEADE SPDLIPGDFL LHVRTQPHDR
FSRRENGVDL DTTLTVTLKE ALLGFERSFP HLDGKEFTVR AEGVTPYGTV LKLSGKGMPR
HHVPSERGDL YVKVLFDMPA FLTDSQRKEL EEHL

>TvivJ50 (tviv1189h03.q1k_1) |||chaperone protein DNAJ, putative|Trypanosoma vivax|chr 9|||Manual
MFRFTATMGM RRRCCANVRP CSVPGVAMRF SSEGKDYYKI LGVSQSASPS EIKKAYRKRA
LETHPDQGGN KEDFAEVAEA YECLSNEDRR RVYDQYGSEA ASNMNAGGNM GGFGGRTAED
IFAEFFKGGM GGFGGETRRG PPQVPPLEVT LNLTLEEVYK GVVKSPRVMR PTICSECRGF
GTKSQKRKPK CAHCDGNGHV VQQRHFGPGM VQQTVTQCPR CNGEGTMAKA DDQCSKCRGM
GYRTVIQNVS IDVPAGVPPD VTLVVRGEGG TMPEAQPGDL HVHIQVASHE TFKRRGNDLL
VKKKITLSEA LLGFHLTLKM LDGRSICVEA PKEAVLQPSS VLKVPNEGMP DAHGGRGDLY
ILTRLKLPRK LTEAQRNAIK QAFGDSKEGS TDTSGKTVTA SVMNETVDEL ESTMASQWEA
QQRGGFTQDG GSKRGKRRQQ TAEGVKD

>TvivJ2 (tviv623d01.q1k_23) |||chaperone protein DnaJ, putative|Trypanosoma vivax|chr 2|||Manual
MVKETKYYDA LGVPPSASED DIKRAYRRLA LKYHPDKNKE PGANEKFKEV SVAYECLSDP
EKRKRYDQFG EKGVEMDGAG VDPTDIFASF FGGRRARGEP KPKDITYEHP VPLETFYSGK
TIKLSIVRDR LCSKCNGSGS SLPNSSTKCR ECDGRGVKLI TRSIGPGFIQ QMQVTCPRCS
GKGTDIREED KCQGCKGAQI TKDKKVFEVV VEKGMQRGDH VTFQGEGDQI PGVRLAGDII
IIFDEKPHPV FTRKGDHLIL EHPISLSEAL TGFVLNIKHL DNRQLSIQST GIIDPTKLWC
VSREGMPVPH TGGVERGDLI VKFKVMYPAA QSLPNEDAVT LRRILGYPQQ HEPHPDAMVL
GLTESSIDLE KLKSTRQQFA DDDDDDGHQT HTGATGAHG

>TvivJ46 (Tviv1689e09.p1k_6) |||heat shock protein DnaJ, putative|Trypanosoma vivax|chr unknown|||Manual
MRPATLAKAI SVALPCLLAT LAWRMAAAEA DEEDYYSVLG LGKEREEANE RDIKSAWRKL
SKKYHPDLSG ESNRKRYQRI QQAYEVLGDR RKRKIYDILG VEGVKDFDKT RERKGRGGSL
LDSFASFFGG SSHEQHRGND EELPLVVPLE DLYTGASWTV KLPRIKLCRA CRGTGARSKT
DVVSCPRCKG KGRVVQRFEI IPGFVQQVER ECDRCGGHGH TIKERCPVCQ GRRMVRGISS
ISIDIEQGTP NGHKLTYELE GDQRPGIVPG DVIFTVSTAP HPQFRRTSDG ASDKADDLAT
TLTLTLKEAL LGFNRTIKHL DGRAVELSES GVTKYGETRR VKGEGMPRHH VPSERGDLLV
TYLVMLPKTL TRSQREAVER ALG

Trypanosoma congolense Type I Hsp40s

Protein doesn't end with a stop codon Send to GeneDB omniBLAST Send to GeneDB BLAST Send to BLAST at NCBI

>TconJ27 (congo92c09.q1k_8) |||chaperone protein DNAJ, putative|Trypanosoma congolense|chr 9|||Manual
MRQRSVLLRH HMKRLAAFAP FTLSERQCIV EKRCIGVCST GSSAASCAAL VVARRFSATN
AKKDLYSVLG VARNATQEEI KTAYKKKAKQ LHPDVNPNPR AAEDFADVKQ AFDVLSDPQK
RSMYDMTGNS GAADRFGSGP GFNPFAQAGG ENPFGAGANP FAGAAGAQGF SFNDFEEIFQ
KMSGTGKDKT RKPQGPEPGA DIHFKLRLSF LESVNGCSKE ISFNTFRRCG SCTGSGPQDS
GSKAKCVHCG GRGKKVMSTG FFSMCQDCTH CGGTGETGRS SCTNCSGKGI VKDRSVQTLP
VPKGVDNKER LKVSGKGEAG VRNGPPGNLY VEISVDDHPV FHREGCDIHV ITPITLATAV
LGGTVRVPTL TGEVETKVPV GTQQGDKLVL RGRGVHRPNQ NKTGDL

>TconJ2 (congo365g12.p1k_2) |||chaperone protein DnaJ, putative|Trypanosoma congolense|chr 2|||Manual
MVKETKYYDA LGVSPDASED DIKRAYRKLA LKYHPDKNKE PGANEKFKEV SVAYECLSDP
EKRRRYDQFG EKGVEADGVG IDPTDIFSSF FGGRRARGEP KPKDIVHEQS ISLDAFYNGK
TIKLSISRDR LCSSCNGSGS KVPNASVRCR DCDGRGVRLI TRSIGPGFVQ QMQVSCSRCQ
GKGTDIREED KCGNCRQQQI VKDKKVFDVV VEKGMQRGDH VTFQGEGDQI PGIHLPGDII
IIFDEKPHHM FTRKGDHLLM EHTISLAEAL TGFTINIKHL DGRELSLQSN DVIDPQKLWS
VSREGMPVPR TGGIEKGDLV IKFHVVYPTA GSLPASSVTP LRSILGYPQQ SEPHPDATVC
TVAENTIDLD KEAKRRRVAT ADDDEDMGQH AHTGATGTGG

>TconJ46 (congo1293a06.q1k_4) |||chaperone protein DnaJ, putative|Trypanosoma congolense|chr 3|||Manual
MWVVRDPVTL SAWIFISVTL AITVVGGDEE EDFYEILGLE KEREDASERD IKSSWRKLSK
KHHPDLAGES QRVRYQRIQR AYEVLGDRRK RKIYDILGVE GLKKYERPDE GQRMNQGIFS
TFFSFVGGSG GNDRGEDEEV TLLVPLEDMY NGAAHTVRMP RMKICRKCRG TGAKSKEDYQ
QCPYCRGSGR MVRRVQIVPG FVQQVEHVCD HCEGRGRVIK KVCPVCGGHR VVQGTSSISI
DIEQGTPDKH KLTYELEADQ KPNQVPGDIV FTITTLPHPR FVRVSSGKPD KPDGLATTVE
LTLREALLGF NKTLEHLDGR VLSLTETGIT KHGAVRRYAG EGMPRHHVPS ERGSLRVVYE
VHLPTSLTEE QRRVIEQALG

Protein doesn't end with a stop codon Send to GeneDB omniBLAST Send to GeneDB BLAST Send to BLAST at NCBI

>TconJ3 (congo541b10.q1k_1) |||chaperone protein DNAJ, putative|Trypanosoma congolense|chr 10|||Manual
MVKETEYYEL LGVAVDASEN DIKRAYRRLA LRYHPDKNPG NEEAADMFKK IGHAYETLSD
TEKRHIYDQH GKDGLSGSGG DADFDASDIF SMFFGGGRRP RGERKPKDLV HELAISLEDM
YNGRVKRVTV VRDRICDICN GNGMRPGAQQ QTCGSCGGHG VQMFVQNVIP GVRQQVQVTC
QSCGGCGKYA LESDLCPRCH GRRKVKSEKV LEVVIERGMK ADDALRFEGE GDEIAGMRLK
GDVLIIIAEK PHDMFRRVGS HLIMNYRITL QEALCGCELP L

Leishmania infantum Type I Hsp40s

>LinJ3 (LinJ21_V3.0550) |||DNAJ protein, putative|Leishmania infantum|chr 21|||Auto
MVRETELYEV LNVSVEADEH EIKRSYRRLA LKYHPDKNTG DEAAADMFKK VSNAYEVLSD
PEKRQVYDKY GKEGLERGAG EGGGFHDATD IFSMFFGGGA RERGEPKPKD IVHELEVKLD
DLYNGATKKV MISRDPLCGT CEGSGLKPGG KRTTCAQCRG RGVLLRTQQV PPGFHHQVQM
RCPACQGEGE IVAASDLCTG CRGKRAVREK SVLEVHIDRG ASKSDHFTFT GEGNQEPGIR
LSGDVLIFLS VRPHPVFHRI NDHLMMRCPI TLQEALCGFE VPIEHLDGRQ LVIKASPGQV
VHSDSAWSVY NEGMPVKGTG GLQKGKLFIY FDVEWPETLP REQIDKIVTA LNVPEKPGKL
GGHVVELTEY KAAGKFKGGK NGKRGGAAGS RSAGAGRGRG AARSRQAHAE EDEFEDVTDD
DDDEQQQYGG QQYFRAGPQG FNGSTQTVEE AQQ

>LinJ2 (LinJ27_V3.2350) |||heat shock protein DnaJ, putative|Leishmania infantum|chr 27|||Auto
MVKETGYYNA LGVSPDASED EIKRAYRKLA LKYHPDKNTE PGAQEKFKEV SVAYECLSDP
EKRKRYDQFG KDAVEMQGGG VDPSDIFASF FGGGSRPRGE PKPKDIVHEL PVPLEAFYCG
KTIKLAITRD RLCTQCSGTG SKVAGVSATC KDCGGRGVRM MTRQLQPGFI QQIQTACPVC
KGKGTNLREE DKCVSCRGQQ ITKDKKVFEV MVEKGMHRGD SVTFSGEGDQ IPGVKLSGDI
IIILDQKPHQ AFIRKGDHLF LEQTISLAEA LTGFSLNITQ LDGRELAISS TAGTIIDPAN
MYSVSREGMP VAHTGGMERG DLIIRFKVVF PKTLRQGCVP ELRKMLGYPQ QPPFKDGAEQ
YTLQESHIDL EKEARRNAYD DDGDQPRVQT AGEAEE

>LinJJ46 (LinJ25_V3.1140) |||DNAJ-like protein|Leishmania infantum|chr 25|||Auto
MRGRHITRVV AALLVLVWVA AFVAEVPVHM AGAADPRDED AKAVDAVLRL PEDDFYAVLG
LDEAREDATE RDIKNAFRRL SKKYHPDVAT GDQDSYRLVY QRVQRAYEIL GDRRKRKIYD
ILGIDGVTRL EKPQQQQQQM NPFFAFFGVG QQADAERGKD MELLMVVPLE DIYRGAAHTS
RFAKRKICRA CKGTGARSGE DVVKCPHCQG RGRLVQRVQI APGFVQQMEQ VCPHCQGKGT
HVAHMCPVCR GKKVLPGEAV LSVDIEEGLP EGHVLTYELE ADQAPGQVPG DVLLTVVSAP
HPVFHRSGND LYANVSITLK EALLGFKKTF THLDGHNVEL HWDGVMQNTQ QVRIAGEGMP
RHHVPSERGD LYITYNVLLP AALTVEQRAL FQEHFA

>LinJ50 (LinJ35_V3.3030) |||chaperone protein DNAJ, putative|Leishmania infantum|chr 35|||Auto
MLRYLSASTA RRACAWSAGA SLSAAPGAAS TFCSSTTPAR LYSSGNKDYY KMLGVDRNAD
LKEIKKAYRK RALETHPDQG GNKEEFAEVA EAYEVLSNPE KRKVYDQYGS EAATNPSMGG
AGMGGFGAGG RSAEDIFAEF FRGGMGGMGG FGDMFGGGPG GRQGTPTLQP LEVRTRLTLE
DVYKGVTKTM RVNRPQVCAD CTGFGTKSKT EKPKCTQCNG SGHVVQQHRM GPGMVQQTIS
ECPRCEGTGT VAKPEDQCHK CHGKGYRTVS QDVTIEIPAG VPSNVTLVVR GEGGTIPGAP
PADMHLHVEL SPHRVFQRRG NDLIVNKDVT LQEALLGLHM PLKLLDGRTV NVETSADQIL
KPEGVIKVTG EGLPGTSGER GDVYIFTHLK MPNKLSDEQK AHIKKAFGKP EKDADASPGN
TVRARVLRES REQLEEQKRS LWASQESGGY SGGGGGNSRR PAGSTHGSGA QQVEEATG

>LinJ27 (LinJ04_V3.0940) |||chaperone protein DNAJ, putative|Leishmania infantum|chr 4|||Auto
MKASPCRLAA SRPPAAARLY RAYGPVHLLP RTTAMRRLAA ASVMMATAGS LGETLLSSAA
VSAGGAAHSA TGGSTALLQH RRWQSGGGSK KDLYSVLGVA RNATPEQIKS AYKKRAKALH
PDVNPSPTAA EDFAEAKQAY ETLSDPQKRS LYDMTGNASA AGGFGGPGGT GGGFNPFGAG
GNPFAAGGNP FANMGGANQP GGGQGGFSFN DFEEIFQKMS NSGKDKTRKP QGPEPGADIH
YKLVLSFLDA VNGCQKEISY NTMRRCGACT GSGCQDTGSR TKCPHCGGRG KKVMSTGFFH
MQQDCTHCGG TGELGRTTCT QCSGKGIVKD RSVQTLPVPK GVDNKERLKV TGKGEAGVRN
GPPGNLYIEI SVEDDPVFHR EGSDIHVITP ITLSTAVLGG TARVPTLTGE VETRVPVGTQ
QGDKLVLRGR GVHRPNHNKT GDFYIHFAVM LPKELTEEQK KAIADFSKDE RPLHLNDAQL
QELKGRYRSW FAT

>LinJ45 (LinJ32_V3.3500) |||chaperone protein DNAJ, putative|Leishmania infantum|chr 32|||Auto
MTILHSRRAA VRCACMLVLV LLCVALSSRA FFDFGGGRRA DAPSAEVHHA QEVDYYKVLQ
LEDKREEATE KDIRQQFRRL SRLYHPDVAK TEEDKAKYSQ VNRAYEVLSD KRKRKVYDMR
GERGLEQLEH IDRSKDTPGG GMNPLARLFG MRVDDGLRGP DMELEAKVDL AKLFTGGQET
LQVNKQKVCH ACKGSGADTK AAIVQCRQCG GEGVLRQRIQ FAPGMIQELH QKCPSCQGAG
RRPEPLCSVC RGKKVLDGSS TVTLELEPGM EEGHVLKFEM EAEESPDRLP GDLLVHVHTL
PHPVFSRRRN QLDLDTSLTL TLQEALVGFD RNITHLDGVE QVRVQRLDTV SPYGTVLRLP
GKGMPKMNVA SERGDLYVRL QYDMPAQLTE EQRKLVEMLL

>LinJJ4 (LinJ15_V3.1220) |||hypothetical protein, unknown function|Leishmania infantum|chr 15|||Auto
MVKETDLYDE LGISPDATEP QIRSAYRRKA LQYHPDKNSG DPAAAEKFKK VAEAYEILSD
AERRKQYDTF GRNGLGSAAG GSGGVPGGGF GSSFGAGIDP MDIFSSFFGF ASGGGPGASR
QQRPAKPSFI LVELQCSLEE LYCGTVKVLQ VRRRRLCPSC HGHGTSSGRP SQVCPQCKGN
KIVTKALNTG GLTAYQQCRC DRCEGLGKLP IQSPCAHCGP VYDAHRRSPL SGTHKPDPLP
RGVVVEEKMI KVSIEPGTED SDALHYVGQG DELPGFDAVG DVLVVIEQLP HPYYRRVNST
DLLLSNCRVP LGCLFRDAFS IPIELLDGRV VRLATPLRDG NVPHFLFDSQ HVFVVDNEGM
PLKERKRGGA AADANHDGSE AAEGVAHRKK GKLYLCIDVV FPSSLTPAQV DIITKALGGS
HERDPAATGG DAGRVVTLQS HRGPAPSWYT VDAEGNAWYQ PNALPTAKKK RKATSTAE

Leishmania braziliensis Type I Hsp40s

>LbrJ46 (LbrM25_V2.0990) |||DNAJ-like protein|Leishmania braziliensis|chr 25|||Manual
MLGPRTTRVV VALLVLAWVV TLLAEVPLVA ADAADPRDED VKAVNALLQL PEDDFYALLG
LGEAREDSTE RDIKNAFRRL SKKYHPDVAT GDQDSYRLVY QRVQRAYGVL GDRRKRKVYD
ILGVDGVAKM EKPQQEQHVN PFFAFFGVGQ DTNGERGKDM ELLMVVPLED IYRGAAHTAR
FAKRRICRAC KGTGARSAEE VVKCPHCQGH GRLVQRVQIA PGFVQQVEQA CPHCQGKGTH
VAYMCPVCGG KMVPPGRAVL SVDIEEGLPE GHVLTYELEA DQTPGQVPGD VLVTVVSAPH
PLFRRSGNDL YANVSITLKE ALLGFEKTLA HLDGHEVELH WDGVIQNTQQ VRITGEGMPR
HHVPSERGDL YITYNVVLPS ELTAEQRAFF QEHFA

>LbrJ45 (LbrM32_V2.3590) |||chaperone protein DNAJ, putative|Leishmania braziliensis|chr 32|||Manual
MVSTVVCSHR AVVRCACALA LLLVCVALNS NAFFDFGGGH HTDASKAQVR RGPEVDYYKV
LQLEGKREEV TEKDIRQQFR RLSRLYHPDV AKTEEDKAKY SQVNRAYEVL SDKRKRKIYD
MRGEQGLAQL EELDRTKGSP GGGMNPLAQL FGMRADNGLR GPDMELEAKV DLAKLFTGGQ
EILRINKRRV CQACKGSGAD ATAAVVQCRQ CGGQGVLRQR IQLAPGMIQE VHQRCTSCGG
AGRPPEPVCP VCRGRKVMQQ SSTIVLELEP GMTENSVLKF EMEAEESPDR LPGDVVVRVH
THPHPVFSRR RNQLDLDTSL TLTLKEALVG FDRNITHLDG EEQVRVYRRD AVSPYGTVLR
LPGKGMPKLH VPSERGDLYV RLQYDLPARL TKEQKELVEK LL

>LbrJ27 (LbrM04_V2.0730) |||chaperone protein DNAJ, putative|Leishmania braziliensis|chr 4|||Manual
MNAVSYQLAA FHSTAAARLY RTYGAVHLLP RTAALRRMTA TAGSLGAAHS AMGGSMVLLQ
HRRWQSSGGS KKDLYSVLGV ARNATPEQIK SAYKKRAKAL HPDVNPSPTA AEDFAEAKQA
YETLSDSQKR SMYDMTGNAS AASGFGGSGS GFNPFGAGGN PFAAGGNPFA NMGNANHSGG
GQGGFSFNDF EEIFQKMSNS GKDKTRKPQG PEPGADIHYK LVLRFLDAVN GCQKEISYNT
MRRCGACTGS GCQDTGSRTL CPHCGGRGKK VMSTGFFHMQ QDCTHCGGTG ELGRTTCTQC
SGKGIVKDRS VQTLPVPKGV DNKERLKVTG KGEAGVRNGP PGNLYIEISV EEDPVFHREG
SDIHVITPIT LSTAVLGGTV RVPTLTGEVE TRVPVGTQQG DKLVLRGRGV HRPNHNKTGD
FYIHFAVMLP KELTEEQKKA VADFAKDEKP LDLNDAQLQE LKGRYRSWFA A

>LbrJ2 (LbrM27_V2.2610) |||heat shock protein DnaJ, putative|Leishmania braziliensis|chr 27|||Manual

```
MVKETGYYNA LGVSPDAGED EIKRAYRKLA LKYHPDKNTE PGAQEKFKEV SVAYECLSDP
EKRKRYDQFG KDAVEMQGGG VDPSDIFASF FGGGGRPRGE PKPKDIVHEL PVPLEAFYCG
KTIKLAITRD RLCSQCNGTG SKVAGVSATC KDCGGRGVRM VTRQLQPGPI QQIQTACPAC
KGKGTSLREE DKCLSCRGQQ IHKDKKVFEV VVEKGMHRGD SVTFSGEGDQ IPSVKLSGDI
IIILDQKPHP NFIRKGNHLL MEHTISLAEA LTGFSLNITQ LDGRELAVSS SAGTVIDPAT
MYSVNREGMP VAHTGGMERG DLILHFRVVF PKTLRPTAVP ELRKMLGYPQ QPPTKDGAEM
HTLQESHVDL EKEARRNAYD DDGDQPRVQT AGCAQQ
```

>LbrJ3 (LbrM21_V2.0550) |||DNAJ protein, putative|Leishmania braziliensis|chr 21|||Manual

```
MVRETELYEV LNVSVDADEH EIKRSYRRLA LKYHPDKNTG DEAAADMFKK VSNAYEVLSD
AEKRQVYDKY GKEGLEKGMG EGGGFHDATD IFSMFFGGGA RERGEPKPKD IVHELEVTLD
DLYNGATKKV MISRNRFCGT CKGSGLKPGG KRTTCPQCRG RGVLLRTQQV FPRVPSSGAD
ALHCLRWRGR NRCGH*ICTG CRGKRAAREK SVLEVHIDRG TSKSDHFTFT GEGNQEPGIR
LSGDVLIFLS VRSHPVFHRI NDHLMIRCPI TLQEALCGFD VPIEHLDGRE LIIKASPGQV
VHGDSAWSVY NEGMPVKGTG GLQKGKLFVY FDVQWPETLP RVQIDKIVTA LNVPEKPNKL
GGQVVELSEY RAAGKFKGGK KGMRAGVAGA GSRSAGAGRG REAARGRQAY AEEDEFEDVT
DDDGDDREHQ QHSGQQYFRA GPRGFSSNTQ TVECAQQ
```

***CONTAINS STOP CODONS**

>LbrJ50 (LbrM34_V2.2890) |||chaperone protein DNAJ, putative|Leishmania braziliensis|chr 34|||Manual

```
MLRQLSALAT RRACAFSASA SVLTGSTASS TFCSTTTPVR LYSSGNKDYY KTLGVDRNSD
LKEIKKAYRK RALETHPDQG GNKEEFAEVA EAYEVLSNPE KRKVYDQYGS EAATNPNMGG
PGMSGFGAGG RSAEDIFAEF FRGGMGGMGG FGDMFGGGPG GRQGTPTLQP LEVRTRLTLE
DVYKGVTKTM RVNRPQVCSE CRGFGSKSKT EKPRCTQCNG SGHVVQQHRM GPGMVQQTIS
ECPRCRGTGS MAKPEDQCPK CHGKGYETVS QDVTIEIPAG VPSNVTLVVR GEGGTIPGAP
PADMHLHVEL SPHRLFQRRG NDLIVNKDVT LQEALLGLHM PLKLLDGRTV NVETTADQVL
KPEGVIKITG EGLPGTSGER GDIYIFTHLK MPNKLNEDQK ENIKSAFGAP GRDASASPGN
TVRARVMRES REQLEEQKRS LWASQESGSY SGGSGGSSRR SAGSTPGGGA QQVECATQ
```

>LbrJ4 (LbrM15_V2.1170) |||hypothetical protein, unknown function|Leishmania braziliensis|chr 15|||Manual

```
MNNLLLPPPP PSPSFHTTPN PFSSPSPVPA RLLFGCVACI LSKHCVEGLD EIHTHTRTRR
PPLSSLPRIR LRRRCPPTPP PLTPVAILQL GQMVKETVLY DELGISPDAT EQQIRSAYRS
KALQYHPDKN NGDLAAAEKF KKVSEAYEIL SDADRRKQYD TFGRNGPGGA AGGSGGFPGG
FGHGFGPGMD PMDIFSSFFG FPAGGGPGKS QQQRPAKPSF ILVELQCSLE ELYCGTTKVL
LVRRRRRCPT CQGHGTSNGR PSPVCPQCKG NKTVMKSLSF GGLMTYQQCR CDRCQGLGKL
PIQFPCARCG PIYDAHQRAQ LSGTLPRGVV EEEKKMKVTI APGTEDSDAL NYPGQGDELP
GFDAAGDILI VVVQLPHPHY HRLNSTDLLL SNCRLPLWSL FRDAFSIPVE LLDGRVVRLA
SPLREDNVPH FLFDSHHVFV VENEGMPLKA QTKGGAAADE SGDGSGSAAG AAHLKKGKLY
VCINVVFPHS LTPAQMDAIT KVLGGSGECD RDAIAGDADS VVTLQPHHGP APSWHRVDAE
GNAWYKPRTA AAAKRKKPTR AA
```

Type IV Hsp40s of various Trypanosomatid species

Trypanosoma brucei Type IV/I Hsp40 proteins

>Tbj27 (Tb09.211.0330) |||chaperone protein DNAJ, putative|Trypanosoma brucei|chr 9|||Manual
MYDMTGNSGA ADKFGAGPGF NPFAQAGGEN PFGAGANPFA GAAGGQGFSF NDFEEIFQKM
SGAGKDKTRK PQGPEPGADI HFKLRLSFLE AVNGCSKETS FNTPRRCGSC NGSGFQDSGS
KAKCVHCGGR GKKVMSTGFI HMQQDCTHCG GTGETGESSC THCSGKGIVK DRSVQTLPVP
KGVDNKERLK VSGKGEAGVR NGPPGNLYVE ISVDEHPVFH REGCDIHVIT PITLATAVLG
GAVRVPTLTG EVETKVPVGT QQGDKLVLRG RGVHRPNQNK TGDFFIHFAV LLPKSLTEKQ
KNLIADFAKD EKPLELTDPQ LQELKGRYKS WFAA

>Tbj47 (Tb927.1.1230) |||chaperone protein DNAJ, putative|Trypanosoma brucei
|chr 1|||Manual (Tbj47)
MRGITLALAP PSLLFVPKIQ RRCFNVIQRG NDREVDSLFA LLGFAGDNEA HRIRRTRAEL
RQGFMREAMK LKDPQNDKSD AAKLEKLREA YQLLSNDRFR VQYAAHHYAS PDASLHLLVD
GGQVAANFNP EHQSFNFVDH AISRAAMSPS SRSSSDKQRS FSDFTGQYNS VIGNTGCSTD
ARPYNAPEAR AAINGAGINF MLRISFDESV LGCTKTAVYE KNVSCQRCSG NGRMVLKRPR
KCPQCRGRGS THLPSATYHI ERSCTYCNGD GVTPPPKCSG CRGAGVVPGH TVQVPVDIRP
GTTNMTACRL RGMGHDGVRG GVAGDLIVTV LVQEHRVFHR DGLDLHMVLP ITLSTALLGG
MVSVPLLHGP FCTRVPPCVR NGQQIRLSGR GVTLDGSGVL TNAEEGIDTD SSASKQEQQQ
RGDLYIHLLV VIPKGEELTG AQRSALEQFV VEQDGNGAEG VDDITPTALK RRFRHWLPGT

Leishmania major Type IV/I Hsp40s

>LmJ47 (LmjF20.0550) |||chaperone protein DNAJ-like protein|Leishmania major|chr 20|||Manual
MLAPRSVRWC GGNRRSSGSA SVRVGPHVRT TGPWVPWHTL LYQRRCFQVI QRGHDRPQES
LYNKLGFAKD TEARLVRRST TELRQALLRR VEALTDPVND PADKKRLSEL KNAYTHLQDD
RFRTNYASHY YASNDARLHV LCDGGQVAAN FNPEHQQFTY VDHAIDREST RSGVGGCLPT
GKAGGSASCS GPSGRSEVPT AEALGDAFRN ATGVGESAGG TTSSSATAYK AADAAGALNG
TDITHLLRVT LEQSLAGCEV EVAIHKNVRC ATCRASGRQQ LRTPRKCPQC LGRGSAHMFS
ATYHIERRCL YCGGEGTVPA PPCRACDGRG VQLHQAVQVP VSVPAGTLSN SLFRLRHQGH
DGVRGGHAGD LLLTVLVSEH RYFYRSPERS HELHAMLPLP LSVALLGGRV KVPTLNGFGT
VHVPPCVRSG EVLPLNVYGV PDATNRASAA AAAPHTILYH ALVMIPKGDA LSGRQKAALQ
LYEAHHSYPS ATAAEVEAVG RPRTFGAEAS QQQTLSSSAG TVSREQLMAS CAALKSSYKH
WFHAD

Trypanosoma cruzi Type IV/I Hsp40s

>Tcj47 (Tc00.1047053511423.170) |||chaperone DnaJ protein, putative|Trypanosoma cruzi|chr unknown61|TIGR||Auto
MRQSASAVLL PLLQRHAQSI PRRHFNVIQR GNDREVESLF ALLGFGKESE VERVRRTRDE
LRQGYLREAM KLKDPQRNAK DAAKLADLRR AYTLLSDDQF RARYASHHCV SPDATLHVHV
DGGQKAANFN PEHQSFEFVD HALSFSSSSL LSSCAQRSFG DFTGPFQSAT GVYSSPADAH
PYQPRESSTP QNGNSINFML RISFDESILG CTKSAVYEKD ITCSQCSGDG RQVLNRPRKC
PQCRGRGSTH LPSATYHIER SCGYCGGKGV APPPKCGRCG GAGVIRGHTV SVPIDVRPGT
TTMTVRRFRG KGHDGVRGGN AGDLIVTVLV QEHRLFHRDG IDLHMVLPIP LSVALLGGVV
SVPTLHGAQT LRIPPCVRNG QRLTMDGQGV CLDGESDAAE KNNLEACEGL ANPKQQHCQQ
QRRRGHLYVH LLVVIPRREE LTGAQRCALE SFAAEGEGGI NQENGEREEQ ETPASLKQRF
RHWMTSL

>Tcj47 (Tc00.1047053507949.10) |||chaperone DnaJ protein, putative|Trypanosoma cruzi|chr unknown75|TIGR||Auto
MRQSALAVLL PLLQRHAQSI PRRHFNVIQR GNDREVESLF ALLGFGKESE VERVHRTRDE
LRQGYLREAM KLKDPQRNAK DAAKLADLRR AYTLLSDDQF RARYASHHCV SPDATLHVHV
DGGQTAANFN PEHQSFEFVD HAISFSSSSL LSSCAQRSFG DFTGPFQSAT GVYSSPADAH
PYQPRESSTP QNGNGINFML RISFDESILG CTKSAVYEKD ITCSHCSGDG RQVLNRPRKC
PQCRGRGSTH LPSATYHIER SCQYCQGKGV APFPKCGRCG GAGVIPGHTV SVPIDVRPGT
TTMTVRRFRG KGHDGVRGGN AGDLIVTVLV QEHRLFHRDG IDLHMVLPIP LSVALLGGVV
SVPTLHGAQT LRIPPCVRNG QRLTMDGQGV CLDGESDAAE KNELEACEGS ANPKQQHCQQ
QRRRGHLYVH LLVVIPRGEE LTGAQRCALE SFAAEGEGGI NQENGEKGEQ ETPASLKQRF
RHWMTSL

Trypanosoma brucei gambiense Type IV/I Hsp40s

>TbgambJ47 (Tbgamb.0216) |||chaperone protein DNAJ, putative|Trypanosoma brucei gambiense|chr 1|||Manual
MRGIALALAP PSLLFVPKIQ RRCFNVIQRG NDREVDSLFA LLGFAGDNEA HRIRRTRAEL
RQGFMREAMK LKDPQNDKSD AAKLEKLREA YRLLSNDRFR VQYAAHHYAS PDASLHLLVD
GGQVAANFNP EHQSFNFVDH AISRAAMSPS SRSSSDKQRS FSDFTGQYNS VIGNTGCSTD
ARPYNAPEAR AAINGASINF MLRISFDESV LGCTKTAVYE KNVSCQRCSG NGRMVLKRPK
KCPQCRGRGS THLPSATYHI ERSCTYCNGD GVTPPPKCSG CRGAGVVPGH TVQVPVDIRP
GTTNMTACRL RGMGHDGVRG GVAGDLIVTV LVQEHRVFHR DGLDLHMVLP ITLSTALLGG
MVSVPLLHGP FCTRVPPCVR NGQQIRLSGR GVTLDGSGVL TNAEEGIDAD SSASKQEQQQ
RGDLYIHLLV VIPKGEELTG AQRSALEQFV VEQDGNGAEG VDDITPTALK RRFRHWLPGT

Trypanosoma vivax Type IV/I Hsp40s

>TvivJ47 (tviv796e07.p1k_2) |||chaperone protein DNAJ, putative|Trypanosoma vivax|chr 1|||Manual
MRYCHAVEYA HTLIQQVQKR QFNVIQRGND RQLDSLFTLL GFGKDSEARR AQRSRSELRQ
GFIHEAMKLK DPKNSVEDAA KMAKLCEAYK LLSDDRFRSQ YASHHYASPD AMLHVRVDGG
QVAANFNPEH QSFEFVNHAL SFASTAASQR VSSDAQRSFG DFTGPYQSAV GATSEFTEVR
QHNPREAKTV LNGSNVNFTL RISFDESVLG CTKTVVYEGN VLCTHCGGHG RTKLARHRKC
PQCRGRGSIN LPSATYHIER QCTYCSGDGK APAPKCSVCG GVGVITGHTV SVPVDVRPGT
TSMTVRRIRG KGHDGARGGA SGDLVVTVLV QEHRLFHRDG MNLHLVMPIP LSVALLGGVV
NVPLLDGQYP MRVPPCVRSG QHIVLVGKGV CPDNSACTFK VSSAEGDAEN YKEQQQVTAE
KRGDLYVHLL VVIPRGDELT GAQRCALEAF MVKRDCDHEA AQQEITPVEL KKQFRHWLTS
KC

Trypanosoma congolense Type IV/I Hsp40s

Protein doesn't end with a stop codon Send to GeneDB omniBLAST Send to GeneDB BLAST Send to BLAST at NCBI

>TconJ47 (congo520e01.p1k_13) ||||Trypanosoma congolense|chr 1|||Manual
MRCFAVNSVS PFVPLLPNAQ RRFFNVIQRG NDREVKSLFA LLGFDKDSES HRVHRTRAEL
RQGFMREAMK LTDPQHNKED ASKAAKLKKA YELLSNDRFR AQYAAFHYAS PDASLHLLVD
GGQVAANFNP EHQSFRFVDH ALSGAAVSSS QRSSLKEQRS FSDFTSQYRA VVGAVGTSAD
AKPYDAREAR SPVSGASINF MLRISFDESV MGCTKTAVYE KDVTCQHCSG SGKLPVKMPR
RCPQCRGRGS THLPSATYHI ERTCTYCGGE GKLPPPRCTF AEGRAVRVSV

>TconJ47 (congo1350g02.q1k_11) |||chaperone protein DNAJ, putative|Trypanosoma congolense|chr 1|||Manual
MRCVAVNSVS PSVPLLPNVQ RRFFNVIQRG NDREVKSLFA LLGFDKDSES HRVHRTRAEL
RQGFMREAMK LADPQHNKED ASKAAKLKKA YELLSNDRFR AQYAAFHYAS PDASLHLLVD
GGQVAANFNP EHQSFRFVDH ALSGARVSSS QRSSLKEQRS FSDFTNQYRA VVGAVGTSAD
AKPYDAREAR SPVNGASINF MLRISFDESV MGCTKTAVTE KDVTCQHCSG SGRLPVKHPR
RCPQCRGRGS THLPSATYHI ERTCTVCGGE GKLPPPRCTF CRGAGVVAGQ TVHVPVDVRP
GTTSMTVYRL RGKGNDGTRG GAAGDLIVTV LVQEHRIFHR DGLDLHMVLP ISISTALLGG
VVPVPLPHAA CSVRVPPCAR NGQQIRLAGK GVGLDGSRPA VVGDDGGGTG NSPHGSTRSG
CGDLYVHLLV VIPKGEELTG RQRCVIEKFV ECAESNESDT SDEATPASLK ERFRHWFES

No HPD – probably a Type IV

Leishmania infantum Type IV/I Hsp40s

>LinJ47 (LinJ20_V3.0620) |||chaperone protein DNAJ-like protein|Leishmania infantum|chr 20|||Auto
MLAPRSVRWC GGNRRSSGSA SVGVGPHVRT TGPWMPWHTL LYQRRCFQVI QRGHDRPQES
LYNKLGFAKD TEARLVRRST TELRQALLRR VEELKDPAND PADKKRLSEL KDAYTRLQDD
CFRTNYASHY YASNDARLHV LCDGGQVAAN FNPEHQQFTY VDHAIDREST GSGVGRCLPT
GKAGGSALSS GPPGRSEVPT AEALGDAFRN ATGVGESASG ATSSNATAYK AADAAGALNG
TDITHLLRVT LEQSLAGCEV EVAIHKNVRC VTCRGSGRQQ LRTPRKCPQC LGRGSAHMFS
ATYHIERRCL YCGGEGTVPA PPCRACDGRG VQLHQAVQVP VSVPAGTLSN SLFRLRHQGH
DGVRGGHAGD LLLTVLVSEH RYFYRSPERS HELHAMLPLP LSVALLGGRV NVPTLNGFGT
VHVPPCVRSG EVLPLDVYGA PDATNRASAA AAAPHTILYH ALVMIPKGDA LSGRQKAALQ
LYEAHHSYPS ATAAEVEAVG RSRTFGAEAS QQQAPASSAG TVSREQLMAS CAALKSSYTH
WFHAD

Leishmania braziliensis Type IV/I Hsp40s

None found

Other Hsp40s containing Zinc Finger motif, but no N-terminal J-domain:

Trypanosoma brucei brucei

>Tbj66 (Tb927.7.2070) |||heat shock protein DnaJ, putative|Trypanosoma brucei|chr 7|||Manual
MFAFSDHMDD VFNAFFSGGD MFSGGDMFSG GGGRGRRRQP KDTVHGLPVT LKDLYNGRSI
EIPHTRTTPC VGCDGRGAKS RKDVTCTACR GAGRRMLARQ MGMMIQQVTV PCDACGGEGR
PMDPRDICPV CDGRRVNQTE SSLTVVVEPG MEHREQIVFH GEGSYQPAAD AAGDIVIVLE
QMKDDRFERE GDDLLYTHTI SLAESLCGFQ LVLTHLDGRQ LVVRRERGEI TRPGERKVVL
GEGMPIRGRK GKFGDLVIKF AVSFPERIEE AQVEILRQAL PAPRSVDLSH CDMAQECYVS
RKELDHLRQE LENDVEEQET TSVGCTAQ

Trypanosoma cruzi

Protein doesn't start with a start codon Send to GeneDB omniBLAST Send to GeneDB BLAST Send to BLAST at NCBI

>Tcj66 (Tc00.1047053510165.10) |||heat shock protein DnaJ, putative|Trypanosoma cruzi|chr unknown83|TIGR||Auto
RNQKASYALP VTLSDLYNGK TFELPHSRAV ACPNCEGRGT NSRKNNVCRS CGGNGSRLIV
PQMGMSMCQM SAPCDACGGS GLFVDPKDVC TACHGQRTTE VESFLTVPVE RGMRHHDEVV
FRGEGSCDPY TGEPGDIVIV LEQMKDERFV REEDDLHMNH TITLAESLCG FQFVFKHLDG
RELIVRRERG EITQPGEVKV VLGEGMPRRQ RPGQHGDLVI KFNVTFPNRL ESSQVDALRK
ALPPPKSVDL HQCDDAEVCY VTREELDHLR RELEEEAKEE DEGPSVGCAA G

>Tcj66 (Tc00.1047053511807.70) |||heat shock protein DnaJ, putative|Trypanosoma cruzi|chr unknown87|TIGR||Auto
MFGFSDEVGS MINAFFGGMP DGLHHVGGRR RNQKASYALP VTLSDLYNGK TPELPHSRAV
ACPNCEGRGT NSRKNNVCRS CRGNGSRLIV RQMGMMMQQM SAPCDACGGS GLKVDPKDVC
TACRGQKTTR VESFLTVPVE RGMRHHDEVV FRGEGSCDPY TGEPGDIVIV LEQVKDERFV
REEDDLHMNH TITLAESLCG FQFVFKHLDG RELIVRRERG EITQPGEVKV VLGEGMPRRQ
RPGQHGDLVI KFNVTFPNRL ESSQVDALRK ALPPPKSVDL HQCDDAEVCY VTREELDHLR
RELEEEAKEE DEGPSVGVAA G

Leishmania major

>LmJ66 (LmjF22.0080) |||heat shock protein DnaJ, putative|Leishmania major|chr 22|||Manual
MFGGGMDDML NAMLNGGMAS FGGGRGGRVQ RSRRGRDAAY ALPVTLEDLY NGKMVQVERK
RTVMCPDCKG TGSKRRNLPR CGMMCPVCRG SGSRVMVRQM GMMVQQMQVV CDACQGSGEH
IDPRNRCGRC SGNKTVEVDA SVQVVVEKGM AHRQRITFPR MANEELGVER AGDFVVVIQQ
VKHNVFTRDD CDLHMQHHLS LAEALCGFQF KFTHLDGREL VVRQARGTIT KPGDVKCVIG
EGMPVHKQAN KFGNLIIEFN VKYPDRIEAE QLQLLREALP PPKSVDVAAD NEASDVCYVT
REDLSVLEEE IKKDEEAEEE NEGPQTGCAT Q

Trypanosoma brucei gambiense

>TbgambJ3 (Tbgamb.26876) |||chaperone protein DNAJ, putative|Trypanosoma brucei gambiense|chr 10|||Manual
MYNGRVERVA VTRDRLCSQC DGSGVRPGAQ QQMCEACNGQ GIQVLVQHII PGVRQQVQLT
CQNCGGCGKY VRESDVCQRC HGKQMVRDEK VLEVPIERGM KADDAIRFEG EGDEVLGVRL
KGDVLIILAE KPHDVFRRVG DHLIMNYRIT LQEALCGFEL PVQHLDKRML LIKIPAGQVI
DPEAGWVVHR EGMPLPNTSG IERGNLIIHF EVEYPTKLSS RQIDLIADAF HVSEGFPHVG
GQKVVLRDET ARRQRRNTAS ARQAQQRRSR DTRGFDNPDV FGMGFGGGQT AHQEEG

>TbgambJ66 (Tbgamb.13956) |||heat shock protein DnaJ, putative|Trypanosoma brucei gambiense|chr 7|||Manual
MFAFSDHMDD VFNAFFSGGD MFSGGGGRGR RRQPKDAVHG LPVTLKDLYN GRSIEIPHTR
TTPCVGCDGR GAKSKKNVTC TACRGAGRRM LAQQMGMMIQ QVTVPCDACG GEGRRMDPRD
ICPVCDGRRV NQVESSLTVV VKPGMEHREQ IVFHGEGSYQ PAADAAGDIV IVLEQMKDDR
FEREGDDLLY THTISLAESL CGFQLVLTHL DGRQLVVRRE RGEITRPGER KVVLGEGMPI
RGRKGKFGDL VIKFAVSFPE RIEEAQVEIL RQALPAPRSV NLSHCDMAQE CYVSRKELDH
LRQELENDVE EQETTSVGEE AG

>TbgambJ2 (Tbgamb.2166) |||chaperone protein DnaJ, putative|Trypanosoma brucei gambiense|chr 2|||Manual
MWRKRRRYDQ FGEKGVESEG VGIDPSDIFS SFFGGRRARG EAKPKDIVHQ QPVPLETFYN
GKTIKLAIIR DRLCDSCNGS GSKDPKVSSR CVECDGRGVK IITRSIGPGF VQQMQVACPR
CGGKGTIIKE EHKCQSCRGQ QIVKDKKVFD VVVEKGMQHG DSVTFQGEGD QIPGVRLSGD
IIIILDEKPH PVFTRKGDHL LIHHKISLAE ALTGFTMNIK HLDERAISIR STNVIDPQKL
WSVSREGMPI PGTGGTERGD LVIKFDVVYP SAQSLSGDGI EPLRRILGYP KQEEPAPEAT
EHTLAVTYVD LDREARRRRT AANDDDDDAG QHVHTGATTT GG

Trypanosoma vivax

>TvivJ66 (tviv777g10.p1k_16) |||heat shock protein DnaJ, putative|Trypanosoma vivax|chr 7|||Manual

MFGFPDDMVN MLFEGMGGFT DSMLGRHARR PRATTHALPV TLRDLYVGRT IQIPRTQNIP
CPGCDGRGVR SRRNTVCSAC RGSGVRKTVR QMGLMMQETR VTCNCCDGHG SIIDPRDMCH
VCNGQKTITS ESPLQVEVEP GMENEEKIFF PGEEGGDSDD VVIVLKQVKD EMFERRGADL
HYIHTLTLAE ALCGFQFVLE HLDHRQLVVR RERGELTKHV DIKIVAGEGM PVHRRPGVFG
DLIIEFRVAF PSTIEPPLVE VLRRTLPGPK SVDTCKYENA EECYVTRVEM DSLRSMLAAE
AKESEREENP GFTGAAG

Leishmania infantum

>LinJ66 (LinJ22_V3.0009) |||heat shock protein DnaJ, putative|Leishmania infantum|chr 22|||Auto

MFGGGMDDML NAMLNGGMGG FGGGRGGRVQ RSRRGRDAAY ALPVTLEDLY NGKMVQIERK
RTVMCPDCKG TGSKPRSLPR GGNMCPVCRG SGSRVMVRQM GMNVQQMQVV CDACQGSGEH
IDPRNRCGRC SGNKTVEVDA SVQVVVEKGM AHRQRITFPR MADEELGVER AGDFVVVLQQ
VKHDVFTRED CDLHMQRHLS LAEALCGFQF KFTHLDGREL VVRQARGTIT KPGDVKCVIG
EGMPLHKQAN RFGNLIIEFN VKYPDRIEAG QLQLLREALP PPKSVDVAAD NEAGDVCYVT
REDLSVLEEE IKKDEEAEEE NEGPQTGQAAG

Leishmania braziliensis

>LbrJ66 (LbrM22_V2.0080) |||heat shock protein DnaJ, putative|Leishmania braziliensis|chr 22|||Manual

MCPVCRGSGS RVLVRQMGMM VQQMQMVCDA CQGSGEHIDP RNRCSRCSGN KTVEVDAAVQ
VIVERGMAHR QRITFPRMAD EEVGVERTGD FVVALQQVKH DIFTRDDCDL HMRHHLSLAE
ALCGFQFKFT HLDGRELVVR QARGTITKPG DVKCVIGEGM PVHRQPSKFG NLVIEFEVTY
PDRIESAQLQ LLREALPPPK SVRATADEET GEVCYVTRED LSILEEEIKK DEEAEEDNES
PQHGGAAG

Appendix 2B: *Homo sapiens* Type I Hsp40s according to Qiu et al., 2006. Sequences obtained from NCBI (http://www.ncbi.nlm.nih.gov/; Sayers *et al.*, 2008).

NCBI full content search of Type I Human Hsp40s

>DnaJA1 (NP_001530)

MVKETTYYDV LGVKPNATQE ELKKAYRKLA LKYHPDKNPN EGEKFKQISQ AYEVLSDAKK
RELYDKGGEQ AIKEGGAGGG FGSPMDIFDM FFGGGGRMQR ERRGKNVVHQ LSVTLEDLYN
GATRKLALQK NVICDKCEGR GGKKGAVECC PNCRGTGMQI RIHQIGPGMV QQIQSVCMEC
QGHGERISPK DRCKSCNGRK IVREKKILEV HIDKGMKDGQ KITFHGEGDQ EPGLEPGDII
IVLDQKDHAV FTRRGEDLFM CMDIQLVEAL CGFQKPISTL DNRTIVITSH PGQIVKHGDI
KCVLNEGMPI YRRPYEKGRL IIEFKVNFPE NGFLSPDKLS LLEKLLPERK EVEETDEMDQ
VELVDFDPNQ ERRRHYNGEA YEDDEHHPRG GVQCQTS

>DnaJA2 (NP_005871)

MANVADTKLY DILGVPPGAS ENELKKAYRK LAKEYHPDKN PNAGDKFKEI SFAYEVLSNP
EKRELYDRYG EQGLREGSGG GGGMDDIFSH IFGGGLFGFM GNQSRSRNGR RRGEDMMHPL
KVSLEDLYNG KTTKLQLSKN VLCSACSGQG GKSGAVQKCS ACRGRGVRIM IRQLAPGMVQ
QMQSVCSDCN GEGEVINEKD RCKKCEGKKV IKEVKILEVH VDKGMKHGQR ITFTGEADQA
PGVEPGDIVL LLQEKEHEVF QRDGNDLHMT YKIGLVEALC GFQFTFKHLD GRQIVVKYPP
GKVIEPGCVR VVRGEGMPQY RNPFEKGDLY IKFDVQFPEN NWINPDKLSE LEDLLPSRPE
VPNIIGETEE VELQEFDSTR GSGGGQRREA YNDSSDEESS SHHGPGVQCA III

>DnaJA2b (AAB69313)

MANVADTKLY DILGVPAGAS ENELKKAYRK LAKEYHPDKN PQMQETNFKE ISFAYEVLSN
PEKRELYDRY GEQGLREGSG GGGWHGLIFS LTVFCGGLFG FMGNQSRSRN GRRRGEDMMH
PLKVSLEDLY NGKTTKLQLS KNVLCSACSG QGGKSGAVQK CSACRGRGVR IMIRQLAPGM
VQQMQSVCSD CNGEGEVINP KDRCEKCEGK KVIKEVKILE VHVDKGMKHG QRITFTGEAD
QAPEWNPETL FFLLPGEKNM EVFQRDGNDL HMTYKIGLVE ALCGFQFTLS HLDGRQIVVK
YPPGKVIEPG CVRVVRGEGM PQYRNPFEKG GLYIKFDVQF PENNWINPDK LSELEDLLPS
RPEVPNIIGE TEEVELQEFD STRGSGGGQR REAYNDSSDE ESSSHHGPGV QCAAQ

>DnaJA3 (NP_005138)

MAARCSTRWL LVVVGTPRLP AISGRGARPP REGVVGAWLS RKLSVPAFAS SLTSCGPRAL
LTLRPGVSLT GTKHYPFICT ASFHTSAPLA KEDYYQILGV PRNASQKEIK KAYYQLAKKY
HPDTNKDDPK AKEKFSQLAE AYEVLSDEVK RKQYDAYGSA GFDPGASGSQ HSYWKGGPTV
DPEELFRKIF GEFSSSSFGD FQTVFDQPQE YFMELTFNQA AKGVNKEFTV NIMDTCERCN
GKGNFPGTKV QHCHYCGGSG METINTGFFV MRSTCPRCGG RGSIIISPCV VCRGAGQAKQ
KKRVMIPVPA GVEDGQTVRM PVGKREIFIT FRVQKSPVFR RDGADIHSDL FISIAQALLG
GTARAQGLYE TINVTIPPGT QTDQKIRMGG KGIPRINSYG YGDHYIHIKI RVPKRLTSRQ
QSLILSYAED ETDVEGTVNG VTLTSSGGST MDSSAGSKAR REAGEDEEGF LSKLKKMFTS

>DnaJA4 (NP_061072)

MVKETQYYDI LGVKPSASPE EIKKAYRKLA LKCHPDKNPD EGEKFKLISQ AYEVLSDPKK
RDVYDQGGEQ AIKEGGSGSP SFSSPMDIFD MFFGGGGRMA RERRGKNVVH QLSVTLEDLY
NGVTKKLALQ KNVICEKCEG VGGKKGSVEK CPLCKGRGMQ IHIQQIGPGM VQQIQTVCIE
CKGQGERINP KDRCESCSGA KVIREKKIIE VHVEKGMKDG QKILFHGEGD QEPELEPGDV
IIVLDQKDHS VFQRRGHDLI MKMKIQLSEA LCGFKKTIKT LDNRILVITS KAGEVIKHGD
LRCVRDEGMP IYKAPLEKGI LIIQFLVIFP EKHWLSLEKL PQLEALLPPR QKVRITDDMD
QVELKEFCPN EQNWRQHREA YEEDEDGPQA GVQCQTA

>DnaJA4 (EAW99177) (Extra ~30 aa at n terminus compared to **>NP_061072 DnaJA4)**

MARGGSQSWS SGESDGQPKE QTPEKPRHKM VKETQYYDIL GVKPSASPEE IKKAYRKLAL
KYHPDKNPDE GEKFKLISQA YEVLSDPKKR DVYDQGGEQA IKEGGSGSPS FSSPMDIFDM
FFGGGGRMAR ERRGKNVVHQ LSVTLEDLYN GVTKKLALQK NVICEKCEGV GGKKGSVEKC
PLCKGRGMQI HIQQIGPGMV QQIQTVCIEC KGQGERINPK DRCESCSGAK VIREKKIIEV
HVEKGMKDGQ KILFHGEGDQ EPELEPGDVI IVLDQKDHSV FQRRGHDLIM KMKIQLSEAL
CGFKKTIKTL DNRILVITSK AGEVIKHGDL RCVRDEGMPI YKAPLEKGIL IIQFLVIFPE
KHWLSLEKLP QLEALLPPRQ KVRITDDMDQ VELKEFCPNE QNWRQHREAY EEDEDGPQAG
VQCQTA

>DnaJA5 (NP_919259) (NOT A TYPE I according to my definition)

MKCHYEALGV RRDASEEELK KAYRKLALKW HPDKNLDNAA EAAEQFKLIQ AAYDVLSDPQ
ERAWYDNHRE ALLKGGFDGE YQDDSLDLLR YFTVTCYSGY GDDEKGFYTV YRNVFEMIAK
EELESVLEEE VDDFPTFGDS QSDYDTVVHP FYAYWQSFCT QKNFAWKEEY DTRQASNRWE
KRAMEKENKK IRDKARKEKN ELVRQLVAFI RKRDKRVQAH RKLVEEQNAE KARKAEEMRR
QQKLKQAKLV EQYREQSWMT MANLEKELQE MEARYEKEFG DGSDENEMEE HELKDEEDGK
DSDEAEDAEL YDCLYCPACD KSFKTEKAMK NHEKSKKHRE MVALLKQQLE EEEENFSRPQ
IDENPLDDNS EEEMEDAPKQ KLSKKQKKKK QKPAQDVPGK DSYLPAAHFQ MAWGKKCVLG
ERRDGESEHK CAKMLLENRQ NYDDNFNVNG PGEGVKVDPE DTNLNQDSAK ELEDSPQENV
SVTEIIKPCD DPKSEAKSVP KPKGKKTKDM KKPVRVPAEP QTMSVLISCT TCHSEFPSRN
KLFDHLKATG HARAPSSSSL NSATSSQSKK EKRKNR

Appendix 2C: *Saccharomyces cerevisiae* Type I Hsp40s according to Walsh et al., 2004. Sequences were obtained from the Saccharomyces Genome Database (www.yeastgenome.org/; Cherry *et al.*, 1998).

>Xdj1 (YeastCytoplasm)
MSGSDRGDRLYDVLGVTRDATVQEIKTAYRKLALKHHPDKYVDQDSKEVNEIKFKEITAA
YEILSDPEKKSHYDLYGDDNGAASSGGANGFGDEDFMNFFNNFFNNGSHDGNNFPGEYDA
YEEGNSTSSKDIDIDISLTLKDLYMGKKLKFDLKRQVICIKCHGSGWKPKRKIHVTHDVE
CESCAGKGSKERLKRFGPGLVASQWVVCEKCNGKGKYTKRPKNPKNFCPDCAGLGLLSKK
EIITVNVAPGHHFNDVITVKGMADEEIDKTTCGDLKFHLTEKQENLEQKQIFLKNFDDGA
GEDLYTSITISLSEALTGFEKFLTKTFDDRLLTLSVKPGRVVRPGDTIKIANEGWPILDN
PHGRCGDLYVFVHIEFPPDNWFNEKSELLAIKTNLPSSSSCASHATVNTEDDSNLTNNET
ISNFRIIHTDDLPEGIRPFKPEAQDSAHQKARSSYCEQ

>ApJ1 (YeastCytoplasm)
MQQNTSLYDSLNVTAAASTSEIKKAYRNAALKYHPDKNNHTEESKRKFQEICQAYEILKD
NRLRALYDQYGTTDEVLIQEQQAQAQRQQAGPFSSSSNFDTEAMSFPDLSPGDLFAQFFN
SSATPSSNGSKSSFNFSFNNSSTPSFSFVNGSGVNNLYSSSAKYNSNDEDHHLDRGPDIK
HNLKCTLKELYMGKTAKLSLNRTRICSVCDGHGGLKKCTCKTCKGQGIQTQTRRMGPLVQ
SWSQTCADCGGAGVFVKNKDICQQCQGLGFIKERKILQVTVQPGSCHNQLIVLTGEGDEV
ISTKGGGHEKVIPGDVVITILRLKDPNFQVINYSNLICKKCKIDFMTSLCGGVVYIEGHP
SGKLIKLDIIPGEILKPGCFKTVEDMGMPKFINGVRSGFGHLYVKFDVTYPERLEPENAK
KIQNILANDKYIKAERSTMETADSDCYCDLEKSYDSVEEHVLSSFEAPNLNNEVIEDDDL
GDLINERDSRKRNNRRFDESNINNNNETKRNKYSSPVSGFYDHDINGY

>MdJ1 (Yeastmitochondria)
MAFQQGVLSRCSGVFRHHVGHSRHINNILYRHAIAFASIAPRIPKSSFHTSAIRNNEAFK
DPYDTLGLKKSATGAEIKKAYYKLAKKYHPDINKEPDAEKKFHDLQNAYEILSDETKRQQ
YDQFGPAAFGGGGAAGGAGGGSGSPFGSQFHDFSGFTSAGGSPFGGINFEDLFGAAFGGG
GRGSGGASRSSSMFRQYRGDPIEIVHKVSFKDAVFGSKNVQLRFSALDPCSTCSGTGMKP
NTHKVSCSTCHGTGTTVHIRGGFQMMSTCPTCNGEGTMKRPQDNCTKCHGEGVQVERAKT
ITVDLPHGLQDGDVVRIPGQGSYPDIAVEADLKDSVKLSRGDILVRIRVDKDPNFSIKNK
YDIWYDKEIPITTAALGGTVTIPTVEGQKIRIKVAPGTQYNQVISIPNMGVPKTSTIRGD
MKVQYKIVVKKPQSLAEKCLWEALADVTNDDMAKKTMQPGTAAGTAINEEILKKQKQEEE
KHAKKDDDNTLKRLENFITNTFRKIKGDKKN

>Scj1 (YeastEndoplasmicreticulum)
MVIRCSTDKTWWIGQKSVHWLAKRSRTMIPKLYIHLILSLLLLPLILAQDYYAILEIDKD
ATEKEIKSAYRQLSKKYHPDKNAGSEEAHQKFIEVGEAYDVLSDPEKKKIYDQFGADAVK
NGGGGGGPGGPGAGGFHDPFDIFERMFQGGHGGPGGGFGQRQRQRGPMIKVQEKLSLKQF
YSGSSIPFTLNLMDECDACHGSGSADGKLAQCPDCQGRGVIIQVLRMGIMTQQIQQMCGR
CGGTGQIIKNECKTCHGKKVTKGNKFFHVDVPPGAPRNYMDTRVGEAEKGPDFDAGDLVI
EFKEKDTENMGYRRRGDNLYRTEVLSAAEALYGGWQRTIEFLDENKPVKLSRPAHVVVSN
GEVEVVKGFGMPKGSKGYGDLYIDYVVVMPKTFKSGQNMLEQSI
HDEL **ENDOPLASMIC RETICULUM RETENTION SIGNAL**

>Ydj1 (YeastCytoplasm)
MVKETKFYDI LGVPVTATDV EIKKAYRKCA LKYHPDKNPS EEAAEKFKEA SAAYEILSDP
EKRDIYDQFG EDGLSGAGGA GGFPGGGFGF GDDIFSQFFG AGGAQRPRGP QRGKDIKHEI
SASLEELYKG RTAKLALNKQ ILCKECEGRG GKKGAVKKCT SCNGQGIKFV TRQMGPMIQR
FQTECDVCHG TGDIIDPKDR CKSCNGKKVE NERKILEVHV EPGMKDGQRI VFKGEADQAP
DVIPGDVVFI VSERPHKSFK RDGDDLVYEA EIDLLTAIAG GEFALEHVSG DWLKVGIVPG
EVIAPGMRKV IEGKGMPIPK YGGYGNLIIK FTIKFPENHF TSEENLKKLE EILPPRIVPA
IPKKATVDEC VLADFDPAKY NRTRASRGGA NYDSDEEEQG GEGVQCAAQ

Appendix 2D: *Plasmodium falciparum* Type I Hsp40s according to Botha *et al.*, 2007. Sequences were obtained from PlasmoDB (http://plasmodb.org/plasmo/; Bahl *et al.*, 2003; Aurrecoechea *et al.*, 2009).

>PfJ1 (PF11_0359)

MFFSSGFPFDSMGGQQARRKREVNNNKFYEVLNLKKNCTTDEVKKAYRKLAIIHHPDKGG
DPEKFKEISRAYEVLSDEEKRKLYDEYGEEGLENGEQPADATDLFDFILNAGKGKKKRGE
DIVSEVKVTLEQLYNGATKKLAISKDIICTNCEGHGGPKDAKVDCKQCNGRGTKTYMRYH
SSVLHQTEVTCNTCRGKGFIFNEKDKCANCKGMCVLKTRKIIEVYIPKGAPNKHKIVFNG
EADEKPNVITGNLVVILNEKQHPVFRREGIDLFMNYKISLYESLTGFVAEVTHLDERKIL
VNCTNSGFIRHGDIREVLDEGMPTYKDPFKKGNLYITFEVEYPMDLIITNENKEVLKILK
KQNEVEKKYDLENSELEVVSCSPVDKEYIKVRVTKQQQQQQEAYDDEDHQPEMEGGRVA
CAQQ

>Hypothetical *P. falciparum* protein (PFD0462w)

MLALRILRRKVCSEHFLFERSFFTQSIKGKNGCLVTRYDKNKLLFYYKRNINTSRKCLNQDPYTVLGLSR
NATTNDIKKQFRLLAKKYHPDINPSPDAKQKMASITAAYELLSDPKKKEFYDKTGMTDDSNYQNHSSNFE
GAFSGFGDASFMFTDFAEMFTNMAGGNKNTSTRGEDIQSEITLKFMEAIKGCEKNIRLNVKVSCNNCNGS
GKKPGTNLTICKVCNGSGIQRMERGPIIIGVPCRNCSGNGQIINNPCKHCSGSGVKFQTKNITLDIPPGI
KKGMQMRIPNQGHCGYRGGKSGHLFVTINIEPHKIFKWVDDNIYVDVPLTIKQCLLGGLVTVPTLNGDMD
LLIKPKTYPNSEKILKGKGPCKVDSHNNGDLIIKFSLKIPEKLTPRQVELIEEFNTIELNLPNPQTNVKQ
KKNIYETKGNINENIFSMNNTYNNMKGPEGETSNTQAKSMKNQNWNNEKSVNNKGTISKDEKKLNMKNNH
INEKSNLKNSSHMDTNKNEENMSDDEKKKIKKIIPEPPMPHTHKIVNNLESKNSCNIPIPPPPPKSSSKP
ISENQNISNREHNGVTNNSAKLDNNINMNYSCDPYKNVTQNDLNNNDNIKNKIYKDNTNISNHHIFKNDN
INQQQFHCADNSSENNNESDMNTTSTFSFAKKWISDKLKPKN

Appendix 2E: *Arabidopsis thaliana* Type I Hsp40 sequences according to Miernyk , 2001. Sequences were obtained from NCBI (http://www.ncbi.nlm.nih.gov/; Sayers *et al.*, 2008).

A Thaliana Type I Hsp40s:

```
>AAB86799 (A2)                    419 aa (Nt acc L36113)
1    MFGRGPSRKS DNTKFYEILG VPKTAAPEDL KKAYKKAAIK NHPDKGGDPE KFKELAQAYE
61   VLSDPEKREI YDQYGEDALK EGMGGGGGGH DPFDIFSSFF GSGGHPFGSH SRGRRQRRGE
121  DVVHPLKVSL EDVYLGTTKK LSLSRKALCS KCNGKGSKSG ASLKCGGCQG SGMKISIRQF
181  GPGMMQQVQH ACNDSKGTGE TINDRDRCPQ CKGEKVVSEK KVLEVNVEKG MQHNQKITFS
241  GQADEAPDTV TGDIVFVIQQ KEHPKFKRKG EDLFVEHTIS LTEALCGFQF VLTHLDKRQL
301  LIKSKPGEVV KPDSYKAISD EGMPIYQSPF MKGKLYIHFT VEFPESLSPD QTKAIEAVLP
361  KPTKAAISDM EIDDCEETTL HDVNIEDEMK RKAQAQREAY DVDEEDHPGG AHRVQCAQQ
```

```
>AAB49030  (A3)                   420 aa (Nt acc U22340)
1    MFGRGPSKKS DNTKFYEILG VPKSASPEDL KKAYKKAAIK NHPDKGGDPE KFKELAQAYE
61   VLSDPEKREI YDQYGEDALK EGMGGGGGGH DPFDIFSSFF GGGPFGGNTS RQRRQRRGED
121  VVHPLKVSLE DVYLGTMKKL SLSRNALCSK CNGKGSKSGA SLKCGGCQGS GMKVSIRQLG
181  PGMIQQMQHA CNECKGTGET INDRDRCPQC KGDKVIPEKK VLEVNVEKGM QHSQKITFEG
241  QADEAPDTVT GDIVFVLQQK EHPQFKRKGE DLFVEHTLSL TEALCGFQFV LTHLDGRSLL
301  IKSNPGEVVK PDSYKAISDE GMPIYQRPFM KGKLYIHFTV EFPDSLSPDQ TKALEAVLPK
361  PSTAQLSDME IDECEETTLH DVNIEDEMRR KAQAQREAYD DDDEDDDHPG GAQRVQCAQQ
```

```
>CAB80659.1 (A24) Acc AL161596
MFAQGSLPFLSLTGVSPNTHSRRGARFTVRADTDFYSVLGVSKNATKAEIKSAYRKLA
RSYHPDVNKDAGAEDKFKEISNAYEILSDDEKRSLYDRYGEAGVKGAGMGGMGDYSNP
FDLFESLFEGMGGMGGMGGGMGSRGSRSRAIDGEDEYYSLILNFKEAVFGIEKEIEIS
RLESCGTCNGSGAKAGTKPTKCKTCGGQGQVVASTRTPLGVFQQVMTCSPCNGTGEIS
KPCGACSGDGRVRRTKRISLKVPAGVDSGSRLRVRGEGNAGKRGGSPGDLFAVIEVIP
DPVLKRDDTNILYTCKISYVDAILGTTLKVPTVDGEVDLKVPAGTQPSTTLVMAKKGV
PVLNKSKMRGDQLVRVQVEIPKRLSKEEKMLVEELADMSKNKVANSRR
```

>AAD22362 (A26) 442 aa (AC006592)

```
1   MAIIQLGSTC VAQWSIRPQF AVRAYYPSRI ESTRHQNSSS QVNCLGASKS SMFSHGSLPF
61  LSMTGMSRNM HPPRRGSRFT VRADADYYSV LGVSKNATKA EIKSAYRKLA RNYHPDVNKD
121 PGAEEKFKEI SNAYEVLSDD EKKSLYDRYG EAGLKGAAGF GNGDFSNPFD LFDSLFEGFG
181 GGMGRGSRSR AVDGQDEYYT LILNFKEAVF GMERRIRTSR LSSCGTCEGS GAKPGTKPTK
241 CTTCGGSGQV VSAARTPLGV FQQVMTCSSC NGTGEISTPC GTCSGDGRVR KTKRISLKVP
301 AGVDSGSRLR VRGEGNAGKR GGSPGDLFVV IEVIPDPILK RDDTNILYTC KISYIDAILG
361 TTLKVPTVDG TVDLKVPAGT QPSTTLVMAK KGVPVLNKSN MRGDQLVRVQ VEIPKRLSKE
421 EKKLIEELAD MSKNKTANST SR
```

>BAB11067.1 (A30) ACC AB017064

```
MVPSNGAKVLRLLSRRCLSSSLIQDLANQKLRGVCIGSYRRLNTSVGNHANVIGDYAS
KSGHDRKWINFGGFNTNFGSTRSFHGTGSSFMSAKDYYSVLGVSKNAQEGEIKKAYYG
LAKKLHPDMNKDDPEAETKFQEVSKAYEILKDKEKRDLYDQVGHEAFEQNASGGFPND
QGFGGGGGGGFNPFDIFGSFNGDIFNMYRQDIGGQDVKVLLDLSFMEAVQGCSKTVTF
QTEMACNTCGGQGVPPGTKRETCKACNGSGMRAYSDLTEECYVCHPNNLPESICKSCR
GARVVPGQKSVKVTIDPGVDNSDTLKVARVGGADPEGDQPGDLYVTLKVREDPVFRRE
GSDIHVDAVLSVTQHLFWTSGAVSAILGGTIQVPTLTGDVVVKVRPGTQPGHKVVLRN
KGIRARKSTKFGDQYVHFNVSIPANITQRQRELLEEFSKAEQGEYEQRTATGSSQ
```

>AAD55483 (A52) 499 aa (AC009322)

```
1   MAALASPSLI PSSLCFAAAA DGPRSLSSNF SAFSDGGSNF RYHKSFLSLS SSSSSSTPYR
61  NRRGRSLVVF ATSGDYYATL GVSKSANNKE IKAAYRRLAR QYHPDVNKEP GATEKFKEIS
121 AAYEVLSDEQ KRALYDQYGE AGVKSTVGGA SGPYTSNPFD LFETFFGASM GGFPGMDQAD
181 FGRTRRSRVT KGEDLRYDIT LELSEAIFGS EKEFDLTHLE TCEACAGTGA KAGSKMRICS
241 TCNGSGRCQK WHLIGASELM SMCSLLKLIC LWLVVSICPN CGGDGEVISE MCRKCSGEGR
301 VRIKKSIKVK IPPGVSAGSI LRVAGEGDSG PRGGPPGDLY VYLDVEDVRG IERDGINLLS
361 TLSISYLDAI LGAVVKVKTV EGDTELQIPP GTQPGDVLVL AKKGVPKLNR PSIRGDHLFT
421 VKVSVPNQIS AGERELLEEL ASLKDTSSNR SRTRAKPQQP STLSTAPSGS ENKKDEVKEE
481 NEEPEQENYL WNNIKEFAG
```

>BAB02706 (A54) 438 aa (AB019230)

```
1   MAAMARCALI PSINPAHSFR HQFPQPNASF YLPPTLPIFS RVRRFGISGG YRRRVITMAA
61  GTDHYSTLNV NRNATLQEIK SSYRKLARKY HPDMNKNPGA EDKFKQISAA YEVLSDEEKR
121 SAYDRFGEAG LEGDFNGSQD TSPGVDPFDL YSAFFGGSDG FFGGMGESGG MGFDFMNKRS
181 LDLDIRYDLR LSFEEAVFGV KREIEVSYLE TCDGCGGTGA KSSNSIKQCS ECDGKGRVMN
241 SQRTPFGLMS QVSTCSKCGG EGKTITDKCR KCIGNGRLAA RKKMDVVVPP GVSDRATMRI
301 QGEGNMDKRS GRAGDLFIVL QVDEKRGIRR EGLNLYSNIN IDFTDAILGA TTKVETVEGS
361 MDLRIPPGTQ PGDTVKLPRK GVPDTDRPSI RGDHCFVVKI SIPKKLSERE RKLVEEFSSL
421 RRSSSSTGPT GTMLSQSN
```

>AAF07843 (A63) 572 aa (AC010871)

```
1   MVRTRLAISV VLVSTLLLLN VKAKSVDPYK VLGVSKDAKQ REIQKAFHKQ SLKYHPDKNK
61  DKGAQEKFAE INNAYEILSD EEKRKNYDLY GDEKGQPGFD SGFPGGNGGY SYSSSGGGFN
121 FGGPGGWQNM GGGGGSKSFS FSFGGPSESS FGFGMDDIFS MFSGGSSKGK EQFGGFGSSS
181 NAESKSKSST VAAIKTINSQ VYKKDVVDQG MTWLLLSYLP SQRGSQYHES IIEEVAESLQ
241 GALKVGRLNC ETESSLCKQL GIVPRRAPRM FVYSYTSSGK ATLAEYTEEL VAKKVKSFCQ
301 EHLPRFSKKI DLNTFDVSAV SSQKTPKVLL LSTKKDTPVI WRVLSGLYNG RFVFYNTEVH
361 DTSDPKIQKL GVDKFPAIVG WLSNGEKQVL KTGITVKNLK SAVQELGKLL EGLEKKNKKV
421 SSKSQAGQAP NESSEKIPLL SRPNFDSICG ENTPVCIIGA FRSSNGKEKL QSIMSKVSQK
481 SLSRRQASTT GSQDTVSYSL LDATKQSAFL SSLDKSEFKT SSDKLLIAYK PRRGKFATFK
541 GDMTIEEVEK FVAAVLNGDI QFTKTRQKPQ IK
```

>CAB86083.1 AL163002 (A36)

```
MDLFRVWSGMDFLAWRGMAYTLLLLNFVFACQLLLLQPLVSALDGQSVDAAELFERASQSIKVKRYSDA
LDDLNAAIEADPALSEAYFKRASVLRHFCRYEDSENSYQKYLEFKSGDSNAEKELSQLHQAKSALETAS
TLYESKDIAKALEFVDKVVLVFSPACSKAKLLKVKLLMVSKDYSGAISETGYILKEDENNLEALLLRGR
AYYYLADHDIAQRHYQKGLRLDPEHSELKKAYFGLKKLLKKTKSAEDNANKGKLRVSAEEYKEAIALDP
EHTANNVHLYLGLCKVSVRLGRGKDGLNSCNEALNIDAELIEALHQRGEAKLLLEDWEGAVEDLKQAAQ
NSQDMEIHESLGKAEKALKMSKRKDWYKILGISRTASISEIKKAYKKLALQWHPDKNVGNREEAENKFR
EIAAAYEILGDDDKRARFDRGEDLEDMGGGGGGGYNPFHGGGGGGQQYTFHFEGGFPGGGGGFGGFGF
```

Appendix 2F: *Escherichia Coli* (K12) DnaJ Sequence

> NP_414556

MAKQDYYEIL GVSKTAEERE IRKAYKRLAM KYHPDRNQGD KEAEAKFKEI KEAYEVLTDS
QKRAAYDQYG HAAFEQGGMG GGGFGGGADF SDIFGDVFGD IFGGGRGRQR AARGADLRYN
MELTLEEAVR GVTKEIRIPT LEECDVCHGS GAKPGTQPQT CPTCHGSGQV QMRQGFFAVQ
QTCPHCQGRG TLIKDPCNKC HGHGRVERSK TLSVKIPAGV DTGDRIRLAG EGEAGEHGAP
AGDLYVQVQV KQHPIFEREG NNLYCEVPIN FAMAALGGEI EVPTLDGRVK LKVPGETQTG
KLFRMRGKGV KSVRGGAQGD LLCRVVVETP VGLNERQKQL LQELQESFGG PTGEHNSPRS
KSFFDGVKKF FDDLTR

Appendix: Chapter 4:
Appendix 4.1: Plasmid map of pQE Tchsp70

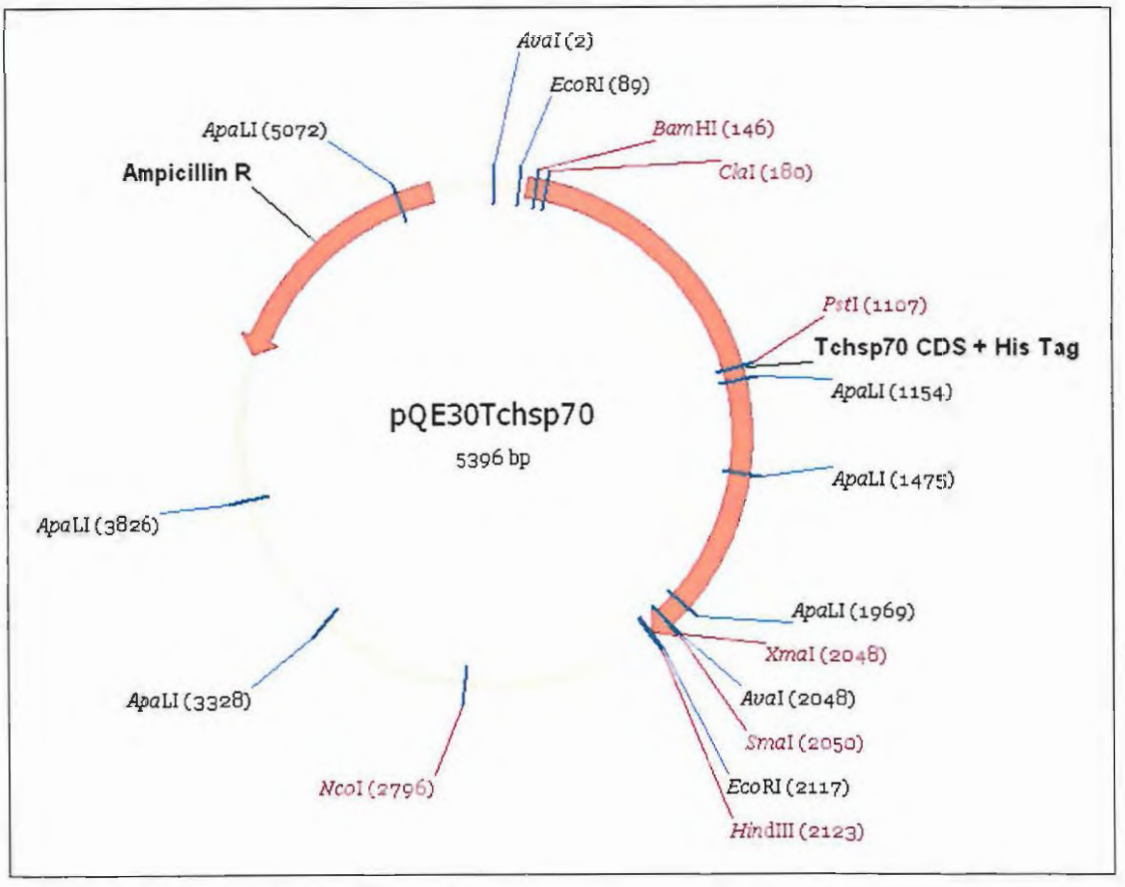


Figure A4.1.: Plasmid map of pQE30Tchsp70. The Tchsp70 coding sequence was ligated into the pQE30 vector using *Bam*HI and *Hind*III. Tchsp70 produced from this expression vector has a Histidine tag at the N-terminus. There are only 3 amino acids between the Histidine tag and the start of the Tchsp70 coding sequence and this Histidine tag containing peptide adds a total of 13 extra amino acids to the N-terminus of Tchsp70. In contrast the pET14bTchsp70 vector used in Edkins *et al.* (2004) had a total of 27 amino acids added to the N-terminus of Tchsp70 and 17 amino acids between the N-terminus and the Histidine tag. The ampicillin resistance gene of pQE30 acts as a selectable marker in the *E. coli* system.

Appendix: Chapter 5:

Appendix 5.1: Modified HMI-9 culture medium for *T. brucei* (Hirumi and Hirumi, 1994)

Per 5 litres:

1X Iscoves Modified Dulbeccos Media

0.05 mM Bathocuproine disulphonic acid

1.50 mM L-cysteine

1.00 mM hypoxanthine‡

1.00 mM sodium pyruvate

0.16 mM thymidine

15.12g NaHCO3

71.5 µl of 2-mercaptoethanol

5 ml Penicillin/Streptomycin (5mg/ml)

Adjust pH to 7.5 and filter sterilize into 500ml aliquots and store at 4°C.

Add 10% Foetal bovine serum (heat inactivated) prior to using the medium.

‡Hypoxanthine – Dissolve 0.4g of Sodium Hydroxide in 100ml of H_2O, then add 1.36g Hypoxanthine (0.1M). (50ml of this concentration per 5 litres of medium).

Appendix 5.2: dsRNA target sequences for Tbj2 RNAi knockdown

Primers for Tbj2 RNAi target PCR:

Intragenic RNAi (2)F: TTTCCAGTTTCTTTGGTGG

Intragenic RNAi (2)R: TGCAACTCTGGCACTTATGC

UTR (d)F: TAATTCCAAGCATGAAAAGG

UTR (d)R: GAAGTAAAAGGTTCGCATTG

Tb927.2.5160 DnaJ putative

Tb927.2.5170 Hypothetical protein

Figure A5.2: Primers shown on the Tbj2 sequence

The Tbj2 (Tb927.2.5160) CDS is highlighted in Blue, and the downstream gene (Tb927.2.5170) is highlighted in purple. The primers used for amplifying the RNAi dsRNA target sequences are highlighted in red. The fragments resulting from amplification are underlined in black.

References

REFERENCE LIST

Abdul, K.M., Terada, K., Gotoh, T., Hafizur, R.M. and Mori, M. (2002) Characterisation and functional analysis of a heart-enriched DnaJ/Hsp40 homolog dj4/Dja4. *Cell Stress and Chaperones.* **7(2):** 156-166.

Agranovsky, A.A, Boyko, V.P., Karasev, A.V., Koonin, E.V. and Dolja, V.V. (1991) Putative 65 kDa protein of beet yellows closterovirus is a homologue of Hsp70 heat shock proteins. *Journal of Molecular Biology.* **217**, 603-619.

Acebron, A.P., Fernandez-Saiz, V., Taneva, S.G., Moro, F. and Muga, A. (2008) DnaJ recruits DnaK to protein aggregates. *The Journal of Biological Chemistry.* **283(3):** 1381-1390.

Alberti, S., Esser, C. and Höhfeld, J., (2003) BAG-1 a nucleotide exchange factor of Hsc70 with multiple cellular functions. *Cell Stress and Chaperones.* **8:** 225-231.

Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K. and Watson, J.D. (1997) Chapter 12: Intracellular Compartments and Protein Sorting pg 551-598: in The molecular Biology of the Cell (3rd edition). Garland Publishing, New York.

Alcina, A., Urzainqui, A. and Carrasco, L. (1988) The heat-shock response in *Trypanosoma cruzi*. *European Journal of Biochemistry.* **172:** 121-127.

Alibu, V.P., Storm, L., Haile, S., Clayton, C. and Horn, D. (2005) A Doubly inducible system for RNA interference and rapid RNAi plasmid construction in *Trypanosoma brucei. Molecular and Biochemical Parasitology.* **139:** 75-82.

Alsford, S. and Horn, D. (2008) Single-locus targeting constructs for reliable regulated RNAi and transgene expression in *Trypanosoma brucei. Molecular and Biochemical Parasitology.* **161:** 76-79.

Alsford, S., Kawahara, T., Glover, L. and Horn, D. (2005) Tagging a *T. brucei* RRNA locus improves stable transfection efficiency and circumvents inducible expression position effects. *Molecular and Biochemical Parasitology.* **144:** 142-148.

Altin, J.G., White, F.A.J. and Easton, C.J. (2001) Synthesis of the chelator lipid nitriloacetic acid ditetradecylamine (NTA-DTDA) and its use with the IAsys biosensor to study receptor-ligand interactions on model membranes. *Biochimica et Biophysica Acta.* **1513:** 131-148.

Andrade, J.D. and Hlady, V. (1986) Protein adsorption and materials biocompatibility: A tutorial review and suggested hypotheses. *Advances in Polymer Science.* **79:** 1-63.

Anfinsen, C.B. (1973) Principles that Govern the Folding of Protein Chains. *Science*, **181** no. 4096: 223-230.

Angelidis, C.E., Nova, C., Lazaridis, I., Kontoyiannis, D., Kollias, G. and Pagoulatos, G.N. (1996) Overexpression of Hsp70 in transgenic mice results in increased cell thermotolerance. *Transgenics* **2**: 111-117.

Anckar, J. and Sistonen, L. (2007) Heat shock factor 1 as a coordinator of stress and developmental pathways. *Advances in Experimental Medicine and Biology.* **594:** 78-88.

Aranda, M.A., Escaler, M., Wang, D., and Maule, A. (1996) Induction of HSP70 and polyubiquitin expression associated with plant virus replication. *Proceedings of the National Academy of Science, USA.* **93**: 15289-15293.

Argaman, M., Aly, R. and Shapira, M. (1994) Expression of heat shock protein 83 in Leishmania is regulated post-transcriptionally. *Molecular and Biochemical Parasitology.* **64:** 95-110.

Arner, E., Kindlund, E., Nilsson, D., Farzana, F., Ferella, M., Tammi, M.T. and Andersson, B. (2007) Database of *Trypanosoma cruzi* repeated genes: 20 000 additional gene variants. *BMC Genomics.* **8:** 391-406.

Atencio, D.P. and Yaffe, M.P. (1992) Mas5, a yeast homologue of DnaJ involved in mitochondrial protein import. *Molecular and Cellular Biology.* **12(1):** 283-291.

Atwood 3rd, J.A., Weatherly, D.B., Minning, T.A., Bundy, B., Cavola, C, Opperdoes, F.R., Orlando, R. and Tarleton, R.L. (2005) The *Trypanosoma cruzi* Proteome. *Science.* **309:**473-476.

Auger, I. and Roudier, J. (1997) A function for the QKRAA amino acid motif: mediating binding of DnaJ to DnaK. *Journal of Clinical Investigation.* **99(8):** 1818-1822.

Auluck, P.K., Chan, H.Y., Trojauowski, 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**: 809-810.

Aurrecoechea C, Brestelli J, Brunk BP, Dommer J, Fischer S, Gajria B, Gao X, Gingle A, Grant G, Harb OS, Heiges M, Innamorato F, Iodice J, Kissinger JC, Kraemer E, Li W, Miller JA, Nayak V, Pennington C, Pinney DF, Roos DS, Ross C, Stoeckert CJ Jr, Treatman C. and Wang H. (2009) PlasmoDB: a functional genomic database for malaria parasites. *Nucleic Acids Research.* **37:** D539-D543.

Bahl, A., Brunk, B., Crabtree, J., Fraunholz, M.J., Gajria, B., Grant, G.R., Ginsburg, H., Gupta, D., Kissinger, J.C., Labo, P., Li, L., Mailman, M.D., Milgram, A.J., Pearson, D.S., Roos, D.S., Schug, J., Stoeckert, C.J. and Whetzel, P. (2003) PlasmoDB: the *Plasmodium* genome resource. A database integrating experimental and computational data. *Nucleic Acids Research.* **31(1):** 212-215.

Baltz, T., Baltz, D., Giroud, C.H. and J. Crockett (1985) Cultivation in a semi-defined medium of animal infective forms of *Trypanosoma brucei, T. equiperdum, T. evansii, T. rhodesiense* and *T. gambiense. European Moelcular Biology Organisation (EMBO) Journal.* **4:** 1273-1277.

Bandey, H.L., Hillman, A. R., Brown, M.J. and Martin, S.J. (1997) Viscoelastic characterization of electroactive polymer films at the electrode/solution interface. Faraday discussions. **107:** 105-121.

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

Bangs, J.D., Brouch, E.M., Ransom, D.M. and Roggy, J.L. (1996) A soluble secretory reporter system in *Trypanosoma brucei* – Studies on endoplasmic reticulum targeting. *The Journal of Biological Chemistry.* **271(31):** 18387-18393.

Bangs, J.D., Uyetake, L. Brickman, M.J., Balber, A.E. and Boothroyd, J.C. (1993) Molecular cloning and cellular localization of a BiP homologue in *Trypanosoma brucei. Journal of Cell Science.* **105:** 1101-1113.

Bannai, H., Tamada, Y., Maruyama, O., Nakai, K. and Miyano, S. (2002) Extensive feature detection of N-terminal protein sorting signals. *Bioinformatics.* **18:** 298-305.

Barrett, M.P., Burchmore, R.J.S., Stich, A., Lazzari, J.O., Frasch, A.C., Cazzulo, J.J. and Krishna, S (2003) The Trypanosomiases. *The Lancet.* **362:** 1469-1480.

Barrett, M.P. (2003) Drug resistance in sleeping sickness. *WHO expert committee on African Trypanosomiasis (sleeping sickness).* **96 - 111**

Barrett, M.P., Boykin, D.W., Brun, R. and Tidwell, R.R. (2007) Human African Trypanosomiasis: pharmacological re-engagement with a neglected disease. *British Journal of Pharmacology.* **152:** 1155-1171.

Barry, J.D., Marcello, L., Morrison, L.J., Read, A.F., Lythgoe, K., Jones, N., Carrington, M., Blandin, G., Bohme, U., Caler, E., Hertz-Fowler, C., Renauld, H., El-Sayed, N. and Berriman, M. (2005) What the genome sequence is revealing about trypanosome antigentic variation. *Biochemical Society Transactions.* **33:** 986-989.

Bateman, A., Coin, L., Durbin, R., Finn, R.D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E.L., Studholme, D.J. and Yeats, C., Eddy, S.R.(2004) The Pfam protein families database. *Nucleic Acids Research.* **32 (database Issue):** D138-141.

Becker, J., Walter, W., Yan, W. and Craig, E.A. (1996) Functional interaction of cytosolic Hsp70 and a DnaJ-related protein, Ydj1p, in protein translocation *in vivo. Molecular and Cellular Biology.* **16(8):**4378-86.

Benharoudj, N., Triniolles, F. and Ladjimi, M.M. (1996) Effect of nucleotides, peptides and unfolded proteins on the self-association of the molecular chaperone Hsc70. *The Journal of Biological Chemistry.* **271(31):** 18471-18476.

Bennett-Lovsey, R.M., Herbert, A.D., Sternberg, M.J. and Kelley, L.A. (2008) Exploring the extremes of sequence/structure space with ensemble fold recognition in the program Phyre. *Proteins.* **70(3):** 611-625.

Ben-Zvi, A.P. and Goloubinoff, P. (2001) Review: Mechanisms of Disaggregation and Refolding of Stable Protein Aggregates by Molecular Chaperones. *Journal of Structural Biology* **135**: 84-93.

Berdoy, M., Webster, J.P. and Macdonald, D.W. (2000) Fatal Attraction in rats infected with Toxoplasma gondii. *Proceedings of the Royal Society, Biological Sciences.* **267(1452):** 1591-1594.

Berjanskii, M.V., Riley, M.I., Xie, A., Semenchenko, V., Folk, W.R. and Van Doren, S.R. (2000) NMR structure of the N-terminal J domain of murine polyomavirus T antigens. Implications for DnaJ-like domains and for mutations of T antigens. *The Journal of Biological Chemistry.* **275:** 36094-36103.

Berriman, M., Ghedin, E., Hertz-Fowler, C., Blandin, G., Renauld, H., Bartholomeu, D.C., Lennard, N.J., Caler, E., Hamlin, N.E., Haas, B., *et al.*, (2005). The genome of the African Trypanosome *Trypanosoma brucei. Science.* **309:** 416-435.

Bhattacharyya, TKarnezis, A.N., Murphy, S.P., Hoang, T., Freeman, B.C., Phillips, B. and Morimoto, R.I. (1995) Cloning and subcellular localization of human mitochondrial Hsp70. *The Journal of Biological Chemistry.* 1705-1710.

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

Blattner, F.R., Plunkett, G. (3rd), Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J., Mau, B. and Shao, Y. (1997) The complete genome sequence of *Escherichia coli* K-12. *Science.* **277(5331):** 1453- 1474.

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

Blond-Elguindi, S., Fourie, A.M., Sambrook, J.F. and Gething, M.H. (1993) Peptide dependent stimulation of the ATPase activity of molecular chaperone BiP is the result of conversion of oligomers to active monomers. *The Journal of Biological Chemistry.* **268:** 12730-12735.

Blumberg, H. and Silver, P.A. (1991) A homologue of the bacterial heat-shock gene DnaJ that alters protein sorting in yeast. *Nature.* **349:** 637-630.

Borchiellini, C., Boury-Esnault, N., Vacelet, J. and Le Parco, Y. (1998) Phylogenetic analysis of the Hsp70 sequences reveals the monophyly of metazoan and specific phylogenetic relationships between animals and fungi. *Molecular Biology and Evolution.* **15**: 647:655.

Borges, J.C., Fischer, H., Craievich, A.F. and Ramos, C.H.I. (2005) Low resolution structural study of two human Hsp40 chaperones in solution. *The Journal of Biological Chemistry.* **280(14):**13671-13681.

Borges, J.C. and Ramos, C.H.I. (2005) Protein Folding Assisted by Chaperones. *Protein and Peptide Letters* **12**:257-261.

Botha, M., Pesce, E.R. and Blatch, G.L. (2007) The Hsp40 proteins of Plasmodium falciparum and other apicomplexa regulating chaperone power in the parasite and the host. *International Journal of Biochemistry and Cell Biology.* 39(10): 1781-1803.

Bradford, M.M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry.* **271**: 25989-25993.

Braun, R. and Evans, T.E. (1969) Replication of nuclear satellite and mitochondrial DNA in the mitotic cycle of Physarum. *Biochimca et Biophysica Acta.* **182:** 511-522.

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

Bridges, D.J., Pitt, A.R., Hanrahan, O., Brennan, K., Voorheis, H.P., Herzyk, P., de Koning, H.P. and Burchmore, R.J.S. (2008) Characterisation of the plasma membrane subproteome of bloodstream form *Trypanosoma brucei*. *Proteomics.* **8:** 83-99.

Broadhead R., Dawe, H.R., Farr, H., Griffiths, S., Hart, S.R., Portman, N., Shaw, M.K., Ginger, M.L., Gaskell, S.J., McKean, P.G. and Gull, K. (2006) Flagellar motility is required for the viability of the bloodstream trypanosome. *Nature.* **440(7081):**224-227.

Buchner, J., Grallert, H. and Jackob, U. (1998) Analysis of chaperone function using citrate synthase as nonnative substrate protein. *Methods in Enzymology.* **290:** 323-338.

Buczynski, G., Slepenkov, S.V., Sehorn, M.G. and Witt, S.N. (2001) Characterisation of a lidless form of the molecular chaperone DnaK: deletion of the lid increases peptide on- and off-rate constants. *The Journal of Biological Chemistry.* **276:** 27231-27236.

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

Cajo, G.C., Horne, B.E., Kelley, W.L., Schwager, F.,, Georgopoulos, C. and Genevaux, P. (2006) The role of the DIF motif of the DnaJ (Hsp40) co-chaperone in the regulation of the DnaK (Hsp70) chaperone cycle. *The Journal of Biological Chemistry.* **281(18):** 12436-12444.

Caplan, A.J. (2003) What is a co-chaperone. *Cell Stress and Chaperones.* **8(2):** 105-107.

Caplan, A.J. and Douglas, M.G. (1991) Characterisation of Ydj1: a yeast homologue of the bacteria dnaJ protein. *Journal of Cell Biology.* **114:** 609-621.

Caplan, A.J., Tsai, J., Casey, P.J. and Douglas, M.G. (1992a) Farnesylation of Ydj1p is required for function at elevated growth temperatures in *Saccharomyces cerevisiae. The Journal of Biological Chemistry.* **267(26):** 18890-18895.

Caplan, A.J., Cyr, D.M. and Douglas, M.G. (1992b) YDJ1p facilitates polypeptide translocation across different intracellular membranes by a conserved mechanism. *Cell.* **71(7):**1143-55.

Caplan, A.J. and Douglas, M.G. (1991) Characterization of YDJ1: a yeast homologue of the bacterial dnaJ protein. *Journal of Cell Biology.***114(4):** 609-21.

Carmichael, J., Chatellier, J., Woolfson, A., Milstein, C., Fersht, A.R. and Rubinzstein D.C. (2000) Bacterial and yeast chaperones reduce both aggregate formation and cell death in mammalian cell models of Huntington's disease. *Proceedings of the National Academy of Science, USA* **97**: 9701-9705.

Carreira, M.A., Tibbetts, R.S., Olson, C.L., Schuster, C., Renz, M., Engman, D.M. and Goldenberg, S. (1998) TcDJ1, a putative mitochondrial DnaJ protein in *Trypanosoma cruzi. FEMS Microbiology. Letters.* **166 (1):** 141-6.

Casey, P.J., Thissen, J.A. and Moomaw, J.F., (1991) Enzymatic modification of proteins with a geranylgeranyl isoprenoid. *Proceedings of the National Academy of Science, USA.* **88:** 8631 – 8635.

Cedano, J., Aloy, P., Pérez-Pons, J.A. and **Querol, E.** (1997) Relation between amino acid composition and cellular location of proteins. *Journal of Molecular Biology.* **17:** 594-600.

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

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

Cheetham, M.E. and Caplan, A.J. (1998) Structure, function and evolution of DnaJ: conservation and adaptation of chaperone function. *Cell Stress &* Chaperones. **3(1):** 26-36.

Cherry, J.M., Adler, C., Ball, C., Chervits, S.A., Dwight, S.S., Hester, E.T., Jia, Y. Juvik, G., Roe, T., Schroeder, M., Weng, S. and Botstein, D. (1998) SGD: *Saccharomyces* Genome Database. *Nucleic Acids Research.* **26(1):** 73-79.

Chifflet, S., Torriglia, A., Chiesa, R and Tolosa, S. (1988) A method for the determination of inorganic phosphate in the presence of labile organic phosphate and high concentrations of protein: application to lens ATPases. *Analytical Biochemistry.* **168:**1-4.

Choi, H-I., Lee, S.P., Kim, K.S., Hwang, C.Y., Lee, Y-R., Chae, S-K., Kim, Y-S., Chae, H.Z. and Kwon, K-S. (2006) Redox-regulated cochaperone activity of the human DnaJ homolog Hdj2. *Free Radical Biology and Medicine.* **40:** 651-659.

Chou, P.Y. and Fasman, G.D. (1978) Prediction of the secondary structure of proteins from their amino acid sequence. *Advances in Enzymology.* **47:** 45-147.

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. *The The Journal of Biological Chemistry.* **278(32):** 30311-30316.

Clarke, G., Heon, E. and McInnes, R.R. (2000) Recent advances in the molecular basis of inherited photoreceptor degeneration. *ClinicalGenetics.* **57**: 313-329.

Claros, M.G. and Vincens, P. (1996) Computational method to predict mitochondrially imported proteins and their targeting sequences. *European Journal of Biochemistry.* **241:** 779-786.

Clayton, C., and Shapira, M. (2007) Post-transcriptional regulation of gene expression in trypanosomes and leishmanias. *Molecular and Biochemical Parasitology.* **156:** 93-101.

Clayton, C.E., Esteacutevez, A.M., Hartmann, C., Alibu, V.P. and Field, M.C., (2005) Down-regulating gene expression by RNA interference in *Trypanosoma brucei. Methods in Molecular Biology.* **309**:39-60.

Clayton, C., Hausler, T. and Blattner, J. (1995) Protein Trafficking in Kinetoplastid Protozoa. *Microbiological Reviews.* **59 (3):** pg 325-344.

Colon, W. and Kelley, J.W. (1992) Partial denaturation of transthyretin is sufficient for amyloid fibril formation in vitro. *Biochemistry* **15(31)**:8654-8660.

Cosgrove, W.B. and Skeene, M.J. (1970) The cell cycle in *Crithidia fasciculata.* Temporal relationships between synthesis of deoxyribonucleic acid and the nucleus and in the kinetoplast. *Journal of Protozoology.* **17**:172-177.

Cripe, T.P., Delos, S.E., Estes, P.A. and Garcea, R.L. (1995) *In vivo* and *in vitro* association of hsc70 with polyomavirus capsid proteins. *Journal of Virology.* **69**: 7807-7813.

Cui, Q., Jiang, T., Liu, B. and Ma, S. (2004) Esub8: a novel tool to predict protein subcellular localizations in eukaryotic organisms. *Bioinformatics.* **5:** 66-72.

Curie, J. and Curie, P. (1880) Développement, par pression, de l'électricité polaire dans les cristaux hémièdres à faces inclinées. *Comptes Rendus de l'Académie des Sciences.* **91**:294–5.

Cummings, L., Riley, L., Black, L. Souvorov, A. Resenchuk, S., Dondoshansky, I. and Tatusova, T. (2002) Genomic BLAST: custom-defined virtual databases for complete and unfinished genomes. *FEMS Microbiology Letters.* **216(2):** 133-138.

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

Cyr, D.M. (1995) Cooperation of the molecular chaperone YdJ1 with specific Hsp70 homologs to suppress protein aggregation. *FEBS letters.* **359(2-3):** 129-132.

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

Cyr, D.M., Lu, X. and Douglas, M.G. (1992) Regulation of Hsp70 function by a eukaryotic DnaJ homologue. *The The Journal of Biological Chemistry.* **267:** 20927-20931.

DaRocha, W.D., Otsu, K, Teixeira, S.M. and Donelson, J.E. (2004) Tests of cytoplasmic RNA interference (RNAi) and construction of a tetracycline-inducible T7 promoter system in *Trypanosoma cruzi. Molecular and Biochemical Parasitology.* **133:** 175-186.

Davis, A.R., Alevy, Y.G., Chellaiah, A., Quinn, M.T. and Mohanakumar, T. (1998) Characterisation of Hdj-2, a human 40 kD heat shock protein. *The International Journal of Biochemistry and Cell Biology.* **30:** 1203-1221.

DeLano, W.L. (2002) The PyMOL molecular graphics system. Delano Scientific, San Carlos, Ca, U.S.A. http://www.pymol.org

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

De Souza, W. (2002) Basic Cell biology of *Trypanosoma cruzi*. *Current Pharmaceutical Design.* **8:** 269-285.

Dill, K.A., Chan, H.S. (1997) From Levinthal to pathways to funnels. *Nature Strucural Biology.* **4**: 10-19.

Dimitrievski, K., Reimhult, E., Kasemo, B and Zhdanov, V.P. (2004) Simulation of temperature dependence of the formation of a supported lipid bilayer via vesicle adsorption. *Colloids and Surfaces B: Interfaces.* **39:** 77-86.

Djikeng, A. Shi, H., Tschudi, C. and Ullu, E. (2001) RNA interference in *Trypanosoma brucei* cloning of small interfering RNAs provides evidence for retroposon derived 24-26-nucleotide RNAs. *RNA.* **7:** 1522-1530.

Dobson, C.M. and Karplus, M. (1999) The Fundamentals of Protein Folding: bringing together theory and experiment. *Current Opinion in Structural Biology* 9(1):92-101.

Dobson, C.M. (2003) Protein Folding and Misfolding. *Nature* **426**: 884-890.

Dobson, C.M. (2004) Principles of protein folding, misfolding and aggregation. *Seminars in Cell & Developmental biology* **15**: 3-16.

Dolatshahi-Pirouz, A., Rechendorff, K., Hovgaard, M.B., Foss, M., Chevalier, J. and Besebacher, F. (2008) Bovine serum albumin adsorption on nano-rouph platinum surfaces studied by QCM-D. *Colloids and Surfaces B: Biointerfaces.* 66: 53-59.

Donelson, J.E., Gardner, M.J. and El-Sayed, N.M. (1999) More Surprises from the Kinetoplastida. *Proceedings of the National Academy of Science, USA.* **96:** 2579-2581.

Dorn, I.T., Pawlitschko, K., Pettinger, S.C. and Tampé, R. (1998) Orientation and two-dimensional organization of proteins at chelator lipid interfaces. *Biological Chemistry.* **379:** 1151-1159.

Duszenko, M., Ferguson, M.A.J., Lamont, M., Rifkin, R. and Cross, G.A.M. (1985) Cysteine eliminates the feeder cell requirement for cultivation of *Trypanosoma brucei* bloodstream forms *in vitro*. *Journal of Experimental Medicine.* **162:** 1256-1263.

Eastman, R.T., Buckner, F.S., Yokoyama, K., Gelb, M.H. and Van Voorhis, W.C. (2006) Fighting parasitic disease by blocking protein farnesylation. *Journal of Lipid Research.* **47:** 233-240.

Edkins, A.L., Ludewig, M.H. and Blatch, G.L. (2004) A *Trypanosoma cruzi* heat shock protein 40 is able to stimulate the adenosine triphosphate hydrolysis activity of heat shock protein 70 and can substitute for a yeast heat shock protein 40. *The International Journal of Biochemistry and Cell Biology.* **36:** 1585-1598.

El-Sayed, N.M., Myler, P.J., Bartholomeu, D.C., Nilsson, D., Aggarwal, G., Tran, A-N., Ghedin, E., Worthey, E.A., Delcher, A.L., Blandin, G., et al., (2005). The genome sequence of *Trypanosoma Cruzi*, etiologic agent of Chagas disease. *Science*. **309:** 409-415.

Eisenberg, D., Weiss, R.M. and Terwilliger, T. C. (1984) The hydrophobic moment detects periodicity in protein hydrophobicity. *Proceedings of the National Academy of Sciences, USA*. **81(1):** 140-144.

Elias, M.C. da Cunha, J.P., de Faria F.P., Mortara, R.A., Freymuller, E. and Schenkman, S (2007) Morphological events during the *Trypansoma cruzi* cell cycle. *Protist*. 158:147-157.

Elliott, J.T. and Prestwich, G.D. (2000) Maleimide-functionalised lipids that anchor polypeptides to lipid bilayers and membranes. *Bioconjugate Chemistry*. **11:** 832-841.

Ellis, R.J. and Minton, A.P. (2003) Join the Crowd. *Nature* 425:27-28.

Ellis, R.J. (2001) Macromolecular crowding: obvious but underappreciated. *Trends in Biochemical Sciences* **26**(no 10): 597-604.

Ellis, R.J. and Hemmingson, S.M. (1989) Molecular Chaperones: Proteins essential for the biogenesis of some macromolecular structures. *Trends in Biochemical Sciences* **14**: 339-342.

Ellis, R.J. (1987) Proteins as Molecular Chaperones. *Nature* **328**: 378-379.

Emanuelsson, O. and von Heijne, G. (2001) Prediction of organellar targeting signals. *Biochemica et Physica Acta*. **1541:** 114-119.

Emini, E.A., Hughes, J.V., Perlow, D.S. and Boger, J. (1985) Induction of hepatitis A virus-neutralising antibody by a virus-specific synthetic peptide. *Journal of Virology*. **55(3):** 836-839.

Engelman, D.M., Steitz, T.A. and Goldman, A. (1986) Identifying nonpolar transbilayer helices in amino acid sequences of membrane proteins. *Annual Review Biophysics and Biophysical Chemistry*. **15:** 321-353.

Engman, D.M., Henkle-Dührsen, K., Kirchhoff, L.V. and Donelson, J.E. (1995) *Trypanosoma cruzi*: Accumulation of Polycistronic hsp70 RNAs during severe heat shock. *Experimental Parasitology*. **80:** 575-577.

Fan, C-Y., Lee, S. and Cyr, D.M. (2003) Mechanisms for regulation of Hsp70 function by Hsp40. *Cell Stress and Chaperones*. **8 (4):** 308-316.

Fan, C-Y., Lee, S. and Cyr, D.M. (2004) Type I and type II Hsp40s contain exchangeable chaperone modules that specify Hsp70 function. *Molecular Biology of the Cell*. **15:** 761-773.

Fan, C-Y., Ren, H-Y., Lee, P., Caplan, A.J. and Cyr, D.M. (2005) The Type I Hsp40 zinc finger-like region is required for Hsp70 to capture non-native polypeptides from YdJ1. The *The Journal of Biological Chemistry*. **280(1):** 695-702.

Fant, C., Elwing, H. and Höök, F. (2002) The influence of cross-linking on protein-protein interactions in a marine adhesive: The case of two byssus plaque proteins from the blue mussel. *Biomacromolecules*. **3:** 732-741.

Feng, Z.P. and Zhang, C.T. (2001) Prediction of the subcellular location of prokaryotic proteins based on the hydrophobic index of the amino acids. *International Journal of Biological Macromolecules*. **14:** 255-261.

Fersht, A.R. (2000) Transition-state structure as a unifying basis in protein-folding mechanisms: contact order, chain topology, stability and the extended nucleus mechanism. *Proceedings of the National Academy of Sciences, USA* **97**:1525-1529.

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

Field, M.C. and Carrington, M. (2004) Intracellular Membrane Transport Systems in *Trypanosoma brucei*. *Traffic*. **5:** 905-913.

Fink, A.L. (1999) Chaperone-Mediated Protein Folding. *Physiological Reveiws* **79(2)**:425-449.

Fire A, Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E. and Mello, C.C. (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature*. **391:**806-811.

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

Flom, G.A., Lemieszek, M., Fortunato, E.A. and Johnson, J.L. (2008) Farnesylation of Ydj1 is required for in vivo interaction with Hsp90 client proteins. *Molecular Biology of the Cell*. **19:** 5249-5258.

Folgueira, C. and Requena, J.M. (2007) A postgenomic view of the heat shock proteins in kinetoplastids. *FEMS Microbiological Review*. **4:** 359-377.

Fowler, D.M., Koulov, A.V., Alory-Jost, C., Marks, M.S., Balch, W.E. and Kelley, J.W. (2006) Functional Amyloid Formation within Mammalian Tissue. *PLOS Biology*, **4(1)**: 100-107.

Fraumann, R. (2003). "Glossina morsitans" (On-line), Animal Diversity Web. Accessed September 17, 2008 at http://animaldiversity.ummz.umich.edu/site/accounts/information/Glossina_morsitans.html.

Freeman, B.C., Myers, M.P., Schumacher, R. and Morimoto, R.I. (1995) Identification of a regulatory motif in Hsp70 that affects ATPase activity, substrate binding and interaction with HDJ-1. *European Moelcular Biology Organisation (EMBO) Journal*. **14(10):** 2281-2292.

Fridberg, A., Buchanan, K.T. and Engman, D.M. (2007) Flagellar membrane trafficking in kinetoplastids. *ParasitologyResearch*. **100**:205-212.

Frydman, J. (2001) Folding of Newly Translated Proteins *in vivo*: The Role of Molecular Chaperones. *Annual Review of Biochemistry* **70**: 603-647.

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

Galanti, N., Galindo, M., Sabaj, V., Espinoza, I. and Toro, G.C. (1998) Histone genes in trypanosomatids. *Parasitology Today*. **14:** 64-70.

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

Gamez, A., Perez, B., Ugarte, M. and Desviat, L.R. (2000) Expression analysis of phenylketoneuria mutations. Effect on folding and stability of the phenylalanine hydroxylase protein. *The The Journal of Biological Chemistry*. **275**: 29737-29742.

Garcia, A., Courtin, D., Solano, P., Koffi, M. and Jamonneau, V. (2006) Human African Trypanosomiasis: connecting parasite and host genetics. *TRENDS in Parasitology.* **22 (9):** 405-409.

Garimella, R., Liu, X., Qiao, W., Liang, X., Zuiderweg, E.R.P., Riley, M.I. and Van Doren, S.R. (2006) Hsc70 contacts helix III of the J-Domain from polyomavirus T antigens: addressing a Dilemma in the Chaperone Hypothesis of how they release E2F from pRb. *Biochemistry* **45**: 6917-6929.

Gässler, C.S., Wiederkehr, T., Brehmer, D., Bukau, B. and Mayer, M.P. (2001) Bag1-M accelerates nucleotide release for human Hsc70 and Hsp70 and can act concentration dependent as positive and negative cofactor. *The Journal of Biological Chemistry.* **276:** 32538-32544.

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

Garnier, J., Osguthorpe, D.J. and Robson, B. (1978) Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. *Journal of Molecular Biology.* **120:** 97-120.

Gelb, M.H., Van Voorhis, W.C., Buckner, F.S., Yokoyama, K., Eastman, R., Carpenter, E.P., Panethymitaki, C, Brown, K.A. and Smith, D.F. (2003) Protein farnesyl and N-myristoyl transferases: piggy-back medicinal chemistry targets for the development of antitrypanosomatid and antimalarial therapeutics. *Molecular and Biochemical Parasitology.* **126:** 155-163.

Genevaux, P., Georgopoulos, C. and Kelley, W.L. (2007) The Hsp70 chaperone machines of *Escherichia coli*: A paradigm for the repartition of chaperone functions. *Molecular Microbiology.* **66(4):** 840-857.

Genevaux, P., Lang, f., Schwager, F., Vartikar, J.V., Rundell, K., Pipas, J.M., Georgopoulos, C., and Kelley, W.L. (2003) Simian virus 40 T antigens and J-domains: Analysis of Hsp40 cochaperone functions in *Escherichia coli. Journal of Virology.* **77(19):** 10706-10713.

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

Genevaux, P., Wawrzynáw, A., Zylicz, M., Georgopoulos, C., and Kelley, W.L. (2001) DjlA is a third DnaK co-chaperone of *Escherichia coli*, and DjlA-mediated induction of colonic acid capsile requires DjLA-DnaK interaction. *The The Journal of Biological Chemistry.* **276:** 7906-7912.

Gething, M-J. and Sambrook, J. (1992) Protein Folding in the cell. *Nature* **355**: 33-45.

Ghaemmaghami, S., Hu, W-K., Bower, K., Howson, R.W., Belle, A., Dephoure, N., O'Shea, E.K. and Weissman, J.S. (2003) Global analysis of protein expression in yeast. *Nature.* **425:** 737-741.

Ghedin, E., Bingaud, F., Peterson, J., Myler, P., Berriman, M., Ivens, A., Andersson, B., Bontempi, E., Eisen, J., Angiuoli, S., Wanless, D., Von Arx, A., Murphy, L., Lennard, N., Salzberg, S., Adams, M.D., White, O., Hall, N., Stuart, K., Fraser, C.M. and El-Sayed, N.M.A. (2004) Gene Synteny and evolution of genome architecture in trypanosomatids. *Molecular and Biochemical Parasitology.* **134:** 183-191.

Gillespie, J.R., Yokoyama, K., Lu, K., Eastman, R.T., Bollinger, J.G., Voorhis, W.C.V., Gelb, M.H. and Buckner, F.S. (2007) C-terminal proteolysis of prenylated proteins in trypanosomatids and RNA interference of enzymes required for the post-translational processing pathway of farnesylated proteins. *Molecular and Biochemical Parasitology.* **153:** 115-124.

Glasmäster, K., Larsson, C., Höök, F. and Kasemo, B. (2002) Protein adsorption on supported phospholipids bilayers. *Journal of Colloid and Interface Science.* **246:** 40-47.

Glover, J.R. and Tkach, J.M. (2001) Crowbars and ratchets: Hsp100 chaperones as tools in reversing protein aggregation. *Biochemistry and Cell Biology.* **79**: 557-568.

Goldberg A.L. (2003) Protein degradation and protection against misfolded or damaged proteins. *Nature* **426**: 895-899.

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

Goloubinoff, P. and de los Rios, P. (2007) The mechanism of Hsp70 chaperones (entropic) pulling the models together. *Trends in the Biochemical Sciences.* **32(8):** 372-380.

Goloubinoff, P., Mogk, A., Zvi, A.P.B., Tomoyasu, T., and Bukau, B. (1999) Sequential mechanism of solubilisation and refolding of stable protein aggregates by a bichaperone network. *Proceedings of the National Academy of Science, USA* **96**: 13732-13737.

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

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

Guda, C. (2006) pTARGET: A web server for predicting protein subcellular localization. *Nucleic Acids Research.* **35:**W210-213.

Guda, C., Subramaniam, S. (2005) pTARGET: A new method for predicting protein subcellular localization in eukaryotes. *Bioinformatics.* **21:** 3963-3969.

Guidon, P.T. and Hightower, L.E. (1986) Purification and initial characterization of the 71-kilodalton rat heat-shock protein and its cognate as fatty acid binding proteins. *Biochemistry* **25(11):** 3231-3239.

Gull, K. (1999) The Cytoskeleton of Trypanosomatid Parasites. *Annual Review of Microbiology.* **53:** 629-655.

Gull, K. (2002) The Cell Biology of Parasitism in *Trypanosoma brucei*: insights and drug targets from genomic approaches. *Current Pharmaceutical Design.* **8:** 241-256.

Gull, K. (2003) Host-Parasite interactions and trypanosome morphogenesis: a flagellar pocketful of goodies. *Current Opinion in Microbiology.* **6:** 365-370.

Guttes, E. W., Hanawalt, P.C. and Guttes, S. (1967) Mitochondrial DNA synthesis and the mitotic cycle in *Physarum polycephalum. Biochimca et Biophysica Acta.* **142:** 181-194.

Hall, T.A. (1999) Bioedit, a user-friendly biological sequence alignment editor and analysis programme for Windows 95/98/NT. *Nucleic Acids Symposium Series.* **41:** 95-98.

Hammarton, T.C. (2007) Cell cycle regulation in *Trypanosoma brucei. Molecular and Biochemical Parasitology.* **153:** 1-8.

Hammarton, T.C., Monnerat, S. and Mottram, J.C. (2007) Cytokinesis in trypanosomatids. *Current Opinion in Micorbiology.* **10:** 520-527.

Hammarton, T.C., Clark, J., Douglas, F., Boshart, M. and Mottram, J.C. (2003) Stage-specific differences in cell cycle control in *Trypanosoma brucei* revealed by RNA interference of a mitotic cyclin. *The Journal of Biological Chemistry.* **278:** 22877-22886.

Hannon, G.J. (2002) RNA Interference. *Nature.* **418:** 244 – 251.

Hartl, F.U. and Hayer-Hartl, M. (2002) Molecular Chaperones in the Cytosol: from Nascent Chain to Folded Protein. *Science* 295: 1852-1858.

Hartl, F.U. (1996) Molecular chaperones in cellular protein folding. *Nature.* **381(6583):** 571-579.

Hausler, T., Stierhof, Y.D., Blattner, J. and Clayton, C. (1997) Conservation of mitochondrial targeting sequence function in mitochondrial and hydrogenosomal proteins from the early-branching eukaryotes *Crithidia*, *Trypanosoma* and *Trichomonas*. *European Journal of Cell Biology.* **73:** 240-251.

Hendrick, J.P. and Hartl, F.U. (1993) Molecular Chaperone Functions of Heat Shock Proteins. *Annual Review of Biochemistry* **62**:349-384.

Hennessy, F., Boshoff, A. and Blatch, (2005a) Rational mutagenesis of a 40 kDa heat shock protein from *Agrobacterium tumefaciens* identifies amino acid residues critical to its *in vivo* function. *The International Journal of Biochemistry and Cell Biology.* **37:** 177-191.

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

Hennessy, F., Cheetham, M.E., Dirr, H.W. and Blatch, G.L. (2000) Analysis of the levels of conservation of the J-domain among the various types of DnaJ-like proteins. *Cell Stress and Chaperones.* **5(4):** 347-358.

Hertz-Fowler, Christiane, Peacock, C.S., Wood, V., Aslett, M., Kerhornou, A., Mooney, P., Tivey, A., Berriman, M., Hall, N., Rutherford, K., Parkhill, J., Ivens, A.C., Rajandream, M-A. and Barrell, B. (2004) GeneDB: a resource for prokaryotic and eukaryotic organisms. *Nucleic acids research.* **32 (Database issue):** D339-343.

Hill, K.L. Hutchings, N.R., Russell, D.G. and Donelson, J.E. (1999) A novel protein targeting domain directs proteins to the anterior cytoplasmic face of the flagellar pocket in African trypanosomes. *Journal of Cell Science.* **112:** 3091-3101.

Hill, R.B., Flanagan, J.M. and Prestegard, J.H. (1995) 1H and 15N magnetic resonance assignments, secondary structure, and tertiary fold of Escherichia coli DnaJ (1-78). *Biochemistry.* **34:** 5587-5596.

Hirumi, H. and Hirumi, K. (1994) Axenic Culture of African Trypaonsome Bloodstream Forms. *Parasitology Today.* **10(2):** 80-84.

Hirumi, H. and Hirumi, K. (1989) Continuous cultivation of Trypanosoma brucei bloodstream forms in a medium containing a low concentration of serum-protein without feeder cell layers. *Journal of parasitology.* **75:** 985-989.

Hirumi, H, Doyle, J.J. and Hirumi, K. (1977) African trypanosomes: cultivation of animal-infective *Trypanosoma brucei in vitro*. *Science.* **196:** 992-994.

Hobbs, H.H., Russell, D.W., Brown, M.S., and Goldstein, J.L. (1990) The LDL receptor locus in familial hypercholesterolemia: mutational analysis of a membrane protein. *Annual Review of Genetics.* **24**: 133-170.

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

Höök, F., Vörös, J., Rodahl, M., Kurrat, R., Boni, P., Ramsden, J.J., Textor, M., Spencer, N.D., Tengvall, P., Gold, J. and Kasemo, B. (2002) A comparative study of protein adsorption on titanium oxide surfaces using in situ ellipsometry, optical waveguide lightmode spectroscopy, and quartz crystal microbalance/dissipation. *Colloids and Surfaces B-Biointerfaces*. **24:**155-170.

Höök, F., Kasemo, B., Nylander, T., Fant, C., Sott, K. and Elwing, H. (2001) Variations in coupled water, viscoelastic properties, and film thickness of a Mefp-1 protein film during adsorption and cross-linking: a quartz crystal microbalance with dissipation monitoring, ellipsometry, and surface plasmon resonance study. *Analytical Chemistry*. **73:** 5796-5804.

Höök, F., Rodahl, M., Brzezinski, P. and Kasemo, B. (1998a) Energy dissipation kinetics for protein and antibody-antigen adsorption under shear oscillation on a quartz crystal microbalance. *Langmuir*. **14:** 729-734.

Höök, F., Rodahl, M., Kasemo, B., and Brzezinski, P. (1998b) Structural changes in hemoglobin during adsorption to solid surfaces: Effects of pH, ionic strength, and ligand binding. *Proceedings of the National Academy of Sciences, USA*. **95:** 12271-12276.

Hopp, T.P. and Woods, K.R. (1981) Prediction of protein antigenic determinants from amino acid sequences. *Proceedings of the National Academy of Sciences, USA*. **78(6):** 3824-3828.

Horton, P., Park, K-J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C.J. and Nakai, K. (2007) WoLF PSORT: protein localization predictor. *Nucleic Acids Research*. **35:** Web server issue W585-W587.

Horton, P., Park, K-J., Obayashi, T. and Nakai, K. (2006) Protein Subcellular Localisation Prediction with WoLF PSORT. *Proceedings of Asian Pacific Bioinformatics Conference, APBC06, Imperial College Press, London*. Pp. 39-48.

Hsu, M.P. Muhich, M. and Boothroyd, J.C. (1989) A developmentally regulated gene of trypanosomes encodes a homologue of rat protein, ELP-1. *Cell*. **69:**625-635.

Hu, J., Wu, Y., Li, J., Qian, X., Fu, Z. and Sha, B. (2008) The crystal structure of the putative peptide-binding fragment from the human Hsp40 protein Hdj1. *BMC Structural Biology*. **8(3)** : in press.

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

Hulo, N., Bairoch, A., Bulliard, V., Cerutti, L., Cuche, B., De Castro, E., Lachaize, C., Langendijk-Genevaux, P.S. and Sigrist, C.J.A. (2007) The 20 years of PROSITE. *Nucleic Acids Research*. **36(Database issue):** D245–D249.

Hunt, C and Morimoto, R.I. (1985) Conserved Features of Eucaryotic Hsp70 genes revealed by the comparison with the nucleotide sequence of human Hsp70. *Proceedings of the National Academy of Science, USA*. **82:** 6455-6459.

Huth, J.R., Norton, S.E., Lockeridge, O., Shikone, T., Hsueh, A.J.W. and Ruddon, R.W. (1993) Bacterial expression and *in vitro* folding of the beta-subunit of human chorionic gonadotropin (hCG beta) and functional assembly of recombinant hCG beta with hCG alpha. *Endocrinology*. **135**(no.3):911-918.

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

Ivens, A.C., Peacock, C.S., Worthey, E.A., Murphy, L., Aggarwal, G., Berriman, M., Sisk, E., Rajandream, M-A., Adlem, E., Aert, R., et al. (2005). The genome of the kinetoplastid parasite, Leishmania major. *Science*. **309:** 436-442.

Jahn, T.R. and Radford, S.E. (2005) The Yin and Yang of Protein Folding. *The FEBS Journal* **272**: 5962-5970.

James, P., Pfund, C. and Craig, E.A. (1997) Functional specificity among Hsp70 molecular chaperones. *Science*. **275:** 387-389.

Jameson, B.A. and Wolf, H. (1988) The antigenic index: a novel algorithm for predicting antigenic determinants. *Computer Applications in the Biosciences (CABIOS)*. **4(1):** 181-186.

Janin, J., Wodak, S., Levitt, M. and Maigret, B. (1978) Conformation of amino acid side-chains in proteins. *Journal of Molecular Biology*. **125:** 357-386.

Jiang, J., Maes, E.G., Taylor, A.B., Wang, L., Hinck, A.P., Lafer, E.M. and Sousa, R. (2007) Structural basis of the J cochaperone binding and regulation of Hsp70. *Molecular Cell*. **28:** 1-12.

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

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

Johnson, J.L. and Craig, E.A. (2001) An essential role for the substrate-binding region of Hsp40s in *Saccharomyces cerevisiae*. *The Journal of Cell Biology*. **152(4):** 851-856.

Johnson, J.L. and Craig, E.A. (2000) A role for the Hsp40 Ydj1 in repression of basal steroid receptor activity in yeast. *Molecular and Cellular Biology*. **20:**3027-3036.

Jones, A., Faldas, A., Foucher, A., Hunt, E., Tait, A., Wastling, J.M. and Turner, C.M. (2007) Visualisation and analysis of proteomic data from the procyclic form of *Trypanosoma brucei*. *Proteomics*. **6(I):** 259-267.

Jones, A., Faldas, A., Foucher, A., Hunt, E., Tait, A., Wastling, J.M. and Turner, M. (2006) Visualisation and analysis of proteomic data from the procyclic form of *Trypanosoma brucei*. *Proteomics*. **6:** 259-267.

Kabani, M., Beckerich, J.-M. and Brodsky, J.L. (2003) The yeast Sis1p and Fes1p proteins define a new family of Hsp70 nucleotide exchange factors. *Current Genomics*. **4:** 263-273.

Kampinga, H.H., Hageman, J., Vos, M.J., Kubota, H., Tanguay, R.M., Bruford, E.A., Cheetham, M.E., Chen, B. and Hightower, L.E. (2009) Guidelines for the nomenclature of the human heat shock proteins. *Cell Stress and Chaperones*. **14:** 105-111.

Kanazawa, M., Terada, K., Kato, S. and Mori, M. (1997) Hsdj, a human homolog of DnaJ, is farnesylated and is involved in protein import into mitochondria. *The Journal of Biochemistry*. **121:** 890-895.

Karasev, A.V., Nikolaeva, O.V., Mushegian, A.R., Lee, R.F. and Dawson, W.O. (1996) Organization of the 3'-terminal half of beet yellow stunt virus genome and implications for the evolution of closteroviruses. *Virology.* **221**: 199-207.

Karplus, P.A. and Schultz, G.E. (1988) Prediction of chain flexibility in proteins. *Naturwissenschaften.* **72:** 212-213.

Katayama, T., Imaisumi, K., Sato, N., Miyoshi, K., Kudo, T., Hitomi, J., Morihara, T., Yoneda, T., Gomi, F., Mori, Y., Nakano, Y., Takeda, J., Tsuda, T., Itoyama, Y., Murayama, O., Takashima, A., St George-Hyslop P., Takeda, M. and Tohyama, M. (1999) Presenilin-1 mutations downregulate the signaling pathway of the unfolded-protein response. *Nature Cell Biology* **1**: 479-485.

Keller, C.A., Gläsmaster, K., Zdhanov, V.P. and Kasemo, B. (2000) Formation of supported membranes from vesicles. *Physical Review Letters.* **84(23):** 5543-5546.

Keller, C.A. and Kasemo, B. (1998) Surface specific kinetics of lipid vesicle adsorption measured with a quartz crystal microbalance. *Biophysical Journal.* **75:** 1397-1402.

Kelley, L.A. and Sternberg, (2009) Protein structure prediction on the web: a case study using the Phyre server. *Nature Protocols.* **4(3):** 363- 371.

Kelley, L.A., MacCallum, R.M. and Sternberg, M.J. (2000) Enhanced genome annotation using structural profiles in the program 3d-PSSM. *Journal of Molecular Biology.* **299:** 499-520.

Kelley, W.L. (1999) Molecular Chaperones: how J-domains turn on Hsp70s. *Current Biology.* **9:** R305-R308.

Kelley, W.L. (1998) The J-domain family and the recruitment of chaperone power. *Trends in the Biochemical Sciences.* **23(6):** 222-227.

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

Kennedy, P.G.E. (2006) Diagnostic and neuropathogenesis issues in human African trypanosomiasis. *International Journal for Parasitology.* **36:** 505-512.

Kiang J.G. and Tsokos, G.C. (1998) Heat shock protein 70kDa: Molecular biology, biochemistry and Physiology. *Phamacology and Therapeutics.* **80**: 183-201.

Kienberger, F., Moser, R., Schindler, H., Blass, D. and Hinterdorfer, P. (2001) Quasi-crystalline arrangement of human rhinovirus 2 on model cell membranes. *Single Molecules.* **2:** 99-103.

Kim, H-Y., Shn, B-Y. and Cho, Y. (2001) Structural basis for the inactivation of retinoblastoma tumor suppressor by SV40 large T antigen. *European Molecular Biology Organisation(EMBO) European Moelcular Biology Organisation (EMBO) Journal.* **20:** 295-304.

Kim, D., Lee, Y.J. and Corry, P.M. (1992) Constitutive Hsp70: oligomerization and its dependence on ATP binding. *Journal of Cellular Physiology.* **153:** 353-361.

Kinsella, B.T., Erdman, R.A. and Maltese, W.A., (1991) Postranslational modification of Ha-ras p21 by farnesyl versus geranylgeranyl isoprenoids is determined by the COOH-terminal amino acid. *Proceedings of the National Academy of Science, USA.* **188:** 8934-8938.

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

Koonin, E.V. (2005) Orthologues, paralogs and evolutionary genomics. *Annual Review of Genetics.* **39:** 309-338.

Kraus, J.M., Verlinde, C.L., Karimi, M., Lepesheva, G.I., Gelb, M.H. and Buckner, F.S. (2009) Rational modification of a candidate cancer drug for use against Chagas disease. *Journal of Medicinal Chemistry.* **52(6):** 1639-1647.

Krautz, G.M., Peterson, J.D., Godsel, L.M., Krettli, A.U. and Engman, D.M. (1998) Human antibody response to Trypanosoma cruzi 70 kDa Heat-shock proteins. *The American Journal of Trpical Medicine and Hygiene.* **58:** 321-324.

Kumar, N. and Zheng, H. (1992) Nucleotide sequence of a plasmodium falciparum stress protein with similarity to mammalian 78-kDa glucose-regulated protein. *Molecular and Biochemical Parasitology.* **56(2):** 353- 356.

Kyte, J. and Doolittle, R.F. (1982) A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology.* **157:** 105-132.

LaCount, D.J. and Donelson, J.E. (2001) RNA interference in African Trypanosomes. *Protist.* **152(2):** 103-111.

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

Lamande, S.R., Chessler, S.d., Golub, S.B., Byers P.H., Chan, D., Cole, W.G., Sillence, D.O. and Bateman, J.F. (1995) Endoplasmic reticulum-mediated quality control of type I collagen production by cells from osteogenesis imperfecta patients and mutations in the pro alpha 1(1) chain carboxyl-terminal propeptide which impair subunit assembly. *Journal Biological Chemistry.* **270**: 8642-8649.

Landfear, S.M. and Ignatushchenko, M. (2001) The Flagellum and flagellar pocket of trypanosomatids. *Molecular and Biochemical Parasitology.* 115: 1-17.

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

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

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

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, A.W., Lopez, R., Thompson, J.D., Gibson, T.J. and Higgins, D.G. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics.* **23 (21):** 2947-2948.

Larsson, C., Rodahl, M. and Höök, F., (2003) Characterisation of DNA immobilization and subsequent hybridization on a 2D arrangement of streptavidin on a biotin-modified lipid bilayer supported on SiO_2. *Analytical Chemistry.* **75:** 5080-5087.

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

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. *Journal of Applied Crystallography.* **26:** 283-291.

Lau, M.M. and Neufeld, E.F. (1989) A Frameshift mutation in a patient with Tay-Sachs disease causes premature termination and defective intracellular transport of the alpha-subunit of a beta-hexosaminidase. *The Journal Biological Chemistry.* **35**:21376-21380.

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

Lee, S., Fan, C-Y., Younger, J.M., Ren, H. and Cyr, D.M. (2002) Identification of essential residues in the type II Hsp40 Sis1 that function in polypeptide binding. *The The Journal of Biological Chemistry.* **277(24):** 21675-21682.

Lee, G.J. and Vierling, E., (2000) A small heat shock protein cooperates with heat shock protein 70 systems to reactivate a heat-denatured protein. *Plant Physiology.* **122 (1):** 189-197.

Lee, M. G-S. and Van der Ploeg, L.H.T. (1990) Transcription of the heat shock 70 locus in *Trypanosoma brucei. Molecular and Biochemical Parasitology.* **41**:221-232.

Lee, M. G-S., Polvere, R.I. and Van der Ploeg, L.H.T. (1990) Evidence for segmental gene conversion between a cognate hsp70 gene and the temperature-sensitively transcribed hsp70 genes of *Trypanosoma brucei. Molecular and Biochemical Parasitology.* **41:** 213-220.

Leifso, K., Cohen-Freue, G., Dogra, N., Murray, A. and McMaster, W.R. (2007) Genomic and proteomic expression analysis of *Leishmania* promastigote and amastigote life stages: the *Leishmania* genome is constitutively expressed. *Molecular and Biochemical Parasitology.* **152(1):** 35-46.

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

Leone, G., Coffrey, M.C., Gilmore, R., Duncan, R., Maybaum, L. and Lee, P.W. (1996) C-terminal trimerisation, but not N-terminal trimerisation, of the reovirus cell attachment protein is a posttranslational and Hsp70/ATP-dependent process. *The The Journal of Biological Chemistry.* **271:** 8466-8471.

Li, G.C., Li, L., Liu, R.Y., Rehman, M. and Lee, W.M.F. (1992) Protection from thermal stress by human Hsp70 with or without its ATP-binding domain. *Proceedings of the National Academy of Science, USA.* **89**:2036-2040.

Li, G.C., Li, L., Liu, Y-K., Mak, J.Y., Chen, L. and Lee, W.M.F. (1991) Thermal response of rat fibroblasts stably transfected with the human 70-kDa heat shock protein-encoding gene. *Proceedings of the National Academy of Science, USA.* **88**: 1681-1685.

Li, J., Qian, X. and Sha, B. (2009) Heat Shock protein 40: structural studies and their functional implications. *Protein and Peptide Letters.* **16(6):** 606-612.

Li, J., Wu, Y., Qian, X. and Sha, B. (2006) Crystal structure of yeast Sis1 peptide-binding fragment and Hsp70 Ssa1 C-terminal complex. *Biochemical Journal.* **398:** 353-360.

Li, J., Qian, X. and Sha, B. (2003) The crystal structure of the yeast Hsp40 Ydj1 complexed with its peptide substrate. *Structure*. **11:** 1475-1483.

Li, J., and Sha, B. (2005) Structure-based mutagenesis studies of the peptide substrate binding fragment of Type I heat shock protein 40. *Biochemical Journal*. **386:** 453-460.

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

Liang, P. and MacRae, T.H. (1997) Molecular Chaperones and the Cytoskeleton. *Journal of Cell Science* **110**:1431-1440.

Lindquist, S. and Craig, E.A. (1988) The Heat Shock proteins. *Annual Review of Genetics*. **22**: 631-677.

Litt, M., Kramer, P., LaMorticella, D.M., Murphey, W., Lovrien, E.W., and Weleber, R.G. (1998) Autosomal dominant congenital cataract associated with a missense mutation in the human alpha crystalline gene CRYAA. *Human Molecular Genetics* 7: 471-474.

Lopez, N., Aron, R. and Craig, E.A. (2003) Specificity of class II Hsp40 Sis1 in maintenance of yeast prion [RNQ+]. *Molecular Biology Cell*. **14:** 1172 -1181.

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

Loo, M.A., Jensen, T.J., Cui, L., Huo, Y., Chang, X.B. and Riordan, J.R. (1998) Perturbation of Hsp90 interaction with nascent CFTR prevents its maturation and accelerates its degeneration by the proteasome. *European Molecular Biology Organisation(EMBO) Journal*. **23**: 6879-6887.

Louw, C.A., Ludewig, M.H. and Blatch, G.L. (2010) Overproduction, purification and characterization of Tbj1, a novel Type III Hsp40 from *Trypanosoma brucei*, the African sleeping sickness parasite. *Protein Expression and Purification*. **69:**168-177.

Louw, C.A. (2009) Characterisation of Trypansomal Type III and Type IV Hsp40 Proteins. *PhD Thesis, Rhodes University.*

Lu, Z., and Cyr, D.M. (1998a) The conserved carboxyl terminus and zinc finger-like domain of the co-chaperone YdJ1 assist Hsp70 in protein folding. *The Journal of Biological Chemistry*. **273(10):** 5970-5978.

Lu, Z., and Cyr, D.M. (1998b) Protein folding activity of Hsp70 is modified differentially by the Hsp40 co-chaperones Sis1 and YdJ1. *The Journal of Biological Chemistry*. **273(43):** 27824-27830.

Lucklum, R., Behling, C. and Hauptmann, P. (1999) Role of Mass Accumulation and Viscoelastic Film Properties for the Response of Acoustic-Wave-Based Chemical Sensors. *Analytical Chemistry*. **71:** 2488-2496.

Luke, M.M., Sutton, A. and Arndt, K.T. (1991) Characterisation of SIS1, a Saccharomyces cerevisiae homologue of bacterial DnaJ proteins. *The Journal of Cell Biology*. **114(4):** 623-638.

Lukes, J., Hashimi, H. and Zikova, A. (2005) Unexplained complexity of the mitochondrial genome and transcriptome in kinetoplastid flagellates. *Current Genetics*. **48:** 277-299.

Lustig Y., Sheiner, L. Vagima, Y., Goldschmidt, H., Das, A., Bellofatto, V. and Shulamit, M. (2007) Spliced-leader RNA silencing: a novel stress induced mechanism in *Trypanosoma brucei*. *European Molecular Biology Organisation(EMBO) Reports*. **8:** 408-413.

Magee, T. and Seabra, M.C. (2005) Fatty acylation and prenylation of proteins: what's hot in fat. *Current Opinion in Cell Biology.* **17:** 190-196.

Maia, C., Rolao, N., Nunes, M., Goncalves, L. and Campino, L. (2007) Infectivity of Five different types of macrophages by *Leishmania infantum. Acta Tropica.* **103(2):** 150-5.

Maier, A.G., Cooke, B.M., Cowman, A.F. and Tilley, L. (2009) Malaria parasite proteins that remodel the host erythrocyte. Nature Reviews, Microbiology. **7(5):** 341-354.

Marcotte, E.M., Xenarios, I., van Der Bliek, A.M. and Eisenberg, D. (2000) Localising proteins in the cell from their phylogenetic profiles. Proceedings of the *National Academy of Sciences, USA.* **97:** 12115-12120.

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

Massad, E. (2008) The Elimination of Chagas' Disease from Brazil. *Epidemiology and Infection.* **136:** 1153-1164.

Matthews, K. R. (2005) The Developmental Cell Biology of *Trypanosoma brucei. Journal of Cell Science.* **118:** 283-290.

Maurer-Stroh, S., Koranda, M., Benetka, W., Schneider, G., Sirota, F.L. and Eisenhaber, F. (2007) Towards complete sets of farnesylated and geranylgeranylated proteins. *PLoS Computational Biology.* **3(4):** 634-648.

Maurer-Stroh, S. and Eisenhaber, F. (2005) Refinement and prediction of protein prenylation motifs. *Genome Biology.* **6:** R55.

Maurer-Stroh, S., Washietl, S. and Eisenhaber, F. (2003) Protein prenyltransferases: anchor size, pseudogenes and parasites. *Biological Chemistry.* **384:** 977-989.

Mayer, M.P. and Bukau, B. (2005) Hsp70 Chaperones: Cellular functions and Molecular mechanism. *Cellular and Molecular Life Sciences.* **62:** 670-684.

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

Mayer, M.P., Rudiger, S. and Bukau, B. (2000a) Molecular basis for interactions of the DnaK chaperone with substrates. *Biological Chemistry* **381**:877-885.

Mayer, M.P., Schröder, H., Rüdiger, S., Paal, K., Laufen, T. and Bukau, B. (2000b) Multistep mechanism of substrate binding determines chaperone activity of Hsp70. *Nature Structural Biology.* 7:586-583.

Mayer, M., Kies, U., Kammermeier, R. and Buchner, J. (2000c) BiP and PDI cooperate in the oxidative folding of antibodies *in vitro. The Journal of Biological Chemistry.* **275:**29421-29425.

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

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

McConville, M.J., Mullin, K.A., Ilgoutz, S.C. and Teasdale, R.D. (2002) Secretory Pathway of Trypanosomatid Parasites. *Microbiology and Molecular Biology Reviews.* **66(1):** 122-154.

McGuffin LJ, Bryson K and Jones DT. (2000) The PSIPRED protein structure prediction server. *Bioinformatics.* **16:** 404-405. http://bioinf.cs.ucl.ac.uk/psipred/psiform.html

McKean, P.G. (2003) Coordination of cell cycle and cytokinesis in *Trypanosoma brucei*. *Current Opinion in Microbiology.* **6:** 600-7.

McTaggart, S.J. (2006) Isoprenylated proteins. *Cellular and Molecular Life sciences.* **63:** 255-267.

Meacham, G.C., Browne, B.L., Zhang, W., Kellermayer, R., Bedwell, D.M. and Cyr, D.M. (1999) Mutations in the yeast Hsp40 chaperone protein Ydj1 cause defects in Axl1 biogenesis and pro-a-factor processing. *The The Journal of Biological Chemistry.* **274(48):** 34396-34402.

Meyer, R.R. and Simpson, M.W. (1968) DNA biosynthesis in mitochondria: Partial purification of a distinct DNA polymerase from isolated rat liver mitochondria. *Biochemistry.* **61:** 130-137.

Miernyk, J.A. (2001) The J-domain proteins of Arabidopsis thaliana: an unexpectedly large and diverse family of chaperones. *Cell Stress and Chaperones.* **6(3):** 209-218.

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

Minton, A.P. (2000) Implications of Macromolecular Crowding for Protein Assembly. *Current Opinion in Structural Biology* **10**:34-39.

Misselwitz, B., Staeck, O., Matlack, K.E. and Rapoport, T.A. (1999) Interaction of Bip with the J-domain of the Sec63p component of the endoplasmic reticulum protein translocation complex. *The Journal of Biological Chemistry.* **274:** 20110-20115.

Mohler, P.J. Hoffman, J.A., Davis, J.A., Davis, J.Q., Abdi, K.M., Kim, C.R., Jones, S.K., Davis, L.H., Roberts, K.F. and Bennett, V. (2004) Isoform specificity among ankyrins. An amphipathic alpha-helix in the divergent regulatory domain of ankyrin-b interacts with the molecular co-chaperone Hdj1/Hsp40. *The Journal of Biological Chemistry.* **279(24):** 25798-25804.

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

Moore, J.H. (2007) Bioinformatics. *Journal of Cellular Physiology.* **213:** 365-369.

Morano, K.A. and Thiele, D.J. (1999) Heat shock factor function and regulation in response to cellular stress, growth, and differentiation signals. *Gene Expression.* **7:** 271-282.

Moreira, D., López-Garcia, P. and Vickerman, K. (2004) An updated view of kinetoplastid phylogeny using environmental sequences and a closer outgroup: proposal for a new classification of the class Kinetoplastea. *International Journal of Systematic and Evolutionary Microbiology.* **54:** 1861-1875.

Morimoto, R.I. (1988) Regulation of the heat shock transcriptional response: cross talk between a family of heat shock factors, molecular chaperones, and negative regulators. *Genes and Development.* **12:** 3788-3796.

Morimoto, R.I. (2008) Proteotoxic stress and inducible chaperone networks in neurodegenerative disease and aging. *Genes and Development.* **22:** 1427-1438.

Morris, A. L., MacArthur, M. W., Hutchinson, E. G., and Thornton, J. M. (1992). Stereochemical quality of protein structure coordinates. *Proteins.* **12:** 345-364.

Mott, R., Schultz, J., Bork, P. and Ponting, C.,P. (2002) Predicting protein cellular location using a domain projection method. *Genome Research.* **12:** 1168-1174.

Motyka, S.A. and Englund, P.T. (2004) RNA interference for analysis of gene function in trypanosmatids. *Current Opinion in Microbiology.* **7:** 362-368.

Nair, R. and Rost, B. (2003) Better prediction of sub-cellular localization by combining evolutionary and structural information. *Proteins.* **53:** 917-930.

Nair, R. and Rost, B., (2002) Inferring sub-cellular localization through automated lexical analysis. *Bioinformatics.* **18:** S78-S86.

Nakai, K. and Horton, P. (2007) Computational prediction of subcellular localization. *Methods in Molecular Biology.* **390:** 429-466.

Nakai, K. and Horton, P. (1999) Psort: a program for detecting sorting signals in proteins and detemining their subcellular localization. *Trends in Biochemical Sciences.* **24:** 34- 36.

Nakai, K. and Kanehisa, M. (1992) A knowledge base for predicting protein localization sites in eukaryotic cells. *Genomics.* **14:** 897-911.

Naumann, C.A., Prucker, O., Lehmann, T., Rühe, J., Knoll, W. and Frank, C.W. (2002) The polymer-supported phospholipids bilayer: tethering as a new approach to substrate-membrane stabilization. *Biomacromolecules.* **3(1):** 27-35.

Ngo, H., Tschudi, C., Gull, K. and Ullu, E. (1998) Double-stranded RNA induces mRNA degradation in *Trypanosoma brucei. Proceedings of the National Academy of Sciences, USA.* **95(25):** 14687-14692.

Ngosuwan, J., Wang, N.M., Fung, K.L. and Chirico, W.J. (2003) Roles of cytosolic Hsp70 and Hsp40 molecular chaperones in post-translational translocation of presecretory proteins into the endoplasmic reticulum. *The Journal of Biological Chemistry.* **278(9):** 7034-7042.

Nicoll., W.S., Botha, M., McNamara, C., Schlange, M., Pesce, E.-R., Boshoff, A., Ludewig, M.H., Zimmermann, R., Cheetham, M.E., Chapple, J.P. and Blatch, G.L. (2007) Cytosolic and ER J-domains of mammalian and parasitic origin can functionally interact with DnaK. *The International Journal of Biochemistry and Cell Biology.* **39(4):** 736-751.

Nomura, T. and Okuhara, M. (1980) Determination of micromolar concentrations of cyanide in solution with a biezoelectic detector. *Analytica Chimica Acta.* **115:** 323-326.

Nomura, T. and Okuhara, M. (*1982*) Frequency shifts of piezoelectric quartz crystals immersed in organic liquids. *Analytica Chimica Acta* **142:** 281-284.

Norde, W. (1995) Adsorption of proteins at solid-liquid interfaces. *Cells and Materials.* **5:** 97-112.

Ohlsson, P.A., Tjarnhage, T., Herbai, E., Lofas, S. and Puu, G. (1995) Liposome and proteoliposome fusion onto solid substrates, studied using atomic force microscopy, quartz crystal microbalance and surface plasmon resonance. Biological acitivities of incorporated components. *Bioelectrochemistry and Bioenergetics.* **38:** 137-148.

Olson, C.L., Nadeau, K.C., Sullivan, M.A., Winquist, A.G. Donelson, J.E., Walsh, , C.T. and Engman, D.M. (1994) Molecular and biochemical comparison of the 70-kDa heat shock proteins of *Trypanosoma cruzi*. *The Journal of Biological Chemistry*. **269(5):** 3868-3874.

Outeiro, T.F. and Tetzlaff, J. (2007) Mechanisms of Disease II: Cellular Protein Quality Control. *Seminars in Paediatric Neurobiology* **14**:15-25.

Ozeki, T., Verma, V., Uppalapati, M., Suzuki, Y., Nakamura, M., Catchmark, J.M. and Hancock, W.O. (2009) Surface-bound casein modulates the adsorption and activity of kinesin on SiO_2 surfaces. *Biophysical Journal*. **96(8):** 3305-3318.

Pack, D.W., Chen, G., Maloney, K.M., Chen, C-T. and Arnold, F.H. (1997) A metal-chelating lipid for 2D protein crystallization via coordination of surface histidines. *Journal of the American Chemical Society*. **119:** 2479-2487.

Page, R.D.M. (1996) TREEVIEW: an application to display phylogenetic trees on personal computers. *Computer Applications in the Biosciences*. **12:** 357-358.

Pak, M. and Wickner, S. (1997) Mechanism of protein remodeling by ClpA chaperone. *Proceedings of the National Academy of Science, USA* **94**:4901-4906.

Patrick, K.L., Luz, P.M., Ruan, J-P., Shi, H., Ullu, E. and Tschudi, C. (2008) Genomic rearrangements and transcriptional analysis of the spliced leader-associated retrotransposon in RNA interference-deficient *Trypanosoma brucei*. *Molecular Microbiology*. **67(2):** 435-447.

Palleros, D.R., Reid, K.L. Shi, L., Welch, W.J. and Fink, A.L. (1993) ATP-induced protein-Hsp70 complex dissociation requires K^+ but not ATP hydrolysis. *Nature*. **365:** 664-666.

Pays, E., Vanhamme, L., and Perez-Morga, D. (2004) Antigenic variation in *Trypanosoma brucei*: Facts, Challenges and mysteries. *Current Opinion in Microbiology*. 7: 369-374.

Peacock, C.S., Seeger, K., Harris, D., Murphy, L., Ruiz, J.C., Quail, M.A., Peters, N., Adlem, E., Tivey, A., Aslett, M., Kerhornou, A., Ivens, A., Fraser, A., Rajandream, M.A., Carver, T., Norbertczak, H., Chillingworth, T., Hance, Z., Jagels, K., Moule, S., Ormond, D., Rutter, S., Squares, R., Whitehead, S., Rabbinowitsch, E., Arrowsmith, C., White, B., Thurston, S., Bringaud, F., Baldauf, S.L., Faulconbridge, A., Jeffares, D., Depledge, D.P., Oyola, S.O., Hilley, J.D., Brito, L.O., Tosi, L.R., Barrell, B., Cruz, A.K., Mottram, J.C., Smith, D.F. and Berriman, M. (2007) Comparative genomic analysis of three *Leishmania* species that cause diverse human disease. *Nature Genetics*. **39(7):** 839-847.

Pelham, H.R.B. (1984) Hsp70 accelerates the recovery of nucleolar morphology after heat shock. *European Molecular Biology Organisation (EMBO) Journal*. **3**: 3095-3100.

Pellechia, M., Montgomery, D.L., Stevens, S.Y., Vander Kooi, C.W., Feng, H. and Gierasch L.M. et al., (2000) Structural insights into substrate binding by the molecular chaperone DnaK. *Nature Structural Biology*. **7:** 586-593.

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

Pena-Diaz, J., Montalvetti, A., Flores, C-L., Constan, A., Hurtado-Guerrero, R., De Souza, W., Gancedo, C., Ruiz-Perez, L.M. and Gonzalez-Pcanowska, D. (2004) Mitochondrial Localisation of the mevalonate pathway Enzyme 3-Hydroxy-3_methyl-glutaryl-CoA reductase in the trypanosomatidae. *Molecular Biology of the Cell*. **15:** 1356-1363.

Peremyslov, V.V., Hagiwara, Y. and Dolja, V.V. (1999) Hsp70 homolog functions in cell to cell movement of a plant virus. *Proceedings of the National Academy of Science.* **96(26)**: 14771-14776.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C. and Ferrin, T.E. (2004)"UCSF Chimera - A Visualization System for Exploratory Research and Analysis." *Journal of Computational Chemistry.* **25(13):**1605-1612.

Pfeiffer,I., Seantier, B., Petronus, S., Sutherland, D., Kasemo, B and Zäch, M. (2008) Influence of nanotopography on phospholipids bilayer formation on silicon dioxide. *The Journal of Physical Chemistry.* **112(16):** 5175-5181.

Phillips, B., Abravaya, K. and Morimoto, R.I. (1991) Analysis of the specificity and mechanism of transcriptional activation of the human Hsp70 gene during infection by DNA viruses. *Journal of Virology.* **65**: 5680-5692.

Pike. R.N., Bottomley, S.P., Irving, J.A., Bird, P.I. and Whisstock, J.C. (2002) Serpins: Finely balanced conformational traps. *IUBMB life* **54**:1-7.

Pirkkala, I., Nykanen, P. and Sistonen, I. (2001) Roles of the heat shock transcription factors in regulation of the heat shock response and beyond. *The FASEB Journal.* **15:** 1118-1131.

Ploubidou, A., Robinson, D.R., Docherty, R.C., Ogbadoyi, E.O. and Gull, K. (1999) Evidence for novel cell cycle checkpoints in trypanosomes: Kinetoplast segregation and cytokinesis in the absence of mitosis. *Journal of Cell Science.* **112:** 4641-4650.

Prahlad, V. and Morimoto, R.I. (2008) Integrating the stress response: lessons for neurodegenerative diseases from C. elegans. *Trends in Cell Biology.* **19(2):** 52-61.

Prip-Buus, C., Westerman, B., Schmitt, M., Langer, T., Neupert, W. and Schwarz, E. (1996) Role of mitochondrial DnaJ homologue, Mdj1p, in the prevention of heat-induced protein aggregation. *FEBS letters.* **380(1-2):** 142-146.

Qian, X., Hou, W., Zhengang, L. and Sha, B, (2002) Dinrect interactions between molecular chaperones Heat Shock protein (Hsp) 70 and Hsp40: yeast Hsp70 Ssa1 binds the extreme C-terminal region of yeast Hsp40 Sis1. *Biochemical Journal* **361:** 27-34.

Qian, Y.Q., Patel, D., Hartl, F-U. and McColl, D.J. (1996) Nucleic magnetic resonance solution structure of the human Hsp40 (Hdj-1) J-domain. *Journal of Molecular Biology.* **260:** 224-235.

Qiu, X.-B., Shao, Y.-M., Miao, S. and Wang, L. (2006) The Diversity of the DnaJ/Hsp40 family, the crucial partners for Hsp70 chaperones. *Cellular and Molecular Life Sciences* **63(22)**:2560-2570.

Radford, S.E. (2000) Protein folding: progress made and promises ahead. *Trends in Biochemical Science.* **25(no.12)**:611-618.

Radford, N.B., Fina, M., Benjamin, I.J., Moreadith, R.W., Graves, K.H., Zhao, P., Gavva, S., Wiethoff, A., Sherry, A.D., Malloy, C.R. and Williams, R.S. (1996) Cardioprotective effects of 70-kDa heat shock protein in transgenic mice. *Proceedings of the National Academy of Science, USA.* **93 (6)**: 2339-2342.

Ralston K.S., Lerner, A.G. Diener, D.R. and Hill, K.L. (2006) Flagellar motility contributes to cytokinesis in *Trypanosoma brucei* and is modulated by an evolutionarily conserved dynein regulatory system. *Eucaryotic cell.* **5:** 696-711.

Ramos, C.H., Oliviera, C.L. Yang-Fan, C., Torriani, I.L. and Cyr, D.M. (2008) Conserved central domains control the quaternary structure of type I and type II Hsp40 molecular chaperones. *Journal of Molecular Biology.* **383:** 155-166.

Redmond, S., Vadivelu, J., and Field, M.C. (2003) RNAit: an automated web-based tool for the selection of RNAi targets in *Trypanosoma brucei. Molecular & Biochemical Parasitology.* **128:** 115 – 118.

Reid, T.S., Terry, K.L. Casey, P.J. and Beese, L.S. (2004) Crystallographic analysis of CaaX prenyltransferases complexed with substrates defines rules of protein substrate selectivity. *Journal of Molecular Biology.* **343:** 417-433.

Reimhult, E., Zäch, M., Höök, F. and Kasemo, B. (2006) A multitechnique study of liposome adsorption on Au and lipid Bilayer formation on SiO_2. *Langmuir.* **22:** 3313-3319.

Reimhult, E., Larsson, C., Kasemo, B. and Höök, F. (2004) Simultaneous surface plasmon resonance and quartz crystal microbalance with dissipation monitoring measurements of biomolecular adsorption events involving structural transformations and variations in coupled water. *Analytical Chemistry.* **76:** 7211-7220.

Reimhult, E., Höök, F. and Kasemo, B. (2003) Intact vesicle adsorption and supported biomembrane formation from vesicles in solution: influence of surface chemistry, vesicle size, temperature, and osmotic pressure. *Langmuir.* **19(5):** 1681-1691.

Reimhult, E., Höök, F. and Kasemo, B. (2002) Temperature dependence of formation of a supported phospholipids bilayer from vesicles on SiO_2. *Physical Review (The American Physical Society).* **66:** 051905-1 – 051905-4.

Requena, J.M., Jimenez-Ruiz, A., Soto, M., Assiego, R., Santaren, J.F., Lopez, M.C., Patarroyo, M.E. and Alonso, C. (1992) Regulation of hsp70 expression in *Trypanosoma cruzi* by temperature and growth phase. *Molecular and Biochemical Parasitology.* **53:** 201-212.

Reuner, B., Vasella, E., Yutzy, B. and Boshart, M. (1997) Cell density triggers slender to stumpy differentiation of Trypanosoma brucei bloodstream forms in culture. *Molecular and Biochemical Parasitology.* **90:** 269-280.

Reviakine, I. and Brisson, A., (2001) Streptavidin 2D crystals on supported phospholipid bilayers: Toward constructing anchored phospholipid bilayers. *Langmuir.* **17:** 8293-8299.

Revington, M., Zhang, Y., Yip, G.N.P., Kurochkin, A.V. and Zuiderweg, E.R.P. (2005) NMR investigations of allosteric porcessin in a two-domain Thermus Thermophilus Hsp70 molecular chaperone. *Journal of Molecular Biology.* **27:** 163-183.

Rial, D.V. and Ceccarelli, E.A. (2002) Removal of DnaK contamination during fusion protein purifications. *Protein Expression and Purification.* **25:** 503-507.

Richarme, G. and Kohiyama, M. (1993) Specificity of the Escherichia coli chaperone DnaK (70kDa heat shock protein) for hydrophobic amino acids. *The Journal of Biological Chemistry.* **268**: 24074-24077.

Richter, R., Mukhopadhyay, A. and Brisson, A. (2003) Pathways of lipid vesicle deposition on solid surfaces: a combined QCM-D and AFM study. *Biophysical Journal.* **85:** 3035-3047.

Rist, W., Graf, C., Bukau, B. and Mayer (2006) Amide hydrogen exchange reveals conformational changes in Hsp70 chaperones important for allosteric regulation. *The Journal of Biological Chemistry.* **281:** 16493-16501.

Ritossa, F. (1996) Discovery of the heat shock response. *Cell Stress and Chaperones* **1(2)**: 97-98.

Ritossa, F. (1962) A new puffing pattern induced by temperature shock and DNP in Drosophila. Experentia **18**: 571-573

Robinson, K.A. and Beverly, S.M. (2003) Improvements in transfection efficiency and tests of RNA interference (RNAi) approaches in the protozoan parasite *Leishmania*. *Molecular and biochemical Parasitology*. **128:** 217-228.

Robinson, D.R. and Gull, K., (1991) Basal Body movements as a mechanism for mitochondrial genome segregation in the trypanosome cell cycle. *Nature*. **352:** 731-733.

Rodahl, M., Höök, F., Fredriksson, C., Keller, C.A., Krozer, A., Brzezinski, P., Voinova, M. and Kasemo, B. (1997) Simultaneous frequency and dissipation factor QCM measurements of biomolecular adsorption and cell adhesion. *Faraday discussions*. **107:** 229-246.

Rodahl, M., Höök, F. and Kasemo, B. (1996) QCM operation in liquids: An explanation of measured variations in frequency and Q factor with liquid conductivity. *Analytical Chemistry*. **68:** 2219-2227.

Rodahl, M., Höök, F., Krozer, A., Brzezinski, P. and Kasemo, B. (1995) Quartz crystal microbalance setup for frequency and Q-factor measurements in gaseous and liquid environments. *Review Scientific Instruments*. **66(7):** 3924-3930.

Roditi, I. and Liniger, M. (2002) Dressed for success: the surface coats of insect-borne protozoan parasites. *Trends in Micobiology*. **10:** 128-134.

Roskoski, R. and Ritchie, P. (1998) Role of the carboxyterminal residue in peptide binding to protein farnesyltransferase and protein geranylgeranyltransferase. *Archives of Biochemistry and Biophysics*. **356:** 167-176.

Rosenzweig, D., Smith, D., Opperdoes, F., Stern, S., Olafson, R.W. and Zilbertstein, D. (2008) Retooling *Leishmania* metabolism: From sand fly gut to human macrophage. *The FASEB Journal*. **22(2):** 590-602.

Ruddon, R.W. and Bedows, E. (1997) Assisted Protein Folding. *The Journal of Biological Chemistry*. **272(6)**: 3125-3128.

Rüdiger, S., Schneider-Mergener, J., and Bukau, B. (2001) Its substrate specificity characterizes the DnaJ chaperone as scanningfactor for the DnaK chaperone. *European Molecular Biology Organisation (EMBO) Journal*. **20**: 1-9.

Rüdiger, S., Germeroth, L., Schneider-Mergener, J., and Bukau, B. (1997) Substrate-specificity of the DnaK chaperone determined by screeining cellulose-bound peptide libraries. *European Molecular Biology Organisation (EMBO) Journal*. **16**: 1501-1507.

Rusconi, F., Durand-Dubief, M. and Bastin, P. (2005) Functional complementation of RNA intereference mutants in trypanosomes. *BiomedCentral (BMC) Biotechnology*. **5(6):** 1-13.

Sahi, C. and Craig, E.A. (2007) Network of general and specialty J protein chaperones of the yeast cytosol. *Proceedings of the National Academy of Sciences USA*. **104(17):** 7163-7168.

Šali, A. and Blundell, T. L.. (1993) Comparative protein modelling by satisfaction of spatial restraints. *Journal of Molecular Biology*. **234:** 779-815.

Salmon, D., Montero-Lomeli, M. and Goldenberg, S., (2001) A DnaJ-Like protein homologous to the yeast co-chaperone Sis1 (Tcj6p) is involved in initiation of translation in *Trypanosoma cruzi*. The *Journal of Biological Chemistry*. **276:** 43970-43979.

Santoro, M.G. (2000) Heat shock factors and the Control of the Stress Response. *Biochemical Pharmacology* **59**:55-63.

Sauerbrey, G. (1959) "The use of quartz oscillators for weighing thin layers and for microweighing," Z. Phys. *155: 206-222*. **(Sauerbrey, G.** (1959) Verwendung von Schwingquarzen zur Wägung dünner Schichten und zur Mikrowägung. *Zeitschrift für Physik*.**155:**206-222.

Sayers, E.W., Barrett, T., Benson, D.A., Bryant, S.H., Canese, K., Chetvernin, V., Church, D.M., DiCuccio, M., Edgar, R., Federhen, S., Feolo, M., Geer, L.Y., Helmberg, W., Kapustin, Y., Landsman, D., Lipman, D.J., Madden, T.L., Maglott, D.R., Miller, V., Mizrachi, I., Ostell, J., Pruitt, K.D., Schuler, G.D., Sequeira, E., Sherry, S.T., Shumway, M., Sirotkin, K., Souvorov, A., Starchenko, G., Tatusova, T.A., Wagner, L., Yaschenko, E. and Ye, J. (2008) Database resources of the national center for biotechnology information. *Nucleic Acids Research*. **37:** D5-D15.

Schlenstedt, G., Harris, S., Risse, B., Lill, R. and Silver, P.A. (1995) A yeast DnaJ homologue, Scj1p, can function in the endoplasmic reticulum with BiP/Kar2p via a conserved domain that specifies interactions with Hsp70s. *The Journal of Cell Biology*. **129:** 979-988.

Schönfeld, H-J., Schmid, D., Schröder, H. and Bukau, B. (1995) The DnaK chaperone system of *Escherichia coli*: quaternary structures and interactions of the DnaK and GrpE components. *The Journal of Biological Chemistry*. **270:** 2183-2189.

Schröder, H., Langer, T., Hartl, F.U. and Bukau, B. (1993) DnaK, DnaJ and GrpE form a cellular chaperone machinery capable of repairing heat induced protein damage. *European Molecular Biology Organisation (EMBO) Journal*. **12(11):** 4137-4144.

Schubert, U., Anton, L.C., Gibbs, J., Norbury, C.C., Yewdell, J.W. and Bennink, J.R. (2000) Rapid degradation of a large fraction of newly synthesized proteins by proteasomes. *Nature*. **404:** 770-774.

Schultz, S.R. and Nass, S. (1969) DNA-nucleotidyl transferase activity in supernatant and membrane fractions from normal and regenerating rat liver mitochondria. *FEBS Letters*. **4:** 13-15.

Schmid, D., Baici, A., Gehring, H. and Christen, P. (1994) Kinetics of molecular chaperone action. *Science*. **263:** 971-973.

Seantier, B., Breffa, C., Felix, O. and Decher, G. (2005) Dissipation-enhanced quartz crystal microbalance studies on the experimental parameters controlling the formation of supported lipid bilayers. *The Journal of Physical Chemistry*. **109:** 21755-21765.

Sha, B.D., Lee, S. and Cyr, D. (2000) The crystal structure of the peptide-binding fragment from the yeast Hsp40 protein Sis1. *Structure*. **8:** 799-807.

Sheader, K., Vaughan, S., Minchin, J., Hughes, K., Gull, K. and Rudenko, G. (2005) Variant surface glycoprotein RNA interference triggers a precytokinesis cell cycle arrest in African trypanosomes. *Proceedings of the National Academy of Sciences USA*. **102:** 8716-8721.

Shi, YY., Hong, X.G. and Wang, C.C. (2005) The C-terminal (331-376) sequence of *Escherichia coli* DnaJ is essential for the dimerisation and chaperone activity: A small angle X-ray scattering study in solution. *The Journal of Biological Chemistry*. **280(24):**22761-22768.

Shi, J., Blundell, T.L. and Mizuguchi, K. (2001) Fugue: Sequence-structure homology recognition using environment-specific substitution tables and structure dependent gap penalties. *Journal of Molecular Biology.* **310:** 243-257.

Siegel, T.N., Hekstra, D.R. and Cross, G.A.M. (2008) Analysis of the *Trypanosoma brucei* cell cycle by quantitative DAPI imaging. *Molecular and Biochemical Parasitology.* **160:** 171-174.

Sigrist, C.J.A., Cerutti, L., Hulo, N., Gattiker, A., Falquet, L., Pagni, M., Bairoch, A. and Bucher, P. (2002) *PROSITE: a documented database using patterns and profiles as motif descriptors. Briefings in Bioinformatics.* **3:**265-274.

Singer, M.A. and Lindquist, S. (1998) Multiple effects of Trehalose on protein folding *in vitro* and *in vivo. Molecular cell* **1**: 639-648.

Simpson, A.G.B., Lukes, J. and Roger, A.J. (2002) The Evolutionary History of Kinetoplastids and Their Kinetoplasts. *Molecular Biology and Evolution.* **19(12)**: 2071-2083.

Simpson, A.G.B., Stevens, J.R. and Lukes, J. (2006) The evolution and diversity of kinetoplastid flagellates. *TRENDS in Parasitology.* **22 (4)**: 168-174.

Slepenkov, S.V. and Witt, S.N. (2002) Kinetic analysis of interdomain coupling in a lidless variant of the molecular chaperone DnaK: DnaK's lid inhibits transition to the low affinity state. *Biochemistry.* **41:** 12224-12235.

Sondheimer, N., Lopez, N., Craig, E.A. and Lindquist, S. (2001) The role of Sis1 in the maintenance of the [RNQ+] prion. *European Molecular Biology Organisation (EMBO) Journal.* **20(10):** 2435-2442.

Sprenger, J., Fink, J.L. and Teasdale, R.D. (2006) Evaluation and comparison of mammalian subcellular localization prediction methods. *BMC Bioinformatics.* **7(suppl 5):** S3.

Su, X., Wu, Y-J. and Knoll, W. (2005) Comparison of surface plasmon resonance spectroscopy and quartz crystal microbalance techniques for studying DNA assembly and hybridization. *Biosensors and Bioelectronics.* **21:** 719-726.

Subramaniam, C., Veazey, P., Redmond, S., Hayes-Sinclair, J., Chambers, E., Carrington, M., Gull, K., Matthews, K., Horn, D. and Field, M.C. (2006) Chromosome-wide analysis of gene function by RNA interference in the african trypanosome. *Eukaryotic Cell.* **5(9):** 1539-1549.

Stirling, P.C., Lundlin, V.F. and Leroux, M.R. (2003) Getting a grip on non-native proteins. *European Molecular Biology Organisation (EMBO) Journal* **4(6)**: 565-570.

Stubdal, H., Zalvide, J., Campbell, K.S., Scheitser, C., Roberts, T.M. and DeCaprio, J.A. (1997) Inactivation of pRB-related proteins p130 and p107 mediated by the J-domain of simian virus 40 large T antigen. *Molecular and Cellular Biology.* **17:** 4979-4990.

Subramaniam, C., Veazey, P., Redmond, S., Hayes-Sinclair, J., Chambers, E., Carrington, M., Gull, K., Mathews, K., Horn, D. and Field, M.C. (2006) Chromosome-wide analysis of Gene Function by RNA Interferance in the African Trypanosome. *Eukaryotic Cell.* **5(9):** 1539-1549.

Sugimoto, S., Higashi, C., Saruwatari, K., Nakayama, J. and Sonomoto, K. (2007) A gram-negative characteristic segment in *Escherichia coli* DnaK is essential for the ATP-dependent cooperative function with the co-chaperones DnaJ and GrpE. *FEBS Letters.* **581:** 2993-2999.

Suh, W-C., Lu, C. Z. and Gross, C.A. (1999) Structural features required for the interaction of the Hsp70 molecular chaperone DnaK with its cochaperone DnaJ. *The Journal of Biological Chemistry.* **274(43):** 30534-30539.

Sullivan, C.S., Tremblay, J.d., Fewell, S.A., Lewis, J.A., Brodsky, J.L. and Pipas, J.M. (2000) Species-specific elements in ht elarge T-antigen J-domain are required for cellular transformation and DNA replication by simian virus 40. *Molecular and Cellular Biology.* **20:** 5749-5757.

Sullivan, M.A., Olson, C.L., Winquist, A.G. and Engman, D.M. (1994) Expression and localization of *Trypanosoma cruzi* Hsp60. *Molecular and Biochemical Parasitology.* **68:** 197-208.

Summers, D.W., Douglas, P.M., Ren, H-Y and Cyr, D.M. (2009) The Type I Hsp40 Ydj1 utilises a farnesyl moiety and zinc finger-like region to suppress prion toxicity. *The Journal of Biological Chemistry.* **284(6):** 3628-3639.

Svedhem, S., Dahlborg, D., Ekeroth, J., Kelly, J., Höök, F and Gold, J. (2003) In situ peptide- modified supported lipid bilayers for controlled cell attachment. *Langmuir.* **19:** 6730-6736.

Swain, J., Dinler, G., Sivendran, R, Montgomery, D., Stotz, M. and Gierasch, L. (2007) Hsp70 chaperone ligands control domain association via an allosteric mechanism mediated bu the interdomain linker. *Molecular Cell.* **26:** 27-39.

Szabo, A., Langer, T., Schroder, H., H., Flanagan, J., Bukau, B. and Hartl, F.-U. (1994) The ATP hydrolysis-dependent reaction cycle of the *Escherichia coli* Hsp70 system DnaK, DnaJ and GrpE. *Proceedings of the National Academy of Sciences USA.* **91(22):**10345-10349.

Szyperski, T., Pellechia, M., Wall, D., Georgopoulos, C. and Wuthrich, K. (1994) NMR structure determination of the Escherichia coli DnaJ molecular chaperone: Secondary structure and backbone fold of the N-terminal region (residues 2-108) containing the highly conserved J-domain. *Proceedings of the National Academy of Sciences.* **91:** 11343-11347.

Teasdale, R.D. and Jackson, M.R. (1996) Signal-mediated sorting of membrane proteins between the endoplasmic reticulum and the golgi apparatus. *Annual Review of Cell and Developmental Biology.* **12:** 27-54.

Teixeira, A.R.L., Nitz, N., Guimaro, M.C., Gomes, C., and Santos-Buch, C.A. (2006) Chagas Disease. *Postgraduate Medical Journal.* **82:** 788-798.

Terada, K., Kanazawa, M., Bukau, B. and Mori, M. (1997) The human DnaJ homologue dj2 facilitates mitochondrial protein import and luciferase refolding. *Journal of Cell Biology.* **139(5):** 1089-1095.

Theyssen, H., Schuster, H-P., Packschies, L., Bukau, B. and Reinstein, J. (1996) The second step of ATP binding to DnaK induces peptide release. *Journal of Molecular Biology.* **263 (5):** 657-670.

Thomas, P.J., Bao-he, Q. and Pederson, P.L. (1995) Defective protein folding as a basis of disease. *Trends in Biochemical Sciences* **20:** 456-459.

Tibbetts, R.S., Jensen, J.L., Olson, C.L., Wang, F.D. and Engman, D.M. (1998) The DnaJ family of protein chaperones in *Trypanosoma cruzi. Molecular and Biochemical Parasitology.* **91:** 319-326.

Towbin, H., Staehelin, T. and Gordon, J. (1979) Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: Procedure and some applications. *Proceedings of the National Academy of Sciences, USA.* **76:** 4350-4354.

Tsai, J. 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. *The Journal of Biological Chemistry.* **271(16):** 9347-9354.

Tyler, K.M. and Engman, D.M. (2001) The Life cycle of *Trypanosoma cruzi* revisited. *International Journal of Parasitology.* **31:** 472-481.

Ullu, E., Tschudi, C. and Chakraborty, T., (2004) RNA interference in protozoan parasites. *Cellular Microbiology.* 6 :509-519.

Vanhamme, L. and Pays, E. (2004) The trypanosome lytic factor of human serum and the molecular basis of sleeping sickness. *International Journal of Parasitology.* **34(8):** 887-898.

Van Montfort, R.L.M., Basha, E., Friedrich, K.L., Slingsby,, C. and Vierling, E. (2001a) Crystal Structure and assembly of a eukaryotic small heat shock protein. *Nature Structural Biology* 7:1172-1177.

Van Montfort, R., Slingsby, C. and Vierling, E. (2001b) Structure and function of the small heat shock protein/alpha-crystallin family of molecular chaperones. *Advances in Protein Chemistry.* **59:** 105-156

Vasella, E., Reuner, B., Yutzy, B. and Boshart, M. (1997) Differentiation of African trypanosomes is controlled by a density sensing mechanism which signals cell cycle arrest via the cAMP pathway. *Journal of Cell Science.* **110:** 2661-2671.

Vasella, E. and Boshart, M. (1996) High molecular mass agarose matrix supports growth of bloodstream forms of pleomorphic *Trypanosoma brucei* strains in axenic culture. *Molecular and Biochemical Parasitology.* **82:** 91-105.

Vertommen, D., Van Roy, J., Szikora, J.P., Rider, M.H., Michels, P.A. and Opperdoes, F.R. (2008) Differential expression of glycosomal and mitochondrial proteins in the two major life-cycle stages of *Trypanosoma brucei. Molecular and Biochemical Parasitology.* **158(2):** 189-201.

Vogel, V., Schief, W.R. and Frey, W. (1997) Dynamics of two-dimensional protein crystallization at the air/water interface: Streptavidin targeted to surfaces via high-affinity binding or metal coordination. *Supramolecular Science.* 4: 163-171.

Voinova, M.V., Rodahl, M., Jonson, M. and Kasemo, B. (1999) Viscoelastic acoustic response of layered polymer films at fluid-solid interfaces: continuum mechanics. *Physica Scripta.* **59:** 391-396.

Wacker, J.L., Zareie, M.H., Fong, H., Sarikaya, M. and Muchowski, P.J. (2004) Hsp70 and Hsp40 attenuate formation of spherical and annular polyglutamine oligomers by partitioning monomer. *Nature Structural Molecular Biology.* **11(12):**1215-1222.

Walsh, P., Bursac, D., Chern Law, Y., Cyr, D. and Lithgow, T., (2004) The J-Protein family: modulating protein assembly, disassembly and translocation. *European Molecular Biology Organisation (EMBO) reports*: **5 (6):** 567- 571.

Wall, D., Zylicz, M. and Georgopoulos, C. (1995) The conserved G/F motif of the DnaJ chaperone is necessary for the activation of the substrate binding properties of the DnaK chaperone. *The Journal of Biological Chemistry.* 270: 2139-2144.

Wall, D, Zylicz, M. and Georgopoulos, C. (1994) The NH2-terminal 108 amino acids of the *Escherichia coli* DnaJ protein stimulate the ATPase activity of DnaK and are sufficient for λ replications. *The Journal of Biological Chemistry.* **270:** 2139-2144.

Wang, Z., Morris, J.C., Drew, M.E. and Englund, P.T. (2000) Inhibition of *Trypanosoma brucei* gene expression by RNA interference using an integratable vector with opposing T7 promoters. *The Journal of Biological Chemistry.* **275(51):** 40174-40179.

Ward, M.D. and Buttry, D.A. (1990) In situ interfacial mass detection with piezoelectric transducers. *Science.* **249(4972):** 1000-1007.

Westerheide, S.D. and Morimoto, R.I. (2005) Heat shock response modulators as therapeutic tools for diseases of protein conformation. *The Journal of Biological Chemistry.* **280(39):** 33097-33100.

Wickner, S., Hoskins, J. and McKenney, K. (1991) Function of DnaJ and DnaK as chaperones in origin-specific DNA binding by RepA. *Nature.* **350(6314):** 165-167.

Wild, J.E., Altman, T., Yura, T. and Gross, C.A. (1992) DnaK and DnaJ heat shock proteins articipate in protein export in *Escherichia coli. Genes and Development* **6:** 1165-1172.

Wirtz, E., Leal, S., Ochatt, C. and Cross, G.A. (1999) A tightly regulated inducible system for the conditional gene knock-outs and dominant-negative genetics in *Trypanosoma brucei. Molecular and Biochemical Parasitology* **99(1):** 89-101.

Wolynes, P.G., Onichic, J.N. and Thirumalai, D. (1995) Navigating the folding routes. *Science* **267**: 1619-1623.

Wong, J.W., Albright, R.L., Wang and N.H.L. (1991) Immobilised metal-ion affinity-chromatography (Imac) chemistry and bioseparation applications. *Separation and Purification Methods.* **20:** 49-106.

Woodward, R. and Gull, K. (1990) Timing of nuclear and kinetoplast DNA replication and early morphological events in the cell cycle of *Trypansoma brucei. Journal of Cell Science.* **95:** 49-57.

Wu, Y., Li, J., Jin, Z., Fu, Z. and Sha, B. (2005) The crystal structure of the C-terminal fragment of yeast Hsp40 Ydj1 reveals novel dimerisation motif for Hsp40. *Journal of Molecular Biology.* **346:** 1005-1011.

Wu, Y.S., Bevilaqua, V.L. and Berg J.M. (1995) Fibrillin domain folding and calcium binding: significance to Marfan syndrome. *Chemical Biology* **2**: 91-97.

Yan, W., Gale, M.J.J., Tan, S.-L. and Katze, M.G. (2002) Inactivation of the PKR protein kinase and stimulation of mRNA translation by the cellular co-chaperone P58 IPK does not require J-domain function. *Biochemistry.* **41:** 4938-4945.

Yokoyama, K., Trobridge, P., Buckner, F.S., Scholten, J., Stuart, K.D., Van Voorhis, W.C. and Gelb, M.H. (1998) The effects of protein farnesyltransferase inhibitors on trypanosomatids: inhibition of protein farnesylation and cell growth. *Molecular and Biochemical Parasitology.* **98:** 87-97.

Yokoyama, K., Lin, Ying, Stuart, K.D. and Gelbb, M.H. (1997) Prenylation of proteins in *Trypanosoma brucei. Molecular and Biochemical Parasitology.* **87:** 61-69.

Yoshida, N. (2006) Molecular basis of mammalian cell invasion by *Trypanosoma cruzi. Annals of the Brazilian Academy of Sciences.* **78 (1):** 87-111.

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.

Zalvide, J., Stubdal, H. and DeCaprio, J.A. (1998) The J-domain of simian virus 40 large T antigen is required to functionally inactivate RB family proteins. *Molecular Cell Biology.* **18:** 1408-1415.

Zhdanov, V.P., Keller, C.A., Gläsmaster, K. and Kasemo, B. (2000) Simulation of adsorption kinetics of lipid vesicles. *The Journal of Chemical Physics*. **112(2):** 900-909.

Zhang, Y. and Zuiderweg, E.R. (2004) The 70-kDa heat shock protein chaperone nucleotide-binding domain in solution unveiled as a molecular machine that can reorient its functional subdomains. *Proceedings of the National Academy of Science, USA*. **101**: 10272-10277.

Zhu, H-Y., Ling, K-S., Gosczcynski, D.E., Mcferson, J.R. and Gonsalves, D. (1998) Nucleotide sequence and genome organization of grapevine leafroll-associated virus-2 are similar to beet yellows virus, the closterovirus type member. *Journal of General Virology*. 79: 1289-1298.

Zhu, X., Zhao, X., Burkholder, W.F., Gragerov, A., Ogata, C.M., Gottesman, M. and Hendrickson, W.A. (1996) Structural analysis of substrate binding by the molecular chaperone DnaK. *Science* **272**:1606-1614.

Zwanzig, R., Szabo, A. and Bagchi, B. (1992) Levinthal's Paradox. *Proceedings of the National Academy of Sciences, USA*. **89**: 20-22.

Zylicz, M., King, F.W. and Wawrzynow, A. (2001) The Hsp70 interactions with the p53 tumour suppressor protein. *European Molecular Biology Organisation(EMBO) Journal*. **20**:4634-4638.