Characterisation of Trypanosomal Type III and Type IV Hsp40 proteins

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

DOCTOR OF PHILOSOPHY (Science)

of

RHODES UNIVERSITY

By

CASSANDRA ALEXANDROVNA LOUW

May 2009

Abstract

The heat shock protein-70 (Hsp70) family of molecular chaperones are ubiquitous highly conserved proteins that are critical for the viability of cellular homeostasis. The ATPase activity of Hsp70 proteins is critical to their function as the affinity of a given Hsp70 for non-native substrate is modulated by ATP binding and hydrolysis. When bound to ATP, Hsp70s possess a low affinity for a given substrate protein, while the hydrolysis of ATP to ADP causes a conformational change that results in a high affinity for substrate proteins. The basal ATPase activity of Hsp70s is too low to facilitate their function *in vivo*, and co-chaperones are essential to modulate the efficient protein folding by Hsp70. Heat shock protein-40 (Hsp40) heat shock proteins are essential for the *in vivo* function of Hsp70s by stimulating the ATPase activity of these proteins and facilitating transfer of substrates. The Type III class of Hsp40 proteins have not been well characterised due to their poor levels of conservation at the primary sequence level. This is due to the fact that Type III Hsp40s only contain a J-domain and a poorly conserved C-terminal region. The newly identified Type IV class of Hsp40s, contain an abrogated HPD tripeptide motif in the J-domain and have also not been extensively studied.

Trypanosoma brucei (*T. brucei*) is a unicellular flagellated protozoan parasite. It is the causative agent of Human African Trypansomiasis (HAT) which results in thousands of deaths and devastating agricultural losses in many parts of Africa. *T. brucei* undergoes a complex lifecycle that is characterised by the transition from an insect vector to a mammalian host in markedly different conditions of temperature, pH, nutrient availability and respiratory requirements. It has been proposed that molecular chaperones may enhance the survival of these parasites due to their cytoprotective effect in combating cellular stress. Due to the fact that *T. brucei* infection is invariably fatal if left untreated, and that no novel treatment regimens have been developed recently, the identification of potential novel drug targets among proteins essential to the parasite's survival in the host organism is an attractive aspect of *T. brucei* research. Because Type III Hsp40s are poorly conserved with respect to Hsp40s found in the human host, the identification of any of these proteins found to be essential to *T. brucei* survival in humans could

potentially make attractive novel drug targets. An in depth *in silico* investigation into the Type III Hsp40 complement as well as partner Hsp70 proteins in *T.brucei* was performed. *T. brucei* possesses 65 Hsp40 proteins, of which 47 were classed as Type III and 6 of which were identified as being putative Type IV Hsp40s. A small but significant number (5) of Type III TbHsp40s contained tetratricopeptide (TPR) domains in addition to the J-domain. The J-domains of the Type III TbHsp40 complement were found to be conserved with respect to those of canonical Hsp40 proteins, although the mutation of certain residues that play a key role in Hsp40-Hsp70 interaction was noted. Potential partnerships of these proteins in the parasite was also investigated.

The coding regions of three previously uncharacterised TbHsp40s were successfully amplified from *T. brucei TREU927* genomic DNA and cloned into an expression vector. Tbj1, a Tcj1 ortholog, was selected for further study and successfully expressed and biochemically characterised. Tbj1 expressed in *E. coli* was found to be insoluble, but large amounts were recovered with the aid of a denaturing purification followed by refolding elution strategies, and the bulk of the protein recovered was in compact monomeric form as determined by size-exclusion chromatography fast protein liquid chromatography (SEC-FPLC). The addition of Tbj1 to a thermally aggregated substrate resulted in increased levels of aggregation, although Tbj1 was able to assist two Hsp70 proteins in the suppression of aggregation. Tbj1 proved unable to stimulate the ATPase activity of these same Hsp70s, and could not rescue temperature sensitive cells when replacing *E.coli* DnaJ and CbpA. It was concluded that Tbj1 does not possess independent chaperone activity, but could display Hsp40 co-chaperone properties under certain circumstances. This could allude to a specialised function in the *T. brucei* parasite. The lack of human orthologues to Tbj1 could result in the attractiveness of this protein as a novel drug target.

Keywords: *T. brucei*, Hsp40, Hsp70, aggregation suppression, ATPase

Declaration

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

_____

Cassandra A. Louw

4 May 2009

Grahamstown

Republic of South Africa

Table of Contents

List of Figures

List of Tables

Symbols and codes used

List of symbols

Symbol	Meaning
%	Percent
µ	Micro
α	Alpha
β	Beta
λ	Lambda
µl	Microlitre
µM	Micromolar
A360	Absorbance at 360 nm
A580	Absorbance at 580 nm
A595	Absorbance at 595 nm
nmol	Nanomole (nM)
pmol	Picomole (pM)
°C	Degree Celsius
ml	Millilitre
w/v	Weight per volume
v/v	Volume per volume

IUPAC amino acid single letter codes (SLC) and corresponding DNA codons

Amino acid	Single letter Code	Codons
Isoleucine	I	ATT, ATC, ATA
Leucine	L	CTT, CTC, CTA, CTG, TTA, TTG
Valine	V	GTT, GTC, GTA, GTG
Phenylalanine	F	TTT, TTC
Methionine	M	ATG
Cysteine	C	TGT, TGC
Alanine	A	GCT, GCC, GCA, GCG
Glycine	G	GGT, GGC, GGA, GGG
Proline	P	CCT, CCC, CCA, CCG
Threonine	T	ACT, ACC, ACA, ACG
Serine	S	TCT, TCC, TCA, TCG, AGT, AGC
Tyrosine	Y	TAT, TAC
Tryptophan	W	TGG
Glutamine	Q	CAA, CAG
Asparagine	N	AAT, AAC
Histidine	H	CAT, CAC
Glutamic acid	E	GAA, GAG
Aspartic acid	D	GAT, GAC
Lysine	K	AAA, AAG
Arginine	R	CGT, CGC, CGA, CGG, AGA, AGG
Stop codons	STOP	TAA, TGA, TAG

List of abbreviations

Abbreviation	Meaning
Agt	*Agrobacterium tumefaciens*
A	Absorbance
ADP	Adenosine diphosphate
Amp	Ampicillin
AmpR	Ampicillin resistance
APS	Ammonium persulphate
ATP	Adenosine triphosphate
ATPase	Adenosine triphosphatase
BBB	Blood-brain barrier
BLAST	Basic local alignment search tool
BiP	Binding protein
BSA	Bovine serum albumin
BSF	*T. brucei* bloodstream form
CCC	Chronic chagasic cardiomyopathy
CbpA	Curved DNA binding protein A
Cdc37	Cell division cycle 37 homolog
DAPase I	Dipeptidyl aminopeptidase I
DjIa	DnaJ-like A
2DE	Two dimensional gel electrophoresis
DNA	Deoxyribonucleic acid
DnaJ	Prokaryotic Hsp40
DnaK	Prokaryotic Hsp70
DTT	Dithiothreitol
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
E.R.	Endoplasmic reticulum
FPLC	Fast protein liquid chromatography
g	Gram
G / F –rich region	Glycine-phenylalanine rich region
GPI	Glycosylphosphatidylinositol
Grp94	Glucose-regulated protein 94
HAT	Human African Trypanosomiasis
Hdj2	Human DnaJ-like protein 2
Hip	Hsp70-interacting protein
His_6 tag	6 x Histidine tag
HIV / AIDS	Human immunodeficiency virus / acquired immunodeficiency virus
HOP	Hsp70 / Hsp90 organising protein
HPD	Histidine, proline, aspartic acid motif
HRP	Horseradish peroxidase
Hsc	Constitutively expressed heat shock protein
Hsc70	Constitutive heat shock protein 70
Hsp	Heat shock protein
Hsp40	40 kDa heat shock protein
Hsp70	70 kDa heat shock protein

Hsp90	90 kDa heat shock protein
Kan	Kanamycin
kDa	Kilodalton
LIC	Ligation-independent cloning
L. major	*Leishmania major*
LmHsp40 / 70	*Leishmania major* Hsp40 / 70
MD	Middle domain
MDH	Malate dehydrogenase
MOPS	3-(N-morpholino) propanesulphonic acid
Mw	Molecular weight (kDa)
mRNA	Messenger ribose nucleic acid
M. sativa	*Medicago sativa*
MsHsp70	*Medicago sativa* Hsp70
NCBI	National Center for Biotechnology Information
NEB	New England Biolabs
nmol	Nanomole
NMR	Nuclear magnetic resonance
NTD	N-terminal regulatory domain
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
P. falciparum	*Plasmodium falciparum*
pI	Isoelectric point
Pi	Inorganic phosphate
Pro	*T. brucei* procyclic form
RNA	Ribose nucleic acid
SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
TBE	Tris-borate EDTA buffer
TE	Tris EDTA buffer
TPR	Tetratricopeptide
T. brucei	*Trypanosoma brucei*
T. b. brucei	*Trypanosoma brucei brucei*
T. b. gambiense	*Trypanosoma brucei gambiense*
T. b. rhodesiense	*Trypanosoma brucei rhodesiense*
TbHsp40 / 70	Hsp40 / 70 from *T. brucei*
TcHsp40 / 70	Hsp40 / 70 from *T. cruzi*
T. cruzi	*Trypanosoma cruzi*
T. evansi	*Trypanosoma evansi*
T. equiperdum	*Trypanosoma equiperdum*
tRNA	Transfer RNA
URL	Uniform resource locator
VSG	Variable surface glycoprotein
YT	Yeast-tryptone

Acknowledgements

The accomplishment of the present work has often been a journey shared by many. I would like to sincerely thank the following persons for the contributions that they have made during the past two and a half years:

— My supervisor, Prof. G.L. Blatch, for his patience, support, encouragement and for allowing me to steer the project in various directions, not all of which were planned at the outset. Also for financial assistance for the duration of this project

— Dr. David Horn (LSTMH) for supplying *T. brucei* genomic DNA at no cost

— Prof. Theresa Coetzer (UKZN) for kindly providing *T. brucei* procyclic and bloodstream form lysates at no cost

— Joan Miles, for always cheerfully assisting with a myriad of administrative matters

— Mike Ludewig for his friendship, encouragement and technical advice on all aspects of *T. brucei* chaperone research

— Fritha Hennessy for her friendship and encouragement

— Charles Hefer for his ever-patient assistance with all things computer-related and for teaching me the basics of Linux operating systems

— Tjaart de Beer for his assistance with homology modelling

— Eva Pesce, for her friendship and for proofreading of this thesis

— Eva, Earl, Nicodemus, Jess L, Leanne and Taryn for being great lab mates and friends – and for all their encouragement during the stressful times !

— Larey Haughton for her friendship and kindness

— My parents, for giving me so many opportunities throughout life – I can never thank you enough

Audaces fortuna iuvat - *(Virgil)*

Deus Vult

This thesis is lovingly dedicated to

My mother
Elisaveta V. Louw

and

My godmother
Valérie van Zyl (1949-2008)

List of outputs

Publications unrelated to the present work

Louw, C.A., Gordon, A., Johnston, N., Mollat, C., Bradley, G. and Whiteley, C.G. (2007). **Arginine deiminases: Therapeutic tools in the etiology and pathogenesis of Alzheimer's disease**. *Journal of Enzymology and Medicinal Chemistry*, **22 (1)**, 121-126

Publications in preparation

Louw, C.A. and Blatch, G.L. (2009). **Overproduction, purification and characterisation of Tbj1, a novel Type III Hsp40 from *Trypanosoma brucei*, the African Sleeping Sickness parasite**

Louw, C.A., Ludewig, M.H. and Blatch, G.L. (2009). **The Hsp70 proteins of *Trypanosoma brucei*: diverse weapons against a hostile lifecycle (Review).**

Local conference outputs

January 2008
Louw, C.A. and Blatch, G.L. (2008) - Expression and characterisation of two Type III and one Type IV Hsp40 proteins from the African Sleeping Sickness parasite, *Trypanosoma brucei* at the Bio '08 conference hosted by Rhodes University, Grahamstown (Poster presentation)

International conference outputs

Louw, C.A. and Blatch, G.L. (2009) - The cloning, expression and biochemical characterization of Tbj1, a novel type III Hsp40 protein in the African sleeping sickness parasite, *Trypanosoma brucei*. (Abstract accepted): EMBO-FEBS sponsored meeting: The biology of molecular chaperones – Cellular protein homeostasis in disease and ageing (Dubrovnik, Croatia)

CHAPTER 1

Literature Review

1.1. TRYPANOSOMAL PARASITES

Parasitic protozoa are organisms that display highly sophisticated forms of adaptation to a variety of environments during their lifecycles that are often extreme in terms of their contrast to each other (Matthews, 2005). This is due to the fact that the lifecycle of a typical protozoan parasite requires a developmental phase within an insect vector, followed by an infectious and reproductive phase in a mammalian vector (Matthews, 2005). One of the most sophisticated features of the protozoan parasite‟s defence system is the ability to evade the immune system of the mammalian host for long periods of time, thus leading to more success in terms of reproductive capacity and parasite survival (Matthews, 2005). Organisms such as Trypanosoma and Leishmania that belong to the class Kinetoplastida (Eukaryota, Excavata, Euglenozoa) are thought to have separated from the metazoan lineage approximately 3 x 10^9 years ago (Rout and Field, 2001; Simpson *et al*., 2002). Trypanosomes are taxonomically related to *Leishmania* (Sibley and Andrews, 2000). However, the parasitic trypanosome species have with time evolved into two types that differ significantly from *Leishmania* with respect to their biology within invertebrate hosts (Nyalwidhe *et al*., 2003).

The widespread distribution of trypanosomes, their broad host-range and the large divergence between species sequences suggest that they are ancient organisms and it has been proposed that they are at least 100 million years old (Stevens *et al*., 1999). The African trypanosomes are extracellular parasites, while the American trypanosome, *T. cruzi*, develops within cells (Sibley and Andrews, 2000). Trypanosomal species are the cause of devastating disease in both man and animals. *Trypanosoma brucei* (*T. brucei*) and *Trypanosoma cruzi* (*T. cruzi*) infections cause African sleeping sickness and Chagas disease, respectively, while Leishmaniases kill and debilitate hundreds of thousands of persons every year (Simpson *et al*., 2006). *T. brucei* is responsible for approximately 300 – 500 000 new infections per annum, and as the disease is fatal if left untreated an equal amount of deaths are presumed but not confirmed due to irregular monitoring of the disease (Berriman *et al*., 2005). *T. cruzi* infection leads to a far greater number of confirmed cases of illness (an estimated 16 – 18 million people are infected of which

approximately 21 000 die each year) but this could be due to better medical reporting in the South American countries in which it is prevalent (El-Sayed *et al*., 2005).

Table 1.1. Major Trypanosomal groupings and subspecies[1]

ORGANISM	SPECIES	SUBSPECIES
Duttonella	*Trypanosoma vivax*	
	Trypanosoma sp. T78	
Herpetosoma	*Trypanosoma blanchardi*	
	Trypanosoma lewisi	
	Trypanosoma rabinowitschae	
	Trypanosoma rangeli	*Trypanosoma rangeli Choachi*
		Trypanosoma rangeli SC58
Nannomonas	*Trypanosoma congolense*	
	Trypanosoma simiae	
	Trypanosoma simiae Tsavo	
	Trypanosoma sp. Fly9	
Schizotrypanum	*Trypanosoma cruzi*	*Trypanosoma cruzi cruzi*
		Trypanosoma cruzi marinkellei
		Trypanosoma cruzi strain CL Brener
		Trypanosoma cruzi strain Esmeraldo
Trypanozoon	*Trypanosoma brucei*	*Trypanosoma brucei brucei*
		Trypanosoma brucei gambiense
		Trypanosoma brucei rhodesiense
		Trypanosoma brucei TREU927
	Trypanosoma cf. brucei Msubugwe	
	Trypanosome equiperdum	
	Trypanosoma evansi	*Trypanosoma evansi evansi*

[1]**-Information obtained from the NCBI Taxonomy browser (Internet 1)**

The classification of African trypanosomes, especially those within the *Trypanozoon* subgenus is problematic due to the morphological identity of the parasites (Hide *et al.*, 1990). As a result, the five species and subspecies (*Trypanosoma brucei brucei*, *T.b. rhodesiense*, *T.b. gambiense*, *T. evansi* and *T. equiperdum*) that belong to this subgenus are classified according to various factors such as their preferred host, the type of infection / disease that they cause and their geographical distribution (Biteau *et al.*, 2000). This problem of identification is currently being solved by relying on biochemical and genotypic criteria as opposed to classification based on host or disease type (Biteau *et al.*, 2000).

1.1.1. *TRYPANOSOMA BRUCEI*

Trypanosoma brucei (*T. brucei*) is a unicellular flagellated protozoan that belongs to the Salivarian group of the genus *Trypanosoma* (Mhlanga *et al.*, 1997). Three subspecies of *T. brucei* are known to occur, viz. *T.b. brucei*, *T.b. rhodesiense* and *T.b. gambiense* (Matthews, 2005). These parasites are the causative agents of African sleeping sickness in mammals (including humans) in 36 countries in Sub-Saharan Africa (Matthews, 2005). *T. brucei* undergoes cyclic transmission between an insect vector and a mammalian host (Mhlanga *et al.*, 1997). These parasites are transmitted between mammalian hosts by the tsetse fly, Glossina spp., where they reside in the midgut after the fly takes a bloodmeal prior to migrating to the salivary glands in preparation for transmission to another mammalian host (McCulloch, 2004). Because of the relative inefficiency of transmission between the insect vector and the mammalian host, *T. brucei* causes the development of a chronic-type infection as opposed to an acute type of infection in order to allow the parasites sufficient time to be taken up by the vector again upon commencement of infection and to, in so doing, complete their lifecycle (Pays *et al.*, 2001).

1.1.2. VARIABLE SURFACE GLYCOPROTEINS (VSGS)

In the mammalian system, *T. brucei* is extracellular and survives in the bloodstream, where it evades the host‟s immune response by means of antigenic variation (McCulloch, 2004; Pays *et al.*, 2004). This antigenic variation is accomplished by

means of the sequential expression of antigenically distinct variable surface glycoproteins (VSGs) that are linked to the surface membrane of the parasite by means of a glycosylphosphatidylinositol (GPI) anchor (Matthews, 2005). This protective strategy on the part of trypanosomes relies on a protein coat that surrounds the trypanosomal parasite body and which is composed of a single protein (Ferrante and Allison, 1983). This protein is the VSG, which protects the trypanosome‟s plasma membrane from recognition by the host‟s immunoglobulins (Cross, 1975). VSG proteins in *T. brucei* consist of a combination of an N-terminal domain that consists of approximately 340 amino acid residues and either one or two C-terminal domains that consist of 30 – 50 amino acid residues each (Carrington *et al*., 1991). VSG N-terminal domains have been categorised into three different types (A, B and C) while the C-terminal domains have been categorised into six different types (1 – 6). Different combinations of these N- and C-terminal domains allow for a significant amount of variability in the VSG protein make up, and a given infecting population of *T. brucei* expresses a series of VSGs from a large reservoir of sequences found on the genome of the parasite (Cross, 1975; Ziegelbauer and Overath, 1993).

The completion of the *T. brucei* genome sequencing project has elucidated great amounts of information on the VSG complement of the parasite (Berriman *et al*., 2005). Between 1000 – 2000 potential VSG sequences occur on the *T. brucei* genome, although only 7 % of these sequences correspond to open reading frames (Berriman *et al*., 2005; Barry *et al*., 2005). The remaining sequences correspond to atypical VSG sequences that are disrupted, frame-shifted or fragmentary (Berriman *et al*., 2005; Barry *et al*., 2005). The rate at which trypanosomes switch their VSG coat protein is less than 0.01 per cell generation, which is high enough to successfully maintain infection in the host for as long a period as the supply of novel VSGs are not exhausted. It is subsequently proposed that the length of time that parasites are able to sustain an infection is dependent on the size of the VSG reservoir that is available to them (Barry and McCulloch, 2001; Berriman *et al*., 2005). It is worth noting that the VSG complement in *T. brucei* has diverged with time to become strain-specific, and that this could play a direct role in the differences in disease progression between HAT resulting from different strains of *T. brucei* (Hutchinson *et al*., 2007).

1.1.3. THE *T. BRUCEI* LIFECYCLE

The *T. brucei* lifecycle (Figure 1.1) is complex and characterised by the migration of the parasite between starkly contrasting environments.

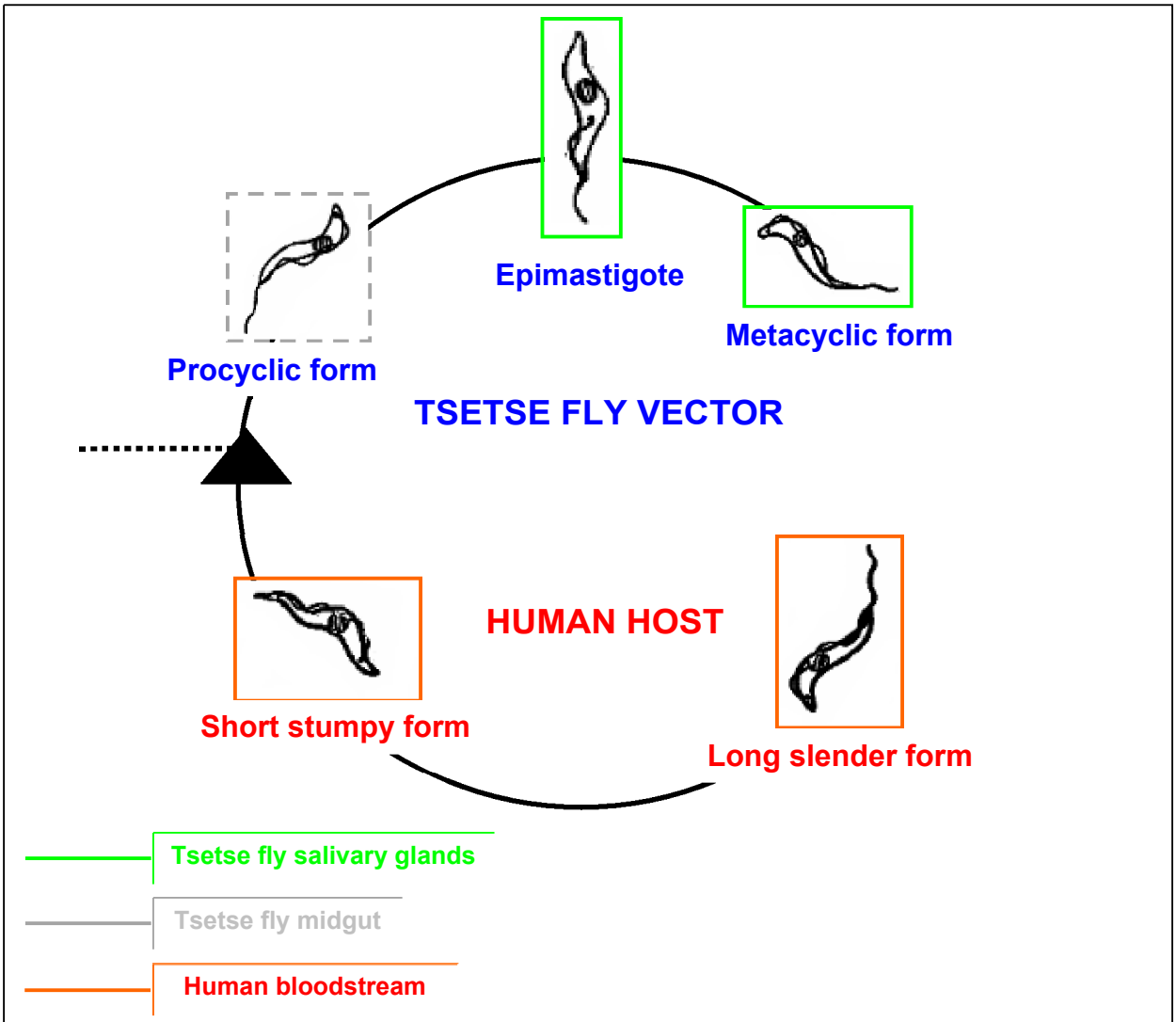


Figure 1.1. Schematic representation of the *T. brucei* lifecycle in the tsetse fly vector and the human host.

Increased parasite numbers upon transmission to the human host result in differentiation from the long slender bloodstream form to morphologically stumpy forms, which are adapted to survive transmission to tsetse flies after ingestion during a bloodmeal. After ingestion by the tsetse fly, procyclic forms that reside in the insect midgut are generated (dashed grey lines) which subsequently migrate to the salivary glands of the tsetse fly. The surface coat of these forms is different to that of bloodstream forms. Once they are established in the insect midgut, the trypanosomes once again arrest in division prior to migrating to the fly's salivary glands (epimastigote stage). These proliferative forms generate non-proliferative forms that have re-acquired a VSG coat in preparation for transmission to a new mammalian host (adapted from Matthews, 2005 and Internet 2).

Approximately 5 000 metacyclic parasites are introduced into the human host‟s dermal connective tissue during a tsetse fly bloodmeal. Inflammation occurs at the site of infection, and is characterised by the formation of a chancre (Barry and Emery, 1984; Mhlanga, 1994). After being transmitted by the bite of an infected female tsetse fly, the parasites proliferate in the infected mammalian host‟s bloodstream as morphologically slender forms. This form of *T. brucei* multiplies by means of asexual binary fission every 6 – 8 hours and is collectively known as long slender trypomastigotes (Mhlanga, 1994). The morphologically slender forms are replaced by stumpy forms that do not proliferate as parasite numbers in the infected mammal‟s bloodstream increase (Roditi and Liniger, 2002). This limiting of parasite numbers serves a two-fold purpose. Firstly, the limiting of parasite numbers ensures prolonged survival of the host, thus ensuring that the *T. brucei* parasites can propagate themselves for a longer period of time and increasing the chances of disease transmission, thus ensuring the long-term survival of the parasite species (Matthews, 2005). Secondly, the arrest of the parasites during a specific developmental phase ensures that the morphological changes that occur during uptake by the tsetse fly vector are co-ordinated with respect to the cell cycle, which is vital for the successful completion of the cell cycle of tsetse fly midgut procyclic forms (Matthews, 2005). Levels of parasitaemia in infected individuals vary with time due to the parasite‟s ability to circumvent the human immune system (Mhlanga, 1994).

1.1.4. KINETOPLAST GENOMES

The *T. brucei* genome has been fully sequenced (Berriman *et al.*, 2005) and is lodged on the GeneDB database online (http://www.genedb.org/genedb/tryp/) which greatly facilitates the study of Trypanosomal proteins. The 26-megabase nuclear genome of *T. brucei,* is found on 11 chromosomes while the mitochondrial (kinetoplast) genome consists of an unspecified number of smaller chromosomes (30 – 700 kb) (Berriman *et al*., 2005). The initial annotation efforts have identified 9068 predicted genes, of which 900 are proposed to be pseudo-genes, and approximately 1700 *T. brucei*-specific genes (Berriman *et al*., 2005). The sequencing of the *T. cruzi* genome was completed alongside that of *T. brucei*, and has revealed a genome far greater in size and plagued by the problem of repeated sequences (El-Sayed *et al*.,

2005). The diploid *T. cruzi* genome contains a predicted 22 750 proteins that are encoded by genes, of which 12 570 are members of allelic pairs (El-Sayed *et al*., 2005). Over 50 % of the *T. cruzi* genome consists of repeated sequences, which include retrotransposons (El-Sayed *et al*., 2005). The number of predicted genes for the TriTryps (as the *T. brucei*, *L. major* and *T. cruzi* parasites are collectively referred to) are summarised in Table 1.2.

Table 1.2[1] Haploid genome size of Kinetoplastid parasites with fully sequenced genomes

Parasite name	No. Genes	No. Species specific genes	Genome size (mb)
Trypanosoma brucei	9068	1392	25
Trypanosoma cruzi	12 000	3796	55
Leishmania major	8311	910	33

[1] **Taken and adapted from Kissinger (2006).**

The kinetoplastid parasites share 6158 ortholog clusters of genes that encode proteins in addition to several highly conserved functional protein domains (Kissinger, 2006). This is made all the more remarkable by the fact that these parasites differ markedly in their interactions with their insect vectors and mammalian hosts (Kissinger, 2006).

1.1.5. *T. BRUCEI* BIOLOGY

The study of *T. brucei* is essential not only because of its role as a disease-causing vector but also due to the rich and often unusual biology that occurs in this parasite which could shed light on early evolutionary mechanisms (Schneider *et al*., 2007). Phylogenetic analyses of the genes that encode small subunit RNA (ssRNA) suggest that Salivarian trypanosomes, including *T. brucei*, separated from other trypanosomal species approximately 300 million years ago (Steverding, 2008). To this end, it has been proposed that trypanosomiasis played a significant role in the evolution of early hominids, as the severity of the disease would have caused the evolutionary selection of trypanosome-resistant early terrestrial hominids (Lambrecht, 1985). Research has revealed several of the unusual characteristics possessed by

kinetoplastid parasites in terms of their cellular, biochemical and genomic properties (Donelson *et al*., 1999). The genomes of trypanosomatid parasites possess high gene density that is linked to polycistronic transcription. Kinetoplastids also have overlapping genes and their genomes are virtually intronless (Myler *et al*., 1999; Liniger *et al*., 2001). One of the most intriguing features of *T. brucei* is the presence of a single mitochondrion in the parasite that has a genome which is physically connected to the basal bodies of it"s flagella by means of proteinaceous connections that span the two mitochondrial membranes (Figure 1.2) (McKean, 2003). The mitochondrial DNA of *T. brucei* occurs in a body known as a kinetoplast (Simpson *et al*., 2002).

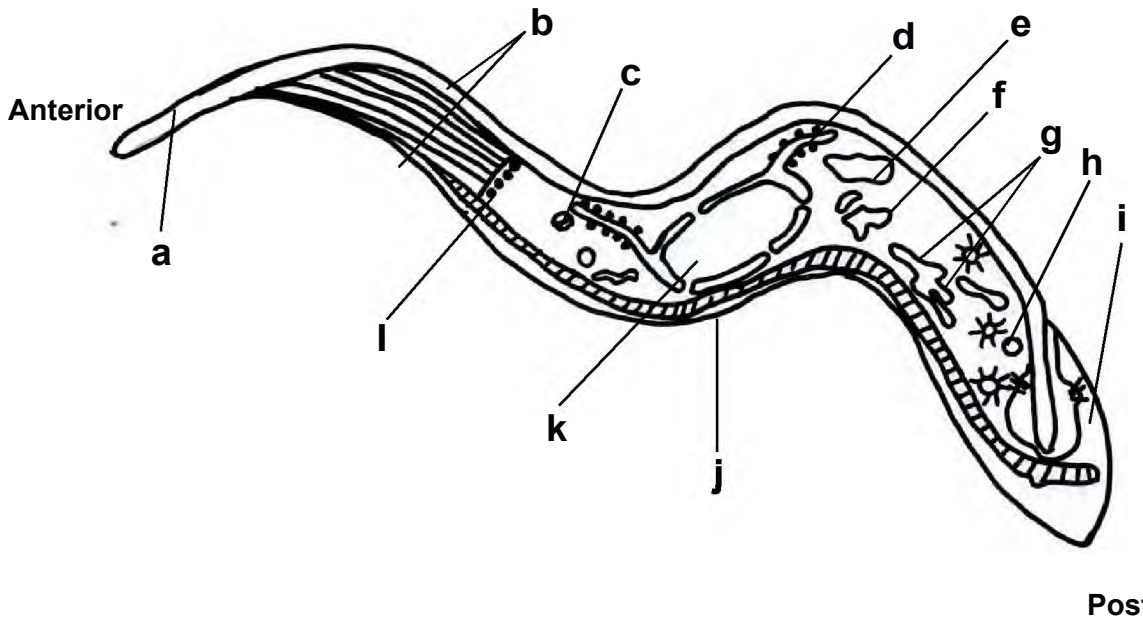


Figure 1.2. Schematic diagram of the *T. brucei* parasite cell body

Schematic representing the cell body of a *T. brucei* parasite. Letters on the figure highlight organelles of interest as follows: [a] – free flagellum; [b] – VSG surface coat (covers the entire surface of the *T. brucei* cell body); [c] – glycosome; [d] – rough endoplasmic reticulum (E.R.); [e] – Golgi apparatus; [f] – lysosome; [g] – endosomes; [h] – vesicle pinched off from the flagellar pocket membrane; [i] – flagellar pocket; [j] – mitochondrion; [k] – nucleus; pellicular microtubules (covers entire cell surface beneath the cell membrane); [l] – Sub-pellicular microtubules . Figure taken and adapted from Internet 3.

This unusual mitochondrial feature is further characterised by the presence of two genetic elements (as opposed to one in most other organisms) which are

topologically intertwined (McKean, 2003). These genetic elements are known as maxicircles and minicircles (Chanez *et al*., 2006). The expression of mitochondrial genes in *T. brucei* is complex due to the fact that many genes are composed of incomplete open reading frames. The primary transcripts obtained from such incomplete reading frames have to be edited in order to be converted into translatable messenger RNAs (mRNAs) (Schneider *et al*., 2007). In addition to unusual gene expression, the *T. brucei* parasite also displays an unconventional way of protein translation in its mitochondrion. No tRNAs are present in the *T. brucei* mitochondrion, and hence mitochondrial protein translation requires tRNAs that are imported from the cytosol in a process that is not well-characterised at present (Schneider, 2001).

1.2. HUMAN AFRICAN TRYPANOSOMIASIS (HAT)

Human African Trypanosomiasis (HAT) is a lethal disease caused by *T. brucei* infection that occurs throughout Africa and which has devastating social and economic consequences. Also commonly known as African sleeping sickness, HAT is endemic in 36 African countries and is one of the leading causes of morbidity on the continent (Luscher *et al*., 2007; Kennedy, 2006). HAT is caused by protozoan *T. brucei* parasites that are transmitted by the blood-sucking tsetse fly of the *Glossina* species. An important factor in terms of disease management and treatment is that these flies remain infective for the duration of their lifespan (Vickerman, 1997). The *T. brucei* parasite includes three subspecies, viz. *T.b. gambiense*, *T.b. rhodesiense* and *T.b. brucei* (Bentivoglio *et al*., 1994; Kennedy, 2006). The first two of these subspecies cause HAT in humans, while the latter sub species causes different forms of trypanosomiasis in animals, viz. Nagana in cattle and milder forms of HAT-like diseases in both wildlife and domestic animals. The *T.b. brucei* parasite is non-infective in humans and highly infective in mice and other domesticated animals (Mulenga *et al*., 2001).

1.2.1. HAT IN HISTORY

The devastating impact of HAT infection has been well-documented throughout the course of history (Steverding, 2008). A disease infecting cattle that strongly

resembles modern descriptions of Nagana (*T. brucei* infection in cattle) is documented in a veterinary papyrus of Kahun Papyri that dates from the second millennium before Christ (B.C.) (Griffith, 1898, reviewed in Steverding, 2008). During the Middle Ages, Arab traders and explorers described HAT-like infections in countries such as Benin, Ghana, Mali and Songhai in West Africa (Steverding, 2008). One of the first written accounts of HAT comes from the Arabian historian, Ibn Khaldun (1332-1406) who described the death of the King of Mali of a disease known as the "sleeping illness" (Williams, 1996). Accounts of HAT in early modern times are mainly derived from ship doctors and medical officers that were contracted to companies operating in the slave-trade. To that effect, the first accurate medical analysis of HAT focussing predominantly on the neurological symptoms of the late-stage disease was published in 1734 by John Atkins (1685-1757) who was an English naval surgeon (Cox, 2004). Another English physician, Thomas Winterbottom (1766-1859) published a report describing the physiological symptoms in more detail in 1803 (Cox, 2004). The connection between the bite of tsetse flies and Nagana infection was first made by the Scottish missionary and explorer, David Livingston (1813-1875) although it took a further 50 years before trypanosomes were recognised as the cause of Nagana and HAT (Bruce, 1895 reviewed in Steverding, 2008). Trypanosomes were first observed in blood in 1901 by Robert Ford (1861-1948) who was a British surgeon attending to cases in the Gambia (Forde, 1902 reviewed in Steverding, 2008). Despite the observation of trypanosomes in the blood of an infected human, it was not until the parasites were discovered in the cerebrospinal fluid of an infected patient in 1902 by Italian physician Aldo Castellani (1878-1971) that the positive connection between the parasites and disease was made (Cox, 2004). The cyclical transmission of *T. brucei* by the tsetse fly was described by the German military surgeon, Friedrich Karl Kleine (1859-1951) during 1909 (Kleine, 1909 reviewed in Steverding, 2008). It is thus evident that the causative mechanisms of HAT have been known since the turn of the last century, and that the disease has had a significant impact on the affairs of Africans for several centuries prior to that.

From 1900 to the 1940"s, HAT reached epidemic proportions, but the disease was thought to have been brought under control by 1960. HAT has, however, re-emerged during the last two decades and is thought to be occurring at levels comparable to

those observed at the turn of the century (Garcia *et al*., 2006). Current estimates propose that approximately 300 000 – 500 000 persons develop HAT in Africa each year, with around 50 000 fatalities resulting from the disease during the same period of time (Bouteille *et al*., 2003; Kennedy, 2006). Over 60 million people are at risk of contracting the disease in Sub-Saharan Africa, but detection and treatment are hampered by the fact that only 10-15 % of the people at risk of the disease are under any form of medical surveillance (Garcia *et al*., 2006; Kennedy, 2006). HAT is currently the most significant cause of mortality in certain provinces of Angola as well as in the Congo and Southern Sudan, surpassing HIV / AIDS in its impact on human health (Matthews, 2005). The re-emergence of the HAT epidemic in the last two decades is thought to be the result of a relaxation of disease control that involved monitoring areas at risk as well as controlling the tsetse fly population. This, coupled to a lack of funding for research, has resulted in a dramatic resurgence of HAT infections across Sub-Saharan Africa (Garcia *et al*., 2006).

1.2.2. AETIOLOGY OF HAT

The symptoms of HAT infection can vary from patient to patient, but it is generally accepted that the disease occurs in two stages, although the distinction between these stages are not always very clear (Kennedy, 2006). *T. brucei* parasites are transmitted to a mammalian host during a bloodmeal by an infected tsetse fly. The parasites proliferate in the area surrounding the bite mark and this area subsequently becomes inflamed and swollen, forming a painful chancre that lasts 3-4 weeks (Burchmore *et al*., 2002). This chancre heals with time, although it is often not observed at all in cases of *T.b. gambiense* infection (Burchmore *et al*., 2002). During the initial stages of infection, the parasites spread to the bloodstream and lymph nodes (Burchmore *et al*., 2002). The first stage of HAT is characterised by numerous (and often non-specific) disease symptoms that could be described as flu-like in nature. Patients may display fever, malaise and facial oedema prior to the development of symptoms that target specific organs (Kennedy, 2006). In addition to these symptoms, patients may also experience myocarditis, pulmonary oedema as well as anaemia and pericardial effusion (Burchmore *et al*., 2002). During the second, and final, stage of the disease (meningo-encephalitic stage), severe central

nervous system malfunctions become apparent (Bentivoglio *et al.*, 1994). This is due to parasite invasion of the brain and spinal cord and occurs within the first month of infection with *T.b. rhodesiense* (Burchmore *et al.*, 2002). This stage is characterised by numerous and complex neuro-psychiatric symptoms that culminates in dementia towards its final stages (Bentivoglio *et al.*, 1994; Kennedy, 2006). The aetiology of HAT is dependent on the Trypanosomal subspecies with which a given patient has been infected. Two forms of HAT that differ in their clinical phenotype are known to occur (Kennedy, 2006). Infection with the *T. brucei gambiense* parasites (occurring largely in West and Central African countries, shown in Figure 1.3) results in a long-term chronic type of infection that can last several years.

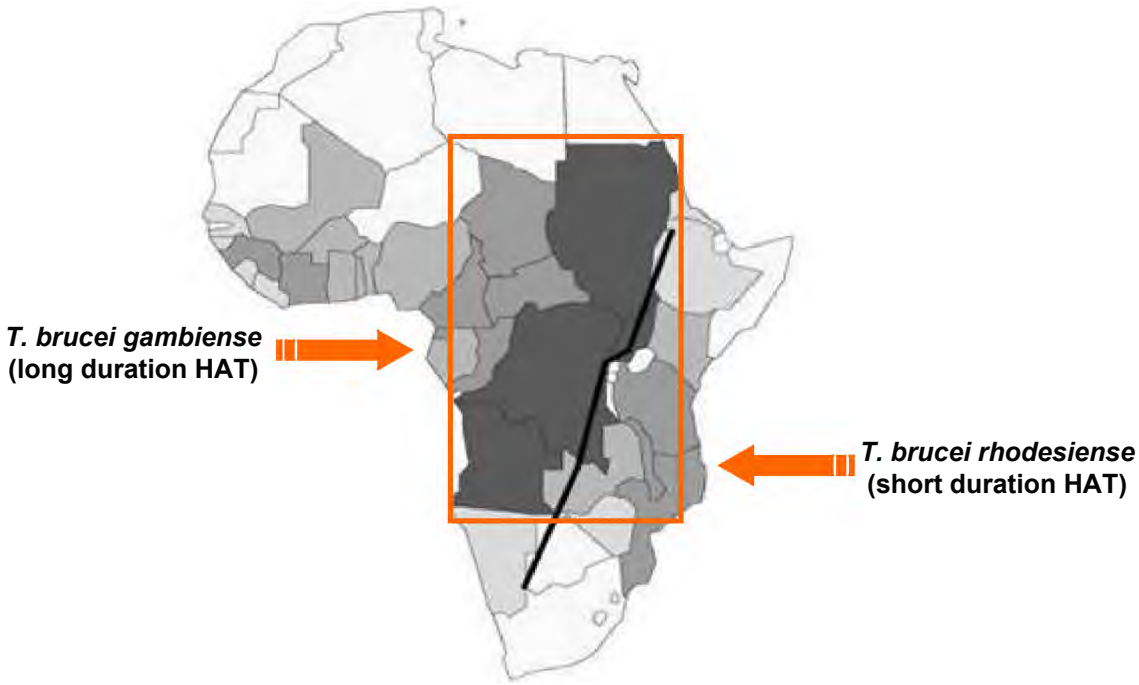


Figure 1.3. Prevalence of Human African Trypanosomiasis across the African continent.

Regions affected by West African Trypanosomiasis (caused by *T. brucei gambiense*) and East African Trypanosomiasis (caused by *T. brucei rhodesiense*) are indicated on the map as shaded regions with areas at greatest risk shaded in dark grey and highlighted in red. Areas with little or no shading are not considered to be at risk for HAT, figure taken and adapted from Internet 4.

In contrast, infection with the *T. brucei rhodesiense* parasite (occurring in East African countries) results in a rapid onset of disease that can be fatal within weeks to

months (Garcia *et al*., 2006; Kennedy, 2006). The cause of this difference in the aetiology of HAT caused by different trypanosomal subspecies is not known, but it has been proposed that the difference in infectivity and disease duration is largely attributable to the fact that *T. brucei gambiense* displays a greater adaptation to its human host (Kennedy, 2006). Elucidation of this adaptation mechanism, which has not been characterised at present, could play a significant role in the discovery of novel drug targets and the development of new treatment strategies for HAT.

1.2.3. CURRENT TREATMENT REGIMENS FOR HAT

At present, successful treatment regimens for HAT are severely lacking, partly due to financial and political constraints, but largely due to the immense complexity of the infectious disease, which circumvents the human host‟s immune defence responses (Bouteille *et al*., 2003). Treatment is also complicated by the fact that, in order to be successful, it has to accurately target the particular strain as well as the disease stage (Garcia *et al*., 2006). Some of the drugs currently used to treat HAT include pentamidine, suramin, melarsopol and nifurtimox (Bouteille *et al*., 2003). A summary of the currently available drugs for HAT is given in Table 1.3. The challenge involving current treatment regimens for HAT is twofold: firstly, the drugs are largely ineffective due to overuse which causes resistance in parasites and secondly, the drugs are known to cause unacceptably high levels of toxicity in patients undergoing treatment (Delespaux and de Koning, 2007). The need for parenteral administration of drugs coupled to an uncertainty in terms of supply and demand also plays a role in hampering effective treatment of HAT (Boykin *et al*., 1996; Ansede *et al*., 2004). Of further concern is the severe levels of resistance to all manner of drugs observed during treatment of livestock for related disorders such as Nagana (Geerts *et al*., 2001).

It is concerning to note that no novel successful treatment regimens for HAT have been developed over the last 50-year period (Bouteille *et al*., 2003). This, coupled with the scope and high levels of morbidity, human suffering and economic hardship caused by HAT, necessitates the discovery of novel drug targets and the development of more successful treatment around these targets.

Table 1.3. Current treatment regimens available for HAT [a]

Drug name	Marketed as	Used for	Drug target	Method of administration	Side effects
Suramin [b]	Germanin® by Bayer	*T. brucei rhodesiense*	Oligopeptidase B	Slow intravenous injection	Neuropathy, rash, fatigue, anaemia, hyperglycemia, hypocalcemia, coagulopathies, neutropaenia, renal insufficiency and transaminitis
Pentamidine [b]	Nebupent® by Sanofi-Aventis	*T. brucei gambiense*	Oligopeptidase B	Intramuscular injection	Hypotension, painful injection, nephrotoxicity, leucopenia and liver enzyme abnormalities
Melarsoprol [b]	Melarsen Oxide-BAL® by Sanofi-Aventis	*T. brucei rhodesiense*	Unknown. Thought to bind to trypanothione	Intravenous injection	Severe. Convulsions, neurological effects, death due to reactive encephalopathy, pyrexia, headache, pruritis, thrombocytopaenia and heart failure
Eflornithine [b]	Ornidyl® by Sanofic Aventis	*T.b. gambiense*	Inhibits ornithine decarboxylase	Intravenous infusion	Fever, headache, hypertension, macular rash, peripheral neuropathy, tremor, gastrointestinal upsets
Pafuramidine maleate [b]	DB289® (Clinical trials in progress)	*T.b. gambiense and rhodesiense*	Unknown. Proposed to be the parasite"s mitochondrion	Oral	Fever and pruritis
Nifurtimox [b]	Lampit® by Bayer	*T. brucei gambiense*	Given in cases of Melarsoprol failure. Generates free radicals that generate reduced oxygen metabolites	Oral	Toxic to central nervous system, increased propensity to develop cancer, gastrointestinal difficulties

[a] - Information condensed from Barrett et al., 2007, Burchmore et al., 2002 and Morty et al., 2007

[b] - Drugs for early stage HAT are shown in red and drugs for late stage HAT are shown in green. Drugs that are currently undergoing clinical trials or that are in the early stages of development are shown in pink. Nifurtimox is also shown in pink as it has a high treatment failure rate and is currently only administered in combination with other drugs.

This is especially important in light of the fact that *T. brucei* infection is not amenable to the generation of vaccines due to the transient nature of the VSG protein coat

(Barrett *et al*., 2007). In addition to the drugs mentioned in Table 3, the veterinary trypanocide Diminazine, despite not being a novel treatment developed for HAT, has been used with some success in the treatment of humans (Pepin and Milord, 1994). The successful development of novel drugs to treat HAT depends greatly on the identification of essential metabolic pathways or proteins in the parasite (Chambers *et al*., 2008). One such essential pathway that is being investigated in terms of novel drug design is the glycolytic pathway, which is essential to bloodstream form (BSF) parasites as glycolysis is the sole method in which energy generation takes place during this stage of the trypanosomal life-cycle (Pepin and Milord, 1991). Lonidamine was initially developed as an antispermatogenic agent, but was later found to be effective against cells that are undergoing high rates of metabolic turnover, such as parasites (Paggi *et al*., 1988; Floridi and Lehninger, 1983). The drug has been found to be toxic to BSF parasites as well as procyclic parasites by targeting *T. brucei* hexokinase 1 (TbHKI), which is the initial enzyme in the *T. brucei* glycolysis pathway (Chambers *et al*., 2008).

More effective means of diagnosis are also essential in improving HAT treatment, as parasite eradication is hard to accomplish once the parasites have migrated across the blood- brain barrier (BBB) as is seen in the second stage of the disease (Masocha *et al*., 2007). Trafficking of parasites across the BBB is regulated by interferon-γ, a cytokine. Several research groups are investigating the possibility of designing novel drugs to impede the movement of trypanosomes across the BBB or that will be able to eliminate trypanosomes from the brain parenchyma once they have passed through the BBB (Masocha *et al*., 2007). Novel drugs for second-stage HAT are especially crucial due to the severe side-effects that have been attributed to drugs for late-stage HAT (Masocha *et al*., 2007).

1.3. *TRYPANOSOMA CRUZI*

Performing a thorough survey of the literature pertaining to the *T. cruzi* parasite proves challenging due to the fact that much of the most significant literature on the subject is in Spanish due to the high rate of prevalence of *T. cruzi* infections in Latin America and the Southern United States of America. In spite of this challenge, a

sufficient amount of literature is available to create a general overview of the *T. cruzi* parasite and the clinical progression of Chagas disease.

1.3.1. T. CRUZI TRANSMISSION AND CHAGAS DISEASE

Chagas disease is resultant from human infection with the *T. cruzi* parasite (Moncayo, 2003). The parasite is transmitted in the faeces of infected triatomine insects that enter the human body through bites and cuts (Moncayo, 2003). As mentioned in the introductory section to this work, Chagas disease affects a very large number of persons (up to 17 million infections in the Southern United States of America and in South America as far south as Argentina) and is perceived a very significant threat to public health in both the areas in which it is endemic and neighbouring regions (Moncayo, 2003; Volk *et al*., 1986). Approximately 300 000 new infections are reported every year (World Health Organisation, 2002). The threat to neighbouring regions is great due to a migration of infected individuals as opposed to migration patterns of insects and / or parasites. *T. cruzi* infection is readily passed from mother to child during pregnancy and birth, as well as during blood transfusion from an infected individual to an uninfected one (Dutra *et al*., 2003; Moncayo, 2003). The latter route explains the occurrence of new Chagas infections in regions where it was previously absent, such as the United States of America (Dias *et al*., 2002).

There are several marked similarities between Chagas disease and HAT. An investigation of different isolates of *T. cruzi* has shown that the population is highly polymorphic, and that infection with distinct parasite populations has an effect on disease progression and outcome in affected persons (Buscaglia and Noia, 2003; Macedo *et al*., 2004; Vago *et al*., 2000). This is similar to the affect of *T. brucei* subspecies on the effect of disease progression in HAT (Kennedy, 2006). Interestingly, as with HAT, the patient‟s immune response to *T. cruzi* infection is crucial in determining the fate of infection (Macedo *et al*., 2004; Vago *et al*., 2000). *T. cruzi*, however, is unable to multiply extracellulary in a vertebrate host, which is not the case in *T. brucei* (Volk *et al*., 1986).

1.3.2. THE AETIOLOGY OF CHAGAS DISEASE

The acute phase of Chagas disease lasts between 2 – 4 months and is difficult to detect (Dutra *et al*., 2005). This acute stage evolves to a chronic infection that can last for the rest of the infected person‟s lifespan (Dutra *et al*., 2005). The most significant challenge of Chagas disease from a clinical and therapeutic perspective is that most patients in the chronic stage of infection remain asymptomatic and the disease is only positively identified after extensive laboratorial examinations (World Health Organisation, 2002). In approximately 20 – 30 % of chronic Chagas infections, patients develop the cardiac clinical form of the disease which results in alterations in heart function (Morris *et al*., 1990; Rocha *et al*., 2003). These cardiac symptoms may range from mild and barely perceptible changes in heart function, to severe symptoms that include the development of clinical chronic chagasic cardiomyopathy (CCC) which is associated with high rates of mortality (Morris *et al*., 1990; Rocha *et al*., 2003).

1.3.3. THE MORPHOLOGICAL CHANGES UNDERGONE BY *T. CRUZI* DURING INFECTION OF HUMANS

The *T. cruzi* parasite undergoes a similar number of dramatic changes during its transition from its insect vector to mammalian host as *T. brucei* (Volk *et al*., 1986; Matthews *et al*., 2005). When *T. cruzi* parasites enter the human bloodstream from the reduviid bug‟s gut, the parasites have to adapt to the drastically different conditions almost instantaneously in order to survive (Volk *et al*., 1986; Wendel *et al*., 1992). Upon entering their human hosts, the *T. cruzi* parasites take on the amastigote form by undergoing a dramatic morphological change which includes the loss of their undulating membrane and flagellum (Volk *et al*., 1986). In their amastigote stage, the *T. cruzi* parasites are able to replicate intracellularly and invade every organ of the body until infected cells rupture and release parasites into the bloodstream (Volk *et al*., 1986; Wendel *et al*., 1992). Other stages and morphological changes in the *T. cruzi* parasite‟s lifecycle, which are similar and in some cases identical to that of the *T. brucei* lifecycle stages, include the infective replicative epimastigotes, the metacyclic trypomastigotes (occurring in the haematophagous

vector) and the bloodstream trypomastigotes in the human host (Wendel *et al*., 1992).

1.4. MOLECULAR CHAPERONES

Molecular chaperones form part of a highly conserved protein family that is found in all studied organisms. Their function is to bind selectively to nascent polypeptides or partially folded intermediate proteins and facilitate the folding of these species into their optimal conformations to ensure functionality (Fink, 1999). Molecular chaperones also prevent aggregation of proteins, which is essential to promote the formation of viable proteins in the crowded cellular environment, where newly translated proteins are prone to enter non-productive folding pathways that lead to aggregate formation (Anfinsen, 1973; Fink, 1999). Although chaperones do not contain any steric information regarding protein folding, they assist proteins to fold by guiding them along suitable folding pathways by binding to protein chains as they emerge from ribosomes (Hendrick and Hartl, 1993). In addition to the role played by molecular chaperones in protein folding, is the important role that these proteins play in facilitating the translocation of proteins across intracellular membranes into organelles (Hendrick and Hartl, 1993). Molecular chaperones also facilitate the proteolytic degradation of unstable proteins that have been targeted for degradation pathways (Hendrick and Hartl, 1993).

Most molecular chaperones are nucleotide-dependent in that protein binding and release occur in response to conformational changes resulting from ATP binding, hydrolysis and exchange (Fink, 1999). Heat shock proteins (Hsps) or stress proteins are a class of molecular chaperones that are upregulated in response to various environmental stresses or a period of heat shock (Fink, 1999; Finkelstein and Strausberg, 1983). Certain chaperones that are involved in regular cellular housekeeping functions can also be constitutively expressed (Hsc proteins) (Finkelstein and Strausberg, 1983). Several classes of Hsps have been described, and these are highly conserved across a wide variety of organisms ranging from bacteria to higher mammals (Santoro, 2000). Hsp proteins are generally classified into various classes based on their subunit molecular mass (Baneyx and Pulambo,

2003; Smith *et al*., 1998). Several major classes of Hsp that play a significant role in cellular homeostasis are highlighted in Table 1.4 and discussed more fully in the text.

Table 1.4. Major classes of heat shock proteins that enable the maintenance of cellular homeostasis during periods of heat or environmental stress [a]

Eukaryote	Prokaryote	Major function of chaperone class
Hsp40	**DnaJ**	Co-chaperone to Hsp70 / DnaK. Some can act as chaperones in their own right.
Hsp60	**GroEL**	Also known as chaperonins. They bind partially folded intermediates and serve to prevent their aggregation. This facilitates the folding and assembly of proteins.
Hsp70	**DnaK**	Prevents premature folding, misfolding and aggregation of nascent polypeptides coming off the ribosome. Their function is regulated by Hsp40 co-chaperones.
Hsp90	**HtpG**	Play an important role in modulating the activity of various proteins; especially signal transduction proteins and transcription factors, the most important of which are the steroid receptors. Diverse functions that include proteolysis and conference of heat tolerance.

[a] – information condensed from Fink (1999)

1.4.1. THE HSP90 CHAPERONE FAMILY AND HOP

Hsp90 proteins form part of a highly conserved group of molecular chaperones that are essential in organisms ranging from unicellular bacteria to mammals (Fink, 1999). Hsp90 is a highly abundant protein; even when cells are not subjected to stress conditions it amounts to approximately 1 % of all soluble cytoplasmic proteins (Welch and Feramisco, 1982). The Hsp90 family includes proteins such as Grp94 (in the E.R.) and HtpG in *E.coli* as well as the cytosolic eukaryotic Hsp90 (Fink, 1999). Hsp90 is essential in eukaryotic cells and is a unique molecular chaperone in that it

serves a specific set of client / substrate proteins as opposed to playing a role in the maturation or maintenance of most proteins in the cell (Picard, 2002). The assembly and maturation of a number of important cellular complexes is dependent on Hsp90. Several of these complexes are involved in signal transduction pathways (Picard, 2002; Zhao *et al*., 2005). Hsp90 facilitates interactions with cytoskeletal elements as well as protein kinases (Fink, 1999). It also aids in the maturation and assembly of a number of complexes that are critical for cellular survival. Several of the other client proteins of Hsp90 are just as important in the cell, most notably the enzyme, telomerase as well as client proteins implicated in signal transduction and the immune response (Holt *et al*., 1999; Pratt and Toft, 2003; Pearl and Prodromou, 2006). In order to effectively perform its functions within the cell, Hsp90 frequently occurs in complexes with other proteins (Fink, 1999). Hsp90 belongs to the GHKL family of proteins, which are dimeric proteins that bind ATP in their N-terminal domains using what is known as a "Bergerat fold" and are ATPases (Dutta and Inouye, 2000). Proteins of the GHKL family include DNA gyrase, Hsp90, histidine kinase and MutL (Dutta and Inouye, 2000).

It has been proposed that Hsp90 operates by maintaining its client proteins in a state of readiness for activation until they are stabilised by a conformational change, or by the binding of a suitable cofactor, ligand or protein partner (Buchner, 1999). In order to function efficiently *in vivo*, Hsp90 utilises a number of other chaperone proteins to assist its function. These proteins are termed co-chaperones, and include: Hsp40, Hsp70, Hsp70 / Hsp90-organising protein (Hop), Cell division cycle 37 homolog (Cdc37), Prostaglandin E synthase 3 (cytosolic) (p23) and high molecular weight immunophilins (Picard, 2002). It has been proposed that many as yet uncharacterised co-chaperones of Hsp90 exist (Zhao *et al*., 2005). Certain co-chaperones have clear roles in the Hsp90 network viz. Hsp40, Hsp70 and Hop act in concert to present substrates to Hsp90, while the role played by others remains unclear at this stage (Wegele *et al*., 2006).

All known Hsp90 proteins exist as obligate homodimers in order to function efficiently; each protomeric species comprising the homodimer consists of an N-terminal regulatory domain (NTD), a charged region that is not well conserved and dispensible to protein function, a middle domain (MD) and a C-terminal domain that facilitates

dimerisation (Dutta and Inouye, 2000). The NTD contains the ATP and inhibitor binding sites. It consists of an eight-stranded anti-parallel β-sheet and nine α-helices (Buchner, 1999). The ATP-binding site is situated in a substrate pocket that is 15 Å deep and the unusual geometry of this binding site has been likened to that of DNA gyrase B (Buchner, 1999). A proposed mechanism of the Hsp90 mode of action suggests that the chaperone activity of Hsp90 is directly linked to a conformational change that depends on an ATP-driven conformational change within the NTD. This conformational change leads to the closure of a mobile "lid-like" component over the ATP binding pocket, which allows transient N-terminal dimerisation, allowing the catalytic residues in the MD and NTD to align and resulting in ATP hydrolysis (Obermann *et al*., 1998; Chadli *et al*., 2000; Johnson *et al*., 2000). This proposed mechanism is supported by new crystals structures of yeast Hsp82 and bacterial HtpG (Shiau *et al.*, 2006). In the case of cytosolic Hsp90s, their chaperone abilities are dependent on the function of co-chaperones such as Cdc37 and p23 that bind to the NTD of Hsp90 (Phillips *et al*., 2007). The MD has been proposed to function as the main binding site for client proteins (Phillips *et al*., 2007). The C-terminal region of Hsp90 contains the interface required for dimerisation as well as the MEEVD motif that binds to co-chaperones possessing a tetratricopeptide (TPR) motif (Buchner, 1999; Phillips *et al*., 2007).

In terms of its operation within the cell, Hsp90 possesses a weak ATPase activity that is essential for it to function (Panaretou *et al*., 1998). Human Hsp90 only hydrolyses one molecule of ATP every 20 minutes (0.04 min^{-1}) while yeast Hsp90 can hydrolyse one molecule of ATP every one or two minutes (Panaretou and Prodroumo, 1998; Scheibel *et al*., 1998; McLaughlin *et al*., 2002). Several co-chaperones have been shown to regulate the ATPase activity of Hsp90 as mentioned above (Phillips *et al*., 2007). Studies that have been performed on the N- and C-terminal domains of Hsp90 have shown that both of these domains can suppress aggregation of non-native proteins, suggesting that Hsp90 may have two chaperone binding sites (Scheibel *et al*., 1998; Young *et al*., 1997). These two sites are different in terms of their specificity for substrates and their nucleotide dependence; the N-terminal site is known to associate with non-native substrates in an ATP-dependent manner while the C-terminal domain site acts as a general chaperone that is ATP-independent (Scheibel *et al*., 1998). It can be proposed that the specificity of Hsp90 for substrate

can also, in part, be attributed to the co-chaperones with which it is known to form complexes. Hsp90 proteins are known to interact with the Hsp40 / Hsp70 chaperone pathway (McLaughlin *et al*., 2002). The interaction between Hsp90 and Hsp40 / Hsp70 is facilitated by the Hsp70-interacting protein (Hip) and Hop (Pearl and Prodroumo, 2000). Hip (an Hsp70 specific cofactor) regulates the ATPase activity of Hsp70 while Hop serves as a linker protein between the Hsp70 and Hsp90 pathways (Buchner, 1999). It has been proposed that Hsp70 interacts with a given protein substrate (possibly acquired by means of an Hsp40-based interaction) and passes this client protein on to Hsp90 by means of interaction with Hop, which binds Hsp90 and Hsp70 simultaneously by means of its N- and C-terminal TPR domains (Odunuga *et al*., 2004).

Due to its ability to regulate essential pathways, Hsp90 has shown great importance in terms of medical research. Research has shown that Hsp90 is essential for the stability and functionality of a number of oncogenic proteins that cause cancer (Powers and Workman, 2006). Inhibitors that halt the function of Hsp90 have been found to cause the degradation of its client proteins via the ubiquitin-proteasome pathway (Connell *et al*., 2001; Demand *et al*., 2001) the concomitant depletion of several oncoproteins as well as the downregulation of signals that are propagated through numerous oncogenic signalling pathways (Maloney and Workman, 2002). This affects all aspects of a given cancerous phenotype and has great promise in future treatments of this disease (Maloney and Workman, 2002; Workman, 2004). One of the main advantages of a treatment for cancer that targets multiple pathways is that resistance is less likely to develop. A further advantage is that a drug that targets a crucial protein such as Hsp90 could facilitate treatment in a number of different types of cancer that are derived from a number of different molecular abnormalities (Powers and Workman, 2006). Several highly effective and specific inhibitors to Hsp90, such as geldanimycin and radicicol have been identified (Roe *et al*., 1999). A point of interest is that both of these inhibitors mimic the ATP binding mechanism of Hsp90 proteins and subsequently inhibit the Hsp90 ATPase function, resulting in a blockage of substrate protein maturation, which eventually leads to substrate degradation (Roe *et al*., 1999).

1.4.2. TPR DOMAIN-CONTAINING PROTEINS

Tetratricopeptide (TPR) repeat domains are protein-protein interaction motifs (Hirano *et al*., 1990; Sikorski *et al*., 1990). A typical TPR domain consists of a degenerate amino acid sequence that is characterized by a 34-residue degenerate sequence ($-W_4-L_7-G_8-Y_{11}-A_{20}-F_{24}-A_{27}-P_{32}$) that usually occurs as multiple tandem repeats (Goebl and Yanagida, 1991; Lamb *et al*., 1995). TPR domains are found in proteins with diverse cellular functions which include cell division, transcription, protein transport and organism development (Blatch and Lässle, 1999; Cliff *et al*., 2005; Goebl and Yanagida, 1991; Lamb *et al*., 1995). TPR domain-containing proteins have been found to occur in virtually all cell types and cellular locations and are present in organisms ranging from bacteria to humans (Sikorski *et al*., 1990).

The TPR domain has been found to consist of antiparallel α-helices in certain cases, such as in protein phosphatase 5 (Das *et al*., 1998). In terms of structure, numerous crystallographic studies have revealed that TPR domains share the common ability to form antiparallel alpha-helical hairpins which, when clustered together forms a domain that possesses a grooved surface, as shown in Figure 1.4 (Das *et al*., 1998; Scheufler *et al*., 2000). This domain has been likened to a hand with curled fingers which can bind to other proteins (Smith, 2004). Due to the high sequence variability among different TPR motifs, however, it is possible that other configurations for these domains exist (Melville *et al*., 2000).

TPR domains further facilitate the specificity of interactions with partner proteins by means of their side-chains which project into the solvent space in the groove that binds partner proteins and interact with side-chains from the capture polypeptide (Scheufler *et al*., 2000). A subclass of the TPR domain-containing protein family that contain a J-domain typically found in DnaJ proteins and their homologs has been identified (Melville *et al*., 2000). The function of these proteins has not been studied in any detail.

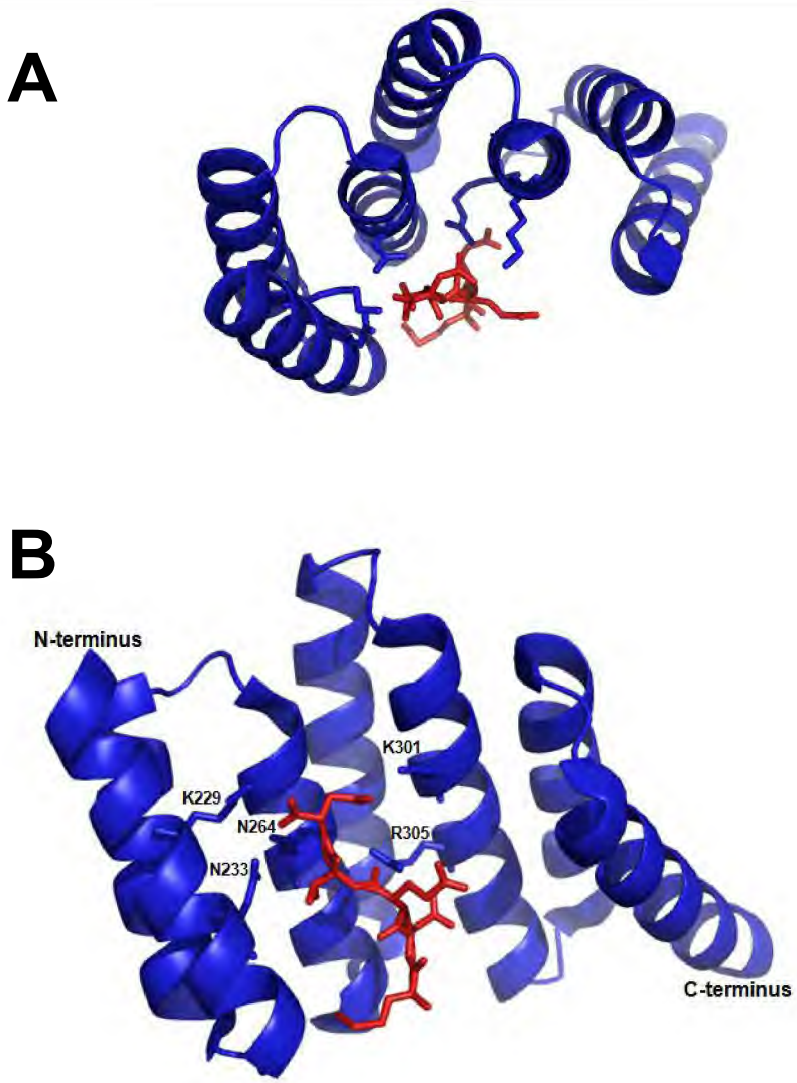


Figure 1.4. Ribbon representation of the crystal structure of the TPR2A domain of Hop complexed to the MEEVD region of Hsp90

The TPR2A TPR domain of Hop (1ELR) is shown in blue cartoon format and the MEEVD region of Hsp90 is indicated with red sticks. The structure was visualised with Pymol software (DeLano, 2002). [A] – TPR2A domain of Hop seen from the top showing region of MEEVD interaction; [B] – the same model seen from the side with key residues in Hop shown on the figure.

1.4.2.1. TPR DOMAIN CONTAINING CO-CHAPERONES

TPR domains occur in a large number of co-chaperones that predominantly interact with Hsp70 or Hsp90 (Odunuga *et al.*, 2003; Odunuga *et al.*, 2004). The most important feature of these TPR-containing co-chaperones is their ability to regulate

the activity of Hsp70 and Hsp90 (Smith, 2004). For example, certain TPR domain-containing co-chaperones also play a vital role in Hsp70 / Hsp90 complexes that are involved in steroid receptor maturation e.g. Hip, Hop, Chip and CyP40 (Smith, 2004). These co-chaperones possess the ability to bind to either Hsp70 (e.g. Hip) or Hsp90 (e.g. Cyp40) (Smith, 2004). Certain TPR-containing co-chaperones, such as Hop and its homologues are able to interact with both Hsp90 and Hsp70. This is accomplished when Hsp70 binds to the N-terminal TPR repeat domain of Hop while Hsp90 binds to the first central TPR repeat domain simultaneously (Lässle *et al*., 1997; Odunuga *et al*., 2003; Odunuga *et al*., 2004).

In co-chaperones such as TPR-containing Hsp40s and Hop, the TPR domain is a key mediator of Hsp binding, although regions separate from or adjacent to the TPR domain can also influence binding (Smith, 2004). It has been shown that the TPR domains of Hop bind to the EEVD consensus sequence at the C-terminus of Hsp70 and Hsp90 (Chen *et al*., 1998; Scheufler *et al*., 2000). General binding is conferred by the side chains of certain conserved residues in the TPR groove, also known as carboxylate clamp residues (Scheufler *et al*., 2000; Figure 1.4). These carboxylate clamp residues interact with the negatively charged side chains of the EEVD region (Scheufler *et al*., 2000; Figure 1.4). Unique amino acids that occur adjacent to the EEVD region as well as side chains of less conserved residues in the TPR domain grooves confer specificity to the interactions between TPR1 and GPTIEEVD as opposed to TPR2A with MEEVD in Hop (Odunuga *et al*., 2003). However, it is important to note that this is not always the case, as some TPR-containing co-chaperones such as Hip do not bind to the EEVD consensus sequence of Hsp70 (Höhfeld *et al*., 1995).

Minimal protein-refolding assays have revealed that several of these TPR-containing co-chaperones have independent chaperone activity, although the physiological relevance of this independent chaperone activity is not known (Freeman *et al*., 1996). Although the TPR domains of TPR-containing co-chaperones are key mediators of binding to Hsp70 and Hsp90, the role of regions adjacent to, or further removed from, the TPR domains themselves must not be ignored (Smith, 2004).

1.4.3. THE HSP70 PROTEIN FAMILY

Proteins of the Hsp70 family are the most highly conserved proteins across species (Lindquist and Craig, 1988; Gupta and Singh, 1994; Hunt and Morimoto, 1985). Hsp70s occur in all known species including archaebacteria, bacteria (referred to as DnaK), plants and humans (Daugaard *et al*., 2007). These proteins are critical to cellular viability and preside over a number of crucial roles in cellular homeostasis. Some of the most important house-keeping functions of Hsp70 proteins include protein transport between different cellular compartments, the degradation of unstable and / or misfolded proteins, prevention of protein aggregation as well as the dissolution of protein aggregates, the folding and refolding proteins and the control of various regulatory proteins (Daugaard *et al*., 2007).

All eukaryotic species possess several genes that encode Hsp70s. *Saccharomyces cerevisiae* (yeast) possesses eight Hsp70 homologs (Werner-Washburne *et al*., 1989). Six of these Hsp70s are localised to the yeast cytosol, while the remaining two are localised to the other cellular compartments, viz. the mitochondria and the E.R. (Werner-Washburne *et al*., 1989). Some of the Hsp70s in yeast (the four Ssa proteins) can compensate for each other‟s function, while their simultaneous deletion has been found to be lethal (Werner-Washburne and Craig, 1989). Similarly, the human genome encodes 13 members of the HSPA (Hsp70) family, excluding the many pseudogenes (Tavaria *et al*., 1997; Brocchieri *et al*., 2008; Vos *et al*., 2008; Kampinga *et al*., 2009). In terms of parasitic Hsp70s, the malaria parasite, *Plasmodium falciparum* (*P. falciparum*) possesses 6 Hsp70 proteins that also include cytosolic, E.R. and mitochondrial homologs (Shonhai *et al*., 2007).

1.4.3.1. DOMAIN ORGANISATION OF HSP70S

The Hsp70 protein family members recognise non-native protein domains as substrates (Mosser *et al*., 2000). These domains may be exposed during protein synthesis and folding, translation, membrane translocation of proteins, oligomerisation of proteins and ultimately during protein degradation pathways (Mosser *et al*., 2000). The constitutive form of this protein family, the heat shock

cognate protein 70 (Hsc70), is highly abundant in the cytoplasm and nucleus of cells and is involved in maintaining cellular homeostasis. Hsc70 is aided in its task by its homolog, Hsp70, which is highly inducible (Bukau and Horwich, 1998). Hsp70 induction and synthesis is controlled by the amount of non-native protein (substrate) that is present in the cell at any given time (Bukau and Horwich, 1998).

The levels of expression of Hsp70 proteins within a given organism can vary considerably depending on the organelle and the stage of the cell cycle during which they are upregulated (Cascardo *et al.*, 2001). Peptides (non-native substrates) bind a 15 kDa region that occurs at the C-terminal region of Hsp70 as indicated in Figure 1.5 (Mosser *et al.*, 2000). In terms of its crystal structure, the substrate binding domain of Hsc70 consists of an α-helical lid and a β-sheet base (Jiang *et al*., 2005). The movement of the lid region is mediated by means of conformational changes triggered by ATP hydrolysis in the ATPase domain (Mayer and Bukau, 2005; Popp *et al*., 2005). Hsp70s are able to bind to extended polypeptides that contain a core of 4-5 residues enriched in leucine, isoleucine, valine, phenylalanine and tyrosine residues (Rudiger *et al*., 1997). In the presence of bound ATP, and in the absence of a partner Hsp40 protein, the Hsp70 lid remains „open" which enables transient interactions with any substrate proteins, but prohibits Hsp70s from actively binding target proteins for any given length of time (Laufen *et al*., 1999). When a given Hsp70 protein interacts with a high-affinity substrate and an adjacent J-domain of a partner Hsp40 protein, ATP hydrolysis is triggered and the substrate becomes tightly bound to the ADP-bound form of Hsp70 with a concomitant 1000-fold decrease in the dissociation rate of binding (Laufen *et al*., 1999). It is important to note that the activity of Hsp70s can be regulated by other co-chaperones in addition to Hsp40s, and that these co-chaperones are able to target Hsp70s to appropriate substrates, i.e. misfolded or specific native proteins (Hennessy *et al*., 2005). The C-terminal domain of Hsp70 also features a 10 kDa region that contains the highly conserved EEVD sequence which is essential for Hsp70 binding and association to co-chaperones of the Hsp40 family as well as Hop (Demand *et al*., 1998). This sequence is conserved in all known eukaryotic Hsc70s and Hsp70s, and mutation thereof prevents the stable association of non-native substrates (Demand *et al*., 1998).

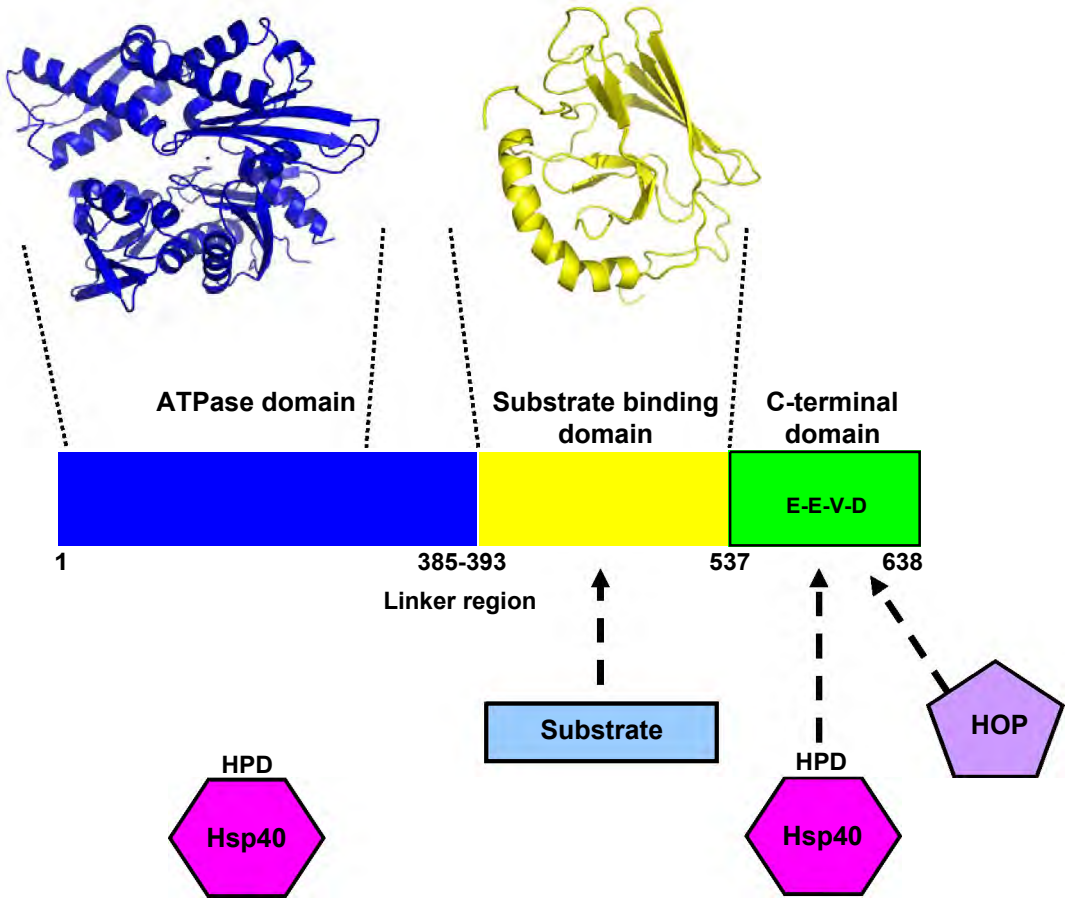


Figure 1.5. Diagram of the domain organisation in Hsp70

Hsp70 proteins consist of three domains: the ATPase domain (44kDa), the substrate binding domain (15kDa) and the C-terminal domain (10kDa). The J domain of Hsp40 co-chaperones interacts with the ATPase domain of Hsp70. The C-terminal domain contains the EEVD region that facilitates the binding of co-chaperones such as Hsp40 and Hop. The linker region (residues 385-395) joins the ATPase domain to the substrate-binding domain. Figure adapted from Bukau and Horwich (1998) and Demand *et al.* (1998). The ribbon representations of the crystal structure were rendered using PyMol software (DeLano, 2002). The ATPase domain model shown (1HJO) is from human Hsp70 and the substrate binding domain (1BPR) is from *E.coli* DnaK.

In addition to these conserved domains, certain Hsp70s contain localisation signals at their N-termini and retention signals at their C-termini which target them to the appropriate organelle for their specific functions; for example BiP which is retained in the E.R. by a C-terminal KDEL sequence (Daugaard *et al.*, 2007).

1.4.4. THE HSP40 PROTEIN FAMILY

As previously mentioned, Hsp40 proteins (DnaJ in prokaryotes) primarily function as co-chaperones to Hsp70 proteins. They stimulate the ATPase activity of their partner Hsp70s by stimulating ATP hydrolysis (Hennessy *et al*., 2000). In addition to acting as Hsp70 co-chaperones, Hsp40s have also been implicated in translocation of proteins (Feldheim *et al*., 1992), protein degradation (Jubete *et al*., 1996), clathrin uncoating (Jiang *et al*., 1997) as well as viral infection (Campbell *et al*., 1997). Despite such a vital role in all the processes above, excessively high amounts of Hsp40 have been found to decrease the viability of cells in *E. coli*, and it has subsequently been proposed that high levels of Hsp40 are toxic to cells (Nicoll *et al*., 2006).

The proteins of the Hsp40 heat shock protein family are characterised by the presence of a highly conserved J-domain of approximately 78 amino acid residues that bears similarity to the first 73 amino acids of *E. coli* DnaJ (Laufen *et al*., 1998). This J-domain is involved in stimulating ATP hydrolysis in Hsp70 proteins, and hence in regulating the chaperone function of Hsp70 proteins (Hennessy *et al*., 2000). Furthermore, the J-domain is proposed to be the main site of interaction between Hsp40 proteins and Hsp70 proteins (Huang *et al*., 1999).

Nuclear magnetic resonance (NMR) studies have assisted in elucidating the structure of two Hsp40 proteins, namely *E. coli* DnaJ and human Hdj1 (Huang *et al*., 1998). These studies have revealed that the J-domain contains 4 helices that are joined by a loop region containing a highly conserved triad of amino acids (histidine, proline and aspartic acid) that is known as the HPD motif (Qian *et al*., 1996). The HPD motif is found between helices II and III and is shown in Figure 1.6. Helices II and III of the J-domain are antiparallel and amphipathic (Qian *et al*., 1996). The His^{33} residue in the HPD motif is a critical residue required for Hsp40 and Hsp70 interaction (Mayer *et al*., 1999). Studies in which this His residue of the J-domain has been substituted for another amino acid have shown abrogation of Hsp40-Hsp70 interaction and subsequent loss of functional interaction between the two proteins (Laufen *et al*., 1999).

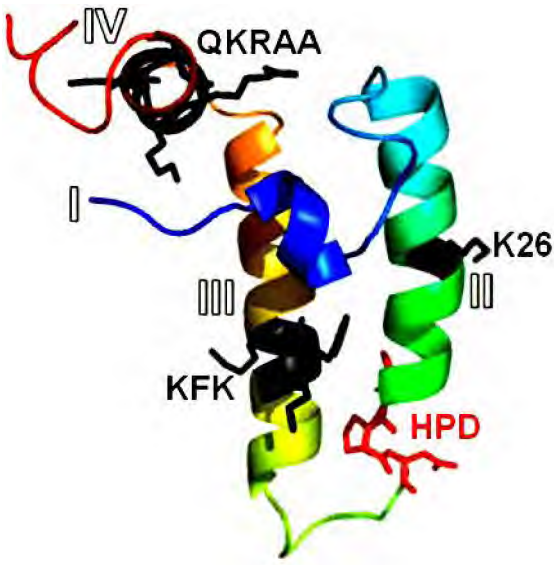


Figure 1.6. Helical structure of the J-domain of E. coli DnaJ

The helices (I-IV) are indicated on the figure and the HPD motif is highlighted in red sticks. The location of residues critical for function is highlighted in black on the figure. This ribbon representation of the J-domain of E. coli DnaJ (1 XBL) was rendered using Pymol software (DeLano, 2002).

The KFK motif in Helix III is also conserved in certain Hsp40 proteins, and an F47A mutation has been found to result in a loss of function *in vivo* (Genevaux *et al*., 2002; Hennessy *et al*., 2005; Nicoll *et al*., 2007). In addition to the highly conserved HPD tripeptide, Helix II also plays a crucial role in the interaction of the J-domain with Hsp70 proteins by binding to an acidic groove that is located in the ATPase domain of a given partner Hsp70 protein (Langer *et al*., 1992; Bukau and Horwich, 1998). Broadly speaking, Hsp40s can be divided into 3 different classes (I-III) (Figure 1.7) based on how many of the canonical *E. coli* DnaJ domains occur in the Hsp40 being studied (Hennessy *et al*., 2000). *E. coli* DnaJ consists of 4 canonical domains: the first 73 amino acids of the protein comprise the J-domain followed by a glycine-phenylalanine (G/F)-rich region, a region containing several repeats of cysteine and glycine in the order CXXCXGXGX where X is any amino acid, and a variable C-terminal domain (Hennessy *et al*., 2000). The cysteine repeat region folds in a zinc-dependent fashion with two repeats bound to one zinc ion, while the highly variable C-terminal domain is known to form a sheet–like structure that is involved in the dimerization of Hsp40 proteins (Szabo *et al*., 1996; Szyperski *et al*., 1994).

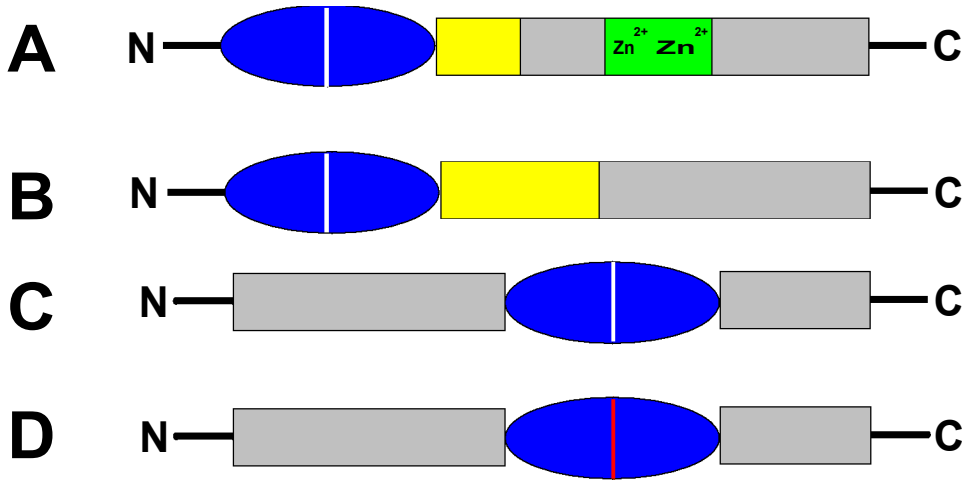


Figure 1.7. The domain organisation of Hsp40 (DnaJ-like) proteins showing the differences in organisation between Type I (A), Type II (B), Type III (C) and Type IV (D) Hsp40 proteins.

The domains of Hsp40s include the well conserved J-domain found in all Hsp40s, the G / F rich region found in Type I and Type II Hsp40s, and the CXXCXGXG region found only in Type I Hsp40s. Type I Hsp40s contain all three domains, while Type III Hsp40s contain only the J-domain, which can occur anywhere on the protein. The white line indicates the conserved HPD motif, while a red line indicates the absence of an HPD motif, which occurs in Type IV Hsp40s. Adapted from Cheetham and Caplan (1998) and Botha et al. (2007).

The cysteine-rich region and C-terminal domain regions of Hsp40 proteins are typically involved in substrate binding and presentation to a partner Hsp70 (Li *et al.*, 2003). The glycine/phenylalanine-rich (G/F) region is found adjacent to the J-domain in both Type I and Type II Hsp40s and is a disordered region that has been suggested to lend flexibility to these proteins (Karzai and MaMacken, 1996).

Cheetham and Caplan proposed a division of Hsp40 proteins that is based on their similarity to *E.coli* DnaJ as shown in figure 1.7 (Cheetham and Caplan, 1998). Type I Hsp40s are similar to DnaJ across all 4 canonical domains, while Type II Hsp40s are similar only in terms of the J-domain and the G/F-rich region (Cheetham and Caplan, 1998). The C-terminal domain of Type I and Type II Hsp40s binds to protein substrates. Type I Hsp40s have also shown independent chaperone activity in that they inhibit the denaturation and aggregation of proteins (Langer *et al.*, 1992). Type III Hsp40s have no similarity to *E.coli* DnaJ other than the J-domain (Cheetham

and Caplan, 1998). They are the most poorly conserved of all the Hsp40 proteins, and vary greatly in terms of their size, sequence and structure. Type III Hsp40s are highly specialised and have specific functions (Nicoll *et al*., 2007).

It is interesting to note that while Type I and Type II Hsp40s only differ in terms of the presence / absence of a zinc finger domain, they exhibit marked differences in terms of their chaperone ability (Banecki *et al*., 1996; Szabo *et al*., 1996). The reason for this difference has been attributed to the observation that Type I and Type II Hsp40s have evolved structurally distinct polypeptide binding domains (Banecki *et al*., 1996; Szabo *et al*., 1996). This observation is based on biochemical and structural studies which suggest that Type I Hsp40s utilize undefined amino acid residues that lie within the zinc-finger-like region as well as an adjacent domain to bind and subsequently deliver non-native proteins to a partner Hsp70 (Lu and Cyr, 1998). X-ray and NMR studies have assisted in the elucidation of the structure of the helices that make up a J-domain (Nicoll *et al*., 2007). Type I Hsp40s have been shown to possess a short Helix I region, and, despite the presence of a number of highly conserved hydrophobic residues, the tertiary structure of Helix I is divergent in Type II and Type III Hsp40s (Nicoll *et al*., 2007). Both Helix II and Helix III are well conserved in terms of their structure across all known Hsp40 J-domains. The largely positively charged Helix II is thought to interact with the negatively charged underside of the ATPase domain of a partner Hsp70 protein (Mayer *et al*., 1999). Mutagenesis studies have revealed a number of residues other than the highly conserved HPD motif that are essential for Hsp40 interaction with Hsp70 (Nicoll *et al*., 2007). In contrast to Type I Hsp40s, crystallographic studies suggest that Type II Hsp40s function as divalent homodimers that possess a clamp-like architecture and utilize a shallow groove that lies on the surface of monomers to bind to non-native proteins (Sha *et al*., 2000). The great diversity in terms of primary sequence identity and secondary structure of Type III Hsp40s renders any predictions of a generalised mode of action difficult.

The HPD motif was initially thought to occur in all known J-domains except for those of the ring-infected erythrocyte surface antigen proteins (RESA) from the malaria parasite, *P. falciparum* (Bork *et al*., 1992). Recent studies, however, have indicated that several Hsp40s in parasitic species have no HPD motif in their J-domains, which led to the description of a new class of Hsp40 protein, the Type IV Hsp40s (Botha *et*

al., 2007). These Type IV Hsp40s display variations in the highly conserved HPD motif that occurs in all known J-domains (Botha *et al*., 2007). As it is known that mutation of the HPD motif causes an abrogation in binding of Hsp40 to Hsp70, it is proposed that Type IV Hsp40s exert their function via different mechanisms to that of the Type I – Type III Hsp40s (Botha *et al*., 2007). Walsh and colleagues proposed that proteins containing J-domains with non-conservative substitutions in their HPD motifs imitate conserved HPD-motif containing J-domains, which results in a more complex regulation of Hsp70s (Walsh *et al*., 2004). Type IV Hsp40s are abundant in the malaria parasite, *P. falciparum*, and are also known to occur in *T. brucei* (Botha *et al*., 2007). Their function in these parasites is currently unknown but is proposed to be highly specialised. Of great interest, then, is the characterisation of the function of these Type IV Hsp40s and how they interact with Hsp70s, if at all.

1.4.4.1. Type III Hsp40s

Type III Hsp40s are proposed to play specialised roles within organisms. These proteins display a high degree of variability in terms of their overall sequences, and often within their J-domains as well (Schlendstedt *et al*., 1995). In spite of the overall variability found within this group of chaperones, the HPD motif is well-conserved throughout the Type III Hsp40s (Schlendstedt *et al*., 1995). Type III Hsp40s with highly specialised functions include the Arg1 (altered response to gravity) protein from *Arabidopsis thaliana*, Sec63 from *Saccharomyces cerivisiae*, auxilin (*Bos Taurus*), T antigens (polyomaviruses) and Djla (*E. coli, C. burnetti* and *H. inlfuenzae*) (Sedbrook *et al*, 1999; Brodsky and Pipas, 1998; Greener *et al*., 2000; Clarke *et al*., 1996). The prediction of function in Type III Hsp40s is rendered difficult by the high degree of sequence variability outside the J-domain that these proteins demonstrate (Schlendstedt *et al*., 1995). Interestingly, the degree of specialisation of Type III Hsp40s appears to extend even to the J-domain level, as experiments by Schlendstedt and colleagues have illustrated (Schlendstedt *et al*., 1995). In these experiments, it was shown that substitution of the J-domain of a Type III E.R. luminal-facing membrane-bound Hsp40 (the yeast Sec63) with that of the cytosolic Sis1 resulted in a non-functional protein (Schlendstedt *et al*., 1995). Similarly, the J-domain of another Type III E.R. luminal-facing membrane-bound Hsp40 (mouse

Erdj1) could not substitute for the J-domain of a cytosolic prokaryotic DnaJ (Nicoll *et al*., 2007).

Biochemical characterisation of Djla, a Type III Hsp40 from *E.coli* has shown that the Djla J-domain can substitute fully for the *E.coli* DnaJ J-domain, despite the low overall sequence similarity of the proteins (Genevaux *et al*., 2001). Genevaux and colleagues also discovered that the purified cytoplasmic fragment of Djla is capable of stimulating the ATPase activity of DnaK to similar levels as found for DnaJ stimulation in addition to being able to act as a DnaK co-chaperone in the aggregation suppression of luciferase *in vitro* (Genevaux *et al*., 2001). Despite the information available on Djla, and other Type III Hsp40s, the lack of overall sequence conservation in these proteins, renders any comparative analyses difficult, especially when studying novel Type III Hsp40s. However, the collective data from the studies outlined above suggests that the J-domains from Hsp40 proteins, irrespective of type, can substitute for one another if they interact with a similar Hsp70 protein.

1.4.5. THE HSP40 / HSP70 INTERACTION

The ATPase activity, and hence affinity for substrate, of Hsp70 proteins is modulated by co-chaperone proteins such as Hsp40s and nucleotide exchange factors (Hennessy *et al*., 2005). In the *E.coli* system, nucleotide exchange is facilitated by GrpE (Liberek *et al*., 1991). GrpE homologs have not been found in the cytosol of eukaryotes, but have been found in the mitochondria of eukaryotes (Bolliger *et al*., 1994). Other nucleotide exchange factors such as Bag1 and HspBP1 also play a role in the eukaryotic Hsp70/Hsp40 partnership, but have different mechanisms of action compared to GrpE (Kabani *et al*., 2003; Shomura *et al*., 2005). In addition to stimulating the ATPase activity of Hsp70, Hsp40 proteins have also been shown to bind to a given substrate and target it towards the appropriate Hsp70, hence conferring specificity to the reaction outlined in Figure 1.8 (Han and Christen, 2003).

The hydrolysis of Hsp70 ATP to ADP facilitates the tight binding of a substrate protein, which leads to the formation of an Hsp70-substrate protein complex that can be further stabilised by the Hsc70-interacting protein (Hip) (Höhfeld *et al*., 1995).

This added stabilisation allows for enough time for the client protein to fold into its optimal native conformation (Höhfeld *et al.*, 1995). Substrates undergo repeated cycles of binding and release during the Hsp70-assisted protein-folding pathway (Bukau and Horwich, 1998). The stoichiometry of this process is frequently one molecule of Hsp70 monomer to one molecule of substrate protein (Bukau and Horwich, 1998). Hsp40s display a certain degree of specificity to a given Hsp70, both within the same organism and in other organisms (Nicoll *et al.*, 2007). An example of this is that mammalian Hsc70 is stimulated by *E.coli* DnaJ but DnaK is not stimulated by the human Hdj1 (Minami *et al.*, 1996). It has been proposed that the J-domain contains sequences or motifs and structural features that confer a certain degree of specificity in its interactions with Hsp70s (Hennessy *et al.*, 2005). As discussed above, it would appear that J-domains are interchangeable if they originate from proteins that are known to bind homologous Hsp70s (Genevaux *et al.*, 2001; Nicoll *et al.*, 2007). J-domains that originate from proteins that are involved in diverse cellular processes are less interchangeable (Schlenstedt *et al.*, 1995; Nicoll *et al.*, 2007). This suggests that two different types of J-domain have evolved to meet the needs of protein folding in the cell. The first type of J-domain (generally in Type I Hsp40s) interacts specifically with Hsp70 that facilitate general protein folding pathways, while the second type (generally Type III Hsp40s) interacts specifically with Hsp70s that facilitate more specialised processes in the cell (Nicoll *et al.*, 2007).

No systematic J-domain swapping experiment of the interchangeability of J-domains from Type I – Type IV Hsp40s within different cellular compartments has been attempted (Nicoll *et al.*, 2007). J-domain swapping experiments aim to indicate which elements in the J-domain are involved in conferring specificity to a given Hsp70 and could provide valuable information when a given Hsp40 is being considered as a drug target in mammalian systems. This is due to the fact that a knowledge of binding partners is essential, so that drugs are developed that can inhibit a vital parasitic or bacterial cycle without compromising the health of the human host.

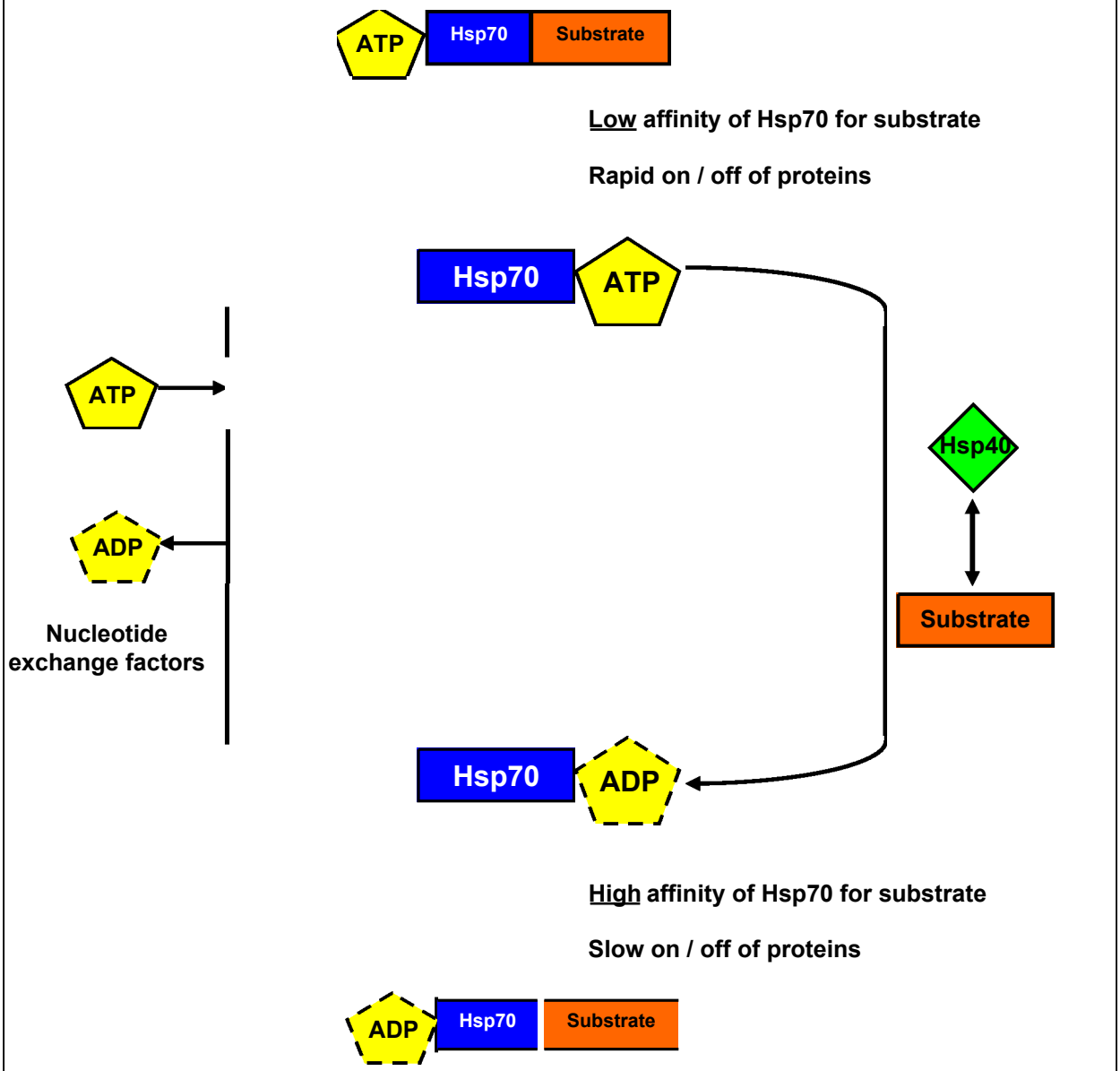


Figure 1.8. The ATP dependent cycle of Hsp70 substrate binding

ATP-bound Hsp70 has a low affinity for protein substrates. Hsp40 co-chaperones mediate the hydrolysis of ATP to ADP, which causes Hsp70 to have a high affinity for substrate and a slower on / off rate for proteins. This process is facilitated by Hsp40 proteins in the presence of nucleotide exchange factors. Figure taken and adapted from Bukau and Horwich (1998).

1.5. HEAT SHOCK PROTEINS IN KINETOPLASTIDS

The study of molecular chaperones in parasitic systems greatly enhances the current understanding of the evolution of these proteins. This is due to the fact that parasitic

protozoa of the order kinetoplastida are thought to be among the most primitive organisms in the eukaryotic kingdom and are as distant (on an evolutionary scale) from yeast as they are from vertebrates (Sogin *et al*., 1989). This places great value on the study of the chaperone machinery in *T. brucei* and *T. cruzi* which belong to the order kinetoplastida.

Parasitic protozoa undergo extensive changes during their lifecycles (Van der Ploeg *et al*., 1985). In the case of *T. brucei*, these changes includes the loss of the protective parasitic coat upon entering the tsetse fly vector, extensive morphological changes, as well as differences in the type of metabolism that is favoured in the vector and host (Van der Ploeg *et al*., 1985). The *T. brucei* parasite undergoes a rapid and drastic change of environmental conditions in terms of temperature, pH, metabolism and respiration during its transition from vector to host. This heat shock phenomenon is also observed in the *T. cruzi* parasite (Requena *et al*., 1992). Heat shock proteins are synthesised in response to these marked changes in environment and are required to assist the parasite to make the transition to its mammalian host successfully whilst remaining viable and infective (Olson *et al*., 1994). A review of the current literature indicates that Hsps, including Hsp70s, occur in various parts of the *T. brucei* parasite. Proteomics analyses indicate that Hsp60, Hsp70, mitochondrial Hsp70 and Hsp90 are upregulated during different stages of the parasite lifecycle (Bente *et al*., 2003; Paba *et al*., 2004). It is not known, however, whether these heat shock proteins are expressed in order to facilitate the parasite‟s differentiation or whether it is simply the natural response of parasites undergoing stress due to rapid environmental changes (Folgueira and Requena, 2007). Very little biochemical characterisation has been performed on trypanosomal heat shock proteins, with the most attention being given to the study of the Hsp70s of *T. cruzi* (Olson *et al*., 1994; Tibbetts *et al*., 1994).

1.5.1. HSP70S IN KINETOPLASTID PARASITES

The Hsp70 family is the most well-studied family of chaperones in *T. cruzi* (Olson *et al*., 1994). One cluster of *T. cruzi* Hsp70 encoding genes is arranged in a tandem array of 10 genes that are transcribed as a polycistronic stretch of mRNA (Requena

et al., 1988). These Hsp70 proteins are some of the major immunogens that are detected during *T. cruzi* infection; their expression is increased 4-fold during the temperature shift from 26°C to 37°C (Engman *et al*., 1990). It is interesting to note that both *T. cruzi* and other kinetoplastid parasites display stage-specific expression of Hsp70 proteins with increased levels of expression being detected during the parasite‟s lifecycle stage in the mammalian host (Olson *et al*., 1994). This observation could prove to have important implications in studying parasite infectivity. In addition, it has been shown that a kinetoplastid parasite closely related to *T. cruzi* and *T. brucei*, *L. major* exhibits increased infectivity when exposed to elevated temperatures (Smejkal *et al*., 1988). Hsp70 transcripts have been found to be more numerous in blood stream trypanosomes (within the human host) than in procyclic trypanosomes (within the tsetse fly) (Lee *et al*., 1990). This could serve to indicate that Hsp70 plays a role in protecting the parasite against the stresses encountered upon entering the markedly different environment that its mammalian host presents after development within an insect vector.

In *T. brucei*, higher levels of Hsp70 expression are observed when the parasite is cultured under conditions that are similar to those occurring in the mammalian host. This upregulation in Hsp70 expression is independent of temperature, and could potentially result from other stress-related influences (Van der Ploeg *et al*., 1985). The *T. brucei* homolog of BiP (E.R. resident of Hsp70 family) has been characterised, in addition to the *T. cruzi* homolog, which is known as Grp78 (Bangs *et al*., 1993; Tibbetts *et al*., 1994). A mitochondrial Hsp70 (mtHsp70) has also been identified and the gene cloned from *T. cruzi* (Engman *et al*., 1989). A cytoplasmic, inducible Hsp70 (TcHsp70) from *T. cruzi* has been identified. This protein was implicated in parasite cytoprotection in response to the heat shock experienced during the transition from an insect vector to a mammalian host (Edkins *et al*., 2004). In addition to its chaperone activity, TcHsp70 has also been identified as one of the immunodominant antigens that occur in patients that are infected with the *T. cruzi* parasite (Arif *et al*., 1998). It is interesting to note that while *T. brucei* has a constitutive Hsp70 protein, *T. cruzi* does not appear to possess an equivalent (Van der Ploeg *et al*., 1985).

The gene for mtHsp70 has also been cloned from *L. major* (Searle *et al*., 1993). As was the case in *T. cruzi* (10 tandem Hsp70s), 5 Hsp70 genes in the *L. major* are

located in tandem on the same chromosome (chromosome 30) and a similar cluster of Hsp70 genes is found in the *T. brucei* genome (Folguiera and Requena, 2007). An Hsp70 gene locus of *T. brucei* contains five genes that are arranged in a direct tandem repeat. Approximately 200 base pairs exist between the polyadenylation site of one gene and the trans acceptor site of the next gene. These five Hsp70 genes are preceded by an open reading frame and a sixth copy of an Hsp70 gene (Glass *et al*., 1986; Bock and Langer, 1993).

Several studies have shown that Hsp70 transcripts vary according to the life cycle stage of the parasite. Interestingly, the most in-depth work on the TriTryp heat shock protein complement to-date, a review by Folguiera and Requena, indicates that *T. brucei* has 12 Hsp70 proteins, while *T. cruzi* possesses the same number of Hsp70s if proteins with partial sequences are not taken into account (Folguiera and Requena, 2007). If, however, one takes into account the partial sequences given on the *T. cruzi* genome, the total number of Hsp70s increases to 28 proteins, with multiple proteins in *T. cruzi* for each canonical Hsp70 subgroup (Folguiera and Requena, 2007). By contrast, the malaria parasite, *P. falciparum*, possesses only 6 Hsp70 proteins (Shonhai *et al*., 2007). *L. major*, the third member of the TriTryp parasite group, possesses 14 Hsp70 proteins, most of which are situated in the mitochondria (Folguiera and Requena, 2007).

1.5.2. HSP40 PROTEINS IN KINETOPLASTID PARASITES

Numerous Hsp40 proteins exist in trypanosomal parasites. Because more than one Hsp40 can interact with a given Hsp70 and confer specificity to it / regulate its function, Hsp40s in a given organism are more numerous than Hsp70s (Cyr *et al*., 1992). Despite the importance of *T. brucei* as a human pathogen, only *T. cruzi* Hsp40s have been biochemically characterised to date (Folgueira and Requena, 2007). Four novel cytoplasmic Hsp40s in *T. cruzi* have been identified which were named Tcj1 – Tcj4, while later studies identified a fifth Hsp40 named Tcj6 as well as a mitochondrial Hsp40 named TcDJ1 (Carreira *et al*., 1998; Salmon *et al*., 2001; Tibbetts *et al*., 1998). Tcj2, Tcj3 and Tcj4 are Type I Hsp40 homologues, Tcj6 is a Type II Hsp40 homologue and Tcj1 is a Type III Hsp40 (Tibbetts *et al*., 1998; Salmon

et al., 2001). Tcj1 lacks the CxxCxGxG zinc finger domains that are involved in the peptide substrate binding of certain Hsp40 proteins which could have marked implications in the mechanism of this protein as well as its function within the *T. cruzi* parasite (Tibbetts *et al*., 1998). Tibbetts and colleagues have suggested that Tcj1 possesses a highly specialized function within *T. cruzi* (Tibbetts *et al*., 1998).

1.6. PROTEOMIC ANALYSES OF *T. BRUCEI* AND OTHER TRYPANOSOMAL PARASITES

Proteomic analyses have been used to analyse trypanosomal proteomes in the past. The proteomes studied include those of *T. cruzi*, *T. brucei* and *L. major*. A whole-organism proteomic investigation of the four morphologically distinct forms of the *T. cruzi* parasite was undertaken in 2005 (Atwood *et al*., 2005). This study revealed that different amounts of protein are recovered from the *T. cruzi* parasites; depending on which stage of the life cycle they are at when sampled. Approximately 5000 unique peptides were matched to over 1000 protein groups that contained 2784 proteins (Atwood *et al*., 2005). The *T. cruzi* genome contains identical copies of numerous genes, which includes a large number of gene families that contain hundreds of distinct members (El-Sayed *et al*., 2005). Many of the proteins that were identified in the study by Atwood and colleagues had orthologs in the *T. brucei* genome (Atwood *et al*., 2005). It is interesting to note that a total of 61 heat shock proteins (molecular chaperones) were identified out of 2123 protein spots. Hsps thus account for around 2.87 % of the *T. cruzi* proteome. These were found mostly in the amastigote stage of parasite development, which can be induced by a low pH environment in the laboratory and corresponds to the parasite‟s development in the insect midgut (Atwood *et al*., 2005). The identification of proteins in trypanosomal parasites is facilitated by the complete genome sequence that has recently been made available (Berriman *et al*., 2005; El-Sayed *et al*., 2005; Ivens *et al*., 2005).

Although the proteome of *T. cruzi* has been well established by means of reproducible two dimensional gel electrophoresis methods, little work has been done on the heat shock proteins. The same applies to the *T. brucei* proteome. This is a major gap in the current knowledge; due to the cytoprotective effect of Hsps, it is

reasonable to propose that they play an important role in the ability of trypanosomal parasites to tolerate the markedly differing conditions experienced during the transition from their insect vectors to mammalian hosts. The *T. brucei* proteome is not as well characterised as the *T. cruzi* proteome although several investigations have been conducted. One study focussed on arsenical drug resistance in *T. brucei* parasites, which is one of the major causes of ineffective treatment of patients affected by HAT (Foucher *et al*., 2006). Another proteomic study has examined the differential protein expression of glycosomes from both bloodstream and procyclic forms of the *T. brucei brucei* parasite (Colasante *et al*., 2006). The differential protein expression of midgut proteins in trypanosome-susceptible and non-susceptible tsetse flies has also been investigated using proteomic techniques (Haddow *et al*., 2005). Other proteomic studies in *T. brucei* have focussed on individual proteins such as PUF1 or protein clusters such as the 20S Proteasome (Huang *et al*., 2001; Luu *et al*., 2006). Despite the existence of a reproducible protocol for two dimensional gel electrophoresis (2DE) in *T. brucei*, no comprehensive proteomic investigation of the trypanosomal chaperone machinery has yet been conducted (van Deursen *et al*., 2003).

1.7. MOLECULAR CHAPERONES AS DRUG TARGETS

The use of molecular chaperones as novel drug targets is becoming increasingly attractive in terms of modern drug design. Recent studies have indicated that a large number of proteins are dependent on molecular chaperones to maintain their activation-competent conformations (Schulte *et al*., 1997). Drug design based on chaperone targets relies on the development of a type of inhibitor that does not interact with the targeted protein of interest, but rather seeks to inhibit the function of the associated chaperone that enables the protein to function. The direct result of this is that disease-causing proteins are unable to maintain their activation-competent conformations and are degraded by the proteasome (Schulte *et al*., 1997). In contrast to other types of inhibitors, chaperone-based inhibitors are able to target a wide range of proteins and subsequently diminish the level of many protein targets at once (Neckers and Neckers, 2005). In certain cases, such as the desired sensitisation of cancer cells to facilitate recognition by natural killer cells of the native

immune system, the upregulation of chaperones such as Hsp60, Hsp70 and Hsp72 forms part of a potential therapeutic regimen (Sõti *et al*., 2005).

Various molecular chaperones are currently being investigated as novel drug targets in a variety of disorders. These include Hsp90, which is one of the most abundant proteins in cells and is essential for cell survival and is currently being investigated as a drug target for cancer (Sõti *et al*., 2005; Cullinan and Whitesell, 2006). Hsp70, Hsp27, Hsp40 and HSF-1 are additional novel drug targets implicated in a variety of disorders ranging from ulcers and viral infections to diabetes and neurodegeneration (Sõti *et al*., 2005). The potential of chaperones as drug targets in parasitic disorders such as HAT should not be ignored. This would especially be the case if the *T. brucei* chaperone machinery is sufficiently different to that of the human host to ensure that novel drugs do not inactivate essential chaperones in affected patients but serve to inactivate those of the parasite instead. The lack of conservation of Type III and Type IV Hsp40s, and their specialised function within the cell, makes them ideal candidates for novel drug target investigation in parasites such as *T. brucei*.

1.8. THE IMPORTANCE OF STUDYING THE MOLECULAR CHAPERONE COMPLEMENT OF THE TRITRYPS

Because the chaperone machinery is so crucial for the parasite‟s survival in its human host, it follows that a study of parasitic Hsps, particularly the Hsp40/Hsp70 partnership could reveal the existence of novel drug targets and potentially facilitate the development of more effective treatment regimens for HAT. A study involving the Hsp40/Hsp70 partnership in different *T. brucei* subspecies would also serve to enhance the current understanding of fundamental chaperone function and interaction within a parasitic organism.

Three TbHsp40s were selected for further study during the course of the present investigation; two type III Hsp40s and one Type IV Hsp40. Type III Hsp40s were focussed on due to the marked difference in their overall sequences which indicate that they could possess specialised functions in the *T. brucei* and *T. cruzi* parasites. In addition to this, both the *T. brucei* and *T. cruzi* make use of an unusually large

number of Hsp40s in general, but of Type III Hsp40s in particular (Folgueira and Requena, 2007).

Tbj1 (Tb11.01.8750), a putative Tcj1 orthologue, is a Type III Hsp40 protein. A comparative analysis of Tbj1 and Tcj1 could assist in an enhanced understanding of the similarities and differences in chaperone families between related parasites such as *T. brucei* and *T. cruzi*. Tbj47 (Tb927.1.1230) is a Type IV Hsp40 that has been shown to be an essential for *T. brucei* survival by means of RNAi studies but has not yet been biochemically characterised (Internet 1). Tbj51 (Tb927.4.2220) is a Type III Hsp40 protein that contains a TPR-repeat containing domain. Co-chaperones containing TPR domains recognize the EEVD structural motif common to both Hsp90 and Hsp70 through a highly conserved clamp domain, indicating that this protein could be a partner to both these chaperones in the *T. brucei* system (Carello *et al*., 2004). A summary of the properties of the three TbHsp40s selected for further study is given in Table 1.5.

Table 1.5. Properties of the TbHsp40s selected for further study

Name	Hsp40 Type	MW (kDa)	pI	Domains
Tbj1	III	35.2	6.09	J-domain
Tbj47	IV	52.3	8.70	Atypical J-domain
Tbj51	III	80.4	8.78	J-domain and TPR domains

PROBLEM STATEMENT

Human African Trypanosomiasis (HAT) and Chagas disease are two of the leading causes of mortality in the developing world. These disorders are caused by infection with the *T. brucei* and *T. cruzi* parasites, respectively. Both parasites have complex lifecycles that are characterised by movement from an insect vector to a mammalian host and the ability to survive drastic changes in their environment in terms of nutrient status, temperature, pH and metabolic pathways. Hsp70 proteins comprise one of the major chaperone families in organisms and are generally induced in response to stress conditions. These proteins are at the centre of a number of fundamental cellular processes that include the folding and assembly of newly synthesised proteins, refolding of aggregated / misfolded proteins, transport of proteins as well as proteolytic degradation of unstable proteins. Hsp70 proteins also control the activity of a number of regulatory proteins. In order to perform all these vital tasks within the cell, Hsp70s require the assistance of co-chaperone partner proteins in the form of Hsp40 proteins.

The partnership between Hsp70 and Hsp40 molecular chaperones is thought to facilitate the transition of *T. brucei* and *T. cruzi* from their insect vectors to their mammalian hosts. However, little information is available on the Hsp40 and Hsp70 complement of the *T. brucei* and *T. cruzi* parasites and most of these proteins have not been biochemically characterised. It is proposed that a study of the Hsp complement of *T. brucei* could result in improved knowledge of *T. brucei* infectivity, as well as the role chaperones play in this aspect. It will also result improved knowledge of the fundamentals of chaperone interactions within parasites.

HYPOTHESIS

Trypanosomal Type III and Type IV Hsp40s are not typical co-chaperones of trypanosomal Hsp70s.

AIMS AND OBJECTIVES OF THE RESEARCH

Broad Aims:

1. **To develop an extensive bioinformatics knowledgebase of the Hsp70 and Type III Hsp40 heat shock protein complement of the TriTryps parasites**
2. **To bioinformatically and biochemically characterise two Type III and one Type IV Hsp40 molecular chaperones from *T. brucei***

Specific aims:

a) Development of a database of TriTryp Type III and Type IV Hsp40s as well as Hsp70 proteins for the identification of novel Hsp40 proteins that are potentially important for parasite survival under stress conditions
b) Cloning of three novel TbHsp40 proteins
c) Establishment of a heterologous expression systems for Tbj1, Tbj47 and Tbj51
d) Purification and biochemical characterisation of Tbj1

CHAPTER 2

Bioinformatic analysis of the TriTryp Type III Hsp40 and Hsp70 protein complement

2.1. INTRODUCTION

2.1.1. Bioinformatics

Bioinformatics is a term coined in 1978 by Paulien Hogeweg for a multidisciplinary field that applies information technology to molecular biology (Moore, 2007). It combines the fields of computer science, biostatistics and biomedical / biological sciences and enhances experimental findings in allowing researchers to make predictions by comparing data obtained in the laboratory to data in public databases (Boguski, 1994; Moore, 2007). Bioinformatics as an independent discipline has developed dramatically since the 1970's when high-throughput DNA sequencing became commonplace, largely due to the advent of novel technologies in the DNA sequencing and computing world (Boguski, 1994). Bioinformatics serves to complement experimental data by facilitating the analysis and interpretation of results in a comprehensive fashion. A thorough bioinformatic analysis also allows for conclusions to be drawn about data and comparisons to be made between different biological systems.

Due to the large volume of sequence data that is currently available, an organized and systematic method of analysis has become necessary; at present this need is being accomplished by means of bioinformatic studies (Moore, 2007). A thorough *in silico* investigation has become a vital part of any molecular biology analysis of protein structure and function; numerous databases exist that can aid in determination of protein function, characterisation of known genes / proteins as well as the generation of three-dimensional homology models of proteins (Baxevanis, 2002; Baxevanis, 2003). The tremendous volumes of data available subsequent to the completion of a genome sequencing project need to be accommodated in a database in order to make it possible to categorise and obtain meaningful output from such data (Moore, 2007). Once a suitable database has been compiled, meaningful data can be obtained from it by searching for biomolecular patterns that correlate to a given phenotypic effect

(Moore, 2007). This can be especially useful when studying disease-causing organisms as molecules that are likely candidates for drug targets can be carefully studied prior to the commencement of *in vitro* and *in vivo* analyses, thus saving a significant amount of time and financial resources and allowing for high throughput research. One of the leaders in the field of Bioinformatics databases and data storage is the National Center for Biotechnology Information (NCBI) at the National Institutes of Health which was created in 1988 in order to develop appropriate information systems for molecular biologists (Sayers *et al*., 2009). The NCBI maintains the GenBank® nucleic acid sequence database which has grown from 606 (1982) sequences to 52 016 762 (2005) (Sayers *et al*., 2009). Table 2.1. shows some of the major Entrez databases and the number of records contained in each as of 30 September 2008 (Sayers *et al*., 2009).

Table 2.1. Number of records contained in selected Entrez databases

Database	Focus of database	Number of entries
Nucleotide	Genes and associated sequences	65 786 674
Protein	Genes and associated sequences	22 337 204
Gene	Genes and associated sequences	4 962 281
UniSTS	Genes and associated sequences	514 624
Taxonomy	Entrez search and retrieval system	460 107
3D domains	Molecular structure and proteomics	246 719
GENSAT	Gene expression	83 553
Structure	Molecular structure and proteomics	52 266
dbGaP	Genotypes and phenotypes	39617
Genome	Genomes	8792
OMIM	Genotypes and phenotypes	19 857

Information taken and adapted from Sayers et al., 2009

2.1.2. Kinetoplastid genome sequence data

The study of kinetoplastid parasites has been greatly aided by the complete genome sequencing of *T. brucei*, *T. cruzi* and *L. major* in 2005 (Kissinger, 2006). A thorough knowledge of the full gene complement of these organisms has facilitated an improved understanding of their biology and subsequently enhances the ability of the modern researcher to identify novel drug targets that are likely to inhibit the parasite survival mechanisms without compromising those of the human host. Despite the clear advantage of having a fully sequenced genome to analyse, various challenges still present themselves during an analysis of kinetoplastid biology. The sheer complexity of trypanosomal parasites and the size of their genomes (Table 2.2.) results in a very large volume of data that needs to be processed, particularly when dealing with proteins that are highly abundant, like molecular chaperones.

Table 2.2. Haploid genome size of Kinetoplastid parasites with fully sequenced genomes

Parasite Name	No. of genes	No. of species specific genes	Genome size (Mb)
Trypanosoma brucei	9068	1392	26
Trypanosoma cruzi	12 000	3736	55
Leishmania major	8311	910	33

Taken and adapted from Kissinger (2006).

The kinetoplastid parasites share 6158 ortholog clusters of genes that encode proteins, in addition to several highly conserved functional protein domains (Kissinger, 2006). This is made all the more remarkable by the fact that these parasites differ markedly in their interactions with their insect vectors and mammalian hosts (Kissinger, 2006). The data and annotations obtained from the whole genome sequencing of the TriTryps (*T. brucei*, *T. cruzi* and *L. major*) can be found at the GeneDB website (http://www.genedb.org/) which comprises a

collection of databases including the full sequence of 27 genomes in various stages of curation (Hertz-Fowler *et al*., 2004).

2.1.3. The GeneDB project

The GeneDB project was funded as part of the Wellcome Trust Functional Genomics Development Initiative, which aimed to develop and maintain databases for several organisms, including the kinetoplastid protozoa, *T. brucei*, *T. cruzi* and *L. major* with the intention of making the annotated sequence data of these organisms available to researchers with a user friendly interface and a number of graphical displays (Hertz-Fowler *et al*., 2004). The GeneDB web database is the result of collaboration between software developers from the Sanger Institute, representatives of research communities and organism specific curators. The data within each organism database are under review with regular updates (Hertz-Fowler *et al*., 2004). Each annotated protein for a given organism has a page accessible through links and a user-friendly search function which contains the following elements: full DNA and protein sequences, link to GeneDB BLAST and NCBI BLAST, predicted properties of the full length protein (signal peptide and transmembrane predictions), similarity information (EMBL, SWISS-PROT including annotations), gene ontology annotation, summary of most recent protein domain and motif searches (InterPro, Pfam, PRINTS, PROSITE, BLOCKS, SMART), literature links (if available) and the full Swiss-Prot annotation (Hertz-Fowler *et al*., 2004).

The availability of some sequence information of the tsetse fly (*Glossina morsitans*) in the GeneDB database is of great significance to researchers in the field of *T. brucei* biology and biochemistry (Hertz-Fowler *et al*., 2004). A summary of some of the other organisms for which annotated genome sequence exists is given in Table 2.3.

Table 2.3. Complete genome sequences of a select number of organisms available on the GeneDB website [1]

Species	Haploid genome size (kb)	Annotation status	Curated
Plasmodium falciparum	22 900	Completed	Yes
Plasmodium chabaudi	30 000	In progress	Yes
Plasmodium berghei	26 000	In progress	Yes
Leishmania major	33 600	Completed	Yes
Leishmania infantum	~ 34 000	In progress	No
Trypanosoma brucei	26 000	Completed	Yes
Trypanosoma cruzi	~ 43 000	Completed	No
Trypanosome congolense	~ 35 000	In progress	No
Trypanosoma vivax	~ 35 000	In progress	No
Glossina morsitans	Unknown	In progress	No

[1]**Information condensed from Hertz-Fowler et al., 2004.**

2.1.4. *In silico* analysis of the TriTryp Hsp40 complement

Upon completion of the genome sequencing of the cerebral malaria-causing parasite, *Plasmodium falciparum*, a number of in-depth works of *in silico* analysis have been published offering a detailed and insightful investigation into the molecular chaperone complement of this organism (Acharya *et al*., 2007; Botha *et al*., 2007; Shonhai *et al*., 2007; Sargeant *et al*., 2006). These studies analysed the heat shock protein complement of *P. falciparum* in terms of their localisation (where data was available), presence in parasite pathways, potential function, interaction with other chaperones, isoelectric point (pI), molecular weight and classification. The information that has been obtained from these investigations

is of great value to researchers in the malaria field and is constantly confirmed and updated with experimental data.

Very few *in silico* investigations have been performed on the heat shock proteins of the TriTryps, and no in depth investigation has been performed on the Hsp40 and Hsp70 proteins of these organisms. A general overview of the heat shock proteins in the TriTryps is given in the work by Folguiera and Requena (2007) which also served to order the *T. brucei*, *T. cruzi* and *L. major* Hsp40s in terms of homologues and provided a less cumbersome nomenclature based on the nomenclature proposed by Tibbetts and colleagues (Folgueira and Requena, 2007; Tibbetts *et al*., 1998; Salmon *et al*., 2001).

Type III Hsp40s in particular have not been well-characterised in any organism due to the high levels of sequence variability displayed by these proteins. A study of Type III Hsp40s is important because of the likelihood that these proteins have specialised functions within organisms. The nature of this specialised function could in part be determined by means of a thorough *in silico* investigation as it is known that subtle structural differences between the J-domains of different Hsp40s determine the nature of their interaction – and indeed their ability to interact with – a given partner Hsp70 protein (Hartl, 1996). In their full genome search of the TriTryps, Folguiera and Requena discovered a very large number of DnaJ-like proteins when a BLAST search was performed, so they restricted their analysis to proteins that had been annotated as DnaJ-like proteins, or having homology to the J-domain sequence (Folgueira and Requena, 2007). This analysis revealed the existence of a large number of distinct Hsp40 proteins in *T. cruzi* (67), *L. major* (66) and *T. brucei* (65) and it was interesting to note that while most of these Hsp40 proteins had orthologues in each species, some were unique to a given species and had no equivalents in the other members of the TriTryps (Folgueira and Requena, 2007). The Hsp40s in *T. brucei* and *L. major* are encoded by single copy genes, but for most proteins in *T. cruzi* this was not the case as the majority of proteins were found to derive

from two genes that have been proposed to represent allelic copies of the CL Brener strain of *T. cruzi* due to the hybrid nature of the strain (El-Sayed *et al*., 2005). Folgueira and Requena did not perform any further *in silico* classification and analysis on the Hsp40s from the TriTryps and did not divide them into types based on the generally accepted chaperone classification given by Cheetham and colleagues (Folgueira and Requena, 2007; Cheetham and Caplan, 1998).

2.1.5. *In silico* analysis of the TriTryp Hsp70 complement

Hsp70 proteins play a fundamental role in coordinating a number of essential cellular processes that include the folding and assembly of newly synthesised proteins coming off ribosomes, the refolding of misfolded and aggregated proteins, membrane translocation of secretory proteins, control of regulatory proteins and the proteolytic degradation of denatured or unstable proteins (Bukau and Horwich, 1998; Mayer and Bukau, 2005). A further indication of their vital importance in maintaining cellular homeostasis is the high levels of conservation of Hsp70s that are found across all organisms (Boorstein *et al*., 1994). The inducible forms of Hsp70 proteins are of the most prominent heat shock proteins that exhibit marked increases in their expression levels upon stress exertion in cells (Lindquist and Craig, 1988). As is the case with their partner Hsp40s, Hsp70s exist in all cellular compartments (Lindquist and Craig, 1988). Folgueira and Requena analysed the Hsp70 complement of the TriTryps in their bioinformatic analysis of trypanosomal heat shock proteins. Hsp70 proteins belonging to all known subdivisions of eukaryotic Hsp70 proteins as defined by their different subcellular localisations (BiP / Grp78 in the E.R., mitochondrial Hsp70, cytosolic Hsp70, Grp170 in the E.R., Hsp110 in the E.R. and others) were identified and the overall gene organisation of Hsp70 proteins in the TriTryps was described (Folgueira and Requena, 2007). *T. brucei* possesses 12 Hsp70 proteins while *L. major* possesses 14 (Folgueira and Requena, 2007). As was the case with the Hsp40 proteins, the situation in *T. cruzi* is more complex due to gene duplication events (Folgueira and Requena, 2007).

The *L. major* mitochondrial Hsp70 complement comprises 5 genes on chromosome 30, four of which clustered in a tandem array (Folgueira and Requena, 2007). A *T. brucei* BiP homologue was characteracterised by Bangs and colleagues and it was shown that this protein localises specifically to the *T. brucei* E.R., while a *T. cruzi* homologue termed Grp78 was characterised in subsequent studies (Bangs *et al*., 1993; Tibbetts *et al*., 1994). The use of the protein sequences that have been characterised as outlined above yielded information about other Hsp70s in the TriTryps when submitted to the BLAST server on the GeneDB database (http://www.genedb.org/genedb/seqSearch.jsp).

Novel Hsp70s predicted to be E.R. Hsp70 orthologues were identified in *L. major* (LmjF28.1200), *T. brucei* (Tb11.02.5450 and Tb11.02.5500) and *T. cruzi* (Tc00.1047053506585.40). These proteins are all predicted to be Grp78 orthologues (Folgueira and Requena, 2007). In addition to the mitochondrial members of the TriTryp Hsp70 family, the cytoplasmic Hsp70s have been a major focus in Hsp70 studies in these organisms (Folgueira and Requena, 2007). The cytosolic Hsp70 locus was characterised in *T. brucei*, which revealed that it comprised 5 genes arranged in tandem with a single gene occurring a few kilobases from the tandemly-arranged genes (Glass *et al*., 1986). The latter was initially proposed to be a pseudogene due to lack of conservation of the characteristic Hsp70 3' residues, but subsequent analyses revealed that the gene is transcriptionally active and it was thus termed the cognate Hsc70 gene in *T. brucei* as the mRNA levels upon transcription of the gene are not affected by perturbances in temperature (Lee *et al*., 1990).

The *T. cruzi* Hsp70 gene complement has been shown to contain a cluster organised in a head to tail fashion with a copy number of 10 (Requena *et al*., 1988). In the *L. major* Hsp70 gene complement, 4 of the genes are found to be arranged in tandem with a fifth gene that is located on the same chromosome but on a separate locus (Lee *et al*., 1988). The complete Hsp70 complement in the TriTryps has been identified in the work by Folguiera and Requena (Folgueira

and Requena, 2007). In spite of the existing information on gene location and copy number for the TriTryps Hsp70s, very little information about domain similarity, localisation and interaction with Hsp40s exists at present. It is clear from an overview of the literature that a lot of work remains to be done in the field of trypanosomal heat shock proteins, both in terms of in *silico* and biochemical characterisation.

2.1.6. Residues implicated in the specificity of Hsp40-Hsp70 partner interactions

The resolution of the tertiary structural elements of the J-domain from *E. coli* DnaJ has suggested that the J-domain structural integrity is stabilised by means of a buried core of hydrophobic amino acids including Tyr^{7}, Ile^{9}, Leu^{10}, Val^{12}, Ile^{21}, Ala^{53} and Leu^{57} according to *E. coli* numbering (Hill *et al*., 1995). An in-depth investigation into the conservation of J-domains across a number of species has also revealed that Tyr^{7}, Ala^{53} and Leu^{57} were conserved in 98 % of J-domains analysed, while Leu^{10} was found to be conserved in every J-domain studied (Hennessy *et al*., 2000a). It has been proposed that Tyr^{7} and Leu^{10} could serve to ensure the stability of the helix-loop structure of Helix II and Helix III that could facilitate presentation of the J-domain to a potential partner Hsp70, and, in so doing, enable the two proteins to interact (Hennessy *et al*., 2005a). Another residue in the J-domain that is proposed to be critical for interaction between Hsp40 proteins and Hsp70 proteins is Arg^{26} (in *Agt* DnaJ) which has been shown to be essential for the functionality of *Agt* DnaJ. It is proposed that Arg^{26} is part of a greater network of residues on Helix II, and potentially Helix III that form an-Hsp70 binding site in concert with Tyr^{25} (Genevaux *et al*., 2002; Hennessy *et al*., 2005a). The equivalent residue in *E. coli* DnaJ is the Lys^{26} residue (Hennessy *et al*., 2005a and 2005b). The Arg / Lys^{26} residue has been found to be critical for Hsp70 binding in all J-domains analysed to date (Nicoll *et al*., 2007).

In terms of ATPase stimulation of Hsp70 proteins, one of the most critical residues involved is His^{33} in the highly conserved HPD motif of the J-domain,

substitution of which results in a loss of functional interaction between a given Hsp40 protein and partner Hsp70s (Genevaux *et al*., 2002; Laufen *et al*., 1999). Further studies on the loop region by means of rational mutagenesis revealed that the HPD loop region in addition to Arg^{36} and Asn^{37} form a critical region that forms the conserved consensus sequence HPD-R/K-N (Genevaux *et al*., 2002).

In terms of critical residues in Helix III, the KFK motif is found to be important for the *in vivo* functionality of *E. coli* DnaJ (Hennessy *et al*., 2000). It has been proposed that this is due to the ability of the Phe residue in the KFK motif to interact with His^{33} in the HPD motif (Hennessy *et al*., 2000; Hennessy *et al*., 2005a; Genevaux *et al*., 2002). It is important to note that while this holds true for *E. coli* DnaJ, it may not be the case for all Hsp40 proteins, as the substitution of the F residue in the J-domain of *Agt* DnaJ did not abrogate its function *in vivo* (Hennessy *et al*., 2005a). Further residues in Helix III and Helix IV of the J-domain implicated in functionality are Leu^{57}, Asp^{59} and Arg^{63} (which form part of the QKRAA motif on Helix IV) among others (Hennessy *et al*., 2005a; Suh *et al*., 1999; Hennessy *et al*., 2005b).

2.1.7. Specificity of the Hsp40-Hsp70 interaction

While it is possible for a single Hsp70 protein to interact with a number of different Hsp40 proteins, it is known that there is a degree of specificity to this interaction and that Hsp40 proteins are not all interchangeable with respect to their interactions with a partner Hsp70 protein (Brodsky *et al*., 1993; Wiech *et al*., 1993). An example of this is the inability of cytosolic Hsp70s in *E. coli* to substitute for Hsp70s located to endoplasmic reticulum (Brodsky *et al*., 1993). It has been proposed that the J-domain contains features in both its primary sequence and secondary structure that may confer specificity to the interaction between Hsp40 and Hsp70 proteins (Hennessy *et al*., 2005b). This hypothesis was developed further based on functional studies, and it was suggested that, in fact, two types of J-domains exist. The first type of J-domain is proposed to have

evolved for a specific interaction with a particular partner Hsp70 protein in order to facilitate assisted protein folding, while the second is proposed to have evolved to interact specifically with Hsp70s that function as guardians of more specialised cellular processes (Nicoll *et al*., 2007).

2.1.8. AIMS AND OBJECTIVES

The *in silico* investigation outlined in the present work aimed to:

- determine the exact number of TbHsp40s in each of the TriTryp species (*T. brucei*, *T. cruzi* and *L. major*)
- establish a database of the Hsp40 and Hsp70 complement in the TriTryp species
- classify the TriTryp Hsp40s by Type (Type I – IV)
- determine potential interactions between Type III Hsp40s and Hsp70s in the TriTryps based on a thorough analysis of the J-domain of Type III Hsp40s in the TriTryps

2.2. EXPERIMENTAL PROCEDURES

2.2.1. Acquiring of primary amino acid sequences of TbHsp40s and TbHsp70s

The primary amino acid sequences and protein sequences of TbHsp40 and TbHsp70 proteins were obtained from the GeneDB database (http://www.genedb.org/) and compared to the data in Folguiera and Requena (2007). A database of the Hsp40 complement of *T. brucei*, *T. cruzi* and *L. major* was created using Microsoft ExcelTM software, in which all major annotations from the GeneDB database were recorded. The TriTryp Hsp40s were grouped into Type I, Type II, Type III and putative Type IV Hsp40s prior to further investigation. This was done based on the classification by Cheetham and Caplan (1998) and Botha et al. (2007). The Basic Local Alignment Search Tool for proteins (BLASTP) was used to search the TriTryps genomes using canonical Hsp40s as query sequences in order to identify any Hsp40 proteins that had been overlooked in the annotation process using the default settings of the program (Altschul *et al*., 1990). Redundancy was eliminated as far as possible by carefully comparing individual sequences. An attempt was made to find homologs for some of the TriTryp Hsp40s among annotated canonical Hsp40s of known function in order to elucidate some information on localisation and functionality. Homologues were defined as proteins having a sequence identity greater than 40 % with another protein while orthologues were defined as proteins having an identity greater than 40 % with a protein in another species. It is important to note that these values pertain to the full length of the protein or domain that has been submitted for analysis and not just a portion thereof.

2.2.2. Alignment and databasing of the J-domains of Type III Hsp40s in the TriTryps

The J-domains in each Type III Hsp40 were identified by submitting full protein sequences to the Prosite database (http://au.expasy.org/prosite/) and confirmed by inspection. The J-domain complement of the Type III Hsp40s of the TriTryps

was aligned and studied using the AlignX component of the Vector NTITM suite. The J-domains were aligned from those of known Type III J-proteins as well as those of canonical Type I and Type II Hsp40s in order to determine the identity of any potential well-conserved residues in the TriTryp Hsp40s that may be crucial to protein function.

2.2.3. Prediction of localisation of the TriTryp Hsp40 complement, and *T. brucei* Hsp70s

An investigation into the potential localisation of Type III TbHsp40 proteins was performed with the aid of Wolf PSORT web-based program (http://wolfpsort.org/) that predicts localisation of proteins by converting protein amino acid sequences into numerical localisation features. These features are based on sorting signals, amino acid composition and functional motifs that include DNA-binding motifs (Horton *et al.*, 2007). The Wolf PSORT program is an extension and improvement of the PSORT II program that has previously been used in other bioinformatic studies (Horton *et al.*, 2007; Botha *et al.*, 2007). Wolf PSORT predicts localisation of proteins into more than 10 localisation sites, and the cross-validation studies of the creators suggest that a sensitivity and specificity of approximately 70% can be obtained for proteins targeting the nucleus, mitochondria, cytosol and plasma membrane (Horton *et al.*, 2007). Although comparatively little is known about the mechanisms of trypanosomal protein trafficking and localisation signals in general, it has been proposed that the targeting sequences are similar to those of mammals (Qiao *et al.*, 2006).

The SignalP 3.0 server (http://www.cbs.dtu.dk/services/SignalP/) was used to enhance the localisation predictions obtained (Bendsten *et al.*, 2004). This server predicts the location of potential signal peptidase I cleavage sites which assists in the prediction of signal peptides on protein primary sequences (Bendsten *et al.*, 2004). The presence of a signal peptide in this case could denote that the protein is extracellular or retained in the E.R. / Golgi apparatus (Bendsten *et al.*,

2004) and needs to be confirmed with the aid of *in silico*, *in vitro* and *in vivo* characterisation.

2.2.4. Homology modeling of *T. brucei* Hsp40s, Hsp70s and their domains

The sequences of proteins of interest were obtained from the GeneDB database (www.genedb.org) and annotated by inspection as well as by submission to Prosite (http://au.expasy.org/prosite/) in order to identify the J-domains and TPR domains where relevant. The J-domain sequences were used as queries in a BLAST analysis of the PDB database (http://www.rcsb.org/pdb/home/home.do), and submitted to the Phyre program (http://www.sbg.bio.ic.ac.uk/~phyre/) to find an appropriate template sequence. A suitable template sequence was defined as one that had a similarity of greater than 30% as this is usually considered to be the minimum requirement for a suitable template sequence for homology modeling (Rost, 1999). The potentially suitable template sequences obtained from the Phyre server were studied in order to obtain the most suitable template sequence based on the integrity of the template sequence itself derived from structural information available in the PDB database. The PDB structure that possessed the highest structural integrity and the highest sequence identity to the proteins of interest was selected as a template for homology modelling. The *E. coli* DnaJ J-domain structure (1 XBL) was used as a template for all J-domain modeling as it fulfilled the criteria outlined above in that it possessed an identity of 30 % or greater to all the J-domains studied and the structure was of sufficient fidelity.

The template sequence and the target sequence were submitted together for alignment to the ClustalW multiple sequence alignment package website (http://www.ebi.ac.uk/Tools/clustalw2/index.html), with a selection indicating that the alignment should be in PIR output format. The PIR alignment was saved, and a command document called model-single.py was written in order to run the Modeler 9v5 software in a Windows computing environment. Homology models

were created in Modeler 9v5 for Windows and the PDB files generated were rendered using Pymol software (DeLano, 2002).

2.2.5. Confirmation of the accuracy of homology models

The integrity and accuracy of all homology models generated were confirmed by means of the Procheck program that was adapted for Windows NT® by Bernhard Rupp (Laskowski *et al*., 1993; Morris *et al*., 1992). The generated models were analysed in order to ensure that all residues in the model were within areas that are considered allowable in terms of steric and bond length angles. For the purposes of the present work, a model was considered to be successfully generated if its planar groups were 100 % within allowable regions as per the Ramachandran plots generated by Procheck, and the accuracy of the M / C bond angles were over 80 %, while the M / C bond lengths were over 85 – 90 % accurate.

2.3. RESULTS

A search of all the Hsp40 proteins in a number of trypanosomal species has revealed a larger Hsp40 complement than is found in well-characterised species such as yeast, *E. coli* and humans (Figure 2.1). The number of Hsp40s in the TriTryps were found to compare favourably to that found in the cerebral malaria-causing parasite, *Plasmodium falciparum* (Botha *et al*., 2007).

2.3.1. The TriTryps possess a large Hsp40 complement in relation to other organisms

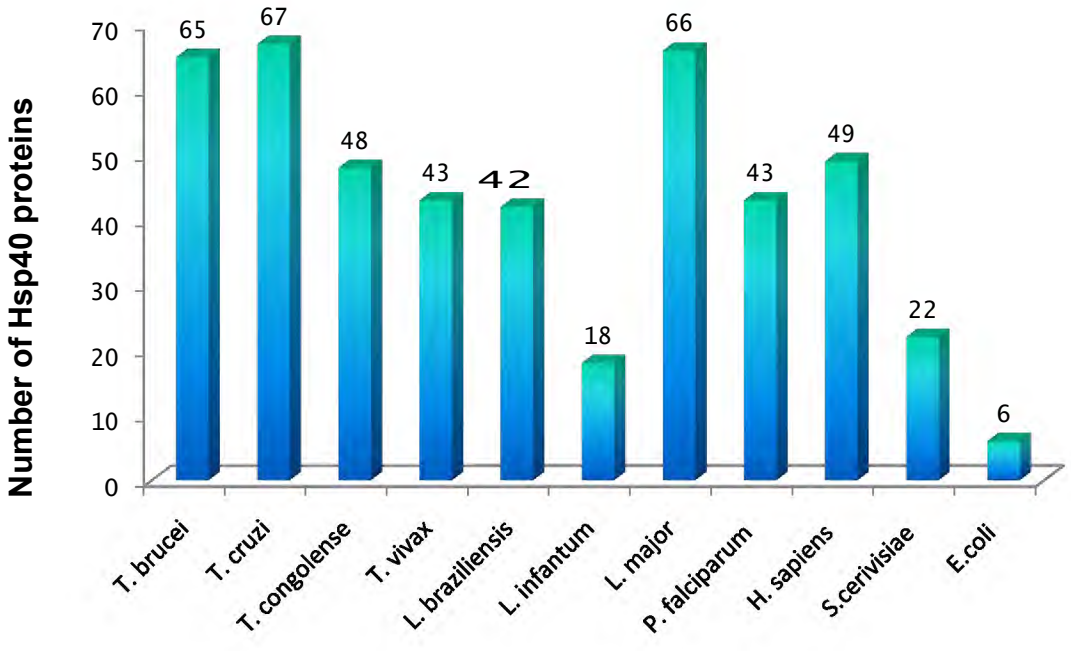


Figure 2.1. Comparison of the total number of Hsp40 proteins in the Kinetoplastida and other well-characterised organisms.

Comparison of the number of annotated Hsp40 sequences available for a number of organisms. Information condensed from Folgueira and Requena (2007), Qiu et al. (2006), Hageman and Kampinga (2009), Botha et al. (2007) and www.genedb.org. The three DnaJ-like proteins (putative Type IV Hsp40s) found in the yeast genome have been omitted from this figure for the purposes of clarity (Walsh et al., 2004).

Both *T. brucei* and *L. major* have a single-copy gene encoding each Hsp40, but *T. cruzi* offers a far more complex situation (Folgueira and Requena, 2007). In most cases, *T. cruzi* contains two genes for each Hsp40 protein, which could represent allelic copies of the CL Brener strain of *T. cruzi* that was used for the genome sequencing project (Folgueira and Requena, 2007). The physiological rationale for such a large Hsp40 complement in the TriTryps is unknown at present, as almost none of these proteins have been biochemically characterised, with the exception of Tcj1-Tcj4, Tcj6 and TcDj1 from *T. cruzi* (Tibbetts *et al*., 1998; Carreira *et al*., 1998; Salmon *et al*., 2001; Edkins *et al*., 2004). Edkins and colleagues have also shown that Tcj2 from *T. cruzi* is able to functionally replace the Ydj1 Hsp40 from yeast, suggesting that these proteins may be functionally equivalent (Edkins *et al*., 2004). At present, no putative interactions between the TriTryp Hsp40s and their partner Hsp70s have been investigated. This is partly due to the large number of proteins involved (ca. 70 different Hsp40s in each TriTryp species and up to 14 potential partner Hsp70s) (Folgueira and Requena, 2007).

2.3.2. *T. brucei* possesses a large and diverse Type III Hsp40 complement

The Hsp40 sequences that were retrieved were analysed and grouped into Type I – IV Hsp40s using the classification outlined by Cheetham and Caplan (1998) and Botha *et al*. (2007) (Figure 2.2.). It was very interesting to note that the number of Hsp40s present in *T. brucei* was due to the very large Type III (47 proteins) and Type IV Hsp40 (6 proteins) complement in the parasite. In terms of the Type I and Type II Hsp40s in *T. brucei*, the numbers compare favourably to that of the yeast and human Hsp40 distribution (Folgueira and Requena, 2007; Qiu *et al*., 2006). Type III Hsp40s display far higher levels of structural and functional divergence than the well-conserved Type I and moderately-conserved Type II Hsp40s (Kelley, 1999). The Type III Hsp40 complement as a whole has not been very well characterised in any organism, largely due to the very low levels of conservation of these proteins which do not readily facilitate

comparisons with other known and well-characterised Hsp40s from different species. It has been proposed, however, that the Type III Hsp40s could serve specialised roles in targeting Hsp70 proteins to suitable organelles where they may be required (Kelley, 1999). Another view is that as Type III Hsp40s are not known to have any ability to bind polypeptide substrates, they are thus not likely to possess any chaperone activity (Walsh *et al.*, 2004). Not enough detailed analysis on Type III Hsp40s has been done to date to prove or disprove either of the hypotheses presented above.

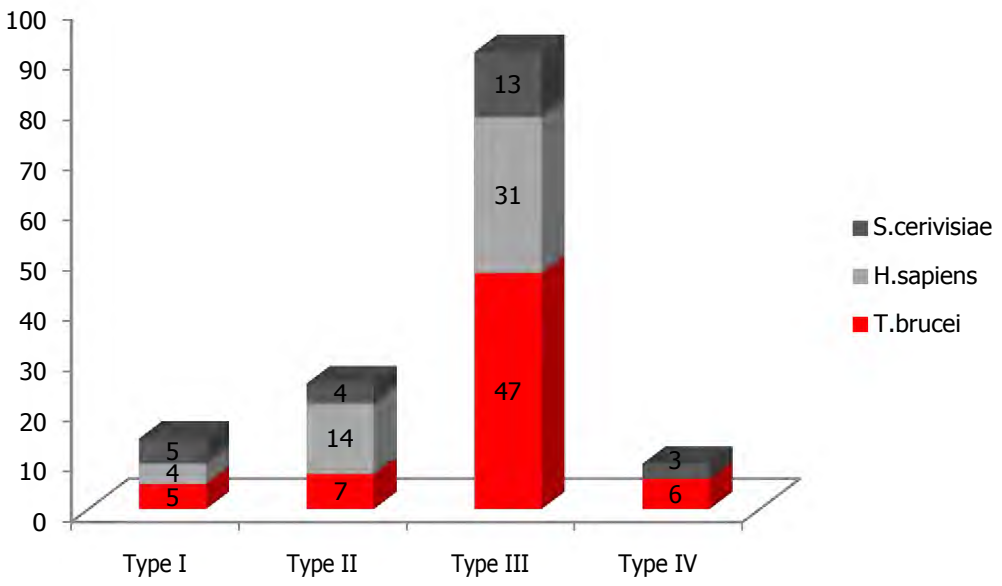


Figure 2.2. Comparison of the proportion of Hsp40s by type in yeast, humans and *Trypanosoma brucei*

The number of Hsp40 proteins by Type (I-IV) was compared for *S. cerevisiae*, *H. sapiens* and *T. brucei*. The Y-axis denotes the number of Hsp40 proteins and the actual number of each Hsp40 type is given on the portion of the bar belonging to that species. Dark grey – *S. cerivisiae*; Lighter grey – *H. sapiens*; Red – *T. brucei*. Information obtained from Folgueira and Requena (2007), Qiu et al. (2006), Hageman and Kampinga (2009) and Walsh et al. (2004).

The size of the Type III Hsp40 complement of *T. brucei* could indicate a specialised function in parasite survival, making these proteins and their functioning unique to *T. brucei*.

2.3.3. The Type III Hsp40 complement in the TriTryp parasites

A detailed investigation of the Type III Hsp40 complement of the TriTryps parasites was undertaken and the data obtained is summarised in Tables 2.4.-2.6.) as well as in the Excel® spreadsheet on the CD attached with this thesis (Appendix AA). The *T. brucei* Hsp40 complement (TbHsp40s) was found to contain 47 Type III Hsp40s, the *T. cruzi* parasite (TcHsp40s) 47, and *L. major* (LmHsp40s) contains 43 Type III Hsp40s. The *T. brucei* Type III Hsp40 complement (Table 2.4.) provides evidence of the poor levels of conservation of these Hsp40 proteins. The Type III TbHsp40s range in size from 21.5 kDa (Tbj14) to 253.8 kDa (Tbj59) and display an even distribution across chromosomes. A total of 10 Type III TbHsp40s were predicted to contain a signal peptide (21%), while a slightly greater proportion (36%) are predicted to contain transmembrane domains with or without signal peptides.

A comparison between proposed targeting signals of *T. brucei* proteins predicted to be localised to the nucleus and mitochondria were compared to known mammalian signals utilised by the Wolf PSORT program, and used to determine the potential localisation of these proteins within the parasite. Of the 47 Type III Hsp40s in *T. brucei*, a total of 13 (27.6 %) were predicted to localise to the mitochondria of the parasite. A multiple sequence alignment of all these proteins was performed to determine if a generalised targeting consensus sequence for localisation could be determined (data not shown). No such sequence was found, which indicates that *T. brucei* could potentially make use of a number of mitochondrial targeting signals, or that some of the localisation predictions made by the Wolf PSORT program were in error. The latter possibility is not unprecedented, as Wolf PSORT is not specific to the kinetoplastids, and the

Table 2.4. The Type III Hsp40s in *T. brucei*

Protein	Gene ID	Chromosome	Mr	pI	Localisation
Tbj1	**Tb11.01.8750**	**11**	**35.2**	**6.4**	**Nuclear cytoplasm**
Tbj5	Tb10.70.1330	10	37.8	9.2	Plasma membrane
Tbj8¤ф	Tb11.02.2880	11	84.7	6.0	Secreted
Tbj9¤	Tb927.6.3850	6	33.3	7.0	Mitochondria
Tbj10 ф	Tb927.7.6660	7	29.8	9.2	Mitochondria
Tbj11 ф	Tb09.211.0140	9	47.8	6.4	Plasma membrane
Tbj12	Tb927.4.2970	4	62.8	7.5	Nucleus
Tbj14	Tb927.5.2880	5	21.5	9.6	Nucleus
Tbj15	Tb10.61.1940	10	52.5	7.6	Nucleus
Tbj16	Tb927.1.1960	1	75.7	6.8	Cytoplasm
Tbj18 ф	Tb11.03.0110	11	41.7	10.5	Plasma membrane
Tbj19 ф	Tb927.4.650	4	28.5	8.4	Mitochondria
Tbj20	Tb10.70.2160	10	30.8	4.4	Cytoplasm
Tbj21	Tb927.7.540	7	50.6	7.1	Cytoplasm
Tbj23 ф	Tb10.61.3100	10	27.2	10.7	Mitochondria
Tbj25¤	Tb927.7.680	7	88.7	9.8	Mitochondria
Tbj26¤ ф	Tb927.7.6200	7	31.5	10.1	Secreted
Tbj28 ф	Tb927.7.740	7	61.9	10.1	Plasma membrane
Tbj29 ф	Tb11.02.3430	11	47.7	9.7	Nucleus
Tbj30 ф	Tb927.8.1010	8	55.3	10.4	Mitochondria
Tbj33	Tb10.6k15.0460	10	29.8	10.2	Nucleus
Tbj34¤ ф	Tb09.211.1550	9	56.7	9.0	E.R.
Tbj35	Tb927.7.4590	7	61.6	8.9	E.R.
Tbj36¤	Tb927.3.1760	3	29.4	4.7	Mitochondria
Tbj37	Tb10.389.1020	10	72.9	11.1	Nucleus
Tbj38 ф	Tb927.6.3730	6	35.0	9.6	Plasma membrane
Tbj39	Tb09.211.4720	9	36.8	10.0	Plasma membrane
Tbj40	Tb927.8.4470	8	26.8	10.6	Mitochondria
Tbj41	Tb927.4.3980	4	54.6	9.5	Nucleus
Tbj42	Tb10.389.1350	10	65.3	7.3	Cytoplasm
Tbj43	Tb09.211.2290	9	31.6	9.3	Nucleus
Tbj44¤	Tb927.8.7010	8	25.5	8.8	Secreted
Tbj48 ф	Tb927.8.1030	8	33.4	10.9	Mitochondria
Tbj49	Tb927.6.2480	6	43.7	10.1	Mitochondria
Tbj51♦	**Tb927.4.2220**	**4**	**80.4**	**8.5**	**Cytoplasm**
Tbj52♦	Tb10.70.2300	10	55.9	7.9	Nucleus
Tbj53♦	Tb927.7.3630	7	60.1	8.0	Secreted
Tbj55¤	Tb11.01.0135	11	14.6	4.4	Secreted
Tbj56	Tb927.2.3960	2	25.8	5.4	Cytoplasm
Tbj59	Tb927.6.3500	6	253.8	6.6	Plasma membrane
Tbj60¤ ф	Tb11.01.5190	11	47.0	8.9	E.R.
Tbj62 ф	Tb10.70.5230	10	83.1	7.9	Plasma membrane
Tbj63 ф	Tb11.01.5390	11	36.1	10.7	Mitochondria
Tbj65♦	Tb927.4.880	4	72.0	8.0	Cytoplasm
Tbj67¤♦	Tb10.70.1950	10	107.2	8.2	Mitochondria
Tbj69	Tb11.01.2720	11	36.8	10.7	Mitochondria
Tbj70 ф*	Tb927.8.8310	8	51.7	5.6	Plasma membrane

" – Information obtained from the *T. brucei* GeneDB database (http://www.genedb.org/genedb/tryp/)
¤ - Proteins with a signal peptide
ф – Proteins with transmembrane domains
♦ - Proteins with TPR domains
Mr – Molecular mass in kDa
ξ - Localisation as predicted by WOLF Psort (http://wolfpsort.org/)
*** - Protein with no orthologs in *T. cruzi* or *L. major***
Sec63 homologue is underlined (Goldschmidt *et al.*, 2008)
Proteins in bold and underlined were of interest in this study

localisation signals of these parasites may differ from that of mammalian systems under certain circumstances. A total of 9 (19 %) of the Type III TbHsp40s were predicted to be membrane bound (Table 2.4.), indicating a potential function as trafficking proteins. Of particular interest was the 253.8 kDa Tbj59 protein, which has been annotated to be a homolog of the receptor- mediated-endocytosis 8 (RME-8) protein that has a broad tissue distribution in mammals and is involved in intracellular trafficking (Girard *et al.*, 2005). Tbj59 has well-conserved orthologs in both *T. cruzi* (Table 2.5.a and 2.5.b) and *L. major* (Table 2.6.).

A further 19 % of Type III TbHsp40s were predicted to be localised in the nucleus of the parasite, including Tbj1, an orthologue of Tcj1 that was identified in *T. cruzi* by Tibbetts and colleagues (Tibbetts *et al.*, 1998). A number of *T. brucei* proteins, most notably those predicted to contain signal peptides were predicted to be secreted (10 %) while the remainder were predicted to be localised to the cytoplasm (14.8 %) and the E.R. (6%). A very large proportion (63 %) of Type III TbHsp40s was found to have basic pI values. Five Type III TbHsp40s (Tbj51, Tbj52, Tbj53, Tbj65 and Tbj67) were found to contain TPR domains, which have very interesting implications for their function. These proteins are discussed more fully in Section 2.3.10 in the present chapter.

2.3.4. The Type III Hsp40 complement of *T. cruzi*

In spite of the completion of the *T. cruzi* genome sequencing project, the gene annotation procedures have been proving more complex than that of *T. brucei* and *L. major*. This is largely due to the number of gene duplications found in the parasite (El-Sayed *et al.*, 2005). For the purposes of the present work, the Hsp40 proteins given in the review by Folgueira and Requena (2007) were analysed and divided into Types I-IV, without taking partial protein sequences into account. It was found that, in spite of nearly double the number of genes than *T. brucei*, *T. cruzi* also possesses 47 Type III Hsp40s (Table 2.5.a and

Table 2.5.a. The Type III Hsp40s in *T. cruzi*

Protein	Gene ID	Contig"	Mr"	pI"	Localisation;
Tcj1	**Tc00.1047053511537.50**	**6146**	**35.4**	**7.1**	**Nucleus**
	Tc00.1047053508707.330	7839	35.3	7.9	Nucleus
Tcj5 ɸ	Tc00.1047053504163.100	4940	36.7	9.4	Mitochondria
	Tc00.1047053510301.70	5885	36.8	9.1	Mitochondria
Tcj8¤ ɸ	Tc00.1047053503981.40	4849	85.1	6.2	Secreted
	Tc00.1047053504131.100	4924	85.0	6.3	Golgi Apparatus
Tcj9	Tc00.1047053511517.44	6139	32.9	7.2	Mitochondria
Tcj10¤ ɸ	Tc00.1047053508911.20	7925	31.3	8.9	Mitochondria
	Tc00.1047053510055.140	8288	31.4	8.9	Mitochondria
Tcj11 ɸ	Tc00.1047053510241.70	8342	49.5	7.9	Plasma membrane
	Tc00.1047053510579.60	5947	49.5	7.5	Plasma membrane
Tcj12	Tc00.1047053506435.50	6943	29.8	10.4	Nucleus
	Tc00.1047053506265.45	6877	30.0	10.8	Nucleus
Tcj14	Tc00.1047053507807.20	7477	20.9	8.1	Cytoplasm
	Tc00.1047053507063.180¤	7192	26.0	8.8	Secreted
Tcj15	Tc00.1047053503833.20	4775	52.8	9.3	Nucleus
	Tc00.1047053511287.130	6090	52.8	9.1	Nucleus
Tcj16	Tc00.1047053506473.20	6959	74.9	7.9	Nucleus
	Tc00.1047053508999.10	7958	75.1	7.9	Nucleus
Tcj18¤ ɸ	Tc00.1047053506925.470	7143	42.0	10.4	Plasma membrane
Tcj19 ɸ	Tc00.1047053507053.120	7189	28.7	9.7	Mitochondria
Tcj20	Tc00.1047053508661.30	7819	31.1	4.6	Nucleus
	Tc00.1047053510293.50	8357	31.1	4.7	Cytoplasm
Tcj21 ɸ	Tc00.1047053506287.90	6885	52.3	8.5	Cytoplasm
	Tc00.1047053506991.10	5424	52.4	8.0	Plasma membrane
Tcj23¤	Tc00.1047053507993.30	7561	26.3	10.4	Mitochondria
	Tc00.1047053511277.480	8643	26.3	10.3	Mitochondria
Tcj25¤	Tc00.1047053508881.10	7913	88.0	9.6	Plasma membrane
	Tc00.1047053506289.74	6886	88.3	9.5	Plasma membrane
Tcj26¤ ɸ	Tc00.1047053506513.30	6979	29.4	10.6	Mitochondria
	Tc00.1047053508919.90	7927	29.4	10.3	Mitochondria
Tcj28¤ ɸ	Tc00.1047053506135.40	5334	65.3	10.0	Plasma membrane
Tcj29¤ ɸ	Tc00.1047053506203.50	6850	55.0	9.6	Secreted
	Tc00.1047053508643.30	5602	47.3	9.9	Mitochondria
Tcj30 ɸ	Tc00.1047053508569.120	7782	59.5	10.0	Mitochondria
	Tc00.1047053503733.30	4725	60.2	10.0	Mitochondria
Tcj33	Tc00.1047053504147.60	4932	30.0	9.8	Nucleus
Tcj34 ɸ	Tc00.1047053510731.80	5982	56.7	9.0	E.R.
	Tc00.1047053506887.90	7128	56.7	9.0	E.R.
Tcj35 ɸ	Tc00.1047053506605.180	7018	57.8	9.1	Plasma membrane
Tcj36 ¤ɸ	Tc00.1047053510091.50	8301	29.7	5.4	Secreted
	Tc00.1047053510421.300	8385	29.8	5.2	Mitochondria
Tcj37	Tc00.1047053503455.10	4586	77.3	9.8	Nucleus
	Tc00.1047053511261.100	6086	77.5	9.8	Nuclear cytoplasm

" – Information obtained from the *T. cruzi* GeneDB database (http://www.genedb.org/genedb/tcruzi/)
¤ - Proteins with a signal peptide
ɸ – Proteins with transmembrane domains
♦ - Proteins with TPR domains
Mr – Molecular mass in kDa
; - Localisation as predicted by WOLF Psort (http://wolfpsort.org/)
Sec63 homologue is underlined (Goldschmidt *et al.*, 2008)
Proteins in bold and underlined were of interest in the present study

Table 2.5.b. The Type III Hsp40s in *T. cruzi* (contd.)

Protein	Gene ID	Contig"	Mr"	pI"	Localisation;
Tcj38¤ ɸ	Tc00.1047053507029.20	7180	35.9	9.5	Secreted
	Tc00.1047053506941.270	7147	35.9	9.5	Secreted
Tcj39 ɸ	Tc00.1047053506999.70	7167	36.5	10.5	Plasma membrane
	Tc00.1047053510759.134	8486	36.4	10.5	Plasma membrane
Tcj40¤	Tc00.1047053509569.150	8134	27.3	11.1	Mitochondria
	Tc00.1047053508989.60	7953	27.4	11.2	Mitochondria
Tcj41¤ ɸ	Tc00.1047053511337.10	6097	55.4	8.0	Plasma membrane
	Tc00.1047053509237.80	8046	50.7	8.8	Plasma membrane
Tcj42¤ ♦	Tc00.1047053507787.90	7470	72.6	6.6	Secreted
	Tc00.1047053507625.110	7407	81.4	6.7	Plasma membrane
Tcj43	Tc00.1047053508461.240	7739	35.9	10.3	Nucleus
Tcj44 ɸ	Tc00.1047053509911.100	8242	23.6	9.5	Mitochondria
	Tc00.1047053507389.60	7315	23.7	9.5	Mitochondria
Tcj48 ɸ	Tc00.1047053508569.140	7782	41.9	10.3	Peroxisome
	Tc00.1047053503733.10	4725	42.1	10.2	Peroxisome
Tcj49	Tc00.1047053509053.130	7980	38.1	10.1	Mitochondria
Tcj51♦	Tc00.1047053506559.430	6996	79.4	8.7	Nucleus
Tcj52♦	Tc00.1047053504203.60	4960	29.4	9.1	Cytoplasm
Tcj53¤ ɸ ♦	Tc00.1047053510407.80	5914	59.6	7.6	Secreted
	Tc00.1047053509937.160	8251	59.6	7.6	Secreted
Tcj55	Tc00.1047053506931.30	5414	17.7	5.3	Mitochondria
	Tc00.1047053510131.20¤ ɸ	8314	21.7	5.3	Secreted
Tcj56	Tc00.1047053511109.90	8600	25.8	5.4	Cytoplasm
	Tc00.1047053504157.50	4937	25.7	5.9	Cytoplasm
Tcj59	Tc00.1047053511511.10	8711	255.1	6.4	Plasma membrane
Tcj60 ɸ	Tc00.1047053506753.160	7082	39.1	9.5	Nucleus
	Tc00.1047053510357.50	8369	52.8	9.4	Plasma membrane
Tcj62 ɸ	Tc00.1047053506193.10	6846	89.0	9.0	Plasma membrane
Tcj63 ɸ	Tc00.1047053509161.110	8019	36.5	10.7	Mitochondria
Tcj65♦	Tc00.1047053508257.160	7660	72.3	9.1	Nucleus
Tcj67♦	Tc00.1047053506009.30	6774	108.6	8.6	Nucleus
	Tc00.1047053510297.30	8358	108.4	8.5	Nucleus
Tcj69¤	Tc00.1047053509823.10	8216	36.6	10.5	Mitochondria
	Tc00.1047053506807.30	5394	36.5	10.6	Mitochondria

" – Information obtained from the *T. cruzi* GeneDB database (http://www.genedb.org/genedb/tcruzi/)
¤ - Proteins with a signal peptide
ɸ – Proteins with transmembrane domains
♦ - Proteins with TPR domains
Mr – Molecular mass in kDa
; - Localisation as predicted by WOLF Psort (http://wolfpsort.org/)

Table 2.5.b). Most of these proteins were found to possess orthologues in *T. brucei* and *L. major*, which appeared to suggest that the TriTryps may possess similar chaperone machinery with functions that are compatible in each parasite. The most notable exception to this was Tbj70 (Tb927.8.8310) which had no orthologs in either *T. cruzi* or *L. major* (shown with an asterisk in Table 2.4). The presence of well-conserved Type III Hsp40 orthologues across the TriTryps was found to have interesting implications in terms of drug discovery. The nuclear

Type III Hsp40 complement in *T. cruzi* comprises 25 % of the total number of Hsp40s, which is a 6 % higher number predicted for *T. brucei*. A large proportion of Type III TcHsp40s were predicted to be localised to the mitochondria (32 %) which was a slightly higher proportion than the 27.6 % predicted in *T. brucei* and the 20.9 % predicted for *L. major*. Only 4 TcHsp40s (8 %) were predicted to be secreted from the parasite, while a slightly higher number were predicted to be membrane bound. Another major difference between the *T. brucei* and *T. cruzi* Type III Hsp40 complement is that only 8 % of *T. cruzi* Hsp40s were predicted to be cytoplasmic, in contrast to 14.8 % of those in *T. brucei*.

2.3.5. The Type III Hsp40 complement in *L. major*

The Type III LmHsp40s were shown to comprise 43 proteins with a range of molecular weights from 24.3 kDa to 273.3 kDa. These proteins were found to be ubiquitous across all organelles of the parasite (Table 2.6.). The gene distribution of the LmHsp40s was very diverse in terms of chromosomal location, which was unlike the case observed in *T. brucei* (Table 2.4.) where clusters of Hsp40 genes were shown to occur on a given chromosome. The LmHsp40s also possessed high (basic) pI values, which was similar to that observed for *T. brucei* and *T. cruzi*. With only a few exceptions (Tbj9, Tbj12, Tbj39 and Tbj70), all TbHsp40s had orthologues in *L. major*. These orthologues were highly similar in size, protein sequence, pI value and localisation prediction for each protein. For example, the *L. major* orthologue of Tbj1 and Tcj, Lmj1, has a molecular weight of 36.7 kDa and a pI of 6.4 with a predicted localisation of the nucleus. This corresponds very well with the information that is known for Tbj1 and Tcj1 leading to the hypothesis that these proteins could have a similar function in their respective hosts, and that elucidation and characterisation of Tbj1 and Tcj1 will shed light not only on their own function, but potentially on that of Lmj1. The endosomal trafficking protein, RME-8 (Lmj59) was also found in *L. major*, and was well-conserved with respect to its orthologues in *T. brucei* and *T. cruzi*. In terms of their localisation, 32.5 %

Table 2.6. The Type III Hsp40s in *L. major*

Protein	Gene ID	Chromosome"	Mr"	pI"	Localisation;
Lmj1	LmjF32.3030	32	36.7	6.4	Nucleus
Lmj5ф	LmjF36.1330	36	34.4	7.8	Plasma membrane
Lmj8¤ф	LmjF24.0520	24	88.9	6.6	E.R.
Lmj10 ф	LmjF17.0460	17	25.3	9.6	E.R.
Lmj11 ф	LmjF04.0780	4	61.7	7.1	Plasma membrane
Lmj14	LmjF08.0990	8	36.4	8.2	Nucleus
Lmj15	LmjF19.0080	19	50.8	10.1	Nucleus
Lmj16	LmjF20.1130	20	74.5	7.1	Nucleus
Lmj18 ф	LmjF27.0410	27	36.0	10.0	Plasma membrane
Lmj19 ф	LmjF34.4080	34	29.3	9.5	Mitochondria
Lmj20	LmjF36.0610	36	28.0	5.7	Nucleus
Lmj21¤ф	LmjF26.1410	26	59.1	8.5	Plasma membrane
Lmj23¤ф	LmjF18.0330	18	27.3	10.6	Secreted
Lmj25	LmjF26.1270	26	93.4	9.5	Nucleus
Lmj26¤ф	LmjF17.0040	17	29.7	10.4	Mitochondria
Lmj28¤ф	LmjF26.1200	26	71.1	10.5	Secreted
Lmj29 ф	LmjF24.1080	24	49.5	10.4	Nucleus
Lmj30 ф	LmjF07.0780	7	33.8	10.2	Plasma membrane
Lmj33	LmjF36.4470	36	30.8	10.2	Nucleus
Lmj34 ф	LmjF35.4630	35	56.1	9.9	E.R.
Lmj35	LmjF14.0110	14	59.1	10.3	Plasma membrane
Lmj36	LmjF25.1690	25	31.2	6.1	Mitochondria
Lmj37	LmjF18.1430	18	121.8	10.6	Nucleus
Lmj38	LmjF30.2450	30	37.8	10.1	Mitochondria
Lmj40	LmjF10.1050	10	30.8	10.0	Mitochondria
Lmj41¤ф	LmjF31.0510	31	63.4	6.2	E.R.
Lmj42¤	LmjF18.1650	18	63.4	7.5	Nucleus
Lmj43	LmjF35.4040	35	42.9	8.2	Nucleus
Lmj44 ф	LmjF31.3100	31	24.3	6.9	Mitochondria
Lmj48 ф	LmjF07.0750	7	33.0	11.3	Nucleus
Lmj49¤	LmjF30.1030	30	34.6	11.5	Secreted
Lmj51♦	LmjF34.2430	34	88.5	8.0	Plasma membrane
Lmj52♦	LmjF36.0500	36	56.1	8.5	Nucleus
Lmj53¤♦	LmjF14.1330	14	63.5	8.6	Secreted
Lmj55	LmjF28.1270	28	51.2	5.4	Nucleus
Lmj56	LmjF33.2690	33	29.2	5.4	Cytoplasm
Lmj59¤	LmjF30.2210	30	273.7	6.5	Plasma membrane
Lmj60¤	LmjF09.1440	9	46.3	9.6	Mitochondria
Lmj62ф	LmjF34.0040	34	76.1	9.4	Plasma membrane
Lmj63ф	LmjF32.0590	32	36.7	10.8	Mitochondria
Lmj65	LmjF34.3870	34	85.0	8.5	Mitochondria
Lmj67♦	LmjF36.0760	36	92.5	8.7	Mitochondria
Lmj69	LmjF36.4970	36	38.5	10.8	Secreted

" – Information obtained from the *L. major* GeneDB database (http://www.genedb.org/genedb/leish/)
¤ - Proteins with a signal peptide
ф – Proteins with transmembrane domains
♦ - Proteins with TPR domains
Mr – Molecular mass in kDa
; - Localisation as predicted by WOLF Psort (http://wolfpsort.org/)
Sec63 homologue is underlined (Goldschmidt *et al.*, 2008)

of LmHsp40s were predicted to be localised to the nucleus, while 20.9 % were predicted to be membrane-bound (Table 2.6). Only 4 LmHsp40s (9.3 %) were predicted to be secreted from the parasite while a further 23 % were predicted to be mitochondrial. *L. major* was found to have 4 Type III Hsp40 proteins, which have orthologues in *T. brucei* and *T. cruzi* (Lmj51, Lmj52, Lmj53 and Lmj67) that have been shown to contain TPR domains in addition to the J-domain (Table 2.6.; annotated with black diamonds). These TPR-domain containing Type III Hsp40s were shown to have a very diverse distribution (plasma membrane, nucleus, secretion pathway and mitochondrion) and occur in virtually every organelle of the parasite with the exception of the cytoplasm.

2.3.6. The levels of J-domain conservation in the Type III Hsp40s from *T. brucei*

Because Type III Hsp40s lack all domains characteristic of typical Hsp40s except for the canonical J-domain, an in-depth study of the nature of their J-domains becomes crucial in the attempt to assign potential partner Hsp70s in addition to determining the nature, if any, of their chaperone function. Once the nature of the chaperone function of these proteins has been determined, predictions about their specialised functions within the TriTryps can be made. The J-domains of all 47 Type III Hsp40s were aligned and studied to determine the extent of sequence variability and conservation in this important region of the proteins. The boundaries of the helices were determined by comparing them to the helices of known and well-characterised Hsp40 proteins (Figure 2.3) and the numbers to the left of the alignment indicate the amino acid number of the starting residue of the J-domain with respect to the total number of amino acids in a given protein. The diverse location of the J-domains in Type III TbHsp40s is shown by the numbers on the left hand side of the alignment in Figure 2.3. The analysis focussed mainly on the residues that have been implicated as being essential for J-domain function in other organisms, most notably *E. coli* DnaJ and *Agt* DnaJ as these J-domains have been extensively characterised.

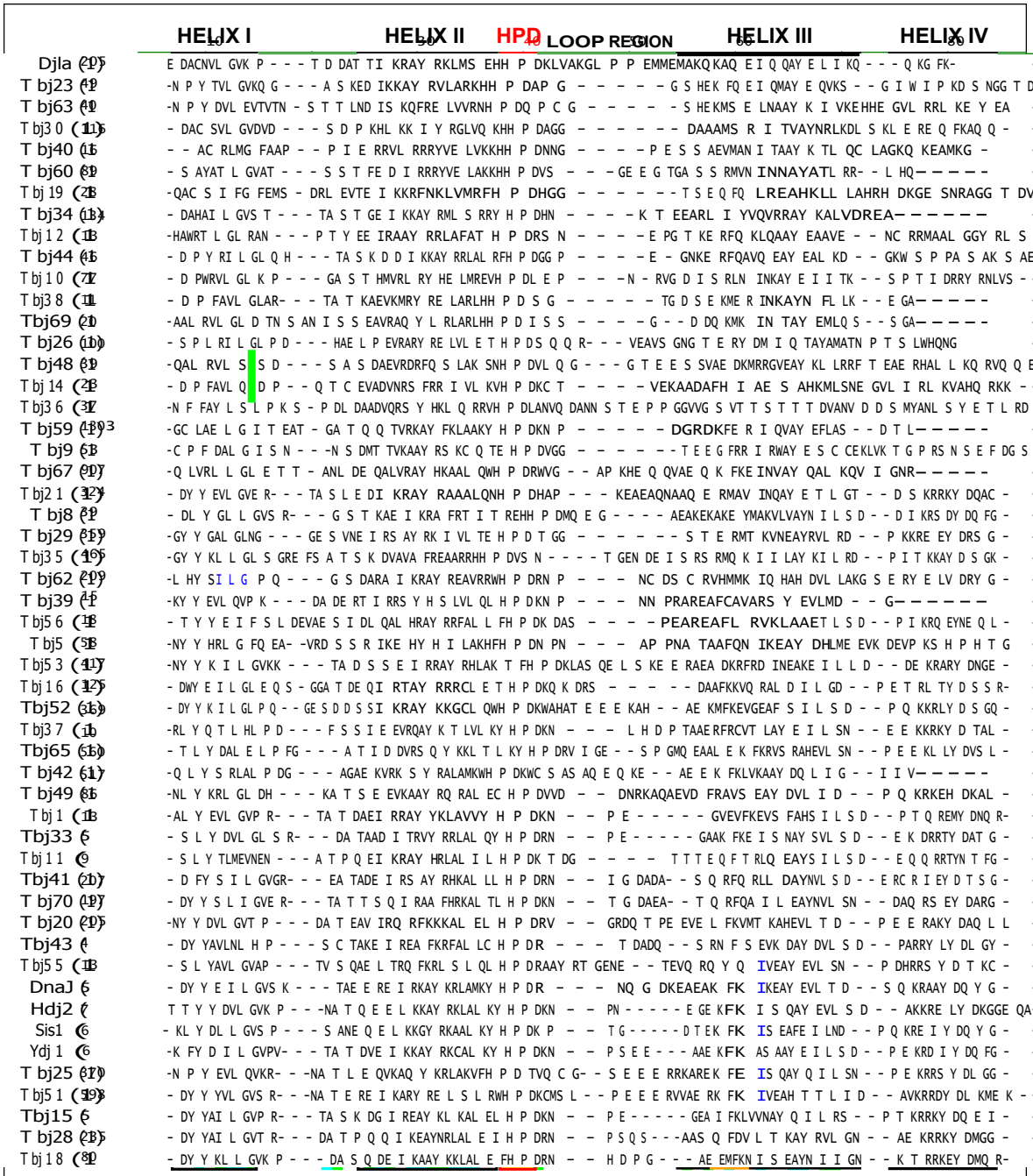


Figure 2.3. Conservation of the J-domain in Type III Hsp40s from *T. brucei*

The sequence alignment of the J-domains of all Type III Hsp40s in the *T. brucei* parasite and those of canonical Hsp40s was generated using the AlignX alignment component of the Vector NTI® suite. Helices are outlined in black and annotated on the figure, the HPD motif is outlined in red and the KFK motif is outlined in brown. The canonical Hsp40s used include: DnaJ – *E. coli* DnaJ (AP 000679); Sis1 – yeast (CAA 95866); Ydj1 – yeast (CAA 95937); Hdj2 – human (AAH08182); Djla – *E. coli* (YP 001742176). [Similar residues] – green; [Identical residues] – yellow background with red writing; [Conservative changes] – turquoise and [Non-similar residues] – black writing on a while background

These residues included Tyr^{7}, Leu^{10}, Lys^{26}, the HPD motif, the KFK motif and the QKRAA motif. Helix I and Helix II were shown to be well-conserved in Type III TbHsp40s. The first amino acid position of the J-domains of Type III Hsp40s was found to be variable (Figure 2.3), but Asn (17 %) and Asp (36 %) were shown to occur most frequently. Position 2 on the Type III J-domain was conserved in several cases, and a number of TbHsp40s displayed either a Tyr residue (32 %) or a Phe residue at this position. The third residue of the J-domains was shown to be highly conserved with 74 % of the J-domains surveyed having contained a Tyr residue at this position in comparison to approximately 60 % of the J-domains of Type III Hsp40s in *E. coli*. With the exception of a few isolated cases, the remainder of the Helix I region of the J-domains of the Type III TbHsp40s were found to be well-conserved in terms of their sequence. The amino acid composition of Helix II was found to be more variable than that of Helix I but it was shown that many of the substitutions found in this helix were conservative substitutions (Figure 2.3). The remainder of the TbHsp40s display higher levels of conservation in the first 5 residues of the J-domain Helix II region than in other regions of this helix.

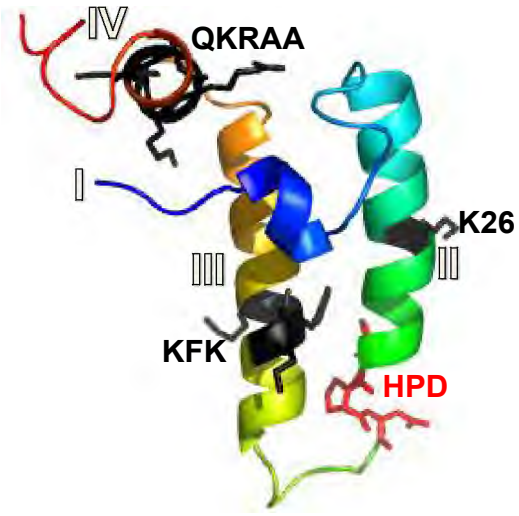
The HPD loop region is highly conserved throughout all the J-domains studied, with a 100 % frequency of occurrence. The prevalence of a Lys residue immediately following the HPD motif (28 %) is marked in a number of TbHsp40s, and is followed in frequency by an Arg residue at this position (25 %). The levels of conservation of amino acids in Helix II is significant for interactions with Hsp70 proteins, as this helix and the highly conserved HPD tripeptide have been proposed to comprise the minimal region of interaction between Hsp40s and Hsp70s (Lu and Cyr, 1998). Helix III was found to be poorly conserved in all the TbHsp40s studied. The exception to this finding is the frequency of occurrence of Ala-Tyr residues (72 %) in the grouping of 5 amino acids at the end of the Helix III. The KFK motif is corrupted in the J-domains of the majority of Type III Hsp40s from *T. brucei* (Figure 2.3), with the exception of those of Tbj33, Tbj42, Tbj51, Tbj65 and Tbj67. Interestingly, three of these proteins (Tbj51, Tbj65 and

Tbj67) are also characterised by the presence of TPR domains in addition to the J-domain. Helix IV was shown to be the most poorly conserved helix in the Type III TbHsp40 J-domains (Figure 2.3). Only two amino acid residues occurring in DnaJ, Sis1 and Ydj1 feature prominently in a number of TbHsp40s. These residues are Arg (46.8 %) and TyrAsp (48.9 %) in the centre and end of the helix, respectively. The well-conserved Arg residue comprises part of the QKRAA motif that characterises Helix IV and has been shown to recognize DnaK in other Hsp40 proteins (Auger and Roudier, 1997). The lack of conservation of this helix could have implications for TbHsp40 function, particularly with respect to its interaction with Hsp70 proteins. The extreme C-terminal regions of the J-domains studied (regions downstream from Helix I) displayed no consensus with respect to one another and these regions, while having formed part of the original alignment, were not displayed in the figure.

2.3.7. Derivation of a J-domain consensus sequence for the *T. brucei* Type III Hsp40s

The multiple sequence alignment presented in Figure 2.3 was used to generate a consensus sequence (Figure 2.4) for the J-domains of the Type III TbHsp40s in order to study the levels of conservation of key residues in more detail. This was performed in a similar way to the study performed by Hennessy et al (2000). For simplicity, the consensus residues were numbered using the *E. coli* numbering, and for the purposes of the present work, all residue numbers given were related to this numbering system. The consensus was derived by taking into account which residues occurred most commonly at a given position, while in cases where the amino acid content of a specific region of the J-domain was too variable, an X denoting the variability at that position was inserted on the consensus sequence (Figure 2.4). The key residues that have been shown to be important for the functionality of the J-domain in *E. coli* DnaJ and other J-domains have been highlighted on the figure of the *E. coli* DnaJ J-domain and the helices of the consensus sequence have been annotated accordingly (Figure 2.4). Tyr^{7} and Leu^{10} in Helix I, which have been implicated in stabilization of the

helix-loop structure of Helix II and Helix III of the J-domain, were found to be conserved in the majority of J-domains (Figure 2.4). The remainder of the residues in Helix I were well-conserved with respect to both the *T. brucei* Type III



HELIX I | HELIX II | HELIX III | HELIX IV

5- xx**Y**x**VLGV**xxx**AT**xx**EIRKAYRKLAL**xx**HPDK**(7x)**A**x**EKF**xx**AYEIL**x**D**xxxxx**R**xx**YD**xx

Figure 2.4. Derived consensus sequence of the J-domains of the Type III TbHsp40 complement

The backbone structure of the canonical *E. coli* DnaJ J-domain (1XBL) is shown with key residues important for function shown as black sticks and annotated on the figure. The model was rendered using Pymol (DeLano, 2002) and the helices are numbered on the figure. The consensus sequence derived for the J-domains of the Type III TbHsp40 complement is shown beneath the model with the consensus residues in bold and where x denotes any residue. Residues in the consensus sequence that correspond to residues important for Hsp40 function and interaction with Hsp70s are shown in black while other residues are shown in green. The helices are annotated above the consensus sequence and the highly conserved HPDK region is shown in red.

Hsp40 complement and the J-domains of Djla, Sis1 and Hdj2 with the Val^{9}LeuGluVal^{12} consensus region formed by this helix (Figure 2.4). In Helix II, the critical Lys / Arg^{26} residue was found to be highly conserved, with a slight majority of the TbHsp40 possessing an Arg^{26} residue at this point (Figure 2.4).

The crucial HPD tripeptide sequence was fully conserved throughout, and a significant number of TbHsp40s have a Lys^{36} residue immediately following this region. As previously mentioned, the Lys^{36} and Asn^{37} residues have been proposed to form a critical region in conjunction with the HPD motif (Genevaux *et al.*, 2002). While a majority of the J-domains studied possessed the Lys^{36} residue, a distinct minority possessed the Asn^{37} residue; at position 37 a number of different residues including Ala^{37}, Pro^{37}, Cys^{37}, $Thre^{37}$ and Glu^{37} were found (Figure 2.4). As can be observed from the derived consensus sequence, the region between Helices II and III was found to be highly variable (Figure 2.4). This variability was observed in the general residue composition of both Helices III and IV (Figure 2.4) although certain residues in Helix III were found to be highly conserved. The residues of the KFK motif were found to be highly variable and were analysed in more detail, due to the importance of this motif in Hsp40 function (Figure 2.5).

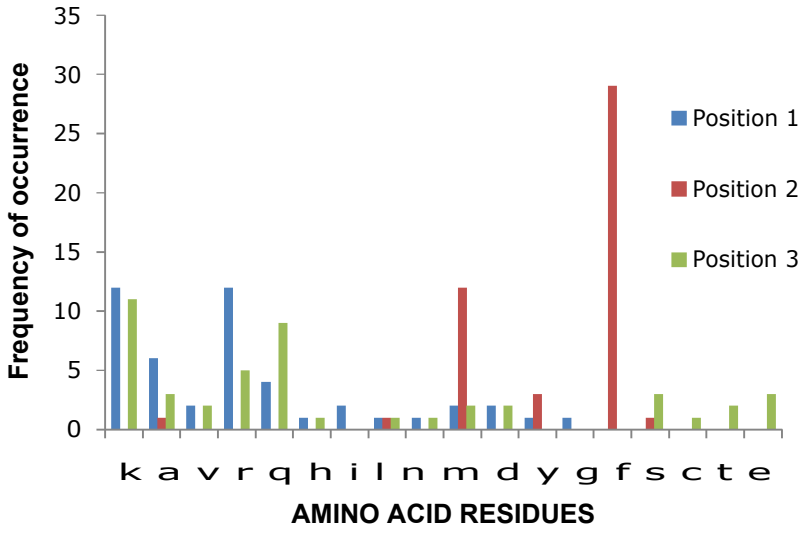


Figure 2.5. Analysis of the levels of conservation of the KFK motif in *T. brucei* Type III Hsp40 J-domains

The frequency of occurrence of the substituted residues in the conserved KFK region is shown. Position 1 refers to the initial Lys^{46} residue of the KFK motif, and all amino acids occurring in this position are indicated with blue. Position 2 refers to the Phe^{47} residue, and substitutions occurring in *T. brucei* are shown in maroon. Position 3 refers to the Lys^{48} residue, and substitutions are shown in green. The frequency of occurrence refers to the actual number of times a residue was counted in the Type III Hsp40 J-domains in *T. brucei*.

The consensus study performed indicated that the Lys^{46} residue of the KFK motif was conserved in approximately half the J-domains surveyed (Figure 2.3, Figure 2.4 and Figure 2.5). The Phe^{47} residue at the second position of the KFK motif was highly conserved (30 % of all J-domains studied) while the third position of the motif tendend to be more variable, and could include residues such as Lys^{48}, Val^{48}, Arg^{48}, Glu^{48}, Cys^{48} and Thr^{48} (Figure 2.5). Previous studies on the KFK motif in *E. coli* have shown that the Phe^{47} residue interacts with His^{33} of the HPD motif, and is crucial for its function, although this was not the case for *Agt* DnaJ (Hennessy *et al.*, 2000; Hennessy *et al.*, 2005a; Genevaux *et al.*, 2002). Because the KFK motif has previously been shown to be essential for the function of certain J-domains, it is unclear what effect the highly variable nature of this region, if any, has on the functionality of the *T. brucei* Type III Hsp40s.

Another motif that is highly conserved in a number of J-domains but partially abrogated in the TbHsp40s is the QKRAA motif (Figure 2.4). The only residue of this motif that was found to occur on the consensus sequence generated in Figure 2.4 was the Arg^{63} residue of the QKRAA motif (Figure 2.4). A Tyr^{66} Asp^{67} region which was shown to occur downstream of the QKRAA motif on the TbHsp40s was found to be highly conserved in all proteins studied. This region also occurred in the J-domains of *E. coli* DnaJ, Ydj1 and Sis1, but not in that of *E. coli* Djla, and the high levels of conservation of these two residues indicate that more investigation of this region by means of mutagenesis studies could play an important role in future investigations of the nature of interactions within the J-domain.

2.3.8. Tbj1 and Tbj51: distinct and diverse Type III TbHsp40s

Tibbetts and colleagues cloned and partially characterised four members of the Hsp40 family in *T. cruzi* (Tcj1 – Tcj4) and found that Tcj1 is a divergent member of the Hsp40 family (Type III Hsp40) that is not heat inducible, and does not contain the G / F-rich region or zinc finger domains that are characteristic of

canonical Hsp40 proteins (Tibbetts *et al*., 1998). Tcj1 (Tc00.1047053511537.50) has a molecular weight of 35.4 kDa, a pI of 7.1 and was predicted to localize to the nucleus based on analysis by the Wolf PSORT. An ortholog of Tcj1 in *T. brucei* was identified and bioinformatically characterised. Tbj1 (Tb11.01.8750) is a Type III TbHsp40 protein that was found to bear an overall sequence identity of 64.3 % to Tcj1. It has a molecular weight of 35.2 kDa, a pI of 6.4 and was predicted to localize to the *T. brucei* nucleus by Wolf PSORT. The localization predicted by Wolf PSORT was confirmed by checking the Tbj1 and Tcj1 protein sequences for the presence of known nuclear localization signals (NLS) from the literature. Previous studies on the HOP homologue in the mouse, mSTI1 have shown that the NLS of this protein corresponds to the region found between residues 222 and 239 of the mSTI1 sequence (KQALKEKELGNDAYKKKD) (Longshaw *et al*., 2004). The NLS sequence of mSTI1 was used as a query to search the sequence of Tbj1 and Tcj1, and was found to match a region from residues 103-118 on the two proteins with respect to conserved substitutions and some identicial residues (data not shown). The putative NLS of Tbj1 and Tcj1 derived from this investigation was shown to be RQFVERKRLEDEAKKK, but further *in vivo* localisation studies are required to confirm the veracity of this data.

Tbj1 was found to possess no known orthologues in humans, yeast or prokaryotes, in cases where an orthologue is defined as a protein that has 40 % or greater sequence identity to the protein submitted for a protein BLAST. The Tbj1 sequence was submitted to the Multi Organism BLAST feature in the GeneDB database, and potential orthologues in other species were identified based on the same criteria above. Three putative Tbj1 orthologues were identified (percentage identity given in brackets): Tviv118g02.p1k_0 (56 %) in *T. vivax*, LinJ32_V3.3220 (49 %) in *L. infantum* and LbM32_V2.3220 (48 %) in *L. braziliensis*. No orthologues to Tbj1 were identified in the malaria parasite, *P. falciparum* or in the *G. morsitans T. brucei* vector.

Tbj51 is an 80.4 kDa TPR-domain containing TbHsp40 protein that was predicted to contain 5 TPR domains and is proposed to play an important role in mediating chaperone interactions in the *T. brucei* cytoplasm. It was selected in addition to Tbj1 and Tcj1 for further study by means of cloning and biochemical characterisation.

The J-domains of Tbj1, Tcj1 and Tbj51 were aligned with the J-domains from Djla, Sis1, DnaJ and Ydj1 (Figure 2.6) and homology models of the Tbj1, Tcj1 and Tbj51 J-domains were rendered in order to illustrate key residue differences (Figure 2.6 B-D). A number of the key residues in Helix I and II discussed in earlier sections of the present chapter were found to be highly conserved in the Tbj1, Tcj1 and Tbj51 J-domains studied (Fgure 2.6). The LGV region of Helix I is conserved throughout the canonical J-domains and Tbj1 / Tcj1 although the rest of the Helix I region of Tbj1 and Tcj1 resembles the yeast Sis1 more than that of the prokaryotic J-domains. The most notable exception to this observation was the Lys / Arg^{26} — Tyr^{26} substitution in Helix II that was observed in both Tbj1 and Tcj1 but not in Tbj51 (Figure 2.6). In contrast, Tyr^{26} was highly conserved amongst all the J-domains studied in this section (Figure 2.6). Arg^{26} in Tbj51 was followed by E^{27}, which adds a negative charge to a region that generally is rich in positive residues. The implications of this substitution are not known, and rational mutagenesis studies could assist with elucidating the function of this residue in the Tbj51 J-domain. Tbj1 and Tcj1 possessed an HPDKNP consensus sequence that was also observed in Ydj1 from yeast, while Tbj51 possessed an HPDKC consensus at this point. As was observed previously in Section 2.3.7, Helix III and IV were found to be more poorly conserved with respect to canonical J-domains. The KFK motif was altered to a VFK motif in Tbj1 and Tcj1, but was conserved in Tbj51 (Figure 2.6). The QKRAA motif was not conserved in Tbj1, Tcj1 or Tbj51 (Figure 2.6), but was substituted for a TQREM (Tbj1), ARRRL (Tcj1) and VKRRD (Tbj51) motif in these J-domains. The J-domains of Tbj1, Tcj1 and Tbj51 were modeled using the J-domain of *E. coli* DnaJ (1 XBL) as a template for reasons described previously.

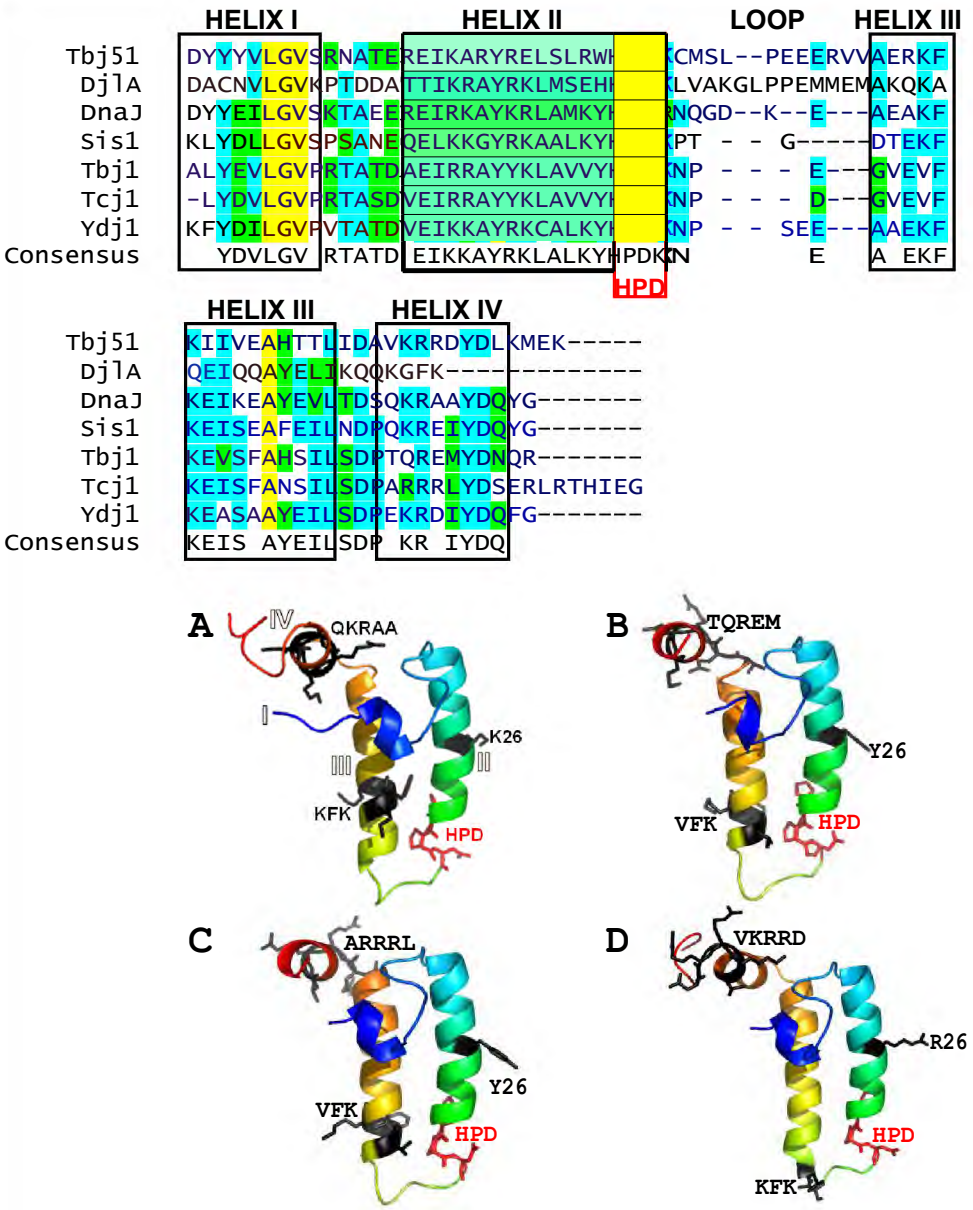


Figure 2.6. Multiple sequence alignment and homology modelling of the J-domains of Tbj1, Tcj1 and Tbj51

The multiple sequence alignment of the J-domains of Tbj1, Tcj1 and Tbj51 with those of a well-characterised Hsp40s in E. coli (Djla and DnaJ) and yeast (Ydj1 and Sis1) is shown. The Helices are highlighted on the figure and the HPD motif is shown in red. [Similar residues] – green; [Identical residues] – yellow background with red writing; [Conservative changes] – turquoise and [Non-similar residues] – black writing on a while background. Ribbon representations of the known or predicted J-domains of all the proteins in the alignment are depicted below the alignment, with the well-conserved HPD motif shown in red sticks. [A] –DnaJ (E. coli; PDB code 1XBL); [B] – Tbj1 [C] – Tcj1; [D] – Tbj51. Models were generated with Modeller (1 XBL as template) and visualised with Pymol (DeLano, 2002). Key residues that differ from DnaJ are shown as black sticks.

Overall, the tertiary J-domain structure of Tbj1 and Tcj1 is well-conserved when compared to the structure of other canonical Hsp40 J-domains although minor differences in structure and sequence may play a crucial role in determining potential binding partners (Figure 2.6). The models of the J-domains generated for Tbj1, Tcj1 and Tbj51 were checked using Procheck software adapted for Windows NT® (Laskowski *et al*., 1993; Morris *et al*., 1992). The models were found to be satisfactory for the purposes of the present investigation and the results are summarised in Appendix K.

2.3.9. Analysis of the C-terminal regions of Tbj1 and Tcj1

Type I and Type II Hsp40s are known to form dimers in solution, but the ability of Type III Hsp40s to do so has not been characterised (Wu *et al*., 2005). To this end, the C-terminal regions of Tbj1 and Tcj1 were compared to the C-terminal regions of Sis1 and Ydj1 which have been resolved by means of X-ray crystallography (Wu *et al*., 2005). The levels of identity between the C-terminal domain of Sis1 and Tbj1 / Tcj1 were very low and the Sis1 C-terminal domain was not included in the alignment (Appendix H; Figure H.1). For the purposes of the present investigation, the C-terminus of Tbj1 and Tcj1 was defined as the region following the J-domain (85 amino acid residues from the N-terminal region). Very few residues were found to be identical over the full length of the C-terminal regions, and while some similar amino acids had been shown, these were the minority (Appendix H; Figure H.1). Most of the amino acids of the Tbj1 and Tcj1 C-terminal domains (which possess a high sequence identity to one another) were found to be conservative substitutions with respect to that of the Ydj1 C-terminal domain. A comparison of the residues in Ydj1 shown to be important for dimerisation with the C-terminal residues of Tbj1 and Tcj1 revealed that only one of the nine residues considered critical for dimerisation in Ydj1 (Leu^{346}) was conserved in both Tbj1 and Tcj1. The C-terminal region of Tbj1 was used as a query in a BLAST search of both the NCBI and GeneDB databases. No homologous sequences (>30 % identity) were observed, with the exception of

the Tbj1 orthologues in *T. cruzi* (Tcj1) and *L. major* (Lmj1) (data not shown). Homology models of the Tbj1 and Tcj1 C-terminal domains could not be generated due to the fact that a suitable template possessing over 30 % similarity could not be found (Kelley and Sternberg, 2009).

2.3.10. TPR domain-containing TbHsp40s in *T. brucei*

A study of the *T. brucei* Type III Hsp40 complement revealed that the parasite possesses a total of 5 Hsp40s that contain TPR domains (Tbj51, Tbj52, Tbj53, Tbj65 and Tbj67). The J-domains of TbHsp40s containing TPR domains were found to be well-conserved with respect to other Hsp40 J-domains, but the TPR domains were found to be highly variable, rendering the development of any general consensus patterns difficult (data not shown). Tbj51, Tbj52 and Tbj53 displayed higher levels of sequence identity with respect to each other in terms of their full-length sequences than to Tbj65 and Tbj67, which could indicate more similar functions in the different organelles of the parasite in which they are located. Tbj51 and Tbj65 were predicted to localise to the cytoplasm, Tbj52 to the nucleus and Tbj67 to the mitochondria while Tbj53 was predicted to be secreted.

2.3.11. Putative Type IV Hsp40s in the TriTryps

Type IV Hsp40 proteins are an extension of the Type I-III class initially proposed by Cheetham and Caplan (1998). These proteins are defined as Hsp40 proteins that contain a corrupted HPD motif in their J-domains and were first identified in the *P. falciparum* parasite (Botha *et al.*, 2007). The *T. brucei* Hsp40 complement includes 6 proteins that have been annotated as Hsp40s despite possessing corrupted J-domains (Table 2.7).

Table 2.7. Putative Type IV Hsp40s in the TriTryps

Protein	Gene ID	Chromosome / contig"	Mr."	pI"	Localisation ξ
Tbj27	Tb09.211.0330	9	35.5	8.7	Cytoplasm
Tbj31	Tb927.7.990	7	86.6	8.5	Nucleus
Tbj47	Tb927.1.1230	1	52.3	8.7	Mitochondria
Tbj58	Tb11.02.3720	11	80.4	5.3	Nucleus
Tbj66	Tb927.7.2070	7	36.3	6.0	Cytoplasm
Tbj68	Tb927.8.6310	8	13.3	9.8	Secreted
Tcj31	Tc00.1047053506729.50	7071	87.9	9.5	Nucleus
Tcj47a	Tc00.1047053507949.10	7450	53.4	8.4	Mitochondria
Tcj47b	Tc00.1047053511423.170	6113	53.6	8.7	Mitochondria
Tcj66	Tc00.1047053511807.70	8792	35.7	5.9	Secreted
Tcj68a	Tc00.1047053503885.70	4801	13.3	9.6	Secreted
Tcj68b	Tc00.1047053511903.270	8822	13.3	9.6	Secreted
Lmj31	LmjF26.0940	26	92.7	10.2	Nucleus
Lmj47	LmjF20.0550	20	58.6	9.6	Mitochondria
Lmj66	LmjF22.0080	22	36.6	6.8	Secreted
Lmj68	LmjF24.1910	24	13.5	9.9	Secreted

" – Information obtained from the GeneDB database (http://www.genedb.org/)
ϕ – Proteins with transmembrane domains
Mr – Molecular mass in kDa
ξ - Localisation as predicted by WOLF Psort (http://wolfpsort.org/)

In *T. brucei*, these putative Type IV Hsp40s range in size from 13.3 kDa to 86.8 kDa and were predicted to localise in the cytoplasm (Tbj27 and Tbj66), the nucleus (Tbj31 and Tbj58), the mitochondria (Tbj47), while one putative Type IV TbHsp40 (Tbj68) was predicted to be secreted (Table 2.7). As was the case with the Type III TbHsp40 complement, the Type IV Hsp40s in *T. brucei* were found to possess orthologues in *T. cruzi* and *L. major* (Table 2.7). It was interesting to note that the *T. cruzi* and *L. major* orthologues of certain *T. brucei* Type IV Hsp40s were Type I Hsp40s with an uncorrupted HPD motif. This was the case for Tbj27 and Tbj58 which were found to be Type IV Hsp40s in *T. brucei* but have Type I orthologues in *T. cruzi* and *L. major*. The *T. cruzi* homolog of Tbj58 possesses two HPD motifs in tandem, although the biological significance of this finding is not known at present. The possibility of sequencing errors in this case cannot be ruled out and further studies in which these coding regions are amplified and sequenced will form part of future investigations into these proteins. *T. cruzi* possesses 4 putative Type IV Hsp40s (Table 2.7) if gene

duplications are not taken into account and *L. major* was found to contain the same number. The *T. cruzi* Type IV Hsp40 complement was predicted to be localised to the nucleus (Tcj31) and the mitochondria (Tcj47). Two Type IV TcHsp40s were predicted to be secreted (Tcj66 and Tcj68). The Type IV Hsp40s in *L. major*, like those of *T. cruzi* were also predicted to localise in the nucleus and mitochondria if they are not secreted. Therefore, this preliminary data suggests that only *T. brucei* possesses Type IV Hsp40s that are located to the cytoplasm. It was interesting to note that in certain cases (eg. Tbj31 and its orthologues) the corruption of the HPD motif was partial (HPD ~ HTD) and the J-domain was well-conserved with respect to key residues, whereas in other cases the corruption was complete which could have novel implications for protein functionality. An example of the latter is Tbj66, which could potentially not be an Hsp40 protein at all.

Tbj31 is an 86.6 kDa molecular chaperone that was predicted to localise to the nucleus (Table 2.7). It has orthologues in both *T. cruzi* and *L. major*. A sequence alignment of the J-domain of Tbj31 with canonical J-domains was performed as elsewhere in this chapter (Figure 2.7). Certain highly conserved J-domain residues such as Tyr^{7} and Leu^{10} were conserved in this protein, but the key Lys / Arg^{26} residue was altered to a Leu^{26} residue followed by a Glu^{27} residue (Figure 2.7; alignment). The HPDK motif that is commonly seen in Type III Hsp40s from *T. brucei*, was replaced by and HTDK motif in Tbj31 and its orthologs. The KFK motif was partially abrogated in Tbj31, and an RFQ motif was found in its place (Figure 2.7). Interestingly, the QKRAA motif in Tbj31 was better conserved with respect to that of *E. coli* DnaJ (GKRAA) than those found in most of the Type III TbHsp40s. The J-domain of Tbj31 was modelled using *E. coli* DnaJ (1XBL) as a template and was tested for accuracy in the same way as previously mentioned (Appendix K). In terms of levels of J-domain conservation, Tbj31and its orthologues are the only proteins that are classified as Type IV Hsp40s which were found to possess any significant similarity to canonical J-domains. At present, no studies are available to determine the effect

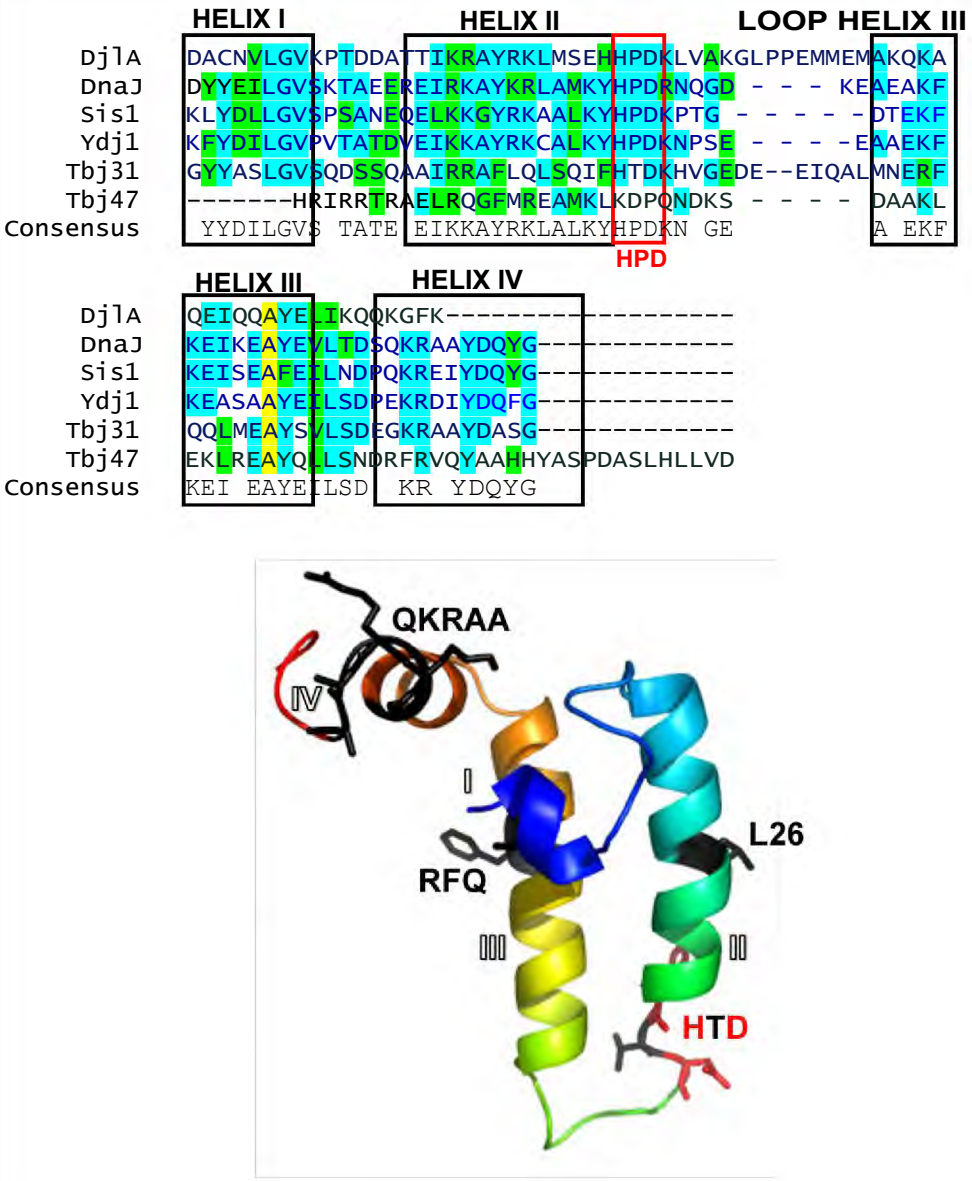


Figure 2.7. Comparison of the J-domains of Tbj31 and Tbj47 with those of canonical Hsp40 proteins

Multiple sequence alignment of the putative J-domains of Tbj31 and Tbj47 with the J-domains of well-characterised Hsp40s. The Helices are annotated on the alignment, with HPD motif in red. [Similar residues] – green; [Identical residues] – yellow background with red writing; [Conservative changes] – turquoise and [Non-similar residues] – black writing on a while background. The figure below is a homology model of Tbj31 based on an *E. coli* DnaJ (1XBL) template produced with Modeller software and visualised with Pymol (DeLano, 2002). The HTD loop region is shown in red sticks and helices are annotated on the figure.

of RNAi knockdown of Tbj31, but it has been detected in both the bloodstream plasma membrane fraction and in the procyclic form and thus appears to be expressed throughout the parasite's life cycle (Bridges *et al*., 2008; Jones *et al*., 2006).

Tbj47 (Tb927.1.1230) is a putative 52.3 kDa mitochondrial Type IV Hsp40 protein that was also shown to possess orthologues in both *T. cruzi* and *L. major*. This protein was of great interest in terms of drug target capabilities due to the availability of RNAi data indicating that the knockdown of Tbj47 in *T. brucei* results in a growth defect in bloodstream form parasites (Subramaniam *et al*., 2006). A general protein BLAST search of the NCBI database (http://blast.ncbi.nlm.nih.gov/Blast.cgi) revealed that, despite high levels of sequence conservation within the TriTryps, Tbj47 did not appear to have any orthologues (sequence identity of 40 % or higher) in other species. Protein sequences displaying any similarity (20-30 %) to Tbj47 were invariably bacterial. A BLAST analysis was also performed against the sequences in the GeneDB database (http://www.genedb.org/genedb/seqSearch.jsp) using the Tbj47 protein sequence as a query. Three potential orthologues were identified (percentage identity given in brackets): Tviv796e07.plk_2 (64 %), LinJ20_v3.0620 (45 %) and congo1350g02.q1k_11 (70 %) in *T. vivax*, *L. infantum* and *T. congolense* respectively. No clear orthologues of Tbj47 (proteins with >40 % identity) were found to exist in parasites such a *P. falciparum*, or in the *G. morsitans* insect vector of the *T. brucei* parasite, indicating that Tbj47 could perhaps elicit a specialised role in trypanosomal parasites. The putative J-domain of Tbj47 and its orthologues in *L. major* and *T. cruzi* was aligned and studied as outlined previously in the present work (Figure 2.7). Unlike the scenario observed in Tbj31, the overall J-domain architecture in Tbj47 was found to be very poorly conserved, with an identity of only 21 % to the *E. coli* DnaJ J-domain as opposed to 41 % for Tbj31. In terms of residues important for function, only Leu^{10} was found to be conserved in Helix I. The key Lys / Arg^{26} residue was also not conserved, and was altered to an Ala^{26} residue in Tbj47 (Figure 2.7), an Asp^{26}

residue in Tcj47 and a Thr^{26} residue in Lmj47 (data not shown). The highly conserved HPD loop region was replaced by LRR (Lmj47), KDP (Tbj47) and LRE (Tcj47a and b). Key residues such as the KFK (KLE) and the QKRAA motif (RFRVQ) in Helix III and IV were also not conserved in the Tbj47 J-domain.

The remainder of the TriTryp putative Type IV Hsp40 complement was analysed with respect to their levels of J-domain conservation (data not shown). Only putative Type IV Hsp40s that were found to possess orthologues in *T. brucei*, *T. cruzi* and *L. major* were analysed. Tbj66 and Tbj68 (and their orthologues) displayed very high levels of sequence variability in the J-domain region (data not shown). In the case of Tbj66, the HPD loop is altered to CDM, while Tcj66 has a CDD motif showing conservation with respect to the sequence in the loop region between the two species. Lmj66, however, possesses a DNE motif at this position, showing similarity with neither canonical J-proteins nor its orthologues in the TriTryps. Tbj68 and its orthologues contain an SED / TED motif, and show a similar lack of overall J-domain conservation to that displayed by other putative Type IV Hsp40s in *T. brucei*. The poor levels of sequence conservation for the Type IV Hsp40s in the TriTryps parasites could serve to indicate their suitability as novel drug targets, especially if found to be essential for parasite survival in conjunction with future RNAi studies.

2.3.12. The Hsp70 complement of *T. brucei*

T. brucei has 12 Hsp70 proteins, all of which have orthologues in *T. cruzi* and *L. major* (Table 2.8).

Table 2.8. The Hsp70 complement of *T. brucei*

T. brucei gene ID	Subfamily ♦	Mr."	pI"	Form / stage"	Localisation ξ
Tb11.02.5500¤	BiP / Grp78[1]	71.4	5.2	BSF & Pro[2]	E.R.
Tb11.02.5450¤	BiP / Grp78[1]	71.4	5.2	BSF & Pro[2,3]	E.R.
Tb927.6.3740¤	Mitochondrial	71.4	5.2	BSF & Pro[2]	Mitochondria
Tb927.6.3750¤	Mitochondrial	71.4	5.8	BSF mem[3]	Mitochondria
Tb927.6.3800¤	Mitochondrial	71.4	5.8	BSF & Pro[2,3]	Mitochondria
Tb11.01.3110	Hsp70	75.3	6.3	BSF & Pro[2,3]	Cytoplasm
Tb11.01.3080	Hsp70.c	73.6	4.9	BSF & Pro[2,3]	Cytoplasm
Tb927.7.710	Hsp70.4	70.2	4.6	BSF & Pro[2,3]	Cytoplasm
Tb10.389.0880	Hsp110	90.8	5.1	BSF & Pro[2,4]	Cytoplasm
Tb09.211.1390¤	Grp170	79.8	8.5	BSF mem[3]	Secreted
Tb927.7.1030	Hsp70.b	93.2	6.0	-	Nuclear cytoplasm
Tb09.160.3090¤	Hsp70.a	90.8	9.2	BSF mem[3]	Secreted

[1] These proteins are 99.8 % identical
[2] Jones *et al.*, 2006
[3] Abnormal endocytosis, golgi organisation, biogenesis (BSF); normal mitochondrial organisation and biogenesis, cell cycle, cell morphogenesis (BSF). RNAi lethal in BSF (Bridges *et al.*, 2008; Subramaniam *et al.*, 2006)
[4] Vertommen *et al.*, 2008
BSF – Bloodstream form
BSF mem – Bloodstream plasma membrane fraction
Pro – Procyclic form
" – Information obtained from the *T. brucei* GeneDB database (http://www.genedb.org/genedb/tryp/)
¤ - Proteins with a signal peptide
Mr – Molecular mass in kDa
ξ - Localisation as predicted by WOLF Psort (http://wolfpsort.org/)
♦ - Classification as per Folgueira and Requena, 2007

The sequences of all TbHsp70s were retrieved from the GeneDB database and sorted into subfamilies as per Folguiera and Requena (2007). The potential

localisation of these proteins was investigated using the Wolf PSORT program in identical fashion to the study of the Hsp40s in Section 2.3.

2.3.12.1 TbHsp70 is a typical eukaryotic Hsp70

The sequence identities of TbHsp70 (Tb11.01.3110), LmHsp70 (LmjF28.2770) and TcHsp70 (Tc00.1047053511211.170) were compared with the sequences of well-characterised Hsp70s (*E. coli* DnaK and *S. cerivisiae* Ssa1p) as well as the cytosolic Hsp70 from *H. sapiens* and the Hsp70 from *Medicago sativa* (Alfalfa) which is commercially available (Table 2.9).

Table 2.9. Percentage identities for Hsp70s from the TriTryps with canonical Hsp70 proteins in other species

	Alfalfa Hsp70	DnaK	LmHsp70	TbHsp70	TcHsp70	Ssa1p	HumHsp70B
Alfalfa Hsp70	-	48	71	70	71	72	71
DnaK	48	-	48	46	46	48	48
LmHsp70	71	48	-	86	87	71	70
TbHsp70	70	46	86	-	88	71	69
TcHsp70	71	46	87	88	-	73	70
Ssa1p	72	48	71	71	73	-	70
HumHsp70B	71	48	70	69	70	70	-

TbHsp70 and TcHsp70 were found to share a high sequence identity (88 %) indicating that the cytosolic Hsp70 protein is well-conserved across the two species. LmHsp70 was equally well conserved with respect to TbHsp70 (85 %) and TcHsp70 (87 %). All three Hsp70s had a low overall sequence identity with respect to *E. coli* DnaK, with an average of 47 % identity for the TriTryp Hsp70s and DnaK. Interestingly, the TriTryp Hsp70s displayed a higher sequence identity when compared to *M. sativa* Hsp70 (average of 71 %) and Ssa1p from *S. cerivisiae* (72 %). Both TbHsp70 and TcHsp70 also shared a high sequence identity to Human Hsp70 (70 % and 69 % respectively). The cytosolic Hsp70 proteins from *T. brucei*, *T. cruzi*, yeast, human and *M. sativa* all possessed a relatively low sequence identity (< 50 %) to DnaK (Table 2.9). A phylogenetic

analysis was performed on TbHsp70, TcHsp70 and LmHsp70 with well-characterised Hsp70s from different species (Figure 2.8).

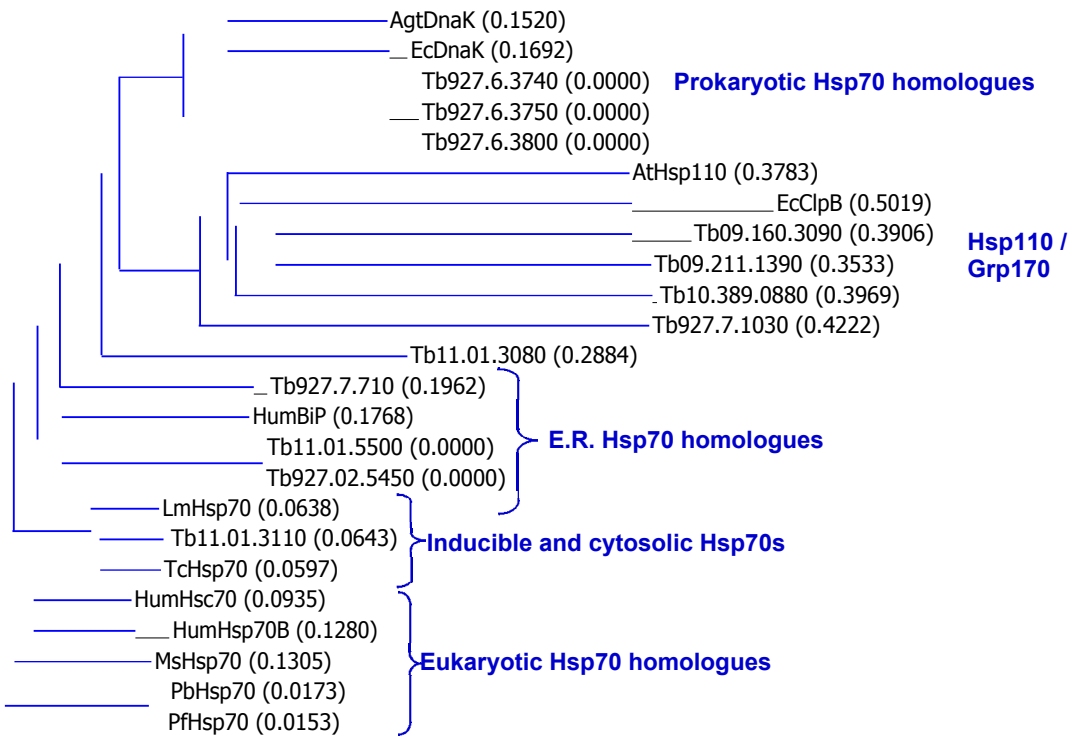


Figure 2.8. Phylogenetic analysis of TbHsp70 and its orthologues in *T. cruzi* and *L. major*

The Hsp70 complement from *T. brucei* was analysed in conjunction with known Hsp70 proteins from other organisms. The NCBI and PlasmoDB accession numbers are as follows: *Plasmodium falciparum* (PfHsp70, PF08_0054); *Plasmodium berghei* (PbHsp70, AAL34314); *E. coli* DnaK (EcDnaK, BAA01595); *Trypanosoma cruzi* Hsp70 (TcHsp70, P05456); *Leishmania major* (LmHsp70; LmjF28.2770); *Homo sapiens* Hsc70 (HumHsc70, AAK17898); *Agrobacterium tumefaciens* DnaK (AgtDnaK, AAR84665); *Homo sapiens* BiP (HumBiP, CAA6120); *Arabidopsis thaliana* Hsp110 (AtHsp110, NP_567510); *E. coli* ClpB (EcClpB, AAB49540); *Medicago sativa* Hsp70 (MsHsp70, AAV98051); *Homo sapiens* inducible Hsp70 (HumHsp70B; 34419635). The Hsp70 proteins were divided into groups based on the alignment as shown on the right side of the figure. The alignment and tree were produced in the AlignX component of the VectorNTI® suite.

TbHsp70 (Tb11.01.3110) clusters with TcHsp70 on the lower branch of the tree in a monophyletic clade that is close to the eukaryotic clade, but distant from the prokaryotic clade (Figure 2.8). PfHsp70, an Hsp70 protein from the *P. falciparum* parasite, clusters with the Hsp70 proteins from *P. berghei* and human under

eukaryotic Hsp70 proteins. Two Hsp70 proteins (Tb11.02.5500 and Tb11.02.5450) were found to cluster with Human BiP that is localised to the E.R., which corresponds to the predicted localisation obtained from the Wolf PSORT program (Table 2.8).

The three TbHsp70s predicted to be mitochondrial by Wolf PSORT (Tb927.6.3740, Tb927.6.3750, and Tb927.6.3800) cluster with DnaK from *E. coli* and *A. tumefaciens* and show similarity to prokaryotic Hsp70 proteins. The remainder of the TbHsp70 complement clusters with ClpB and Hsp110 from *Arabidopsis thaliana (A. thaliana)*. ClpB is a member of the Hsp100 family which are central components of the protein quality control system in prokaryotes, by degrading or unfolding aggregated proteins (Haslberger *et al.*, 2007). ClpB functions by co-operating with the Hsp40-Hsp70 complex to facilitate the solubilisation and reactivation of aggregated proteins, which is crucial to facilitate cellular survival of extreme stress conditions (Weibezahn *et al.*, 2004). It is proposed that Tb927.6.3740, Tb927.6.3750, and Tb927.6.3800 may be responsible for facilitating the survival of the *T. brucei* during the drastic changes experienced during its transition from an insect vector to a mammalian host. Tb09.211.1390 clusters with the ClpB homologues and is proposed to be a Grp170 homologue, which confirms the annotation of Folgueira and Requena (2007).

2.3.12.2. Conservation of the Hsp70 ATPase domain in *T. brucei*

The ATPase domains of TbHsp70 and TcHsp70 were aligned with those of *M. sativa*, *E. coli* DnaK and human Hsp70 proteins (Figure 2.9). Regions that are considered important for Hsp70 function based on the literature and which are normally well-conserved in Hsp70s across species are indicated on the figure. The alignments were annotated using information that has been obtained from studies using *E. coli* DnaJ and DnaK as models (Mount, 1985; Gässler *et al.*, 1998).

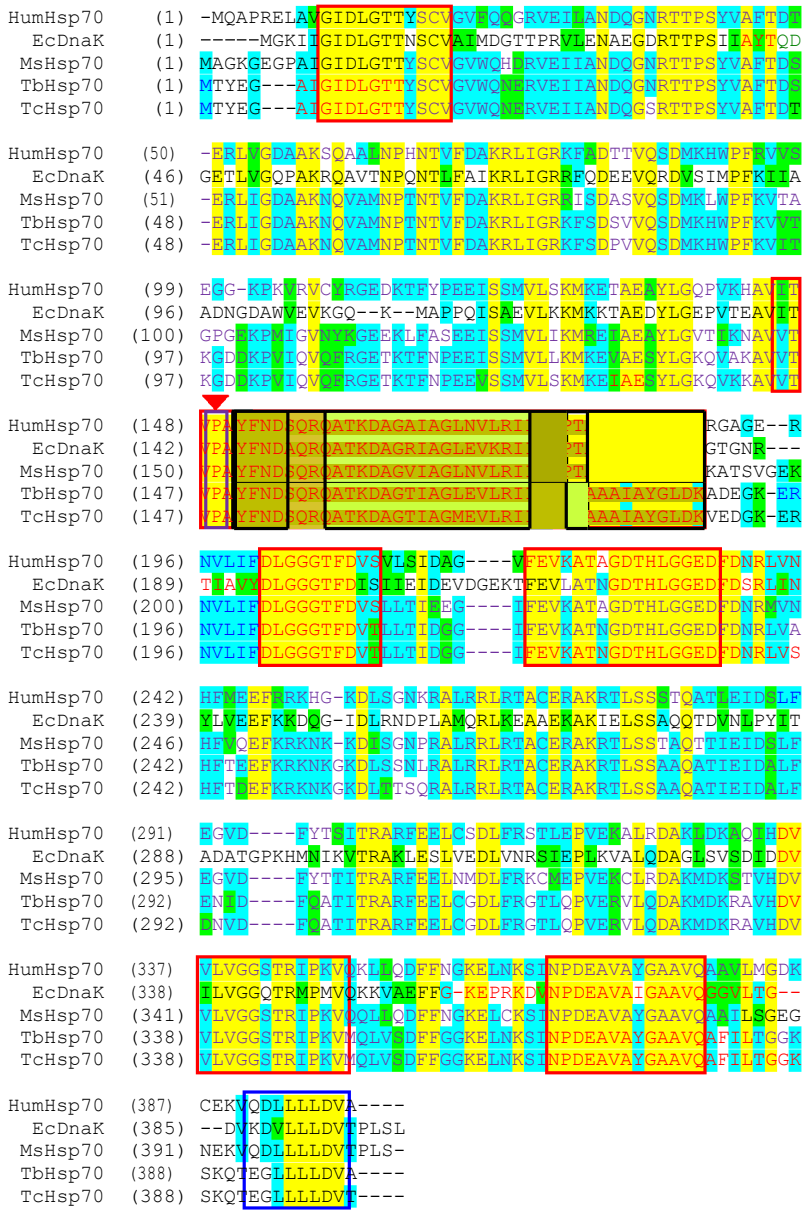


Figure 2.9. Sequence alignment of the ATPase domains of TbHsp70 and TcHsp70 with E. coli DnaK, HumHsp70 and M. sativa ATPase domains

The ATPase domains of TbHsp70 and TcHsp70 were aligned with those of E. coli DnaK (BAA01595), M. sativa (AAV98051) and HumHsp70 (34419635) using the Align X component of the Vector NTI® suite. Key features that are significant to Hsp70 function are highlighted: Red boxes – highly conserved residues associated with Hsps (Mount, 1985); Black boxes – DnaK residues that are proposed to interact with DnaJ viz. Y145, N147, D148, N170 and T173 (Gässler et al., 1998). The red arrow points to the proline allosteric switch (purple box) as well as R151 which is important for interdomain function. The blue box shows the highly conserved linker region (Vogel et al., 2006). [Similar residues] – green; [Identical residues] – yellow background with red writing; [Conservative changes] – turquoise and [Non-similar residues] – black writing on a white background.

As was the case in a study on PfHsp70 in the malaria parasite in our laboratory, it was found that the ATPase domains of TbHsp70 and TcHsp70 were more highly conserved than their substrate binding domains (Shonhai *et al*., 2007). The sequence alignment revealed that the ATPase domain of TbHsp70 possesses a 95 % identity to that of TcHsp70, a 78 % identity to that of Hsp70 from *M. sativa* and a 77 % identity to that of human Hsp70. Both the TbHsp70 and TcHsp70 ATPase domains showed higher sequence identity to the ATPase domain of PfHsp70 (77 % and 75 % respectively) and human Hsp70 than to the ATPase domain of DnaK (51 %) (data not shown). The proline allosteric switch (red arrow in Figure 2.9) was found to be conserved across all the Hsp70s in the present study (Vogel *et al*., 2006). In addition, the highly conserved linker segment (DVLLLD) that occurs between the ATPase domain and the peptide binding domain was shown to be present (Figure 2.9; blue box), with a Gly substituted in place of the first Asp of the sequence in the case of TbHsp70 and TcHsp70.

The peptide binding domains of TbHsp70 and TcHsp70 were also aligned with and compared to, DnaK, MsHsp70 and human Hsp70 (Figure 2.10) as well as PfHsp70 (data not shown). The peptide binding domains of the cytosolic trypanosomal Hsp70s showed lower levels of conservation with respect to other Hsp70 than that of the ATPase domains. The levels of sequence conservation were higher in the β-subdomain of the peptide binding region, which has previously been observed in other studies of a similar nature on different Hsp70s (Shonhai *et al*., 2007). Most of the binding sites deemed important for Hsp70-Hsp40 interaction to take place were shown to be well conserved in both TbHsp70 and TcHsp70, with the notable exception of Met^{404} in DnaK which is Ala^{404} in PfHsp70, TbHsp70, TcHsp70 and MsHsp70, and Ala^{429} which is Tyr^{429} in the same Hsp70 molecules. The absence of the highly conserved EEVD motif in TbHsp70 but its presence in TcHsp70 was of interest, as this could have major implications for a difference in functionality between the two proteins, as well as for the Hsp70-Hsp40 chaperone machinery in the Trypanosomes. The EEVD

motif plays a regulatory role in Hsp40 / Hsp70 function and it is unclear how this affects the interaction between TbHsp70s and the Hsp40 proteins in the parasite.

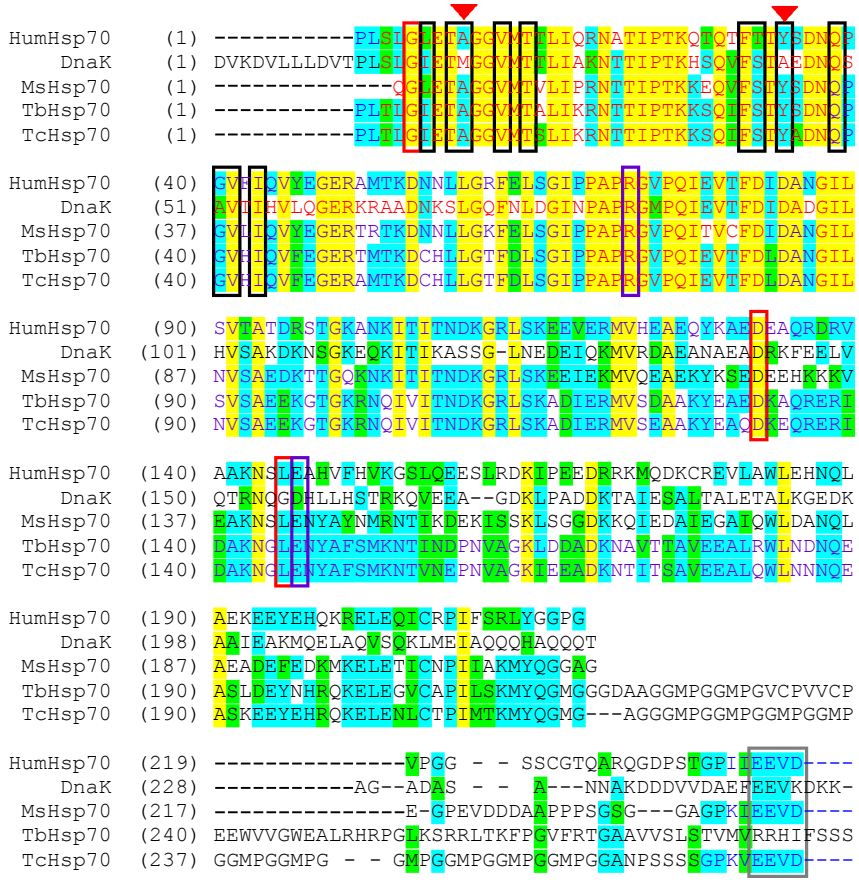


Figure 2.10. Multiple sequence alignment of the peptide binding domains of TbHsp70 and TcHsp70 with those of well-characterised Hsp70 proteins in other species

The peptide binding domains of TbHsp70 and TcHsp70 were aligned with those of E. coli DnaK (accession number BAA01595), M. sativa (accession number AAV98051) and HumHsp70 (34419635) using the Align X component of the Vector NTI® suite. Key features that are significant to Hsp70 function are highlighted: Black boxes – residues in contact with substrate in DnaK (Zhu et al., 1996); Red arrows - residues M404 and A429 comprising the arch; Red boxes – residues (G400, D526, G539) that are important for the DnaK – DnaJ interaction (Suh et al., 1998); Purple boxes – residues implicated in stabilisation of the Hsp70 lid region (Ha et al., 1997; Chang et al., 2001); the grey box shows the highly conserved EEVD region. [Similar residues] – green; [Identical residues] – yellow background with red writing; [Conservative changes] – turquoise and [Non-similar residues] – black writing on a white background. The numbers to the left of the figure are arbitrary, taking the start of the alignment as position one. Residues referred to in the text are defined in the literature and were compared with the above alignment and that of the full Hsp70 sequences by inspection.

The extreme C-terminus of TbHsp70 and TcHsp70 were not well-conserved with respect to DnaK, PfHsp70 or human Hsp70 (Figure 2.10) and could play a role in the specificity of the proteins. The sequence of TbHsp70 and TcHsp70 was used to generate homology models using the structure of bovine Hsc70 resolved by Jiang and colleagues (2005) as a template (Figure 2.11).

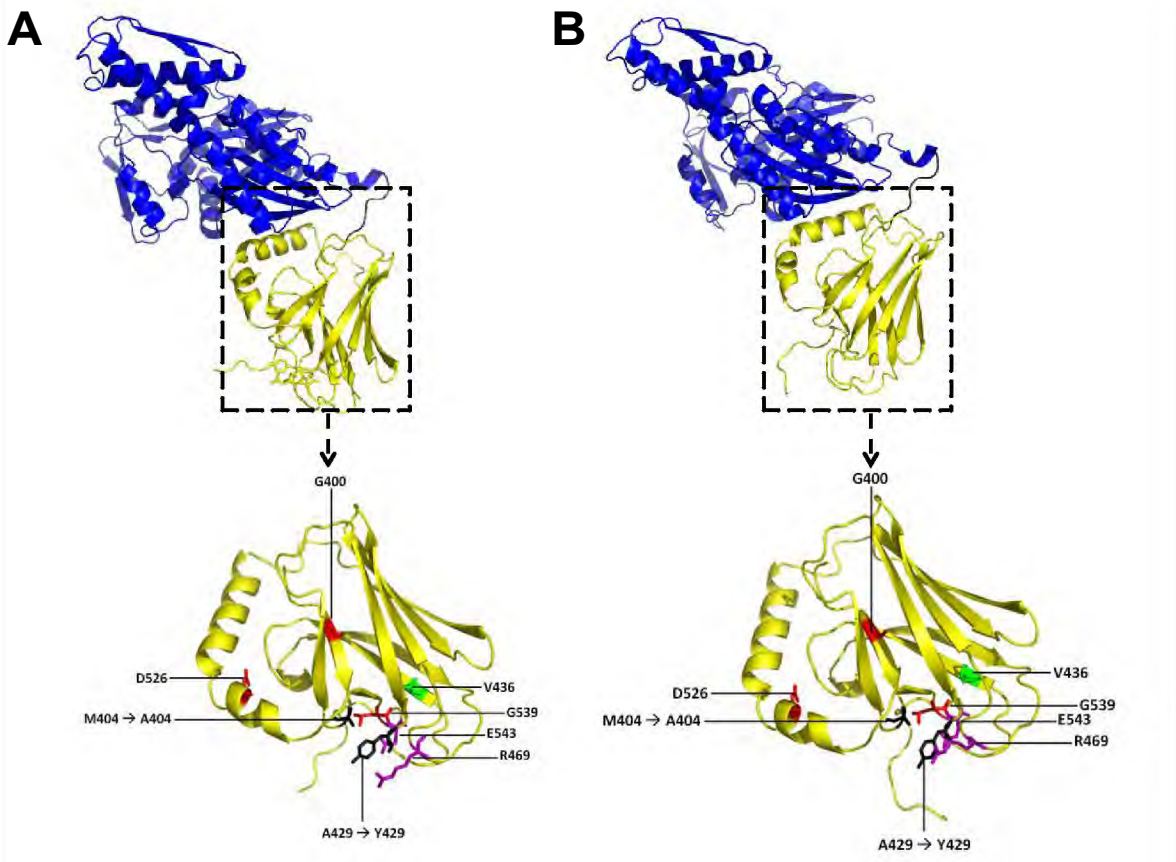


Figure 2.11. Putative TbHsp70 and TcHsp70 residues for interaction with Hsp40 partners

The full length sequences of TbHsp70 [A] and TcHsp70 [B] were modelled using Modeller 9v3 software using the crystal structure of bovine Hsc70 (1YUW; Jiang et al., 2005) as a template. The N-terminal ATPase domains are shown in blue, the linker regions in green and the C-terminal substrate-binding domains in yellow. The substrate binding domains (yellow) are enlarged below the full length structures. Residues that are involved in substrate binding are shown as red sticks (G400, D526, G/L539) with their corresponding amino acids in TbHsp70 and TcHsp70 shown after the arrow. Residues that are involved in forming the arch are shown in black (M/A404 and A/Y429) while the residues implicated in the formation of the latch for the Hsp70 lid region (E543 and R469) are shown in purple. Residues are annotated on the figure using the standard single amino acid code. Models were visualised with Pymol (DeLano, 2002).

The Jiang model does not contain the extreme C-terminus (final 10 amino acids) of the Hsp70 protein which prevented the visualisation of the non-conserved EEVD region in TbHsp70 (Jiang *et al*., 2005). The homology models suggest that, TbHsp70 and TcHsp70 are highly similar on a general structural level (Figure 2.11). The residues involved in peptide binding and Hsp40-Hsp70 interaction in TbHsp70 and TcHsp70 are fully conserved with respect to those of well-characterised Hsp70s such as MsHsp70 (Figure 2.11). Interestingly, a notable exception to this is arch-forming residues M404 and A429, which are A404 and Y429 respectively in both TbHsp70 and TcHsp70, which could have implications for function *in vivo* (Figure 2.11). The accuracy of the homology models was assessed as previously and the results of the analysis are shown in Appendix K.

2.3.12.3. Potential Hsp70-Hsp40 partnerships in the *T. brucei* parasite

The localisation data obtained from Wolf PSORT for the Type III and Type IV Hsp40s in *T. brucei* as well as the Hsp70 complement of the parasite was used to develop a scheme of overall localisation in an attempt to predict potential binding partners for each Hsp70 in a given organelle (Figure 2.12). This data should be regarded as highly preliminary due to the lack of reliable localisation predictors for proteins of the TriTryps to date, and the fact that online databases are error prone. The largest numbers of Hsp70s in the parasite were predicted to be cytoplasmic (4) or mitochondrial (3). In fact, the greatest numbers of Hsps were found to be located to these two organelles which was to be anticipated based on studies performed in other organisms. The putative Type IV Hsp40s had a diverse distribution, with Tbj47 predicted to localise to the mitochondrion and Tbj31 and Tbj58 to the nucleus, while Tbj27 and Tbj66 are proposed to localise in the cytoplasm. The remaining putative Type IV Hsp40, Tbj68, was thought to be secreted from the parasite or membrane-associated, and may interact with an Hsp or other factor in the human bloodstream.

In terms of potential partnerships, the 3 mitochondrial TbHsp70s were found to have a potential total of 13 Type III Hsp40 binding partners. In contrast, the 4 cytoplasmic Hsp70s had a potential complement of 7 Type III Hsp40s for interaction. A total of 8 chaperones (2 Hsp70s, 4 Type III Hsp40s and 1 Type IV Hsp40) were predicted to be secreted or membrane associated, which could suggest that they play a role in parasite defence against the human host.

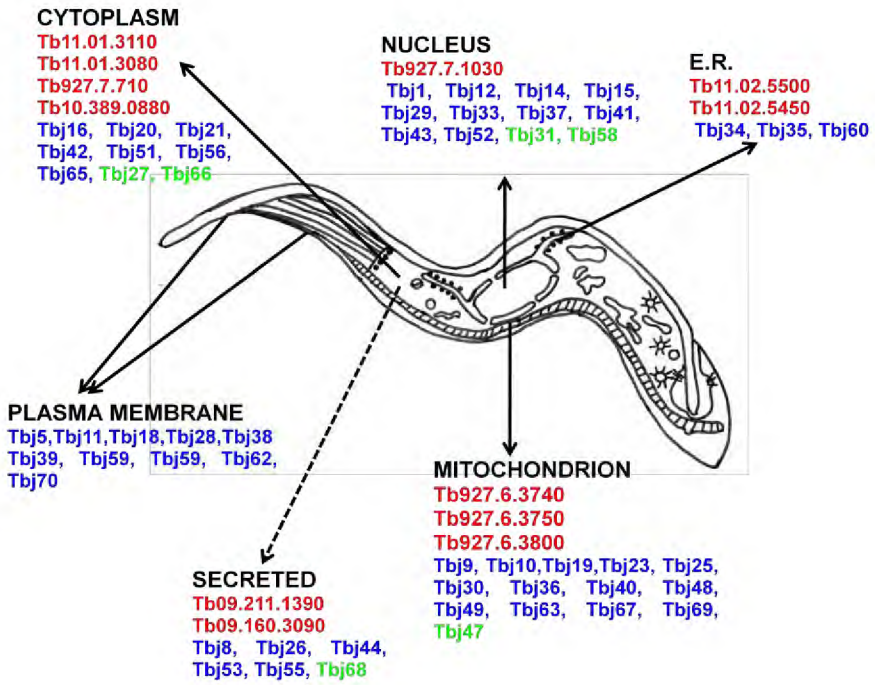


Figure 2.12. Distribution of Type III and Type IV Hsp40s in the *T. brucei* parasite with potential partner Hsp70s

Organelles and their predicted location in the parasite are indicated by black arrows on the trypanosomal body and labelled in black. The Type III Hsp40s predicted by the Wolf PSORT (http://wolfpsort.org/) program to be localised to a given organelle are shown in blue, the putative Type IV Hsp40s are shown in green and the TbHsp70s are shown in red. Secreted Hsp40s are indicated by a dashed line. Image taken and adapted from Internet 3.

This is proposed because both the Hsp70s (Tb09.211.1390 and Tb09.160.3090) were previously detected in the bloodstream form of *T. brucei* and may play a role by interacting with human Hsp40 proteins / proteins in the bloodstream and

aid in parasite survival mechanisms. Interestingly, none of the Type III and Type IV Hsp40s that are proposed to be secreted from the parasite have been detected in the bloodstream form of the parasite according to the information given on the GeneDB database (Appendix AA. on attached CD). It is uncertain at this stage whether or not any of these proteins could act as partners to Tb09.211.1390 or Tb09.160.3090. *T. brucei* also possesses 10 membrane-bound / membrane associated Type III Hsp40 proteins in addition to 3 Hsp40s predicted to be localised to the E.R. along with two potential partner Hsp70s. These Hsp70 proteins (Tb11.02.5500 and Tb11.02.5450) were found to be 99 % identical with respect to their primary sequences and were proposed to be BiP homologues. Interestingly, they were found to possess a PMDDL consensus sequence in place of the more conserved KDEL retention signal. The prediction of Tbj34 (a Sec63 homolog) localising in the E.R. along with the two *T. brucei* BiP homologues was interesting in light of the fact that this partnership is known in other organisms (Goldschmidt *et al.*, 2008). This would appear to indicate that some of the localisation signals of the TriTryps Hsp complement, while being markedly different from those used by other organisms, is still recognised by the localisation prediction software. In order to verify the accuracy of the remainder of the partnerships highlighted in Figure 2.12, an in-depth comparison with known Hsp 40 / Hsp70 partnerships for each protein would have to be made. This will offer an exciting new perspective on the Hsp40 / Hsp70 interactions and partnerships within the *T. brucei* parasite, but falls outside the scope of the present work.

2.4. DISCUSSION AND CONCLUSIONS

The present work aimed to elucidate levels of J-domain conservation in the Type III and Type IV *T. brucei* Hsp40 complement, within the wider context of the Hsp40 complement of the TriTryps parasites. A study of the potential Hsp40-Hsp70 partnership in *T. brucei* was also conducted. The present work is the first of this nature to be completed for the Hsp40 complement of the TriTryps parasites, with the exception of the study involving the identification and nomenclature of general molecular chaperones of *T. brucei*, *T. cruzi* and *L. major* by Folguiera and Requena (2007).

2.4.1. The Hsp40 complement of the TriTryps parasites

The analysis presented here revealed that the TriTryps parasites have a larger Hsp40 complement than several other organisms. *T. brucei*, *T. cruzi* and *L. major* have in excess of 65 Hsp40s each in comparison to the 49 Hsp40s found in *H. sapiens* and the 43 Hsp40s in the malaria parasite, *P. falciparum* (Hageman and Kampinga, 2008; Botha *et al*., 2007). The majority of the Hsp40 complement of these parasites comprises the poorly conserved Type III Hsp40s, which only retain the J-domain of all the canonical J-domain subdomains. The Type III Hsp40s were of interest due to their identity as a functionally distinct group of proteins that are thought to be unable to bind non-native peptides, leading to the hypothesis that they are unable to function as molecular chaperones in their own right (Walsh *et al*., 2004).

2.4.2. The Type III Hsp40 complement in the TriTryps parasites

While possessing few orthologs in species outside the Trypanosomatids, the TriTryps Type III Hsp40s were found to be highly conserved with respect to their orthologs in *T. brucei*, *T. cruzi* and *L. major*. The exception to this observation was Tbj70 which does not possess orthologs in either *T. cruzi* or *L. major*,

indicating that this protein could potentially fulfil a special requirement of the *T. brucei* parasite. A similar divergence in terms of sequences outside the J-domain was observed for the Type III Hsp40 complement of *P. falciparum*, with little conservation existing in terms of overall sequence with the exception of the J-domain region for all protein studied (Botha *et al*., 2007).

2.4.3. Levels of conservation of the J-domains in Type III Hsp40s in the TriTryps

The J-domain plays a vital role in the interaction between Hsp40s and their partner Hsp70s (Kelley, 1998; Tsai and Douglas, 1996; Hennessy *et al*., 2005). The highly conserved HPD motif that forms part of the loop region of the J-domain has been shown to be necessary for the stimulation of ATP hydrolysis in Hsp70; however, this region alone is not sufficient to accomplish the stimulation of the ATPase activity of Hsp70 (Tsai and Douglas, 1996). The identity of additional residues in the J-domain that play a role in the interaction with Hsp70 and the stimulation of its ATPase activity are of great interest in furthering an understanding of the mode of Hsp40-Hsp70 interaction, particularly in parasites of medical significance. The J-domains of the poorly conserved Type III and Type IV TbHsp40s were analysed *in silico* in order to obtain an enhanced understanding of the differences and similarities in this region, which could pave the way for future rational mutagenesis, structural and functional studies in the field of trypanosomal Hsp40 research. A general consensus sequence was derived for the J-domains of the Type III TbHsp40s to determine the relative levels of conservation of the J-domain with respect to the J-domains of well-characterised Hsp40 proteins from the literature.

The general consensus sequence developed for the *T. brucei* Type III Hsp40s (Figure 2.4) indicated that residues previously implicated in the stabilisation of the helix-loop structure between Helices II and III, such as Tyr^{7} and Leu^{10} (Hennessy *et al*., 2000) were conserved in most instances. The stability of the Helix II and Helix III region is critical to the functional ability of a J-domain with an

Hsp70 partner (Hennessy *et al*., 2005a) and the high levels of conservation of these residues in *T. brucei* could indicate that, structurally at least, a number of the Type III Hsp40s found in this parasite behave in a similar fashion to those from other organisms. The Lys / Arg^{26} residue that has previously been shown to be crucial in facilitating Hsp70 binding by the J-domain (Nicoll *et al*., 2007) was conserved in most of the J-domains studied, with a slight bias towards the occurrence of Arg^{26} instead of Lys^{26}. Thus, Helix II and Helix III of the Type III TbHsp40s were found to be well-conserved with respect to certain key residues in canonical Hsp40 proteins (Figure 2.4). The highly critical HPD tripeptide was fully conserved (Figure 2.4) and a significant number of the J-domains analysed contained an HPDK motif, which has previously been identified in other studies (Genevaux *et al*., 2002). This indicates that these Hsp40s should be able to stimulate the ATPase activity of a partner Hsp70 protein.

Interesting differences between canonical J-domains and those of the Type III TbHsp40s were, however, observed in the consensus sequence derived for Helix III and Helix IV (Figure 2.4). The KFK motif, which has previously been implicated as a motif which plays an important structural role in the J-domain (Hennessy *et al*., 2000; Hennessy *et al*., 2005a; Genevaux *et al*., 2002) was shown to be highly variable in the J-domains studied (Figure 2.5). This variability was particularly apparent for the Lys^{46} and Lys^{48} residues (Figure 2.5). These alterations in a region implicated in the structural / conformational integrity of the J-domain could have an effect on the way the Type III Hsp40s of *T. brucei* behave with respect to partner Hsp70 proteins. Homology modelling performed on selected J-domains in the course of the present work did not reveal any significant distortions in the overall structure of the J-domain, but it is important to note that these models do not correspond to the optimal energy minima with respect to conformational space, nor do they present accurate information with regard to the behaviour of the J-domain as part of the greater Hsp40 protein that contains it, or in a complex solution as found *in vivo*.

The QKRAA motif has previously been implicated in J-domain functionality (Suh *et al*., 1999). This motif was found to be abrogated in all the J-domains studied here, with the exception of the Arg^{63} residue (Figure 2.5). The function of these altered residues is unknown at present. The poorly conserved residues in the J-domains studied are proposed to lend specificity to the Hsp40–Hsp70 partnership (Hennessy *et al*., 2000) and the high levels of variability of certain regions within the J-domains of the Type III Hsp40s of *T. brucei* could serve to indicate that highly specific Hsp40-Hsp70 partnerships exist in the parasite.

2.4.4. Identification of a novel ortholog of Tcj1 in *T. brucei*

An ortholog of Tcj1 (Tibbetts *et al*., 1998) has been identified in *T. brucei*. Tbj1 has been predicted to localise to the nucleus of the parasite and was shown to possess a putative NLS (Section 2.3.8). A comparison of the J-domain of this protein to those from canonical Hsp40s indicates that while several of the residues implicated in structural integrity are conserved, a number of key residues including Arg^{26} 4 Tyr^{26} and the KFK motif (VFK) are not. The HPD tripeptide was fully conserved, however, and both Tbj1 and Tcj1 contained an HPDKNP motif which is identical to that observed in the J-domain of Ydj1 from yeast (Figure 2.6). The latter observation would appear to suggest that Tbj1 should be able to stimulate the ATPase activity of a partner Hsp70 protein, providing that it is able to bind successfully due to the Arg^{26} 4 Tyr^{26} mutation (Figure 2.6). Previous studies on Tcj1 have indicated that it is unable to stimulate the ATPase activity of the cytosolic Hsp70 from *T. cruzi*, TcHsp70 (Edkins *et al*., 2004). The impact of minor sequence alterations on the specificity of the J-domain cannot be ignored in this case. The QKRAA motif was found to be completely abrogated in Tbj1, and was replaced with a TQREM motif. The homology models generated indicated that the structure of Helix I and II of the Tbj1 J-domain was not well-conserved and that the helices in this region are abrogated (Figure 2.6; B and C). If this is a true result (which can only be

confirmed by means of advanced *in silico* characterisation) this could have a profound effect on the behaviour of Tbj1 *in vitro* and *in vivo*.

2.4.5. Type III TbHsp40s containing TPR domains

T. brucei was found to possess 5 TPR-containing Hsp40 proteins that were predicted to occur throughout the parasite. These proteins have been proposed to preside over highly specialised processes within the parasite, potentially that of fusing the Hsp70 and Hsp90 interaction pathways. TPR domains are highly degenerate in terms of their sequence conservation but certain TPR domains have been shown to interact with the EEVD motif of Hsp70 and Hsp90 to facilitate an interaction between the two proteins (Odunuga *et al.*, 2003). It is very interesting to note that none of the Hsp70 proteins in the *T. brucei* Hsp70 complement possess a C-terminal EEVD motif, which could suggest that the TPR containing Hsp40s in *T. brucei* may interact preferentially with Hsp90, instead of Hsp70 depending on the nature of the individual TPR domains that they possess. The overall J-domain sequence and structure of these co-chaperones was found to be highly conserved, especially with respect to helices I and III. The loop regions were also found to contain an HPDK motif as was observed in the Type III TbHsp40s that do not contain TPR domains. The localisation for these proteins is diverse, and suggests that they could preside over parallel processes in neighbouring organelles.

2.4.6. Type IV Hsp40s within the TriTryps parasites

The existence of proteins containing corrupted HPD motifs have been confirmed previously, but these proteins have only recently been classified as Type IV Hsp40s (Botha *et al.*, 2007). This phenomenon has been observed in *P. falciparum* and yeast (Botha *et al.*, 2007; Kluck *et al.*, 2002; Walsh *et al.*, 2004). *T. brucei* possesses a number of J-like proteins (or proteins that have been annotated as DnaJ-like molecular chaperones, but possess very low

sequence identity over the J-domain). Some of these have been classified as putative Type IV TbHsp40s, which have very poorly conserved J-domains with respect to each other and to canonical J-proteins. Tbj31 has a partially corrupted J-domain (Figure 2.7) in which the HPD motif is substituted by an HTD motif, but in which all the helices show levels of conservation to canonical J-proteins that are comparable to those observed for the Type III Hsp40s. This change did not appear to make a significant difference in the overall secondary structure of the J-domain of this protein. Tbj31 was the only protein in *T. brucei* which corresponded to the classical definition by Botha and colleagues of a Type IV Hsp40, in that it was shown to possess a fully complete J-domain with the exception of a corrupted HPD motif (Botha *et al.*, 2007).

The loop region of Tbj47 comprises a KDP motif instead of the HPD motif which would abrogate binding to a partner Hsp70 protein unless the two proteins were to interact transiently through other residues. The overall lack of conservation of the "J-domain" of this protein and its homologues (Figure 2.7) suggests that it may not be able to interact with a partner Hsp70 protein in spite of some conserved residues in Helix I and Helix II.

The remainder of the putative Type IV Hsp40 complement of *T. brucei* and their orthologues in *T. cruzi* and *L. major* (J66 and J68) have "J-domains" that can only be regarded as highly speculative at present (data not shown). Both these proteins and their orthologues were found to display insignificant levels of conservation with respect to canonical Hsp40s (< 30 % identity) and fully corrupted loop regions (Tbj66: HPD 4 CDM; Tbj68: HPD 4 LHD). It is proposed that these proteins were either misannotated during the TriTryps genome sequencing project curation period or that these putative J-domains were recruited by non-Hsp40 proteins in order to fulfil a specific task. Once again, this hypothesis would have to be proven experimentally in future investigations. The function of the Type IV TbHsp40 protein complement is not known at present, and further biochemical characterisation is necessary to elucidate whether the

Type IV Hsp40s possess any chaperone properties of their own, whether they function by interacting with a partner Hsp70 or if they have evolved in a specific manner to interact with a specialised Hsp70 that has evolved in tandem. The possibility that the Type IV Hsp40 complement of *T. brucei* could be the result of non-Hsp40 proteins recruiting the J-domain during the evolutionary process without being true Hsp40s cannot be ignored at this point. The corrupted J-domains could also potentially be relics of ancestral J-domains that no longer play a vital role within the *T. brucei* parasite, although the presence of highly conserved orthologues in both *T. cruzi* and *L. major* strongly suggest that this is not the case. The presence of RNAi data for Tbj47, a Type IV Hsp40 in *T. brucei* confirms this (Subramaniam *et al*., 2006). Because abrogation of the conserved HPD motif abolishes the ATPase stimulation of an Hsp70 by a partner Hsp40, it has been proposed that true Type IV Hsp40s exert their function by a different mechanism to that of Type I, II and III Hsp40s, possibly one of negative regulation (Botha *et al*., 2007).

2.4.7 The Hsp70 complement in *T. brucei*

A total of 12 Hsp70 proteins were identified and studied in *T. brucei* (Table 2.9). The *L. major* and *T. cruzi* chaperone complements contain a similar number (data not shown). The number of TbHsp70s is equivalent to that reported for *H. sapiens* (13) and *P. falciparum* (6) although the ratio of TbHsp40s to TbHsp70s is far higher than the situation in *H. sapiens* and *P. falciparum* (Hageman and Kampinga, 2009; Shonhai, 2007). The *T. brucei* parasite possesses Hsp70 proteins from all the conserved subgroups in the Hsp70 protein family including mitochondrial, E.R., cytoplasmic and secreted Hsp70s (Table 2.9). Interestingly, unlike the TbHsp40 complement, most of the TbHsp70s have acidic pI values. Most of the TbHsp70s were found to occur in both the parasite bloodstream form (found in the human host) and the procyclic form (found in the insect vector), while this is not always the case for the TbHsp40s. This could suggest that the TbHsp70s could potentially have certain

partner Hsp40s during the parasite's lifecycle stages in the human host, and different partner Hsp40s during the parasite's lifecycle stages in the insect vector, which would be dependent on which TbHsp40s have suitable binding sites for interaction with TbHsp70s, are localised to the same organelles and are expressed at the same time.

The major cytosolic Hsp70 in the TriTryps was found to be well-conserved with respect to well-characterised Hsp70 proteins such as *M. sativa* Hsp70 (70 % identity) and *S. cerivisiae* Ssa1p (71 % average identity), but not with respect to *E. coli* DnaK (48 % average identity) (Table 2.10). Although a wider study would have to be performed in this regard, it is proposed that minor changes in the TbHsp70 and TcHsp70 sequence and structure could alter function and potential protein partners, preventing them from being interchangeable with other Hsp70s *in vivo*, and possibly, *in vitro*.

The ATPase domains of TbHsp70 and TcHsp70 were found to possess a 95 % identity, and it can reasonably be suggested that these proteins could substitute for one another *in vitro*. Interestingly, despite the low levels of sequence identity of Tb / TcHsp70 with *E. coli* DnaK, a complete conservation of residues identified as being important for Hsp70-Hsp40 interaction was observed in these proteins, indicating that their *in vivo* behaviour may be similar to that of canonical Hsp70 proteins and that they have the ability to interact with well conserved TbHsp40 proteins in the normal mode of chaperone action. A study of the peptide binding domains of TbHsp70 and TcHsp70 revealed that this region of both proteins is poorly conserved with respect to that of Ssa1p, DnaK, PfHsp70 and human Hsp70. This observation has previously been made for Hsp70 proteins in *P. falciparum* and suggests that the key to Hsp70 specificity may lie in this relatively poorly conserved region (Shonhai, 2007).

The most interesting feature of the TbHsp70 complement in general is the complete absence of the highly conserved EEVD motif in all TbHsp70 proteins

studied, in spite of the presence of this motif in all their orthologues in *T. cruzi* and *L. major* (Figure 2.12). This finding has interesting implications for the interaction between TbHsp70s and TPR-containing TbHsp40s as these proteins require the conserved EEVD motif, found in almost all other known Hsp70s, in order to interact (Odunuga *et al*., 2003). Further studies on the TbHsp70s should examine the nature of the protein sequences in detail, and, coupled with a detailed rational mutagenesis strategy and robust *in vitro* and *in vivo* assays, could determine whether any amino acids that are different to those in canonical Hsp70s could result in the binding to / interaction with novel TbHsp40s containing corrupted J-domains.

2.4.8. Potential Hsp40-Hsp70 partnerships in the *T. brucei* parasite

The putatitive localisation data obtained for the Type III and Type IV Hsp40 complement in *T. brucei* was utilised in conjunction with the localisation data for the TbHsp70s to develop a diagram of potential partnerships within the parasite (Figure 2.14). This schematic will provide the basis for future biochemical characterisation of the TbHsp40 and TbHsp70 complement as it identifies potential partnerships that can be experimentally verified at a later stage. It is important to note that some of the more novel and poorly conserved TbHsp40s could potentially interact with only one TbHsp70 partner (if any at all) and that this potential interaction should be regarded as the most preliminary of studies, especially in light of the uncertainly regarding localisation signals in *T. brucei*. It is also important to note that the present work focussed solely on the Type III and Type IV Hsp40 complement, and that the role played by the Type I and Type II Hsp40 complement (investigated in Ludewig, 2009) cannot be ignored when designing such experiments.

The greater majority of TbHsp70s were predicted to localise to the mitochondrion and the cytoplasm (Figure 2.12) which is to be expected based on other chaperone studies. The *T. brucei* plasma membrane potentially contains 10

Type III TbHsp40s but no Hsp70s, suggesting that these proteins may potentially play a role in the regulation of the VSG surface coat of the parasite. This in addition to the secreted Hsp70s (2) and Hsp40s (6) could indicate that *T. brucei* uses an arsenal of molecular chaperones to protect itself from the human host / insect vector that it resides in, and that some of these molecular chaperones may, in fact interact with the human host, as suggested in the case of *P.falciparum* by Botha et al (2007) or that these chaperones may possess independent and as yet uncharacterised functions of their own that are independent of canonical co-chaperone partners.

The *T. brucei* nucleus was found to potentially contain one Hsp70 protein (Tb927.7.1030), 10 Type III Hsp40s and one Type IV Hsp40 (Figure 2.14), while the E.R. potentially contains two Hsp70s (Tb11.02.5500 and Tb11.02.5450) and three Type III Hsp40s. In terms of potential partnerships, the nuclear Hsp70s would have to be able to interact with a number of Hsp40 proteins each, whereas the interaction within the E.R. could be one of greater specificity.

2.4.8. Conclusions

The TriTryps possess a diverse and complex array of Type III and Type IV Hsp40 proteins in greater numbers than that of other species including yeast, humans and the malaria parasite. A complete survey of the Type III, Type IV and TPR-domain containing TbHsp40 complement is given in the present work, with homology models of selected domains of interest. This work serves as a foundation for further biochemical and molecular characterisation of the Hsp40 and Hsp70 complement of the *T. brucei* parasite and enhances the current knowledgebase of trypanosomal molecular chaperone studies.

CHAPTER 3

Cloning and expression analysis of Tbj:L, Tbj47 and Tbj5:L

3.1. INTRODUCTION

3.1.1. Heterologous protein expression

Purification of proteins from their endogenous sources is often not viable due to lack of availability of material, low levels of expression as well as the lack of robust purification protocols for organisms not regularly studied. The *E. coli* bacterium has been a favoured heterologous expression system for a number of years, largely because of its rapid growth and ease of culturing (Hockney, 1994; Weickert *et al*., 1996). The use of *E. coli* as an expression system has not been without challenges, and until recently, the heterologous expression of complex proteins containing more than three disulphide bonds was perceived difficult (Hockney, 1994; Weickert *et al*., 1996). Due to shortcomings such as codon bias and inclusion body formation, the *E. coli* expression system was less extensively used during the early 1980‟s but regained favour in the 1990‟s when optimised strategies resulting in greater protein yield started to appear in the literature (Hockney, 1994). Some of the factors that have once again made large-scale overexpression in *E. coli* possible were:

a) the discovery of the power of molecular chaperones to refold complex and denatured proteins by co-expressing them with a protein of interest
b) the ability of a lowered incubation temperature of a growing *E. coli* culture to increase the amount of heterologous protein being expressed
c) the development of fusion proteins to enhance solubility (Hockney, 1994).

The overexpression and purification of molecular chaperones has become a routine procedure in many laboratories and a wide variety of purification protocols are available (Nicoll *et al*., 2006). In spite of this, the over-expression of novel heat shock proteins can often cause complications in terms of yield in a heterologous expression system, as well as the ease with which they can be purified (Nicoll *et al*., 2006). The use of *E. coli* as a heterologous expression

system is widespread in chaperone research, and often has very satisfactory results (Edkins *et al*, 2004; Nicoll *et al*., 2006; Nicoll *et al*., 2007).

Hsp40 proteins have consistently been shown to be expressed at low levels in organisms (Al-Herran *et al*., 1998), while Hsp70 proteins are far more abundant (Nicoll *et al*., 2006). In the case of *E. coli*, DnaK is produced in 10-fold higher quantities than DnaJ, and even the nucleotide exchange factor GrpE is produced in higher quantities than DnaJ (Bardwell *et al*., 1986; Pierpaoli *et al*., 1998). Over expression of full-length DnaJ in *E. coli* itself has been observed to lead to a dramatic decrease in cell viability, which suggests that excess levels of DnaJ may be toxic to *E. coli* cells, possibly by disrupting the balance of endogenous chaperone pathways in the bacterium (Al-Herran *et al*., 1998). In addition, the localisation of a protein in its endogenous organism is also crucial to the success of heterologous expression, as membrane-bound proteins have consistently been found to be more difficult to purify than those that are localised to organelles such as the cytoplasm, nucleus and endoplasmic reticulum (E.R.) (Al-Herran *et al*., 1998). A further complicating factor in the heterologous expression and purification of Hsp40s in *E. coli*, is the co-purification of bound endogenous DnaK, which is generally remedied by addition of ATP to the purification wash buffers to disrupt the interaction between the Hsp40 of interest and any contaminating Hsp70 proteins (Nicoll *et al*., 2006).

To successfully biochemically characterise proteins, and to study the functionality of their interaction with potential partner proteins, a large amount of highly pure protein (mg quantities) from an endogenous source or a heterologous system is required to perform functional assays (Nicoll *et al*., 2006). The advent of whole genome sequencing coupled with expressed sequence tags (ESTs) and cDNA libraries has greatly enhanced the ease of heterologous protein expression by facilitating the design of highly accurate primers for cloning, peptide directed antibodies and the correct identification of protein size and isoelectric point *in silico* prior to any biochemical investigation. Despite a rapidly growing amount of

bioinformatic data on the Hsps and the recent completion of the TriTryps genome sequencing project, very little expression data are available for trypanosomal proteins. The status of expression studies and the amount of biochemical characterisation done on a number of molecular chaperones of interest from *T. brucei* and *T. cruzi* are outlined below.

3.1.2. Hsp40 and Hsp70 protein expression in the TriTryps: an overview

Most, if not all, cloning and heterologous protein expression studies have been performed in *T. cruzi* (Tibbetts *et al*., 1998; Salmon *et al*., 2001; Carreira *et al*., 1998). The current research focus in *T. brucei* is an *in vivo* approach, largely due to the amenability of the parasite to RNAi and proteomics studies, but this is not supported by an in-depth biochemical characterisation of proteins of interest (Balaña-Fouce and Reguera, 2007; Jones *et al*., 2006; Vertommen *et al*., 2008). *In vitro* and *in vivo* studies have only been performed on Hsp40s from *T. cruzi* to date; no Hsp40-encoding genes from *T. brucei* or *L. major* have been successfully cloned, and no TbHsp40 has been expressed or biochemically characterised (Tibbetts *et al*., 1998; Folguiera and Requena, 2007).

3.1.3. Tcj1, a well characterised Type III Hsp40 from *T. cruzi*

Tcj1 is a Type III Hsp40 from *T. cruzi.* The gene encoding Tcj1 has been successfully cloned by Tibbetts and his group in 1998 (Tibbetts *et al*., 1998), and has been shown to exhibit no up-regulation to heat shock when analysed with Northern blot analysis. Interestingly, the mRNA sequence of Tcj1 was found to be significantly longer than the cDNA sequence due to the presence of an unusually long poly-A tail, but the remainder of the sequence was found to be identical (Tibbetts *et al*., 1998). This suggests that amplification from genomic DNA could also have been utilised to clone the Tcj1 coding region. Tcj1 was successfully overexpressed in *E. coli* by Edkins and colleagues, and biochemically characterised with respect to its ATPase activity and ability to

stimulate the ATPase activity of the cytosolic *T. cruzi* Hsp70 protein, TcHsp70 (Edkins *et al.*, 2004). In this study it was shown that Tcj1 did not possess ATPase activity of its own accord and was not able to stimulate the ATPase activity of TcHsp70, either in the presence or absence of a model substrate (Edkins *et al.*, 2004).

3.1.4. Tbj1, a potential orthologue of Tcj1 in *T. brucei*

Tbj1 (Tb11.01.8750) is a 35.2 kDa Type III Hsp40 that is a putative ortholog of Tcj1 (64.3 % identity over the full-length sequence), as well as Lmj1 in *L. major* (www.genedb.org/). The gene encoding Tbj1 is intronless, and Tbj1 is translated from a 927 b.p. sequence according to information given on the GeneDB database. This protein is predicted to localise to the *T. brucei* parasite nucleus (Chapter 2; Table 2.4). The annotation page for Tbj1 on the GeneDB database states the function of this protein has been inferred from homology, and that it has not been cloned, expressed or characterised in any way prior to the present work. Tbj1 is predicted to possess no signal peptide or transmembrane domains (www.genedb.org) and it is currently unclear whether or not it is expressed in the *T. brucei* bloodstream form (BSF) proteome or in the procyclic (Pro) proteome.

```
MGSDVFELIG NTALYEVLGV PRTATDAEIR RAYYKLAVVY HPDKNPEGVE VFKEVSFAHS
ILSDPTQREM YDNQRLRTHI EGQARKYDPM MDPNVELSAE ELRLFVERKR KEDEEKMRNR
SEFEKQREEE MRRRAEYDAQ NPDFKAEYER MRARAKEEGS QRASAASAMR HLTTAELMQR
LEMKQQEATN SGIGRVRSGD PGKSGANPSS GLSSIKRSML NDFRTRHDSA PPTAESMQLR
AQPPAQSSRL DFVGKQNEKS YTCEMEKLIG KYSNFNYRDF VEKGIVDGDG VMEAAILADA
LGNYDRSR
```

Figure 3.1. Annotated protein sequence of Tbj1 from *T. brucei*

The full protein sequence of the Type III TbHsp40, Tbj1 is given above. The well-conserved J-domain is highlighted in yellow and the HPD motif is underlined.

Tbj1 is classed as a Type III Hsp40 protein because it lacks all the domains occurring in the well-conserved Type I class of Hsp40s as defined by Cheetham and Caplan (1998) with the exception of the J-domain. Tbj1 contains a conserved J-domain (Figure 3.1) which displays significant similarity to

J-domains from canonical Hsp40 proteins as outlined in Figure 2.6 in Chapter 2 of the present work. No RNAi studies have been performed on Tbj1, and at present its function in the parasite is unknown. This, in addition to the lack of biochemical characterisation data available for this protein, makes it an ideal protein for further study.

3.1.5. A novel member of the *T. brucei* Type IV Hsp40 class - Tbj47

Tbj47 (Tb927.1.1230) is a putative Type IV Hsp40 of *T. brucei* that has a molecular mass of 52.3 kDa and is predicted to localise to the *T. brucei* mitochondrion (Chapter 2; Table 2.7). It is predicted to be a molecular chaperone based on homology studies *in silico*, but a survey of the literature reveals that no biochemical verification of its chaperone properties has been performed to date. Tbj47 possesses a highly corrupted J-domain (Figure 3.2) that has very low levels of identity with respect to the well-conserved J-domains of canonical Hsp40s, as outlined in Chapter 2.

```
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
```

Figure 3.2. Annotated protein sequence of Tbj47 from *T. brucei*

The full protein sequence of the putative Type IV TbHsp40, Tbj47 is given above. The putative corrupted J-domain is highlighted in yellow and the predicted motif in place of the well-conserved HPD motif is underlined. The predicted zinc-finger region located towards the central region of the protein is highlighted in green.

Tbj47 is of great interest as a potential novel drug target for therapeutics due to its low levels of conservation with respect to other canonical Hsp40s and the fact that it has no predicted orthologues in humans, which was confirmed by an NCBI

BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Interestingly, despite the poor levels of J-domain conservation in this protein, Tbj47 possesses a fairly well-conserved sequence with respect to other molecular chaperones with a C-terminal zinc finger. This is illustrated when the full protein sequence of this protein is subjected to an NCBI BLAST as alluded to above. Another reason for the attractiveness of Tbj47 as a novel drug target is the result of RNAi studies that have shown that knockdown of this protein results in decreased growth and growth defects of *T. brucei* during the bloodstream stage of its lifecycle (Subramaniam *et al*., 2006). A precursor to any drug design studies would require a full biochemical characterisation of Tbj47 and the development of suitable and highly specific antibodies to the protein in addition to highly accurate homology modelling in order to develop a greater understanding of this unique molecular chaperone. The existence of orthologs of Tbj47 in other Trypanosomal parasites opens up new avenues of multi-organism treatments with minor modifications.

3.1.6. A previously uncharacterised TPR-domain containing Type III TbHsp40, Tbj51

As outlined in Chapter 2 of the present work, the *T. brucei* parasite possesses 5 Type III Hsp40 proteins that contain TPR domains in addition to highly conserved J-domains (Chapter 2; Figure 2.6). Tbj51 (Tb927.4.2220) is a Type III TbHsp40 that contains 5 TPR domains spanning two TPR regions (Figure 3.3) and is predicted to have a molecular mass of 80.4 kDa and is predicted to localise to the *T. brucei* cytoplasm (Chapter 2; Table 2.4).

Unlike Tbj47, Tbj51 is predicted to have an ortholog in humans, with a 51 % identity to SP:Q99615: DnaJ homolog subfamily C member 7 (Tetratricopeptide repeat protein 2)(TPR repeat protein 2) according to the annotation given in the GeneDB database (www.genedb.org). As is the case for Tbj1, Tbj51 is predicted to have no signal peptide and no transmembrane domains, which would be anticipated for a cytoplasmic protein. A search of the current available literature

reveals that Tbj51 has not been cloned, expressed or biochemically characterised to date, and that no RNAi studies have been performed as yet.

```
METTCKEDVP QVRFPSPTSS KSMTTSPQKL SILGDITAEP LSARRSFNDV SKVREPDGRW
CRKFLSTYFR CELCQNVVTD PVQILPNVLL VCRRCALSRR VPSKDLQELP VSLTRAFEDL
YEGQREVQLP SASRMSEIGR ASTTSDEAIC KRRMVRVKRP DAVQEGSDHS QTFPLAASIN
GTLNASGRSS VPLLLNKALC ELENKEYCMR VEIESAEASG ALELKKSYNS DMRKITARSC
GTSKTLKTDA DQKYEQAEYT LALELYTKAI ELQPRDRLTR LTALYGNRSS AYFMAMRYAE
CIADCMKVVE LDPNNVKLFA RAAKAAAIMG DLTAAVSHME SIPEERVTPN IISEREKYKN
GLDTYKRAES SFGKSDSDDA WQMLVAQFSD TIFFRIRYAE SLQNQKRFLK AVEVLDVVPQ
ERRTPKLLYI MAACLFMCGF EHFDKARTCL EDVQQLDENC AQLLKVLNIV DEGKQKGNQY
FQQKKFVAAM EHYTTAIGAA VNNNQILRIL YCNRAASYKE VGKYREAIED CTRTIQLDPA
FSKAYARRAR CHQALSDFAS AIRDFKAAIK YDPNDQELPR ELRSCEQSMA KEGERERDYY
YVLGVSRNAT EREIKARYRE LSLRWHPDKC MSLPEEERVV AERKFKIIVE AHTTLIDAVK
RRDYDLKMEK ERLTRSGGFG GFNGYSSETF RGHSNRFRQG SSGFW
```

Figure 3.3. Annotated protein sequence of Tbj51 from *T. brucei*

The J-domain is highlighted in yellow with the HPD motif underlined. The TPR domains are highlighted in grey. All annotations were performed with the assistance of the ProSite database (http://ca.expasy.org/prosite/)

3.1.7 Genomic data on the TriTryps that could facilitate successful expression in heterologous systems

Like all organisms, the TriTryps exhibit a certain level of codon bias. The *L. major* and *T. cruzi* genomes are more GC-rich on average (59.7% and 51%) than that of *T. brucei* (46.4%) (Chanda *et al.*, 2007). A recent review on the subject of the codon bias displayed by the TriTryps revealed that most of the highly expressing genes from *T. brucei*, *T. cruzi* and *L. major* display a preference for the use of a similar subset of codons that mostly have a G / C – ending (Chanda *et al.*, 2007). In a similar fashion, genes that display low levels of expression in *T. brucei* and *T. cruzi* have been found to prefer codons that end in an A / T (Chanda *et al.*, 2007). An examination of the cDNA sequences of Tbj1, Tbj47 and Tbj51 revealed that most of the codons in Tbj1 end on a G / C residue, indicating that the Tbj1 could result in high levels of protein expression upon translation. Tbj47 and Tbj51 are predicted to offer low expression levels

due to the high number of codons ending with an A / T (Chanda *et al*., 2007). A detailed study of the codon bias phenomenon in the TriTryps by Chanda and colleagues has shown that *L. major* displays the highest extent of bias in terms of synonymous codon usage, while *T. brucei* displays the lowest (Chanda *et al*., 2007). It has thus been proposed that the *L. major* experiences translational selection more strongly than *T. brucei* or *T. cruzi*. The post-transcriptional control of proteins plays a crucial role in the TriTryps; recent studies have revealed that transcriptional control in *T. brucei*, *T. cruzi* and *L. major* is extremely limited when compared to that of other organisms studied in the same amount of detail (Clayton, 2002; Horn, 2008).

3.1.8. Methods of antibody development and production – an overview

Antibodies are glycoproteins that occur in the plasma and intracellular fluid of organisms that are suffering from a specific infection (Lipman *et al*., 2005). Antibodies serve as the initial response factor to infection of living organisms by foreign agents and comprise one of the most important effectors of an immune system that is adaptable to various environmental and pathological challenges (Lipman *et al*., 2005). The ability of antibodies to bind to a given antigen with very high levels of sensitivity and specificity make them invaluable tools in research across a wide spectrum of disciplines (Lipman *et al*., 2005). Antibodies are required for analytical techniques (ELISA), immunohistochemistry and Western blotting (Bellstedt *et al*., 1987). Antibodies can be produced in two different ways; viz. by designing and chemically synthesising a peptide sequence and producing the antibody *in vivo* by inoculation of suitable animal, eg. rabbit, or by purifying the target protein of interest to sufficient levels of purity and inoculating a suitable animal directly with the purified protein (Bellstedt *et al*., 1987). The former method is useful in cases where the protein of interest has not yet successfully been expressed or is a very low expresser, in which case a high enough yield for the second route of antibody production would not be

feasible. Peptide-generated antibodies, if designed with sufficient care should also be more highly specific than antibodies resulting from the direct inoculation route as any contaminants that co-purify alongside the protein of interest will also be minor antigens in the preparation. In cases where a protein has been purified successfully and to a sufficient concentration (>0.35 mg / ml), the use of purified target protein as an antigen is an attractive option (Bellstedt *et al*., 1987). Bellstedt and colleagues developed a method of antibody generation relying on acid-treated („naked") *Salmonella minesota* bacteria as an adsorbant for the target protein of interest which is subsequently inoculated into rabbits for production of high quality antibodies of significant specificity (Bellstedt *et al*., 1987). The development of suitable antibodies to a protein of interest is one of the most important initial stages in any biochemical or immunohistological characterisation and the development of two antibodies to two TbHsp40s (Tbj1 and Tbj47) is outlined in the present work.

3.1.9. AIMS AND OBJECTIVES

The major aims of the present work were to:

- isolate the coding regions for three previously uncharacterised TbHsp40s (Tbj1, Tbj47 and Tbj51)
- successfully express Tbj1, Tbj47 and Tbj51 in a heterologous *E. coli* system
- optimize a purification protocol for Tbj1 and Tcj1
- heterologously express and purify an Hsp70 protein from *T. cruzi* (TcHsp70)
- produce polyclonal antibodies to Tbj1 and Tbj47
- determine whether Tbj1 and Tbj47 are found in both the Pro and BSF lifecycle stages of the *T. brucei* parasite, or whether their expression is stage-specific

3.2. EXPERIMENTAL PROCEDURES

3.2.1. Standard molecular biology procedures

3.2.1.1 PCR-based amplification of coding regions of interest

Primers (Appendix C) were designed to PCR- amplify the coding regions of Tbj1, Tbj47 and Tbj51 with the addition of appropriate restriction sites for cloning into the TAGZyme pQE-1 vector (Invitrogen; U.S.A.) that allows for cleavage of the His_6-tag. The primer design was based on the sequence data for the intronless coding regions of the proteins of interest lodged at *T. brucei* GeneDB (http://www.genedb.org/genedb/tryp/), and restriction sites that cut the vector once without cutting internally in the coding region of interest, were selected. The Tbj1 coding region was PCR amplified with primers containing *Kpn* I and *Pml* I restriction sites (Appendix C). The Tbj47 coding region was amplified with primers containing *Pvu* II and *Hind* III restriction sites (Appendix C) and the Tbj51 coding region was amplified with primers containing *Kpn* I and *Hind* III restriction sites (Appendix C). All primers were synthesised by Whitehead Scientific (IDT; WhiteSci, South Africa). *T. brucei (TREU927)* genomic DNA was kindly donated by Dr. David Horn (London School of Tropical Medicine and Hygiene, U.K.) and served as a template for all PCR reactions unless otherwise specified. The Tbj1 and Tbj51 coding regions were amplified with the Expand High Fidelity PCR kit (Roche, Switzerland) while the Tbj47 coding region was amplified using the Kapa Taq PCR kit (Kapa Biosystems, U.S.A.). All PCR reactions contained primers to a final concentration of 1 µM and 100-150 ng of *T. brucei* genomic DNA as a template. Standard PCR methods were used (Appendix A6).

3.2.1.2. Ligation of *T. brucei* Hsp40 coding regions into pGEM-T® Easy vectors

TbHsp40 coding regions were ligated into the pGEM-T® Easy cloning vector (Promega, U.S.A) subsequent to successful PCR amplification. The standard protocol as outlined by the kit manufacturer was adhered to. An aliquot of the

PCR product (100 – 200 ng; 2 µl) was added to the pGEM-T® Easy vector (50 ng; 1 µl) and 2 x rapid ligation buffer (60 mM Tris-HCl, pH 7.8; 20 mM $MgCl_2$; 20 mM DTT, 2 mM ATP, 10 % polyethylene glycol) prior to the addition of T4 DNA ligase (0.3 U). The ligation reactions were incubated at 4°C for 5 hours prior to transformation into *E. coli* JM109 competent cells. An aliquot of plasmid DNA (1.5 µl) was added to 100 µl of *E. coli* JM109 competent cells that were previously thawed on ice and incubated on ice for 30 minutes. The cells were heat-shocked at 42°C for 45 seconds and immediately placed on ice for 5 minutes. A volume of ice cold 2 x YT broth (900 µl; 1.6 % tryptone, 1.0 % yeast, 0.5 % NaCl) was added to the cells prior to incubation at 37°C for one hour. The cells were pelleted by centrifugation at 12 000g for 2 minutes and re-suspended in 200 µl of the supernatant broth prior to plating on to 2 x YT plates containing appropriate antibiotics and incubating at 37 °C overnight.

An average of 5 colonies were selected from the overnight transformants and screened by means of diagnostic digests using the Alkaline lysis method (Ausubel *et al*., 1991; Sambrook *et al*., 1989) as described in Appendix A.3.B. The presence of an insert in the isolated pGEM-T® Easy constructs was confirmed by means of a restriction digest with *Eco* RI or *Not* I which release inserted DNA fragments in the pGEM-T® Easy vector.

3.2.1.3. Preparation of competent cells and transformation of ligated DNA

Competent cells were prepared from selected *E. coli* strains as described in Appendix A1. A standard transformation reaction was performed by adding an aliquot of plasmid DNA (50 – 100 ng; 1.5 µl) to 100 µl of competent cells that were previously thawed on ice. The plasmid DNA and competent cells were incubated on ice for 30 minutes. The cells were heat-shocked at 42°C for 45 seconds and immediately placed on ice for 5 minutes. A volume of ice cold sterile 2 x YT broth (900 µl) was added to the cells prior to incubation at 37°C for one hour. The cells were pelleted by centrifugation at 12 000 *g* for 2 minutes and

re-suspended in 200 µl of the supernatant broth prior to plating on 2 x YT plates containing appropriate antibiotics.

3.2.1.4. Small-scale preparation of plasmid DNA for confirmation of construct identity

The method outlined below was used in cases where the complete purity of the DNA sample was not required, as in the case of plasmid screening experiments. Plasmid DNA was isolated using the Alkaline lysis method (Birnboim and Doly, 1979; Ausubel *et al*., 1991; Sambrook *et al*., 1989). An overnight culture was prepared from a single transformant as outlined above and pelleted by centrifugation at 12 000 *g* for 2 minutes. The pellet was resuspended in 100 µl Solution 1 (50 mM glucose, 25 mM Tris-HCl, 10 mM EDTA; pH 8.0) prior to the addition of 5 µl RNAse A (10 mg / ml) and vortexing until the pellet was fully resuspended. The mixture was incubated at room temperature for 5 minutes. Solution II (150 µl; 0.2 M NaOH, 1 % w/v SDS; prepared fresh) was added prior to incubating on ice for 2 minutes. After the incubation period, 250 µl ice cold Solution III (3 M potassium acetate; pH 5.0) was added prior to another incubation period on ice for 5 minutes. The sample was centrifuged at 12 000 *g* for 10 minutes and the supernatant was transferred to a sterile eppendorf tube to which 450 µl isopropanol had been added. The supernatant-isopropanol mixture was incubated at room temperature for 2 minutes and centrifuged at 4°C for 30 minutes to yield a white pellet. The supernatant obtained was discarded and the pellet resuspended in 400 µl TE buffer (10 mM Tris-HCl, 1 mM EDTA; pH 8.0), 40 µl sodium acetate (3 M sodium acetate) and 800 µl ice cold ethanol (100 %) before being incubated at – 20°C for 30 – 60 minutes. The plasmid DNA was collected after a further 30 minute centrifugation step at 12 000 *g* (4°C). The pellet obtained was washed with 200 µl ethanol (70 %), pelleted by centrifugation and resuspended in 25 µl TE buffer (10 mM Tris-HCl, 1 mM EDTA; pH 8.0).

3.2.1.5. Small-scale preparation of plasmid DNA for confirmation of sequence identity by means of DNA sequencing

High quality plasmid DNA for DNA sequencing was isolated using the Qiaprep kit (Qiagen, U.S.A) which is based on a modified alkaline lysis method of plasmid isolation (Birnboim and Doly, 1979). A plasmid of interest was transformed into *E.coli* JM109 cells and a single colony was selected and used to inoculate a 5 ml 2 x YT broth treated with 0.1 mg / ml ampicillin. The inoculated culture was grown overnight with shaking at 37°C. After 16 hours growth, the entire bacterial culture was pelleted by centrifugation at 12 000 *g* for 2 minutes. The pellet obtained was re-suspended in buffer P1 (250 µl), prior to the addition of buffer P2 (250 µl) and buffer N3 (350 µl). The sample was centrifuged at 12 000 *g* for 10 minutes before the supernatant was applied to a Qiaprep spin column in a 2 ml collection tube and centrifuged at 12 000 *g* for 60 seconds. The column was washed with 500 µl buffer PB prior to washing with 750 µl buffer PE. After each washing step, the column was centrifuged at 12 000 *g* and the washing solution was discarded. Upon completion of the washing steps, the column was centrifuged for an additional 60 seconds to remove any residual ethanol prior to elution of the bound plasmid DNA after a 1 minute incubation period in buffer EB and centrifugation at 12 000 *g* for 60 seconds.

3.2.1.6. Confirmation of successful insertion of Tbj1, Tbj47 and Tbj51 coding regions into the pGEM T-Easy and pQE-1 plasmid vectors by means of diagnostic restriction digests

Plasmid DNA (8 µl containing 100-300 ng) was added to 10x NEB restriction buffer appropriate to the given restriction enzyme (4 µl), BSA (2 µl) and sterile deionised water (24 µl). The digestion reaction was initiated by the addition of 0.5 U of restriction enzyme and allowed to incubate at 37°C for 1 – 3 hours. The reaction was terminated by the addition of DNA loading buffer (0.5 µl; 0.25 % w / v bromophenol blue, 50 % glycerol v / v).

3.2.1.7. Agarose gel electrophoresis of isolated and restricted plasmid DNA

Agarose [0.8 % (w / v)] was suspended in 0.5 x TBE buffer (45 mM Tris base, 45 mM boric acid, 1 mM EDTA, pH 8.0) and heated in a microwave oven for approximately 90 seconds until fully dissolved. The molten agarose was cooled to 55 - 60°C prior to the addition of ethidium bromide to a final concentration of 0.5 µg / ml. After cooling further, the agarose gel was poured into a casting mould with a spacer comb and left to set for 20 minutes. The gels were resolved in an electrophoresis tank filled with 0.5 x TBE buffer. The samples were resolved at 110 V. For smaller DNA fragments, a 2 % agarose gel was used. Upon completion of the electrophoresis, agarose gels were visualized using the ChemiDoc XRS imaging system (Bio-Rad, U.S.A.).

3.2.1.8. Purification of DNA bands from agarose gels

DNA fragments of interest were gel purified using the DNA gel extraction kit (Fermentas, U.S.A.). Restriction digests of pGEM T- Easy constructs containing the insert of interest as well as pQE-1 plasmids (Qiagen, U.S.A.) were resolved on 1 % agarose gels prepared in 0.5 x TBE buffer containing no ethidium bromide. After the electrophoresis, a solution of 0.1 mg / ml ethidium bromide was prepared and the gels were soaked in this for approximately 20 minutes. The stained gels were subjected to long-wave U.V. for as short a period as possible and the relevant bands were excised and added to 900 µl DNA binding solution (6 M sodium iodide) and 500 µl TBE conversion buffer prior to heating at 55°C for 5-10 mins until the gel slices were completely melted. Silica powder suspension (5 µl) was added and vortexed. The mixture was subsequently incubated at 55 °C for 5 minutes and vortexed every 2 minutes to maintain an even suspension of silica beads. After the 5 minute incubation period, the samples were centrifuged at 12 500 *g* for 30 seconds in order to sediment the silica powder suspension bound to the free DNA in solution. The pellet was washed three times with ice cold wash buffer (concentrated wash buffer

containing Tris-HCl, NaCl and EDTA in 95 % ethanol) and resedimented each time by centrifuging at 12 500 *g* for 20 seconds. After the final wash step, the pellet was air dried for 15 minutes to remove any residual ethanol and resuspended in 25 µl TE buffer (10 mM Tris-HCl, 1 mM EDTA; pH 8.0).

3.2.1.9. Ligation of DNA fragments into the pQE-1 plasmid vector (Qiagen, U.S.A.)

Gel purified plasmid DNA as well as insert DNA in appropriate concentrations were added to 10 x ligation buffer (660 mM Tris-HCl, 50 mM $MgCl_2$, 50 mM dithiothreitol (DTT), 10 mM ATP, pH 7.5), 1 U of T4 DNA ligase and sterile deionised water to make the reaction volume 20 µl. Ligation reactions were allowed to proceed for 1 hour at room temperature or overnight at 4°C prior to transformation and screening of plasmid DNA.

3.2.1.10. DNA sequencing

Plasmids of interest were isolated using the Qiagen miniprep kit (Qiagen, U.S.A.) as outlined above and in A.3.A. Plasmid DNA was eluted in sterile deionised water instead of in TE buffer. A cycle sequencing reaction was prepared using the BigDye V3.1 Terminator cycle sequencing kit (Applied Bioystems, USA) and containing DNA template (500 ng, typically 2 – 5 µl), primer (3.2 pmol), Big Dye Buffer (2 µl), Big Dye mix (4 µl) and sterile deionised water to make up to 20 µl. The following thermal cycling protocol was adhered to over 25 cycles: 96°C for 10 seconds, 50°C for 5 seconds and 60°C for 4 minutes. The extension products obtained were purified from unincorporated dye terminators using the DNA Clean and ConcentratorTM-5 kit (Zymo; U.S.A.) according to the kit manufacturers instructions (Section 3.2.1.11). After clean-up, the samples were eluted in 8 µl sterile deionised water and air dried overnight at room temperature. DNA sequencing was performed on a 3100 ABI Genetic analyser sequencing machine by a qualified technician at the Rhodes University DNA sequencing facility

(Rhodes University, Grahamstown, South Africa). Sequence information obtained was analysed using the AlignX component of the Vector NTI molecular analysis suite (Informax; U.S.A.). DNA sequence identity was also confirmed by submitting isolated clones for sequencing to Inqaba DNA sequencing services (Inqaba, South Africa).

3.2.1.11. Clean up of sequencing reactions prior to sequencing

Unincorporated dye terminators and primers were removed upon completion of cycle sequencing by using the DNA Clean and ConcentratorTM-5 kit (Zymo; U.S.A.). Two volumes (150 µl) of DNA binding buffer was added to every volume of PCR reaction (50 µl). The resulting mixture was vortexed briefly and applied to a Zymo-spinTM column placed in a collection tube. The column was centrifuged for 30 seconds at 10 000 *g* and the flow-through discarded prior to the addition of wash buffer (200 µl) to the column. The wash buffer was eluted from the column by centrifuging at 10 000 *g* for 30 seconds prior to repeating the wash step. The flow-through was discarded after each wash step. The column was centrifuged at 10 000 *g* for 30 seconds in order to remove the residual ethanol. Sterile distilled water (8 µl) was applied to the top of the column matrix and the column was centrifuged at 10 000 g for 15 seconds to elute the clean DNA.

3.2.2. Purification of Histidine-tagged TbHsp40 and TcHsp70 proteins

3.2.2.1. Determination of optimal expression levels for Tbj1, Tbj47, Tbj51, Tcj1 and TcHsp70

Induction studies were performed on all proteins prior to purification in order to determine whether adequate protein expression levels were obtained, and were performed as described in Appendix A.12. The optimal conditions selected from

these expression trials were utilised to obtain a maximal level of expression in order to purify the TbHsp40s of interest.

3.2.2.2. Purification of Tbj1

Due to the fact that the TbHsp40s of interest proved to be insoluble during solubility studies, a denaturing purification protocol with native washes was adopted. A single *E. coli* XL1Blue transformant that was positively identified by means of plasmid isolation and a diagnostic digest was inoculated into a 2 x YT overnight starter culture (30 ml) containing 0.1 mg / ml ampicillin and grown with shaking (180 rpm) at 37°C for 12 -16 hours. A 225 ml flask of 2 x YT broth was inoculated with the overnight starter culture and induced with 1 mM isopropyl-1-thio-β-D-galacopyranoside (IPTG) when the absorbance of the culture at 600 nm (A_{600}) had reached a value of 0.6 – 0.8. The culture was induced for a period of 5 hours and, unless otherwise specified, no additional spiking with ampicillin took place. After 5 hours, the induced cultures were centrifuged at 6 000 *g* for 20 minutes in a JA-14 rotor (Beckman Coulter, U.S.A.) at 4°C and the pellets were resuspended in 10 ml denaturing lysis buffer (8 M Urea, 300 mM NaCl, 100 mM Tris-HCl, 10 mM imidazole; pH 8.0) containing freshly prepared phenylmethylsulfonylfluoride (PMSF) (1 mM). Subsequent to thorough resuspension, the lysate was aliquotted into microcentrifuge tubes (1.5 ml) and stored at -20°C overnight. Sepharose (GE Healthcare, U.K.) beads were charged with 0.1 M nickel sulphate and stored at 4°C until required (Appendix A.15).

Subsequent to thawing the frozen *E. coli* XL1 Blue lysates at room temperature, the protein lysates were centrifuged for 20 minutes at 12 000 *g* in a refrigerated microcentrifuge at 4°C and the cleared lysates were added to a 1 ml nickel-charged sepharose column. The lysate was allowed to bind the nickel sepharose beads for a period of 4 – 6 hours on ice with gentle rocking. The nickel bead-bound lysate was subsequently collected by means of centrifugation at 500 *g* for

3 minutes. Non-specifically bound proteins were eluted by means of 4 wash steps of 2-3 column volumes each with native wash buffer (100 mM Tris-HCl, 300 mM NaCl, 100 mM Imidazole; pH 8.0).

As the target proteins of interest are Hsp40s that are capable of interactions with endogenous *E. coli* Hsp70s, 0.6 mM ATP was added to each wash step to render any native *E. coli* DnaK incapable of such interactions and hence prevent co-purification of these proteins which would interfere with any subsequent *in vitro* studies. When the target protein of interest was sufficiently pure (single band on an SDS-PAGE gel) and a sufficient quantity of contaminants had been washed from the column, the protein was eluted with native elution buffer (100 mM Tris-HCl, 300 mM NaCl, 1 M Imidazole; pH 8.0) and stored at 4°C prior to analysis by SDS-PAGE and Western blotting in order to confirm protein identity and integrity. Samples of each wash and elution step for SDS-PAGE analysis were collected throughout the purification optimization process and analysed until the purification was sufficiently optimised and found to be sufficiently reproducible and robust to render this unnecessary.

3.2.2.3. Expression and purification of Tcj1

The pET23a-Tcj1 construct was a kind donation from Dr. D.M. Engman (Department of Pathology, Northwestern Medical School, U.S.A). Expression of Tcj1 was induced in *E. coli* BL21(DE3) cells under the same conditions mentioned above (1 mM IPTG; 0.1 mg / ml ampicillin) for 4-5 hours at 37°C prior to commencing a denaturing purification with native washes, as outlined for its orthologue, Tbj1.

3.2.2.4. Expression and purification of TcHsp70

TcHsp70 (Tc00.1047053511211.170) is the inducible cytosolic Hsp70 in *T. cruzi*. The pQE30-TcHsp70 plasmid was a kind donation from Michael Ludewig

(Department of Biochemistry, Microbiology and Biotechnology, Rhodes University, South Africa). TcHsp70 was expressed in *E. coli* XL1 Blue cells (1 mM IPTG; 0.1 mg / ml ampicillin) at 37°C for 5 hours prior to commencement of a denaturing purification with native washes as outlined above. The purification regimen adopted was similar to that adopted for Tbj1 purification with the exception that no ATP was added to the wash steps.

3.2.2.5. Preparation of proteins for further *in vitro* analysis

Once protein identity and integrity had been confirmed protein elution fractions were pooled and dialysed in snakeskin dialysis tubing with a molecular weight cut off limit of 10 kDa against dialysis buffer (50 mM Tris-HCl, 200 mM NaCl; pH 8.0) at 4°C for a period of 5-6 hours. The protein samples were quantified using the Bradford assay (see Appendix A.14) both prior and subsequent to dialysis and were concentrated in snakeskin dialysis tubing on a bed of PEG20 000 at 4°C for 1-2 hours where necessary.

3.2.2.6. SDS-PAGE of purified proteins

SDS-PAGE analysis was performed according to a protocol adapted from that of the Laemmli method (Laemmli, 1970). Protein samples were added to SDS-PAGE loading buffer (0.0625 M Tris-HCl, pH 6.8, 10 % glycerol w/v, 2 % SDS w/v, 5 % β-mercaptoethanol v/v, 0.05 % bromophenol blue v/v) in a ratio of 6 : 1 prior to boiling for 10 minutes. The samples were subsequently resolved on 12 % SDS-PAGE separating gels (0.375 M Tris-HCl, pH 8.8; 0.1 % SDS w/v, 12 % acrylamide) with 4 % stacking gels (0.125 M Tris-HCl, pH 6.8; 0.1 % SDS w/v, 4 % acrylamide) that were polymerized by the addition of 0.05 % ammonium persulphate (APS) and 0.005 % N,N,N‟,N‟-tetramethylethylenediamine (TEMED). The gels were resolved in SDS-PAGE running buffer (25 mM Tris-HCl, 192 mM Glycine and 1 % SDS w/v; pH 8.3) at 200 V for 45 minutes prior to staining with Coomassie stain (40 % methanol v/v, 0.7 % acetic acid and 0.075 % Coomassie

brilliant blue G250) and destaining with destain solution (40% methanol v/v, 0.7% acetic acid v/v).

3.2.2.7. Western analysis of purified proteins

The Western transfer of proteins was performed according to the protocol outlined in Towbin *et al*. (1979). The proteins of interest were resolved on 12 % SDS-PAGE gels and transferred to a nitrocellulose membrane (Bio-Rad; U.S.A.) that had previously been soaked in Western transfer buffer (25 mM Tris-HCl, 192 mM Glycine, 20 % v/v Methanol). The transfer was performed in a Bio-Rad Western apparatus (Bio-Rad, U.S.A.) at a constant voltage of 100 V, and a current of no lower than 300 mA for a period of 60 minutes. Upon completion of transfer, the blots were stained with Ponceau Red stain (0.5 % w/v Ponceau S and 1 % glacial acetic acid v/v) to confirm the success of the transfer and placed in block solution [5 % non-fat milk powder in TBS-Tween (Tris-buffered saline with Tween-20; 50 mM Tris-HCl, 150 mM NaCl, 1 % Tween-20; pH 7.5)] overnight at 4°C. The Western blots were developed by incubating with primary antibody for an hour, prior to a wash step with TBS-Tween for an hour and incubation with the secondary antibody for an hour. The blots were washed in TBS-Tween for another hour prior to developing using chemiluminescent reagent (GE Healthcare, U.K.) and the Chemidoc XRS imaging system (Bio-Rad, U.S.A.). The concentration of primary antibody used varied on the antibody titre and the protein being detected.

3.2.3.1. Composite analysis of the Tbj47 protein sequence in order to find suitable antigenic regions

The protein sequence of Tbj47 (Tb927.1.1230) was retrieved from the GeneDB database (http://www.genedb.org/) and analysed. A full composite analysis was performed on the sequence, and several potential antigenic regions were identified based on parameters such as surface probability (Emini plot) (Emini *et*

al., 1985), secondary structure (Garnier-Robson and Chou-Fasman plots) (Garnier *et al*., 1978; Chou and Fasman, 1978), transmembrane regions (Goldman and Eisenberg plots), chain flexibility (Karplus-Schultz plot), antigenic determinants (Hopp-Woods and Jameson-Wolf Index plots) (Jameson and Wolf, 1988) and charge density. Four potential antigenic regions were obtained and were subjected to a BLAST search of the *T. brucei* genome to determine which region possessed the lowest levels of similarity and identity to other *T. brucei* proteins, especially Hsp40 proteins. The antigenic region of interest was submitted to the GenScript U.S.A. (http://www.genscript.com/) for peptide synthesis. The synthesised peptide (EGIDTD SSASKQEQ) was used to generate an antibody against Tbj47 in rabbit by the Genescript corporation (U.S.A.).

3.2.3.2. Production of antibodies against the full length Tbj1 protein

The purification of Tbj1 was optimized as described above in order to obtain a pure protein with levels of endogenous *E. coli* DnaK contamination that were undetectable by Western analysis. The purification was confirmed with both anti-His (1:5000) and anti-DnaK (1:5000) antibodies. The protocol followed was the same as the optimized protocol described in Section 3.2.2.2 above, and in Appendix A.16. The purified Tbj1 was dialysed, concentrated and sent to Prof. D. Bellstedt (University of Stellenbosch, South Africa). The purified protein was applied to acid treated („naked") bacteria and injected into a rabbit from which serum was collected at 0, 14, 28 days. The method used relies on the antigenic determinants of *Salmonella minesota* R595 bacteria and was developed by Dr. D. Bellstedt (Bellstedt *et al*., 1987). The antibodies were tested on both Tbj1 and Tcj1 in concentrations ranging from 1:5000 – 1:50 000 for the primary antibody coupled to 1:5000 secondary anti-rabbit IgG conjugated to horseradish peroxidase (HRP) (Invitrogen) antibody.

3.2.4.1. Detection of Tbj1 and Tbj47 in *T. brucei* procyclic and bloodstream form lysates

T. brucei Pro and BSF lysates were a kind donation from Prof. T. Coetzer (University of KwaZulu-Natal, South Africa). The lysates were resuspended in SDS-PAGE loading buffer to a final concentration of 0.4 mg / ml. An aliquot of each lysate (20 µl) was resolved on duplicate 12 % SDS-PAGE gels (method as described in Section 3.2.2.6) and transferred to a nitrocellulose membrane for Western analysis. The blots were incubated in either rabbit polyclonal anti-Tbj1 primary antibody (1:5000) or rabbit polyclonal anti-Tbj47 primary antibody (1:5000) and anti-rabbit IgG conjugated to horseradish peroxidase (HRP) (1:5000) secondary antibody prior to detection as described in Section 3.2.2.7.

3.3. RESULTS

3.3.1. Strategy for the production of pQE-1 plasmids containing the coding regions of Tbj1, Tbj47 and Tbj51

The sequences of the proteins to be expressed were analysed prior to the design of oligonucleotide primers in order to determine whether introns were present in the coding regions of Tbj1, Tbj47 and Tbj51. This was crucial as the presence of introns would have required *T. brucei* mRNA as the starting template for the amplification of the coding regions, instead of genomic DNA. The absence of introns in *T. brucei* (Huang and van der Ploeg, 1999), as well as a lack of predicted differences between the genomic coding regions and mRNA of the genes of interest allowed for the amplification of the coding regions of interest from *T. brucei TREU927* genomic DNA. While *T. brucei* does display a certain level of codon bias (Horn, 2008), many codons favoured by this organism were found to be widely used in *E. coli*, and due to the ease and versatility of *E. coli* as an expression system, the direct expression and purification from *E. coli* was used as an initial step in our investigation. The use of codon harmonisation / optimisation was not utilized for the proteins of interest.

The pQE-1 plasmid vector is a derivative of the popular pQE-30 expression vector that has found widespread use in the heterologous expression of proteins *in vitro*. The pQE-1 plasmid allows for the in-frame cloning of an N-terminal His_6-tag which is cleavable with the aid of dipeptidyl aminopeptidase I (DAPase I) (Pedersen *et al*., 1999). This allows for versatility in terms of producing proteins with or without the His_6-tag. This plasmid was used in the cloning strategy outlined here, as future work may require the removal of the Tbj1 N-terminal His_6-tag. The strategy adopted for the insertion of the cloning regions of three novel TbHsp40s is outlined in Figure 3.1. Suitable primers based on the cDNA sequences on the GeneDB database were designed to contain restriction sites

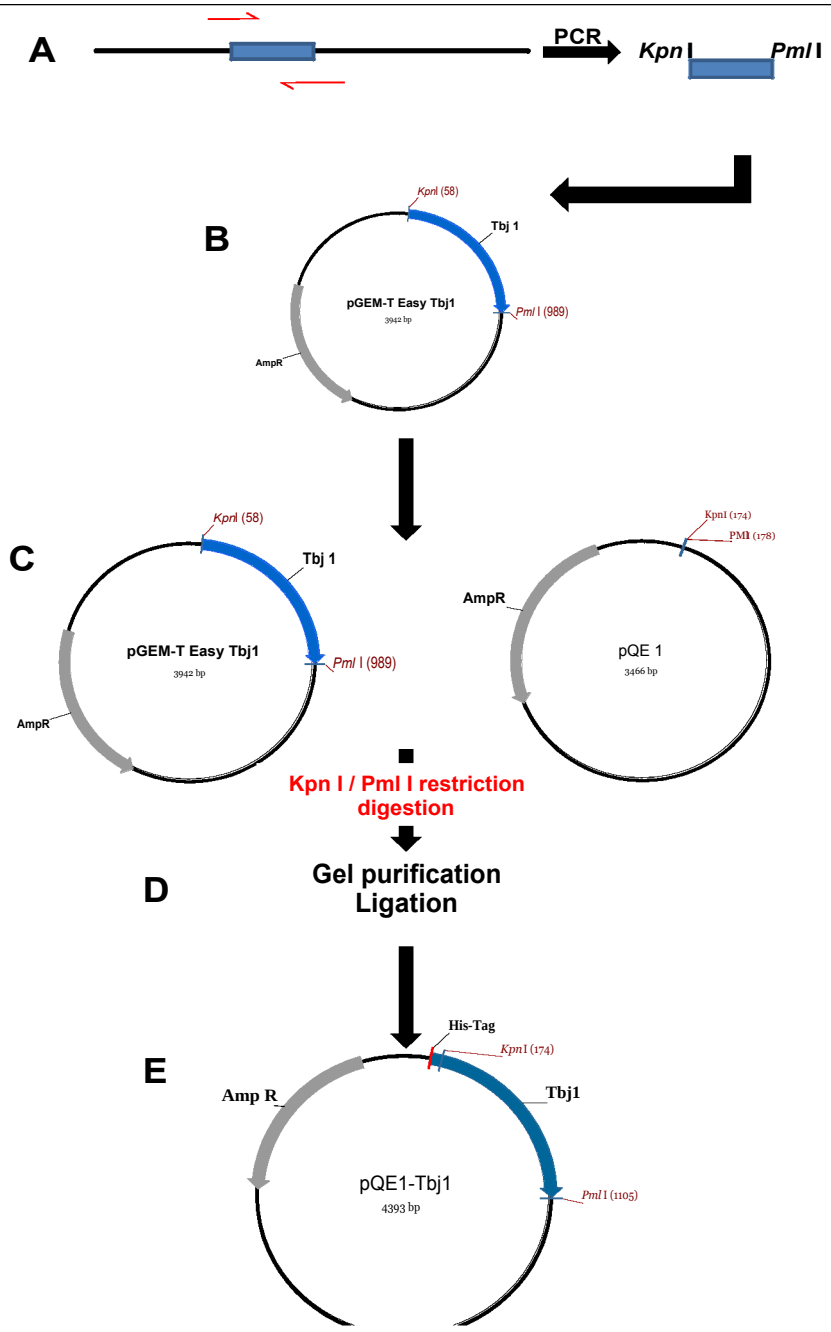


Figure 3.4. Strategy employed for the cloning of three novel TbHsp40 coding regions into the pQE-1 expression vector

Strategy employed for the insertion of TbHsp40 coding regions into the pQE-1 vector (Qiagen, U.S.A.) using the Tbj1 coding region as an example. [A] – Amplification of Tbj1 coding region from *T. brucei* genomic DNA; [B] – Insertion of gene of interest into the pGEM-T® Easy cloning vector; [C] – Gel purification of the fragment of interest from the pGEM-T® Easy cloning vector and restriction of the pQE-1 vector with the same enzymes used to cut the pGEM-T® vector; [D] – Ligation of the fragments of interest into the pQE-1 vector; [E] – Example of Tbj1 in the pQE-1 expression vector showing the N-terminal His_6-tag in red.

not occurring within the coding region of the protein of interest and were used to amplify the coding region of interest directly from *T. brucei* genomic DNA (Figure 3.1). The amplified coding regions of the TbHsp40s of interest were inserted into the pGEM- T Easy® cloning vector prior to digestion with suitable restriction enzymes and insertion into the pQE-1 expression vector (Figure 3.4). The initial cloning step into the pGEM-T® Easy plasmid vector ensured that the inserts were cut to completion facilitating higher levels of success in the experiment.

3.3.2. Successful amplification of Tbj1, Tbj47 and Tbj51 from *T. brucei* genomic DNA by means of the Polymerase Chain reaction (PCR)

The coding regions of three novel TbHsp40 proteins were successfully PCR-amplified from *T. brucei TREU927* genomic DNA (Figure 3.5).

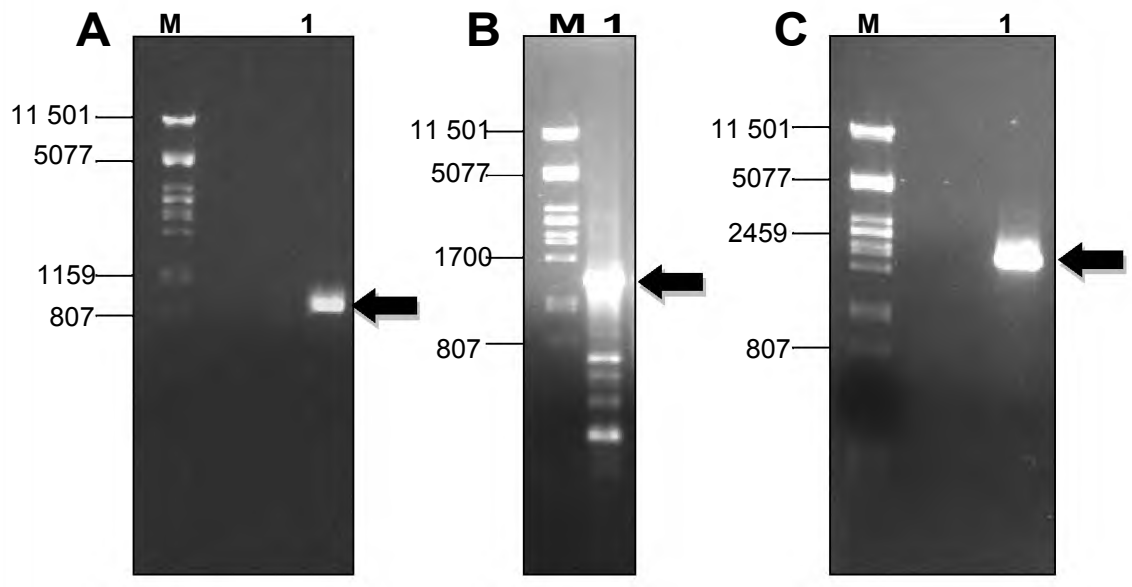


Figure 3.5. PCR amplification of Tbj1, Tbj47 and Tbj51 coding regions

The coding regions of Tbj1, Tbj47 and Tbj51 were successfully PCR-amplified from *T. brucei* genomic DNA and resolved on 1 % agarose gels containing ethidium bromide. The molecular weight marker (Lane M) in all three figures is λ-*Pst* I DNA ladder and Lane 1 denotes the PCR products. Figure A - Tbj1– 927 b.p; Figure B - Tbj47– 1440 b.p. and Figure C Tbj51– 2218 b.p. Amplified products are indicated by a black arrow.

The amplification of the coding region of Tbj1 resulted in the production of a 927 b.p. band displaying no visible contaminants (Figure 3.5.A) while the amplification of the coding region of Tbj51 resulted in the production of a 2218 b.p. band (Figure 3.5.C). The amplification of the Tbj1 and Tbj51 coding regions was successful and did not require optimisation of the PCR-amplification reaction. The amplification of the Tbj47 coding region, however, presented several challenges with respect to optimization. The annealing temperature of the reaction was lowered considerably (50 °C) which resulted in successful amplification of the Tbj47 coding region (Figure 3.5.B) of 1440 b.p., which corresponds to the size predicted by the GeneDB database. However, this amplification, due to the low annealing temperature, also resulted in the amplification of a high number of non-specific contaminating bands (Figure 3.5.B) resulting in the need for optimised resolution by agarose gel electrophoresis prior to ligation of the Tbj47 coding region.

3.3.3. Development of the pQE1-Tbj1, pQE1-Tbj47 and pQE1-Tbj51 plasmid constructs

The coding regions of Tbj1, Tbj47 and Tbj51 were successfully inserted into the pQE-1 expression vector (Qiagen, U.S.A.) and screened by means of diagnostic restriction digests (Figure 3.6) and DNA sequencing (Figure 3.7). The Tbj1 coding region was digested from the pGEM-T® Easy cloning vector and was inserted into the pQE-1 vector by ligation, which readily ligated and was screened as shown in Figure 3.6 (A). A fragment of 927 b.p. was released on digestion of the pQE1-Tbj1 plasmid with *Kpn* I and *Pml* I restriction enzymes, confirming that an insert corresponding to the predicted size of the Tbj1 coding region had been successfully inserted into the plasmid. The plasmid map to the left of Figure 3.6. (A) shows the organisation of the pQE-1 plasmid with the Tbj1 coding region (shown in green) inserted into it, with respect to the location of the cleavable His_6-tag and the ampicillin-resistance gene (grey). The Tbj47 and Tbj51 coding regions were successfully inserted into the pQE1 plasmid vector to

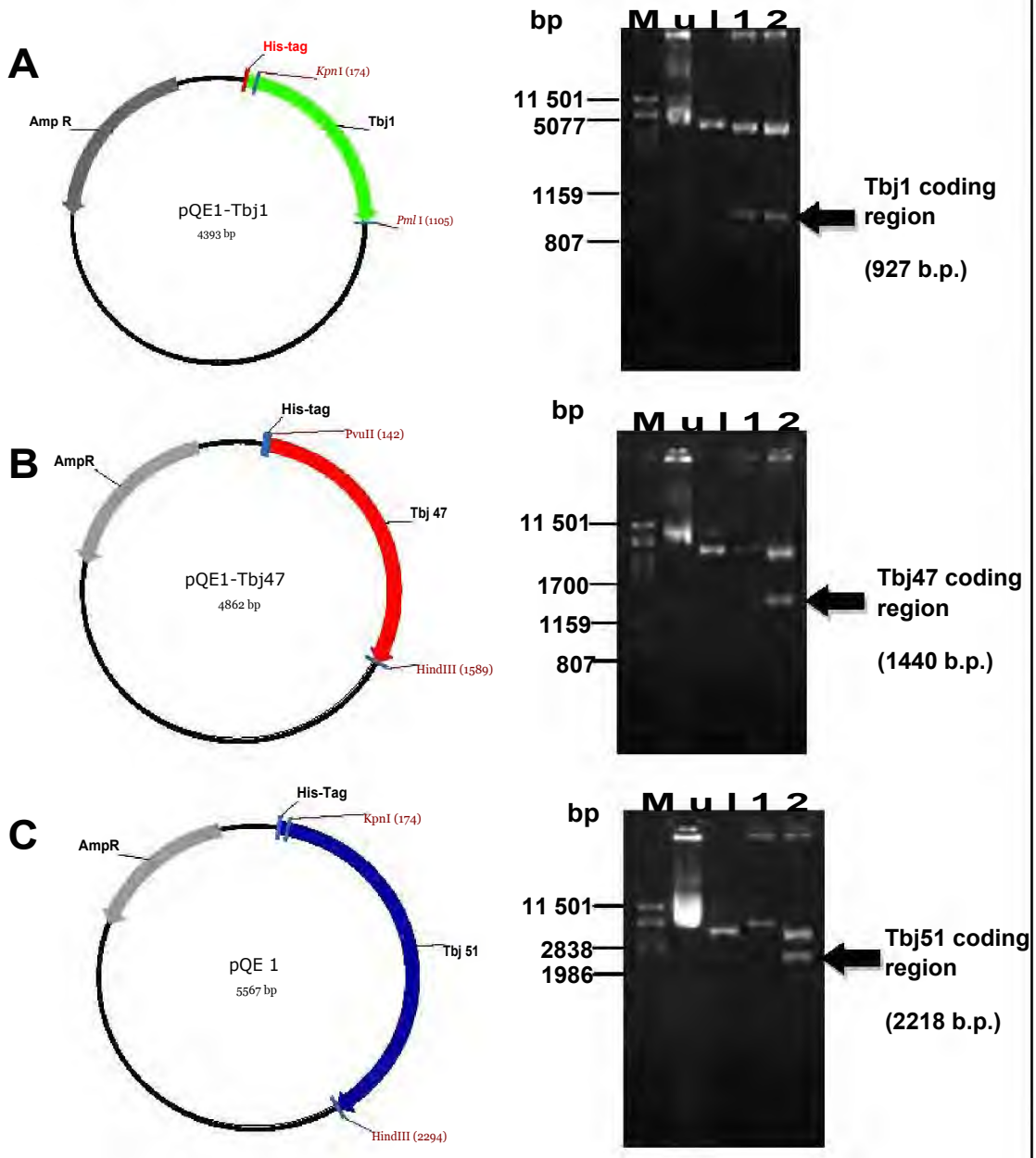


Figure 3.6. Restriction analysis of pQE1-Tbj1, pQE1-Tbj47 and pQE1-Tbj51

Diagnostic restriction digest products of the plasmids were resolved on 1 % agarose gels in TBE buffer with plasmid maps (generated using Vector NTI®) on the left. The linear DNA size markers (Lane M) in all three figures is λ-Pst I restricted DNA, and marker sizes are shown to the left of each figure. Lane U – Uncut plasmid; Lane l – linearised pQE-1 plasmid; Lanes 1 and 2 – digested pQE-1 plasmids containing inserts. The pQE-1 vector is 3 466 bp comprises the uppermost band in lanes 1 and 2 in each figure. AmpR denotes the gene in the pQE1 plasmid that confers resistance to ampicilin. Figure A: pQE1-Tbj1– 927 b.p; Figure B: pQE1-Tbj47– 1440 b.p. and Figure C: pQE1- Tbj51– 2218 b.p.

yield the pQE1-Tbj47 and pQE1-Tbj51 constructs. The identity of the pQE1-Tbj47 construct was confirmed by a diagnostic restriction digestion with *Pvu* II and *Hind* III enzymes, which released a fragment of 1440 b.p., corresponding to the predicted size of the Tbj47 coding region (Figure 3.6.B). The identity of pQE1-Tbj51 was confirmed in a similar fashion, but *Kpn* I and *Hind* III restriction enzymes were used for the diagnostic restriction digestions, releasing a fragment of 2218 b.p. corresponding to the size of the Tbj51 coding region (Figure 3.6.C). Several pQE1-Tbj51 clones screened during cloning and analysis appeared to be partially digested (linearised) when both *Kpn* I and *Hind* III were used to restrict the plasmid (Figure 3.6.C; Lane 1) indicating that one of these enzymes was digesting at suboptimal levels during the screening process.

3.3.4. Tbj1 is a *T. brucei* ortholog of Tcj1

An *in silico* investigation of Tbj1 and Tcj1 was performed (Chapter 2; Section 2.3.7 Figure 3.7) in order to investigate the extent of the similarity and identity between the two proteins, as well as the identity between the Tbj1 coding region amplified in the present study and Tbj1 from the GeneDB database. The coding region amplified from *T. brucei* genomic DNA and inserted into pQE-1 was shown to have an identity of 99.4 % with that of Tbj1 from the GeneDB database and the sequence amplified was submitted to the NCBI database under the accession number **GQ240140** (Figure 3.7). A number of conservative changes were found between the two sequences (Figure 3.7; turquoise residues), but regions proposed to be critical to the *in vivo* function of Tbj1, such as the J-domain, were found to be highly conserved with respect to the two sequences. This served to indicate that the correct protein coding region was amplified from *T. brucei* genomic DNA in the present study. The sequences of Tbj1 and Tcj1 are well-conserved with respect to one another (Figure 3.7), displaying a 64.3 % identity and a 73.7 % similarity with one another over the full length sequence, which leads to the hypothesis that these proteins are indeed orthologs in *T. brucei* and *T. cruzi*, respectively. In both Tbj1 and Tcj1, the J-domain is

N-terminal and conserved with respect to the J-domains of canonical Hsp40s in *E. coli*, yeast and humans (Chapter 2; Figure 2.6).

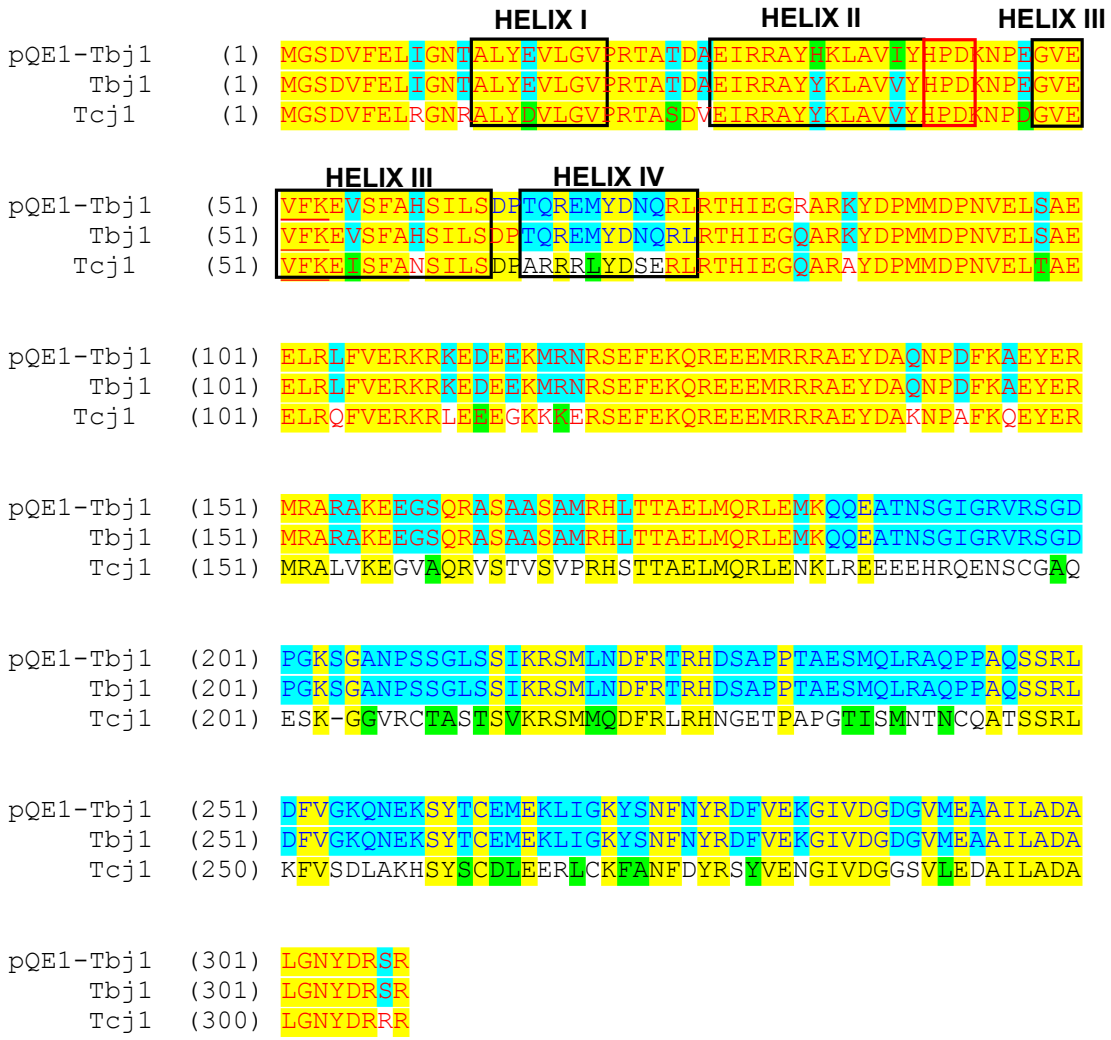


Figure 3.7. Pairwise alignment of Tbj1, Tcj1 and pQE1-Tbj1

The full sequences of Tbj1 and Tcj1 from the GeneDB database were aligned (64.3 % identity and 73 % consensus residues) with the coding region from Tbj1 amplified from *T. brucei TREU927* genomic DNA and inserted into pQE1 (pQE-Tbj1). This sequence had a 99.4 % identity to the Tbj1 sequence from the database, and a 64 % identity to that of Tcj1. The J-domains are annotated, with the position of the helices indicated on the figure in black boxes. The highly conserved HPD motif is shown in red and the modified KFK region is underlined. [Similar residues] – green; [Identical residues] – yellow background with red writing; [Conservative changes] – turquoise. The alignment was performed using the AlignX program of the Vector NTI® suite (Invitrogen, U.S.A.).

In spite of the conservation of several residues that are predicted to be involved in maintaining the structural integrity of the J-domain in both Tbj1 and Tcj1, a number of residues predicted to be important for the interaction of Hsp40s with Hsp70s were found not be conserved in these proteins. These include the Lys / Arg^{26} residue and the KFK and QKRAA motifs (discussed in Chapter 2). In terms of orthologous functionality, it was interesting to note that these J-domain substitutions were common to both Tbj1 and Tcj1. The greatest level of sequence variation between Tbj1 and Tcj1 is in the region from 175 - 250 a.a (Figure 3.7), and this sequence difference could potentially result in slightly different modes of action. Tbj1 and Tcj1 are both Type III Hsp40s and lack a G-F rich region and a zinc finger binding domain. The C-terminal domains of these proteins have been analysed in Chapter 2 (Section 2.3.7).

3.3.5. Tbj1 was successfully overexpressed in *E. coli* and purified

Tbj1 was successfully overexpressed and purified (Figure 3.8). The induction profile of the protein was checked prior to purification (data not shown) and it was determined that one 225 ml culture volume of *E. coli* expressing protein was sufficient to acquire an amount of protein adequate for further biochemical analysis. A solubility study was performed on Tbj1 (data not shown) which revealed that the protein was in the insoluble fraction. In spite of this, a native purification was attempted (data not shown) which was unsuccessful due to the absence of soluble Tbj1. For this reason, a dentaturing lysis purification protocol with native wash and elution buffers was adopted and extensively optimized with respect to wash buffer additives, the volume and number of wash and elution steps and the volume of the nickel-affinity resin used (data not shown).

Tbj1 expressed at high levels in *E. coli* XL1Blue [pQE1-Tbj1] cells (Figure 3.8 A; L) and *E. coli* BL21 [pQE1-Tbj1] cells (data not shown). Because the expression levels of Tbj1 occurring in the former was marginally higher, *E. coli* XL1Blue cells were used for all expression and purification of the protein. A significant amount

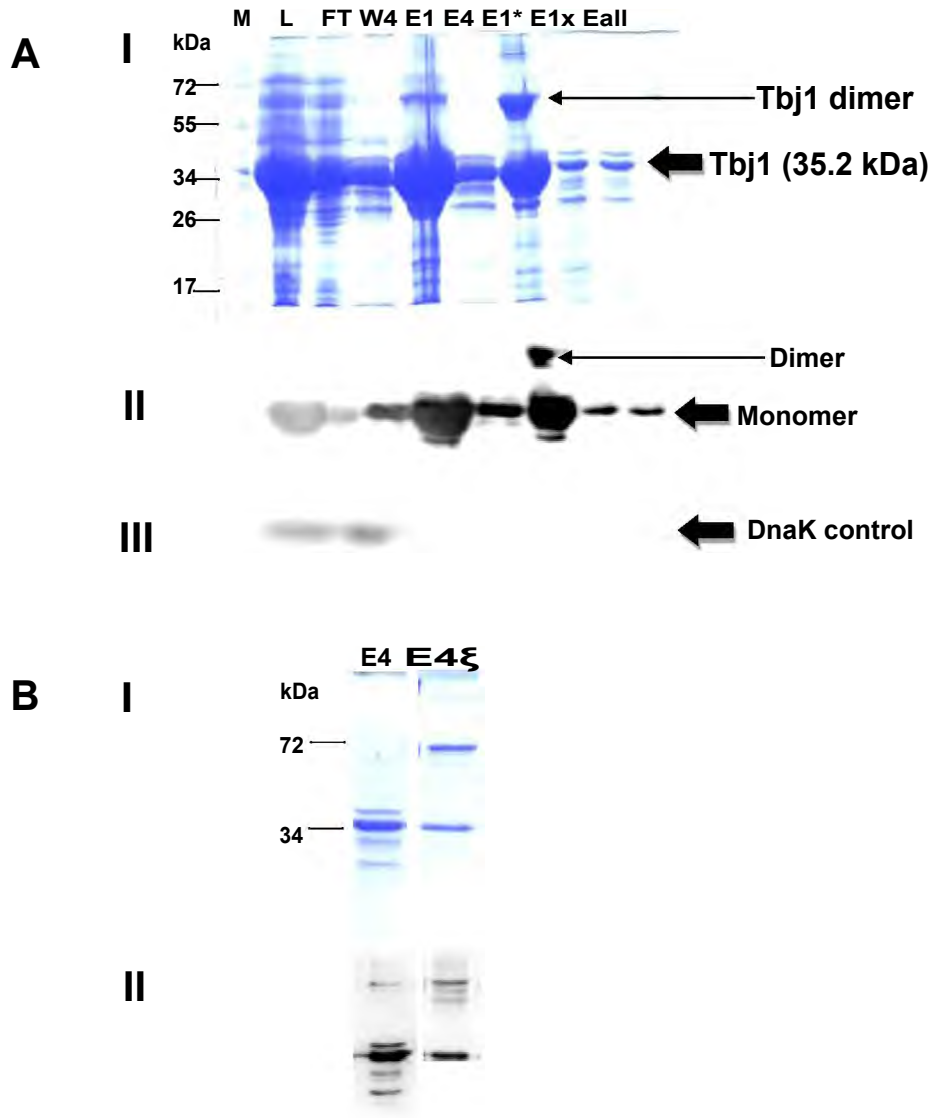


Figure 3.8. SDS-PAGE and Western analysis of Tbj1 purification

SDS-PAGE analysis shows expression and purification of Tbj1 from E. coli XL1Blue [pQE1-Tbj1] cells. Panel A.I: Lanes SDS-PAGE analysis [M] - Protein marker IV (pEQLab), marker sizes indicated next to the gel; [L] – Tbj1 lysate in E. coli; [FT] – Flow through after binding to nickel-sepharose beads; [W4] – Wash 4 during purification (100 mM Imidazole wash); [E1] – Elution 1 (1 M Imidazole); [E 4] – Elution 4 (1 M Imidazole); [E1*] - Elution 1 (unboiled for SDS-PAGE); [E1x] – Elution 1 after dialysis and filtering; [EaII] – Pooled elutions 2 – 4 after dialysis and filtering, Panel A.II: Western analysis using anti-His primary antibodies, lanes as per A.I. Panel A.III: Western analysis using anti-DnaK primary antibodies, lanes as per A.I. Tbj1 monomeric and dimeric species are indicated with arrows. Panel B I: SDS-PAGE analysis of Tbj1 after dialysis and concentration; [E4] – Elution 4 after boiling with β-mercaptoethanol; [E4ξ] – Elution 4 without boiling and β-mercaptoethanol. Panel B II: Western analysis refers to E4 and E4ξ probed with anti-Tbj1 primary antibody.

of protein was lost in the flow-through stage of the purification (Figure 3.8 A; FT) but the amount remaining bound to the column was sufficient to obtain adequate amounts of protein for further purification steps (Figure 3.8 A). The formation of SDS-resistant dimers was observed when Tbj1 that was not boiled was analysed by means of SDS-PAGE analysis (Figure 3.8 A; E1 and E1*). This phenomenon was previously observed by others (Lakhal *et al*., 2008) but was not observed during SEC-FPLC analysis of the protein (Chapter 4; Figure 4.2). This could potentially be attributed to differences in salinity of the buffers used during dialysis as opposed to SEC-FPLC. Approximately 2.5 mg Tbj1 was obtained from a single culture purification, but a large amount of protein was lost during the dialysis step due to aggregate formation, which resulted in a typical yield of 0.9 mg of protein of > 80 % purity by gel analysis and inspection (Figure 3.8 A; I and II). SEC-FPLC analysis confirmed that the Tbj1 protein purification resulted in the successful isolation of a compact monomeric protein that was sufficiently free of contaminants to enable further biochemical characterisation (Chapter 4; Figure 4.2). The purification of Tbj1 was confirmed by means of Western analysis, probing with both anti-His and anti-Tbj1 antibodies (Figure 3.8 A; II and Figure 3.13).

3.3.6. Overexpression and purification of Tbj47 and Tbj51

Tbj47 was not successfully overexpressed in *E. coli* XL1Blue cells or *E. coli* BL21 cells (data not shown). This could potentially be attributed to the fact that the coding region of Tbj47 is rich in codons that end with A or T, which has been shown to be an indicator of low expression ability in a heterologous system for *T. brucei* proteins (Chanda *et al*., 2007). Attempts were made to overexpress the protein in the *E. coli* Rosetta strain, and the RIG plasmid encoding the tRNAs for rare codons was co-transformed with the pQE-1 Tbj47 plasmid. However, none of these attempts proved successful in terms of producing an overexpressed protein. The temperature at which induction took place was also optimized (25 -

37°C) but appeared to have no effect on levels of Tbj47 production (Table 3.1 and data not shown).

Table 3.1. Overexpression and purification parameters investigated for optimization of heterologous expression of Tbj47 and Tbj51

Parameter	Changes made	Effect on expression	Reference
Ampicillin concentration	Range of concentrations from 50 µg / ml – 250 µg / ml	None	- -
"Spiking" with Ampicillin	Addition of 100 µg / ml ampicillin to the growing culture at three-hourly intervals	None	Gräslund *et al*., 2008
Inducer (IPTG) concentration	Range of concentrations from 0.1 mM – 0.5 mM	None	- -
Temperature (°C)	Range of temperatures from 20°C - 37°C	None	Gräslund *et al*., 2008; Vera *et al*., 2007
Cell density prior to induction	Range from A600 of 0.5 – 1.5		Tunac 1989; Brodsky *et al*., 2006
Different *E. coli* strains	Use of *E. coli* XL1Blue, Rosetta, Origami and BL21(DE3)	None	Gräslund *et al*., 2008

The same optimisation attempts to obtain overexpression for Tbj51 as outlined above were made (Table 3.1). This protein also did not overexpress in any of the *E. coli* competent cell types tested, and the presence / absence of the RIG plasmid as well as temperature studies had a negligible effect. For this reason, the main focus of the present work shifted to the biochemical characterisation of Tbj1 and Tcj1, and further studies on the biochemical characterisation of Tbj47 and Tbj51 were not attempted.

The lack of heterologous expression in the present case could not be attributed to mutations in the coding regions amplified for these two proteins as DNA sequencing data indicated that the start and stop codons were in place, and were in-frame with respect to the expression vector used.

3.3.7. Tcj1 was successfully expressed and purified

The pET23b-Tcj1 construct was successfully isolated and its identity was confirmed by means of a diagnostic restriction digest (Figure 3.9). The pET23b-Tcj1 construct yielded two fragments when restricted with *Xho* I and *Eco* RI restriction enzymes, namely the plasmid vector (Figure 3.9; PV) of 3631 b.p. and the Tcj1 coding region of 930 b.p. (Figure 3.9).

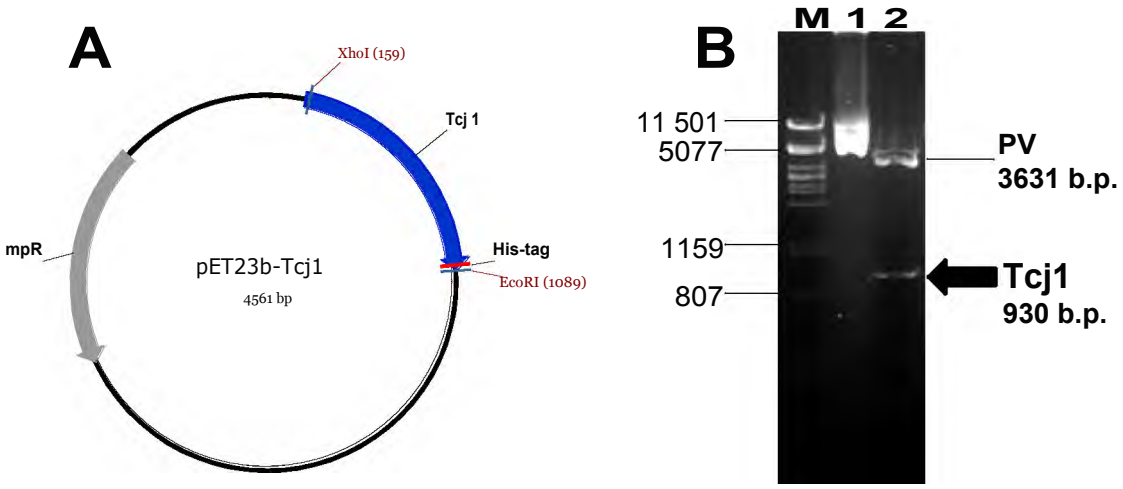


Figure 3.9. Restriction analysis the pET23b-Tcj1 plasmid

[A] – The plasmid map of the pET23b-Tcj1 plasmid generated with Vector NTI® software, highlighting the C-terminal His_6-tag in red. [B] - Diagnostic digestion of the pET23b-Tcj1 construct was conducted with XhoI and EcoRI restriction enzymes and the fragments resolved on a 1 % agarose gel in TBE buffer. The linear DNA size markers (Lane M) are λ-Pst I DNA ladder, and marker sizes are shown to the left of the figure. Lane 1 – Uncut pET23b-Tcj1; Lane 2 – Cut pET23b-Tcj1. The pET23b vector is 3631 bp (PV) in the figure, while the Tcj1 coding region is 930 bp (indicated by black arrows).

Tcj1 was overexpressed successfully in *E. coli* BL21(DE3) cells and purified in adequate quantities (0.6 mg). The protein obtained by means of a denaturing purification with native washes was adequately pure (>80 %) and resolved at a molecular mass of 35.4 kDa which correlates with the molecular mass predicted for Tcj1 (Figure 3.10.A). Sufficient quantities of protein were obtained for further

biochemical characterisation without the need for any additional scale-up procedures.

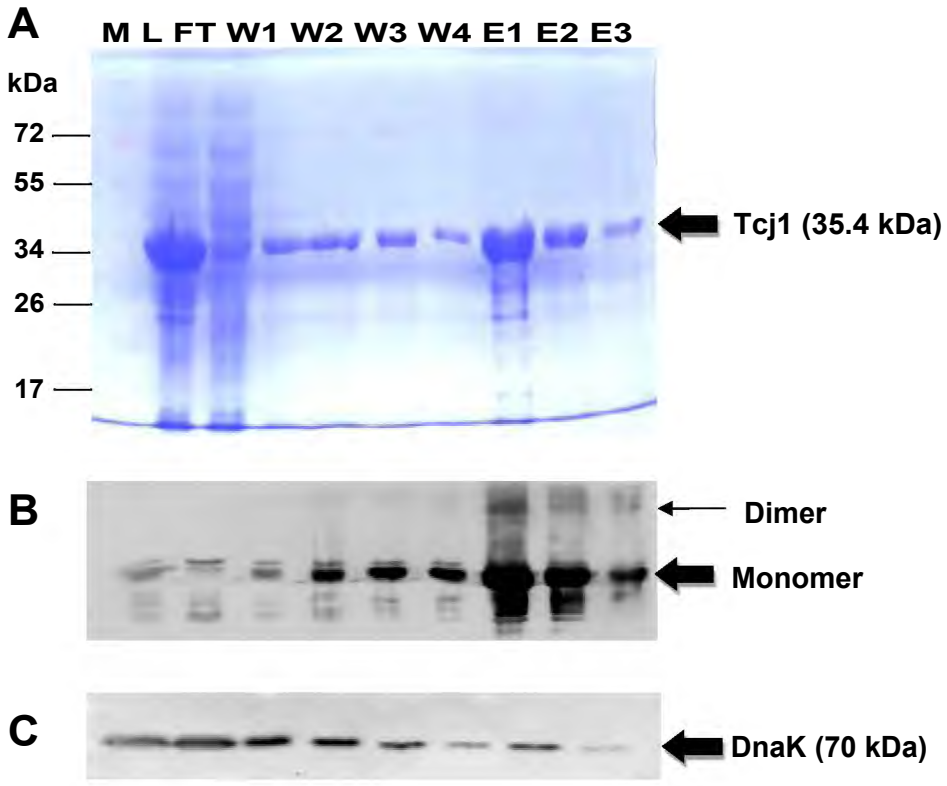


Figure 3.10. SDS-PAGE and Western analysis of Tcj1 purification

Panel A: SDS-PAGE analysis of the expression and purification of Tbj1 from E. coli BL21(DE3) [pET23b-Tcj1] cells. Lanes: [M] – Protein marker IV (pEQLab); [L] – Tcj1-containing lysate; [FT] – Flow through after binding to nickel-sepharose beads; [W1-W4] – Washes 1 to 4 during purification (100 mM imidazole wash); [E1] – Elution 1 (1 M imidazole); [E 2] – Elution 2 (1 M imidazole); [E3]- Elution 3 (1 M imidazole). Panel B: Western analysis using anti-Tbj1 primary antibody. Lanes as for Panel A. Panel C: Western analysis using anti-DnaK primary antibody. Lanes as for Panel A.

The overexpression and purification of Tcj1 could initially not be confirmed by means of Western analysis using detection with anti-His antibodies, a finding which is corroborated by the work of Edkins et al. (2004). It is proposed that the C-terminal His_6-tag is masked in Tcj1. For this reason, an N-terminal His_6-tag was selected for the Tbj1 cloning strategy. An antibody produced to detect Tbj1, was successful in detecting the overexpressed protein (Figure 3.10 B) and

confirmed overexpression and purification of the protein. The detection of Tcj1 by an antibody directed against full-length Tbj1 was anticipated due to the high sequence identity between the two proteins.

In spite of the use of ATP-washes to eliminate DnaK co-purification, Tcj1 invariably eluted with a significant amount of DnaK contamination irrespective of how much ATP was added (Figure 3.7; panel C). Tcj1 was found to adhere strongly to endogenous *E. coli* DnaK, suggesting that the DnaK may recognise misfolded regions of the Tcj1 protein as substrates.

3.3.8. Heterologous expression and purification of an Hsp70 from *T. cruzi*, TcHsp70

An Hsp70 protein from *T. cruzi* (TcHsp70) was successfully overexpressed and purified (Figure 3.11) in order to serve as a partner Hsp70 to Tbj1 and Tcj1 in biochemical characterisation experiments. TcHsp70 (Tc00.1047053511211.170) is the ortholog of TbHsp70 and its structure is described more fully in Chapter 2, Section 2.4. TcHsp70 was found to have slightly lower expression levels than Tbj1 in *E. coli* (Figure 3.11) but was still found to possess adequate levels of expression for the purposes of successful purification. TcHsp70 was found to purify to significant homogeneity in the third elution fraction, although the first two elution fractions were found to contain a number of contaminating proteins (Figure 3.11; E1 and E2) and were discarded prior to dialysis, with only Elution 3 being used for biochemical assays.

TcHsp70 has a molecular weight of 73.2 kDa and was predicted to localise to the cytoplasm of the *T. cruzi* parasite, in the same fashion as its ortholog in *T. brucei* (Chapter 2; Table 2.9). It was found to have an 88 % identity to TbHsp70, and a very similar structural orientation, which suggests that the mode of operation of these two proteins is very similar.

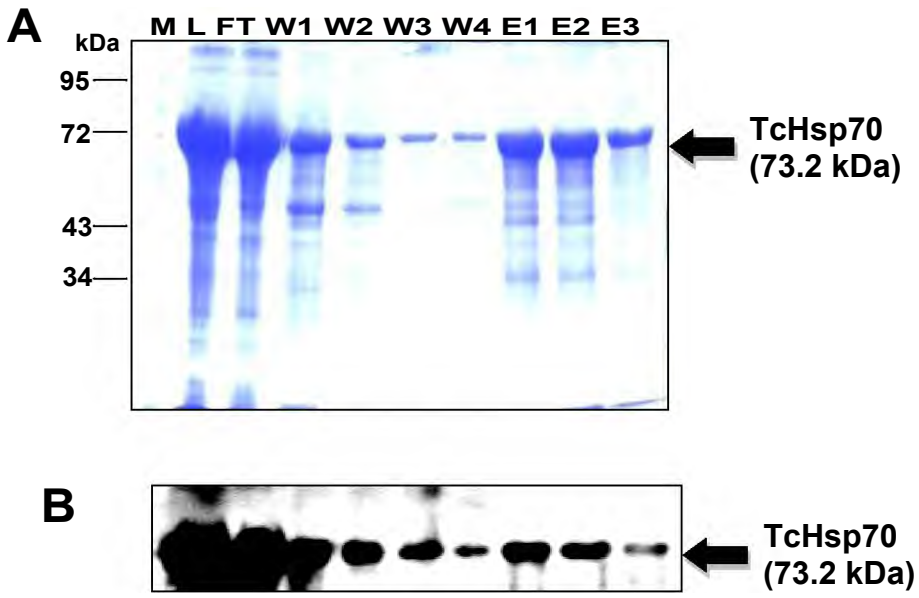


Figure 3.11. SDS-PAGE and Western analysis of TcHsp70 purification

SDS-PAGE analysis of TcHsp70 expression and purification. Panel A: shows expression and purification of TcHsp70 from E. coli XL1 Blue [pQE30-TcHsp70] cells. Lanes: [M] – Protein marker IV (pEQLab) with marker sizes shown on the left hand side of the gel; [L] – TcHsp70 lysate in E. coli; [FT] – Flow through after binding to nickel-Sepharose beads; [W1] – Wash 1 during purification (100 mM Imidazole wash); [W2] Wash 2 during purification (100 mM Imidazole wash); [W3] – Wash 3 during purification (100 mM Imidazole wash); [W4] – Wash 4 during purification (100 mM Imidazole wash); [E1] - Elution 1 (1 M Imidazole); [E2] – Elution 2 (1 M Imidazole); [E3] –Elution 3 (1 M Imidazole). Panel B – Western blot probed with anti-His primary antibody showing TcHsp70 detection.

The purification protocol that was adhered to for TcHsp70 was the same with the exception of the exclusion of ATP from the washes; TcHsp70 was also purified from a single 225 ml culture, and the average amount obtained from each purification was approximately 0.2 – 0.6 mg. The yield for the purification of TcHsp70 was far more varied from batch to batch than that obtained for Tbj1 and Tcj1. As conditions were kept constant, the reason for this variation in expression levels is unknown. To ensure that TcHsp70 of sufficient quality was being used for biochemical characterisation, every batch of protein was resolved using SDS-PAGE and subjected to Western analysis prior to use.

3.3.9. Composite protein analysis of Tbj47 for the design of peptide-directed antibodies

Due to recalcitrance of Tbj47 to heterologous expression in *E. coli* the isolation of Tbj47 for antibody production was not possible. Therefore, a full composite analysis of the Tbj47 protein sequence was performed to identify regions of suitable antigenicity that could be utilized in the design of peptides for peptide-directed antibody synthesis (Figure 3.12). The results of a composite analysis are studied in order to identify potential antigens, which, in the case of Tbj47 were subjected to rigorous BLAST analyses at both the NCBI and GeneDB portals in order to ensure that the antigenic peptide selected would not be suited for the detection of another protein in either humans, *E. coli* or another trypanosomal species (Lipman *et al*., 2005).

The results of the full composite analysis and the antigenic peptide chosen for Tbj47 are shown (Figure 3.12). The region corresponding to amino acids 405-418 was chosen as an antigen for peptide-directed antibody synthesis (EGIDTD SSASKQEQ) as it possessed all the qualities required for a potential epitope. The region is predicted to be surface exposed (Emini Surface Probability Index), hydrophilic (Hopp-Woods analysis), flexible (Karplus-Shultz analysis) and antigenic (Jameson Antigenic Index). The detailed analysis of the peak structure of this region as well as of the rest of the Tbj47 protein are shown in Figure 3.12 overleaf. The peptide region selected was not found to have high levels of identity with any of the proteins it was compared to in a BLAST analysis, which was a special concern for the TriTryps proteins, especially Hsps which are ubiquitous within the parasite. High levels of identity for the purpose of the present work referred to more than 4 identical amino acids in sequence over the total length of the Tbj47 peptide antigen. Subsequent to the successful design of the antigenic peptide to Tbj47, the peptide was synthesised by the GenScript corporation (U.S.A.) where it was used to inoculate two rabbits to produce Tbj47

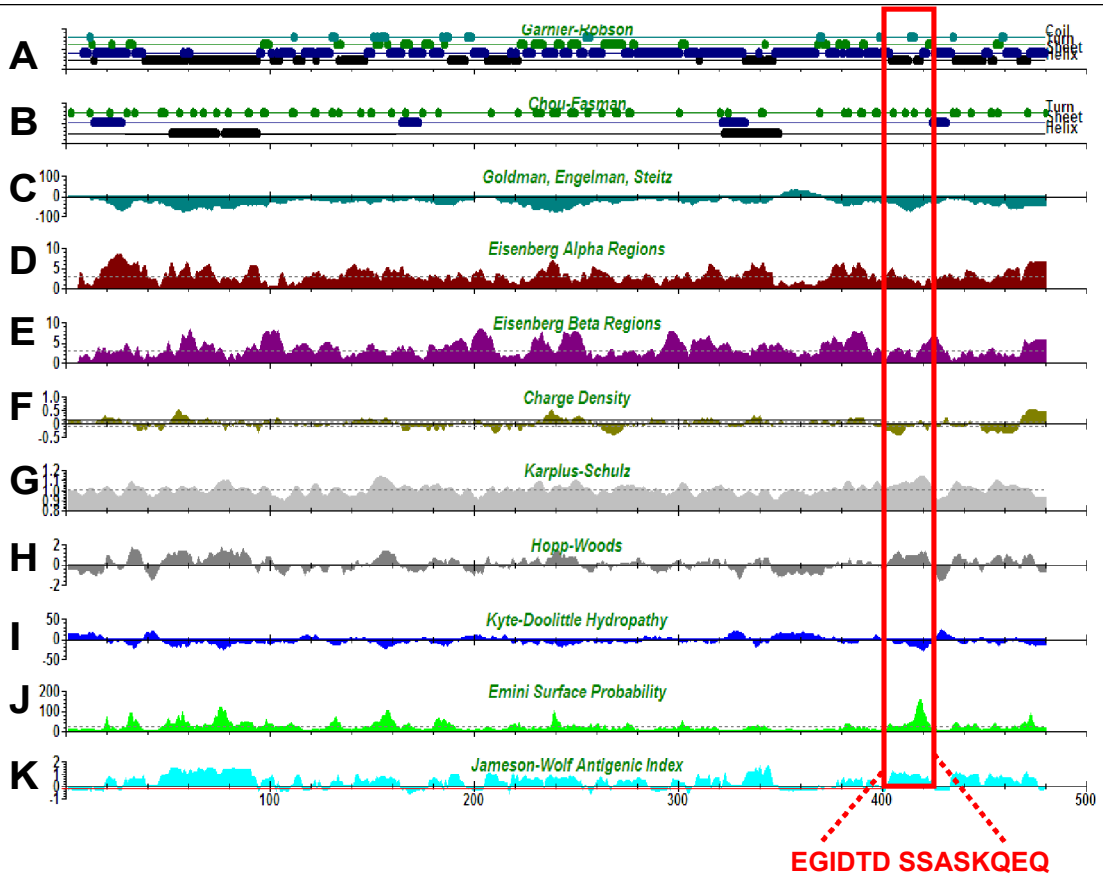


Figure 3.12. Composite analysis of the Tbj47 full-length protein to identify potential antigenic regions

The plots used to determine the optimal antigenic region are shown, with the region chosen outlined in red. The numbers along the X-axis of the graph indicate the relative amino acid position with respect to the N-terminus of the protein. [A] – Garnier-Robson: determines the effect that amino acids have on the conformational distance of others; [B]- Chou-Fasman: predicts secondary structure; [C] – Goldman, Engelman, Schultz: predicts hydrophobicity, where transmembrane regions are shown as positive peaks; [D] and [E] – Eisenberg alpha and beta regions: alpha shows hydrophobic moments characteristic of alpha helices (peaks in graph) whereas beta shows hydrophobic moments characteristic of a beta sheet (peaks in graph); [F] – Charge density: the graph shows the average charge distribution across the length of the protein at a pH of 7.0; [G] – Karplus-Schultz: gives a measure of the flexibility of the protein, with peaks corresponding to more flexible regions; [H] – Hopp-Woods: measure of average hydrophobicity, with positive peaks given as hydrophilic and negative peaks as hydrophobic; [I] – Kyte-Doolittle Hydropathy scale: peaks indicate hydrophobic regions; [J]- Emini Surface probability: predicts the surface exposed regions of a protein (peaks) (Emini et al., 1985; Janin et al., 1978); [K] – Jameson-Wolf antigenic index: yields a combination of analyses [A] – [J] together with secondary structure predictions to generate an overall antigenicity analysis where peaks are considered potential epitopic regions (Jameson and Wolf, 1988; Garnier et al., 1978; Chou and Fasman, 1978; Engelman et al., 1986).

antibody bleeds 3 and 4 in rabbits A and B. The antibodies were received and an attempt was made to detect Tbj47 expressed in *E. coli* (data not shown). This was unsuccessful on numerous occasions as the lack of expression of Tbj47 was consistent throughout the present study.

3.3.10. Development of a polyclonal anti-Tbj1 antibody in rabbits

Recombinant Tbj1 was successfully expressed and purified and a different approach to antibody production was adopted in this case. The Tbj1 expression and purification was optimized extensively (described in Section 3.3.5) in order to obtain the purest protein product possible and the purified protein was dialysed and filtered in order to remove residual imidazole, beads and protein aggregates. Purified Tbj1 was subsequently submitted to Prof. D. Bellstedt‟s research group (Stellenbosch University, South Africa) where it was used to inoculate a rabbit, and in so doing, generate a specific polyclonal antibody. The antibody produced was tested as described in Section 3.3.11.

3.3.11. Testing of the Tbj1 antibody on Tbj1 and Tcj1

The rabbit polyclonal anti-Tbj1 immune serum and pre-immune serum were tested on *E. coli* [pQE-1], [pQE1-Tbj1] and [pET23b-Tcj1] lysates (Figure 3.13) in differing concentrations to determine whether the antibody was successfully produced and whether or not any cross-reactivity with *E. coli* proteins, particularly DnaJ was detected. The pre-immune serum did not cross-react with any proteins in *E. coli* [pQE1-Tbj1] lysates, and the anti-Tbj1 immune serum did not cross-react with or detect any proteins in *E. coli* [pQE-1] lysates (Figure 3.13 A). Therefore, these controls suggested that there were no pre-existing antibodies that cross-reacted with Tbj1 and that the anti-Tbj1 antibodies did not cross-react with any *E. coli* proteins.

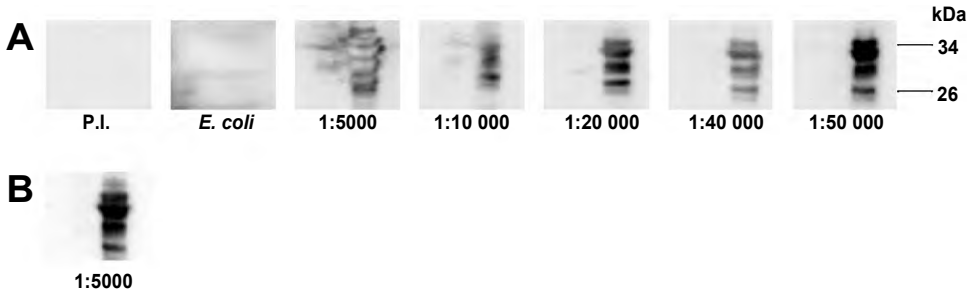


Figure 3.13. Testing of anti-Tbj1 antibody on Tbj1 and Tcj1-containing protein samples

The anti-Tbj1 polyclonal antibody produced was tested on Tbj1 [A] and Tcj1-containing [B] total protein samples. [A] – Western analysis with pre-immune serum (P.I) at a 1:5000 dilution and the 28 day immune serum in differing concentrations where the concentration of primary antibody is given below each figure. *E. coli* refers to the probing of an *E. coli* lysate not expressing Tbj1 with anti-Tbj1 antibody at a 1:5000 dilution; [B] – Detection of Tcj1 with the Tbj1 antibody, at a dilution of 1:5000. The approximate molecular mass (kDa) is given to the right of the figure.

The anti-Tbj1 immune serum successfully detected both Tbj1 and Tcj1 in *E. coli* [pQE1-Tbj1] (Figure 3.13 A) and *E. coli* [pET23b-Tcj1] lysates (Figure 3.13 B) at a range of differing concentrations, and a dilution of 1:5000 was subsequently utilized for convenience (Figure 3.13 A and B).

3.3.12. Detection of Tbj1 and Tbj47 in *T. brucei* lysates

T. brucei lysates were probed with specific anti-TbHsp40 antibodies in order to ascertain which life cycle stage of the parasite Tbj1 and Tbj47 occur in (Figure 3.14). Tbj47 has previously been detected in the *T. brucei* BSF by means of RNAi analysis (Subramaniam *et al.*, 2006) but no data was found on the stage-specific expression of Tbj1. *T. brucei* Pro and BSF lysates were successfully subjected to Western analysis for both Tbj1 (Figure 3.14.B) and Tbj47 (Figure 3.14.C). Tbj47 was detected in the BSF lysate which conforms to what was obtained from literature. Interestingly, Tbj1 also appeared to be expressed in the *T. brucei* BSF. The most important consideration in terms of the detection of Tbj1 in *T. brucei* lysates was the potential detection of a number of Hsp40

proteins because of the conserved nature of the J-domain of Tbj1. A number of TbHsp40s were shown to have a sequence similarity with Tbj1, but only over the J-domain region (data not shown).

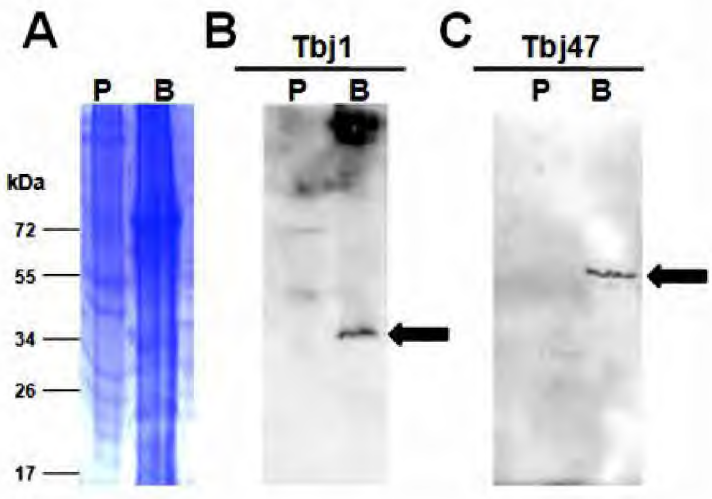


Figure 3.14. Tbj1 and Tbj47 are expressed in the *T. brucei* bloodstream form.

***T. brucei* procyclic and bloodstream form lysates were probed with antibodies specific to Tbj1 and Tbj47. [A] – Procyclic (P) and bloodstream form (B) *T. brucei* lysates resolved on a 12 % SDS-PAGE gel stained with Coomassie; [B] - *T. brucei* P and B lysates probed with anti-Tbj1 antibody (1:5000), the black arrow indicates a band detected at 32-35 kDa in the B sample lane; [C] - *T. brucei* P and B lysates probed with anti-Tbj47 antibody (1:5000), the black arrow indicates a band detected at 50-55 kDa in the B sample lane.**

The absence of multiple bands in the *T. brucei* bloodstream form fraction probed with the anti-Tbj1 antibody produced during the course of the present investigation and the presence of a band corresponding to a molecular mass of 35 kDa (Figure 3.14; B, lane B) would appear to indicate that Tbj1 is being expressed in the bloodstream form of the *T. brucei* parasite. Correspondingly, the absence of a band corresponding to a molecular mass of 35 kDa in the *T. brucei* procyclic form could serve to indicate that Tbj1 is either not expressed in the *T. brucei* procyclic form, or that it is expressed in levels that are undetectable by means of Western analysis (Figure 3.14; B lane P). A single

band corresponding to a molecular mass of 52-53 kDa was detected in the BSF form of *T. brucei* when probed with the anti-Tbj47 antibody produced during the course of the present work (Figure 3.14; C lane B). This is proposed to correspond to Tbj47 and, the absence of such a band in the procyclic form of the parasite (Figure 3.14; C lane P) would appear to suggest that either Tbj47 levels are not detectable in the procyclic form of *T. brucei*, or that Tbj47 is only expressed in the BSF of the parasite.

In terms of general antibody design for the purposes of future work, the results obtained for both Tbj1 and Tbj47 are encouraging in that very little background signal was obtained in the *T. brucei* Pro and BSF lysates, which would indicate that the anti-Tbj1 and anti-Tbj47 antibodies produced in this study would be suitable for future work involving RNAi and two-dimensional gel electrophoresis in the *T. brucei* parasite. The levels of sensitivity and reproducibility obtained with these antibodies are also encouraging for future work.

3.4. DISCUSSION AND CONCLUSIONS

3.4.1. Cloning of the coding regions of three previously uncharacterised TbHsp40 proteins into the pQE-1 expression vector

The successful amplification from *T. brucei* genomic DNA and cloning of coding regions of three novel TbHsp40s has been outlined in the present work (Figures 3.8 and 3.10). This is the first time that the cloning of the coding regions of Type III and Type IV TbHsp40s has been reported.

3.4.2. Heterologous expression of Tbj1, a novel Type III TbHsp40

A previously uncharacterised Type III TbHsp40, Tbj1, was successfully expressed in *E. coli* XL1Blue cells. The *T. cruzi* orthologue of Tbj1, Tcj1 was also purified from *E. coli* BL21 cells. Both proteins were overexpressed and purified in quantities of between 0.6 – 0.9 mg per purification, at levels of purity that exceeded 80 %. The purification protocol developed during the course of this study was robust and highly reproducible and resulted in a maximum loss of contaminating proteins with a compromise in terms of loss of the protein of interest. In all cases, enough protein of sufficient quantity for further biochemical characterisation was obtained.

Although a concerted effort was made to eliminate residual endogenous DnaK contamination from the Tbj1 and Tcj1 purified protein samples by adding ever increasing concentrations of ATP to the purification wash buffers, this was not completely successful as evidenced by residual contamination detected by anti-DnaK antibodies in the final Tcj1 elutions (Figure 3.10). It is proposed that the tendency for Tcj1 to bind strongly to TcHsp70 could potentially be due to denatured exposed surfaces of the protein being able to interact with DnaK by acting as a substrate to the protein. The use of an *E. coli* DnaK knockout strain (*E. coli* BB1994) was initially attempted for the overexpression of all proteins in

the present study (data not shown), but was disregarded as a viable option as the low levels of expression of Tbj1 in this strain resulted in reduced feasibility in a study where large quantities of protein were desired for biochemical characterisation. An Hsp70 protein, TcHsp70, was also successfully expressed from a construct that was previously cloned and the protein was purified in sufficient quantities for further biochemical characterisation.

In contrast to the expression obtained for Tbj1, no expression was obtained for Tbj47 and Tbj51. As several factors complicate successful heterologous expression in a system like *E. coli*, the cause for this lack of expression was not immediately apparent, especially in light of the fact that the sequences for both coding regions have been verified (Appendix H). It is possible that the expression of Tbj47 and Tbj51 require factors not present in the *E. coli* bacterial protein expression machinery, and that these factors may be found in an alternative expression system such as yeast or an insect cell line. Success in expressing and purifying non-glycosylated transferrin receptor in SF9 insect cells (Maier and Steverding, 2008) has opened up a new avenue in the heterologous expression of proteins that cannot be expressed within the standard *E. coli* expression systems. However, because the yeast expression system is a well characterised eukaryotic expression system, future studies will investigate whether it will be sufficient to facilitate the overexpression of these two proteins before attempting to use other systems.

Because the probability of successfully producing soluble protein in a heterologous system decreases significantly for proteins that are larger than 60 kDa (Gräslund *et al*., 2008) it is proposed that the expression of Tbj51 (80.4 kDa) could potentially prove to be problematic in any heterologous system and that the focus on future studies should potentially be limited to *in vivo* analyses that can be performed with the aid of a suitable antibody.

3.4.3. Detection of Tbj1 and Tbj47 in *T. brucei* lysates

Polyclonal antibodies to Tbj1 and Tbj47 were developed. The successful detection of Tbj1 and Tbj47 in *T. brucei* lysates paves the way for future study of these two proteins in the parasite. Tbj47 has previously been detected in *T. brucei* BSF lysates in RNAi analyses (Subramaniam *et al*., 2006) which has been confirmed in the present study. It is interesting to note that Tbj1 was also detected in the BSF of the parasite, but no major band was detected in the procyclic form. This is the first time that Tbj1 has been successfully detected in a *T. brucei* lysate, and this merits attention for future studies of the protein, especially in light of the availability of a suitable antibody that appears not to cross react with other proteins in a *T. brucei* BSF lysate. The BSF of the *T. brucei* lifecycle takes place in the warm-blooded mammalian host, and it is proposed that a number of the highly specialized Type III and Type IV TbHsp40s are predominantly upregulated during this lifecycle stage in order to facilitate the survival of the parasite in the mammalian bloodstream. The use of RNAi and two-dimensional gel electrophoresis studies could greatly enhance the preliminary work begun in the present chapter, and are being proposed for future studies, as described in Chapter 5.

3.4.4. Conclusions

The present work outlines the first known study on the expression and purification of a Type III Hsp40 protein from the *T. brucei* parasite. It has been demonstrated that amplification from genomic DNA is sufficient to produce an accurate protein coding region for the purposes of expression and purification of proteins from *T. brucei*.

CHAPTER 4

Biochemical characterisation of Tbj1 using *in vivo* and *in vitro* assays

4.1. INTRODUCTION

4.1.1. The use of *in vivo* complementation assays in the study of Hsp40-Hsp70 interactions

In vivo complementation assays rely on the ability of an Hsp40 of interest to functionally replace an Hsp40 that has been knocked out in a specialised strain of *E. coli* or yeast. The ability of a given Hsp40 to functionally replace the Hsp40 that has been knocked out in the deficient strain is determined by the ability of the strain to recover from the thermosensitivity that is characteristic of Hsp40 knock out strains and of growth rescue to occur (Edkins *et al*., 2004; Nicoll *et al*., 2007).

The *E. coli* OD259 (*dnaJ*$^{-}$ *cbpA*$^{-}$) thermosensitive strain is frequently used for *in vivo* complementation assays (Nicoll *et al*., 2007). Certain well-conserved Hsp40 proteins such as *Agt* DnaJ are known to be able to reverse the thermosensitivity of *E. coli* OD259 and allow for growth of this strain at 40°C. If an Hsp40 or and Hsp40 chimera are able to functionally replace curved DNA-binding protein A (CbpA) and DnaJ in this thermosensitive strain, then it can be concluded that the protein / chimera in question is functionally equivalent to *E. coli* DnaJ and CbpA (Nicoll *et al*., 2007). CbpA is a Type II Hsp40 protein from *E. coli* that was first described as being an analogue of DnaJ (Ueguchi *et al*., 1994).

The positive control for such an experiment is frequently exogenously expressed *Agt* DnaJ, while the negative control is *Agt* DnaJ-H33Q, because substitution of His^{33} of the HPD motif of *Agt* DnaJ abrogates interaction of the protein with DnaK (Hennessy *et al*., 2005a). Previous studies have shown that a number of Hsp40 chimeras comprising the *Agt* DnaJ protein with diverse J-domains (including those originating from Type III Hsp40s) in place of the *Agt* Dna J-domain were able to functionally replace DnaJ and CbpA in this system, irrespective of their source (Nicoll *et al*., 2007).

4.1.2. The importance of *in vitro* biochemical characterisation of novel proteins

While the study of proteins is only complete upon thorough *in vivo* study, valuable insights into the mechanisms of protein interactions and functions may be obtained by means of careful *in vitro* analyses. Several well-optimised assays are widely used at present to biochemically characterise molecular chaperones, including oligomerisation analysis, aggregation suppression assays and ATPase stimulation assays (Matambo *et al.*, 2004; Boshoff *et al.*, 2008; Nicoll *et al.*, 2006; Schlieker *et al.*, 2002). These techniques provide information on the dimerisation status and chaperoning ability of a given molecular chaperone and often form the cornerstone of further studies on novel proteins (Nicoll *et al.*, 2006). An overview of each of these biochemical assays in chaperone research is presented here.

4.1.3. Oligomerisation analysis using SEC-FPLC

Size exclusion chromatography coupled to fast protein liquid chromatography (SEC-FPLC) allows for the rapid analysis of proteins in solution and can assist in the determination of protein sample purity as well as the predicted molecular mass of a given protein when compared to the retention time of known standards resolved on a SEC-FPLC system (Lathe and Ruthven, 1955). A number of applications for this technique are widely used in protein research, viz. buffer exchange of proteins, and the separation of proteins in purification steps, or the additional purification of a protein that has already been purified (Kunji *et al.*, 2008). The separation of oligomers or aggregates from a purified protein sample is another widely-used application of SEC-FPLC (Kunji *et al.*, 2008).

The study of the oligomeric state of a protein is an important part of any preliminary biochemical characterisation of such a protein. Oligomerisation, and particularly

dimerisation plays a key role in the activity of a number of molecular chaperones. DnaJ has been shown to be a homodimer, and its dimerisation ability is facilitated by the C-terminal 46 amino acids of the protein (Shi *et al*., 2005). The dimerisation of DnaJ is essential for its chaperone activity *in vivo* (Shi *et al*., 2005). Sis 1 from yeast has also been shown to exist as a homodimer (Sha *et al*., 1999). As was the case for DnaJ, the dimerisation of Sis1 plays an important role in its function. Monomeric forms of Sis1 are able to stimulate the ATPase activity of Hsp70, but are unable to assist Hsp70 in the refolding of a denatured substrate (Sha *et al*., 2000). The Type III Hsp40, DjIa from *E. coli*, has previously been shown to dimerise (Toutain *et al*., 2003) although it is not clear whether dimerisation plays a key role in the function of the Type III Hsp40s as a group.

4.1.4. Aggregation suppression assays

In vitro aggregation suppression assays are commonly used to determine whether a chaperone protein of interest possesses the ability to prevent the misfolding of a given model substrate in solution (Basha *et al*., 2004; Schlieker *et al*., 2002; Nicoll *et al*., 2006). A number of different model substrates have been employed in aggregation suppression assays in order to determine the extent of the *in vitro* activity of a given molecular chaperone protein (Schlieker *et al*., 2002). The model substrate used is denatured by means of increased temperature or by the addition of urea / guanidium-HCl and the progress of aggregation measured by means of a change in absorbance (often at 360 nm) (Daugherty *et al*., 1998; Nicoll *et al*., 2006). Molecular chaperones are able to act on aggregating substrates either by preventing the substrate from aggregating or by reversing the aggregation process and allowing the aggregated substrate to refold into its optimal native conformation (Minami *et al*., 1996). Malate dehydrogenase (MDH) is a thermolabile model substrate that is aggregation prone and is frequently used in aggregation suppression assays (Boshoff *et al*., 2008; Goloubinoff *et al*., 1999).

A number of molecular chaperones have previously been shown to be able to bind denatured proteins *in vitro* and prevent their aggregation (Ben-Zvi *et al*., 2001). It is important to note that a distinction exists between the ability to prevent aggregation, and the ability to allow for active refolding of a denatured protein substrate; the latter ability belongs exclusively to chaperones that possess ATPase activity, viz. the DnaK / Hsp70 family of molecular chaperones and is studied using refolding assays as opposed to aggregation suppression assays (Ben-Zvi *et al*., 2001). Hsp70s are able to suppress the aggregation of proteins in the absence of any co-chaperone proteins (Boshoff *et al*., 2008; Shonhai *et al*., 2008). Type I Hsp40s have been shown to possess independent chaperone activity and are able to work in an Hsp70- dependent or independent manner; these proteins are also able to suppress the aggregation of a model substrate in the absence of an Hsp70 partner protein (Borges *et al*., 2005). Type II Hsp40s, which lack the Cys-rich domain of the Type I Hsp40s do not possess autonomous chaperone activity and are unable to suppress the aggregation of proteins in the absence of a partner Hsp70 protein (Borges *et al*., 2005). The C-terminal domain of Type I and Type II Hsp40s are involved in binding of substrate, but it has been shown that despite being structurally similar, these Hsp40 subtypes differ markedly in their ability to suppress the aggregation of a model substrate (Borges *et al*., 2005).

Because Type III Hsp40s do not possess a substrate binding domain similar to that of Type I and Type II Hsp40s, it has been proposed that these proteins are unable to bind to non-native polypeptides and that they are subsequently unable to function as autonomous molecular chaperones and suppress the aggregation of proteins in the absence of an Hsp70 partner protein (Gur *et al*., 2004; Qiu *et al*., 2006). Certain Type III Hsp40s are able to assist an Hsp70 protein in the refolding of denatured substrates, and in the solubilisation of aggregated proteins (Gur *et al*., 2004). In addition to the considerations above, studies on aggregation suppression of model substrates need to consider that the ability of a given molecular chaperone to suppress the aggregation of a model substrate depends in part on the substrate itself. No individual chaperone has been found that is able to prevent the aggregation of any substrate; in fact, some

chaperone families possess enhanced abilities to suppress the aggregation of certain substrates, while being unable to do so with others (Schlieker *et al*., 2002).

4.1.5. ATPase assays: determining the ability of Hsp40s to stimulate the ATPase activity of Hsp70 proteins

The ability of Hsp70 proteins to bind suitable substrates is largely dependent on their ATPase cycle (Hennessy *et al*., 2000; Laufen *et al*., 1999). Despite the fact that Hsp70 proteins have a weak intrinsic ATPase activity, they are largely reliant on partner Hsp40s for their *in vivo* function (Gässler *et al*., 2001; Russell *et al*., 1999; Hennessy *et al*., 2000). ATPase assays often form a useful starting point for the biochemical characterisation of both Hsp40 and Hsp70 proteins (Hennessy *et al*., 2004; Nicoll *et al*., 2006; Hennessy *et al*., 2005). Several methods of determination of ATPase activity of Hsp70 proteins rely on the detection of the levels of phosphate that are released when ATP is cleaved, and a number of colorimetric assays to facilitate this have been developed (Olson *et al*., 1994; Blond-Elguindi *et al*., 1993; Terada and Mori, 2000). The ATPase assay method of Lanzetta et al. (1979) has been used in a number of biochemical characterisation studies, and has been improved and optimized by Lill et al. (1990) and recently in our research unit (M.H. Ludewig, Rhodes University, Pers. Comm.) The original ATPase assay developed by Lanzetta and colleagues utilizes a citrate / arsenite mixture that is added after the molybdate reagent, which allows for enhanced stability and does not detect any newly released phosphate after reactions have been terminated (Lanzetta *et al*., 1979). A modification of this method was developed by Chifflet and colleagues in 1988 (Chifflet *et al*., 1988). This method is characterised by improved sensitivity, so that below 10 nmol quantities of phosphate can successfully be detected, in addition to robustness and low levels of interference by concentrated solutions of different origins, e.g. human serum (Chifflet *et al*., 1988). This assay has recently undergone extensive optimization in our own research group, and the colour stability and sensitivity has been improved by the addition of ascorbic acid to the ammonium molybdate and sodium citrate solutions in the colour development reaction (M.H. Ludewig; Pers. Comm.).

Despite the poor levels of primary amino acid sequence conservation displayed by the Type III Hsp40s, their ability to stimulate the ATPase activity of partner Hsp70 proteins appears to be uncompromised providing that certain key residues in the J-domain implicated in interactions with Hsp70 proteins are conserved (Genevaux *et al*., 2001). An example of this is the Type III Hsp40 from *E. coli*, Djla, which was found to be capable of stimulating the ATPase activity of DnaK in a similar fashion to what was observed for the Type I Hsp40, DnaJ (Genevaux *et al*., 2001). Another Type III Hsp40, $p58^{IPK}$ has also been shown to be able to stimulate the ATPase activity of Hsp70 (Melville *et al*., 1997). Interestingly, Tcj1 from *T. cruzi* has been shown to be unable to do so (Edkins *et al*., 2004).

The low levels of sequence conservation in Type III Hsp40 proteins preclude the prediction of a generalised mode of action or mechanism with respect to the stimulation of the ATPase activity of Hsp70s, and the lack of conservation of key residues in the J-domains of these proteins suggest that they may not functionally associate with a partner Hsp70 protein *in vivo*, as is typically observed for Type I and Type II Hsp40s.

4.1.6. Previous biochemical characterisation of Tcj1

The Type III Hsp40 from *T. cruzi*, Tcj1 has been previously expressed and partially biochemically characterised (Edkins *et al*., 2004). It was first identified and cloned by Tibbetts and colleagues, and characterised as a divergent Hsp40 protein that does not contain a G / F –rich region or a zinc finger domain (Tibbetts *et al*., 1998). Edkins showed that Tcj1 was not able to stimulate the ATPase activity of the major cytosolic Hsp70 protein from *T. cruzi*, TcHsp70 at substoichiometric levels (Edkins *et al*., 2004). Tcj1 was also shown to possess no significant ATPase activity on its own, or in the presence of an artificial substrate such as Reduced carboxymethylated alpha-lactalbumin (RCMLA) (Edkins *et al*., 2004). Aggregation suppression studies have not been performed using Tcj1 and potential partner Hsp70 proteins, hence it is not known whether or not this protein would be able to facilitate the suppression of aggregation of

a model substrate, or assist in its refolding. No biochemical characterisation has been reported for the *T. brucei* ortholog of Tcj1, Tbj1 and the current work presents the first preliminary data on the biochemical properties of this novel Type III TbHsp40 protein.

4.1.7. AIMS AND OBJECTIVES

The investigations outlined in the present work aimed to:

- determine whether Tbj1 was able to functionally replace DnaJ and CbpA in a thermosensitive strain of *E. coli*, and in so doing restore growth at heat shock temperatures by interacting with DnaK
- determine the dominant state (monomeric or dimeric) of Tbj1 and Tcj1 in solution by means of analysis with fast protein liquid chromatography (FPLC)
- determine whether Tbj1 and Tcj1 were able to suppress the aggregation of a model substrate (MDH) in isolation, or in partnership with TcHsp70 and a commercial Hsp70 protein from *Medicago sativa* (*M. sativa* Hsp70)
- determine whether there is a difference in the ability of TcHsp70 and MsHsp70 to suppress the aggregation of a model substrate (MDH)
- determine whether Tbj1 and Tcj1 are able to stimulate the ATPase activity of TcHsp70 and MsHsp70

4.2. MATERIALS AND METHODS

4.2.1. Complementation assays

Competent cells of a temperature sensitive strain of *E. coli*, the OD259 strain (MC4100 *ara*D139 Δ*ara* 714 Δ*cbpA*::*kan dnaJ*::Tn10-42) were prepared using the general competent cell preparation protocol described in Appendix A.1. The heat shock step during the competent cell preparation procedure was performed at 37°C due to the thermosensitivity of the strain. The *E. coli* OD259 strain was previously known as the *E. coli* WKG190 strain and is sensitive to temperatures below 16°C and above 37°C (Kelley and Georgopoulos, 1997; Deloche *et al*., 1997). Freshly prepared *E. coli* OD259 cells were transformed with pQE-1 (vector control), pQE1-Tbj1, pRJ30-Pfj4-J (positive control encoding Pfj4-*Agt*-DnaJ), and pRJ30-H33Q (negative control encoding *Agt*-DnaJ-H33Q) (Nicoll *et al*., 2007).

A single colony from each fresh transformation reaction was used to inoculate an overnight 2 x YT starter culture (5 ml) containing 100 µg / ml ampicillin and 50 µg / ml kanamycin. These cultures were allowed to grow overnight (typically 14 hours) at 30°C. The overnight cultures were diluted (1:10) and used to inoculate 2 x YT cultures that were grown until the $_{A600}$ reached 2.0 (approximately 3 hours). The cultures were diluted to an $_{A600}$ of 0.2 to ensure uniform cell density with respect to the controls and the experiment. The diluted cultures were subsequently utilised to prepare a serial dilution series ($10^{0} - 10^{-10}$) for each culture (controls and experiment) using sterile 2 x YT broth. An aliquot of each dilution (2 µl) was spotted onto 2 x YT plates containing 100 µg / ml ampicillin, 50 µg / ml kanamycin and 50 µg / ml IPTG in order to induce protein expression. An OD259 strain control was performed in which *E. coli* OD259 cells were spotted onto 2 x YT agar containing 50 µg / ml kanamycin in order to determine whether the strain used was functional.

Replicate plates were prepared and sealed with parafilm to prevent dessication prior to incubation at 5 different temperatures (30°C, 37°C, 38.5°C, 40°C and 42°C) for a period of 12-14 hours. The plates were examined and images of the results obtained were captured using a digital scanner. Three independent replicates were performed for each experiment.

4.2.2. Heterologous expression and purification of Tbj1 , Tcj1 and TcHsp70

Tbj1, Tcj1 and TcHsp70 were purified as previously described (Chapter 3; Sections 3.2.2.2. – 3.2.2.5.). Three elutions (1.5 ml) were obtained per batch purification resulting from a 225 ml culture. Once protein identity and integrity had been confirmed protein elution fractions were pooled and dialysed against dialysis buffer (50 mM Tris-HCl, 200 mM NaCl; pH 8.0) at 4°C for a period of 5-6 hours. The protein samples were quantified using the Bradford assay (Appendix A.14) (Bradford, 1976) both prior and subsequent to dialysis and were concentrated down using PEG20 000 if necessary. A final protein concentration of 0.5 mg / ml was maintained where possible in order to facilitate dilutions for *in vitro* characterisation. Three independent protein purification reactions were performed for each investigation, and the success of each purification was confirmed by means of SDS-PAGE and Western blotting in order to ensure protein integrity. The original plasmid construct which was used to transform *E. coli* before each purification was also subjected to diagnostic digests to ensure that the correct Hsp40-encoding plasmid had been transformed prior to protein purification. Proteins were stored at 4°C for a maximum of 2 days prior to use, although both Tbj1 and Tcj1 have been shown to be stable at this temperature for up to 2 weeks, by displaying no visible differences in behaviour with respect to biochemical assays.

4.2.3. Gel filtration chromatography

Gel filtration chromatography was performed on Tbj1 and Tcj1 using a SuperdexTM 20 HR 10 / 30 size-exclusion column (Amersham Pharmacia Biotech; U.S.A.). The column has a separation range for proteins from 10 – 600 kDa. The mobile phase comprised a standard Tris-HCl buffer (50 mM Tris, 150 mM NaCl; pH 7.5) that was filtered and degassed for 45 minutes prior to use, and the flow rate was set at 0.5 ml / min while a constant pressure below 0.25 kPa was maintained for the duration of the analysis. Fractions were collected until a full column volume of buffer had been eluted through the system (approximately 40 minutes each). Proteins of interest were detected at 280 nm. Several molecular mass standards were resolved using the same parameters as described above in order to experimentally determine the molecular mass of the proteins of interest. These include ferritin (450 kDa), catalase (225 kDa), bovine serum albumin (66 kDa), ovalbumin (45 kDa) and lysozyme (14.4 kDa). The column was washed prior to the application of a new protein by passing three column volumes of 0.5 M NaOH and two column volumes of the mobile phase buffer through until a zero reading was obtained on the A_{280}, A_{215} and A_{245} graphs indicating any contaminating proteins that bound were eluted from the column.

The fractions corresponding to a peak on the chromatogram were retained and added to 5 volumes of ice cold acetone and stored at -20°C overnight prior to pelleting of precipitated protein by means of centrifugation at 1000 *g* for 10 minutes at 4°C. The supernatant was discarded and samples were air dried at 37°C for 30 minutes prior to resuspension of the pellets in SDS-PAGE loading buffer (0.0625 M Tris-HCl, pH 6.8, 10 % glycerol v/v, 2 % SDS w/v, 5 % β-mercaptoethanol v/v, 0.05 % bromophenol blue w/v). The resuspended samples were boiled for 10 minutes and resolved on a 12 % SDS-PAGE gel as described previously (Chapter 3; Section 3.2.2.6.) and analysed using Western blotting (Chapter 3; Section 3.2.2.6.) in order to confirm protein identity correlating to a given peak on the chromatogram.

4.2.4. Aggregation suppression assays

An MDH (Roche; Germany) standard was prepared in MDH assay buffer (50 mM Tris-HCl; 100 mM NaCl; pH 7.4). The assay buffer was equilibrated to 48°C in a Helios AlphaDB spectrophotometer with a Peltier-controlled cell before MDH (0.72 µM) was added and its aggregation measured by following the absorbance increase at 360 nm over a 30 minute period (Boshoff *et al*., 2008). The ability of Tbj1, Tcj1, TcHsp70 and MsHsp70 to suppress the aggregation of MDH was tested separately and in combination in differing concentrations. Initial studies focussed on equimolar concentrations of MDH and the Hsp40 / Hsp70 protein of interest but lower concentrations were also utilised in subsequent investigations. Prior to commencement of the assay, an independent study was performed on each chaperone in order to determine whether they would cause an increase in turbidity if incubated alone in heated assay buffer for 30 minutes. None of the proteins studied displayed any increase in turbidity at 360 nm, indicating that these chaperones were not prone to aggregation. The data set obtained for all assays was plotted on an Excel® spreadsheet, and the aggregation of MDH over six independent replicates was taken as the basal aggregation value for all subsequent calculations. BSA (0.72 µM) was added to MDH under the same reaction conditions outlined above as a negative control reaction, while a Type I Hsp40 from *T. cruzi*, Tcj2 was used as a positive control (kindly provided by M.H. Ludewig, Rhodes University).

4.2.5. Determination of the ability of Tbj1 and Tcj1 to stimulate the ATPase activity of TcHsp70 and MsHsp70

Purified, dialysed, concentrated and quantified Tbj1 and Tcj1 were studied with respect to their ability to stimulate the ATPase activity of two different Hsp70 proteins, TcHsp70 and MsHsp70. A modified protocol of the ammonium molybdate / ascorbic acid method of detecting ATP hydrolysis was used (Chifflet *et al*., 1988; M.H. Ludewig, Pers. Comm.). Reaction mixtures set up in triplicate (1000 µl) were comprised of 10 mM HEPES, pH 7.4; 10 mM $MgCl_2$, 20 mM KCl and 0.5 mM DTT (prepared fresh). Tbj1,

Tcj1, TcHsp70 and MsHsp70 were added to the reaction mixture (0.4 µM) where required, depending on what was being assayed in a given reaction mixture. Tbj1 and Tcj1 were tested for their ability to stimulate the ATPase activity of TcHsp70 and MsHsp70 respectively, while the basal ATPase activity of both Hsp70 proteins in the absence of an Hsp40 protein was assessed. The reaction mixtures were incubated at 37°C for 5 minutes prior to the addition of ATP (600 µM) in order to initiate the reaction. Samples (50 µl) were taken in triplicate at different time points (0, 30, 60, 90, 120, 150, 180, 210, 240, 270, 300 minutes) and added to stop solution (10 % SDS; 50 µl) in order to cease further ATP hydrolysis. Control reactions containing each protein to be assessed (0.4 µM) without ATP were also set up in tandem with each experiment. After the samples were taken, the reactions were developed with ammonium molybdate (1 % in 1N HCl; 50 µl), ascorbic acid (6 %; 50 µl) and sodium citrate / acetic acid solution (2 % / 2 %; 125 µl) for 30 minutes at 37°C. The absorbance of the reactions was measured at 850 nm in a KC Junior microplate reader. A phosphate standard curve was prepared from KH_2P0_4 with standards ranging from 0 nmol – 10 nmol and a standard curve was prepared (Appendix B2) to serve as a basis for calculations of phosphate content in the reactions assayed. The specific ATPase activity was calculated for each reaction and reported as nmol Pi / min / mg protein. A negative control containing TcHsp70 or MsHsp70 that had been inactivated by boiling for 20 minutes was prepared as outlined above. Both TcHsp70 and MsHsp70 were assayed in the presence and absence of ATP. Tbj1 and Tcj1 were both assayed in the presence of ATP and the absence of Hsp70 in order to determine if either protein is prone to spontaneous phosphate release. A further control reaction containing only ATP was used in order to correct for the effect of spontaneous ATP hydrolysis in the system for all subsequent assays.

4.3. RESULTS

4.3.1. Tbj1 is unable to functionally replace DnaJ and CbpA in thermosenstive *E. coli* OD259 cells

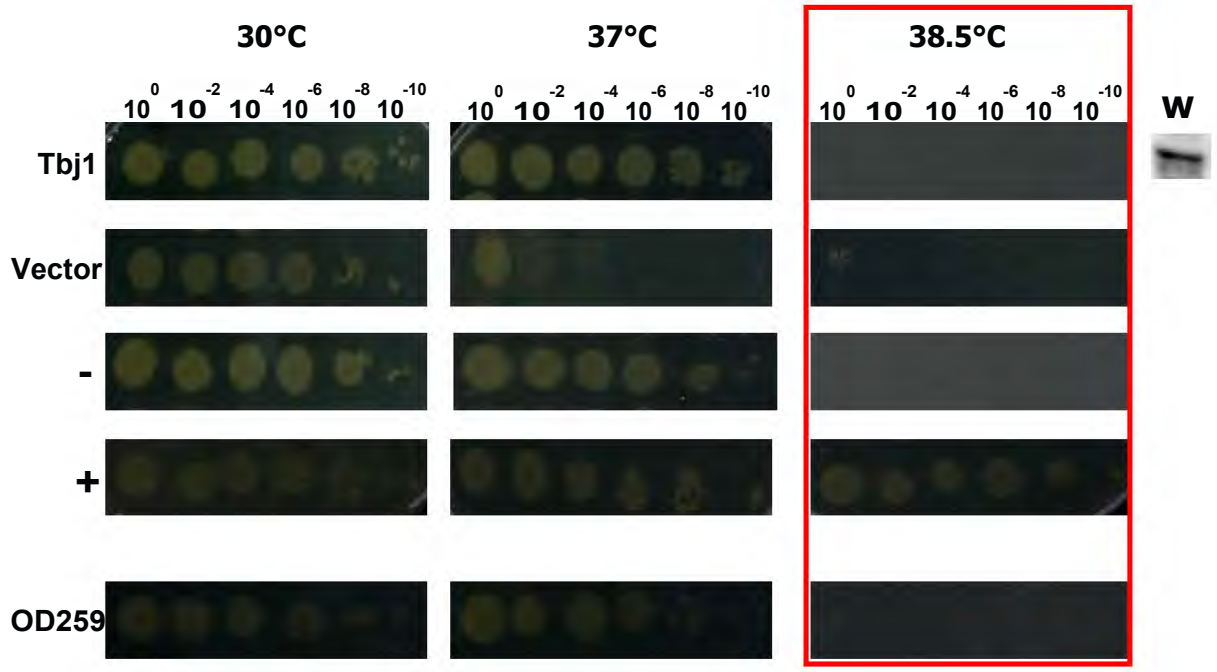


Figure 4.1. Full length Tbj1 is unable to reverse the thermosensitivity of E. coli OD259 cells

The results of the in vivo complementation assay performed with E. coli OD259 [pQE1-Tbj1]. The temperature at which each plate was incubated is given above each panel, and the heat-shock temperature (first temperature at which loss of growth was experienced) is outlined in red. The dilution range of the assay is given above the figure. [Tbj1] – E. coli OD259 [pQE1-Tbj1 cells]; [Vector] – E. coli OD259 [pQE1]; [-] – E. coli OD259 expressing Agt DnaJ-H33Q; [+] – E. coli OD259 [pRJ30-Pfj4-J] cells (Nicoll et al., 2007); [OD259] – E. coli OD259 cells with no added plasmid vector on 2 x YT Kan plates; [W] - Tbj1 band detected in E. coli lysate with anti-Tbj1 antibody.

The *in vivo* complementation assay study of Tbj1 revealed that the protein was unable to reverse the thermosensitivity of *E. coli* OD259 cells and thus could not functionally replace DnaJ and CbpA in *E. coli* OD259 in this strain. Hsp40 proteins that are able to

suppress the thermosensitivity of *E. coli* OD259 cells allow for growth at temperatures in excess of 37 °C (Nicoll *et al*., 2007). The *E. coli* OD259 cells utilised were unable to grow at temperatures above 37 °C (Figure 4.1.) confirming the thermosensitivity of the strain. The negative control for this assay, *E. coli* OD259 expressing *Agt* DnaJ-H33Q, was unable to rescue growth at increased temperatures, and cells expressing *Agt* DnaJ-H33Q were unable to grow at 38.5 °C. This is in direct contrast to the positive control, Pfj4-J-Agt DnaJ, which, when expressed in *E. coli* OD259 cells was able to rescue growth at temperatures ranging from 38.5 °C (Figure 4.1.) to 42 °C (data not shown). The success of the control reactions, and the behaviour of the *E. coli* OD259 system utilised, indicated that the system in use was thermosensitive, and that the thermosensitivity could be reversed by the addition of a suitable Hsp40 protein that is able to functionally replace DnaJ and CbpA in the strain. The transformation of *E. coli* OD259 cells with the pQE-1 expression vector without any Hsp40 encoding insert resulted in no reversal of thermosensitivity (Figure 4.1.), and cells transformed with pQE-1 alone appeared to experience inhibition in growth at temperatures as low as 37 °C, which is in contrast to the *E. coli* OD259 cells alone as well as the *Agt* DnaJ-H33Q expressing *E. coli* OD259 cells.

Tbj1 was unable to reverse the thermosensitivity of *E. coli* OD259 cells in spite of being expressed at easily detectable levels (Figure 4.1). It was interesting to note that the levels of growth displayed by the *E. coli* OD259 cells expressing Tbj1 that were spotted onto solid media that did not contain the IPTG inducer did not differ in any way from that of cells that were placed on media containing IPTG (data not shown). High levels of basal expression of Tbj1 were found in the uninduced control when probed with anti-His and anti-Tbj1 antibodies; this could indicate that the expression of Tbj1 in *E. coli* OD259 is not tightly regulated. The lack of complementation of Tbj1 could be attributed to the lack of conservation of certain key residues in the Tbj1 J-domain (discussed in Chapter 2) which were predicted to be essential for successful interaction with an Hsp70 partner protein. The absence of a conserved substrate binding domain in Tbj1 is also proposed to play a role in this result.

4.3.2. Tbj1 and Tcj1 purify in a compact monomeric form with some aggregate formation

Size-exclusion chromatography (SEC-FPLC) was performed on purified Tbj1 and Tcj1 (Figure 4.2). The initial peak on the chromatogram indicates the point at which the samples were injected onto the SEC-FPLC column. The chromatogram of Tbj1 (Figure 4.2.A) reveals a prominent peak corresponding to approximately 500 mAU at an elution volume of 20 – 22 ml, corresponding to fractions 24 – 27 that were collected during the course of the experiment and an approximate molecular weight of 40 kDa, which is slightly larger than the predicted size of Tbj1, indicating that the protein is in a slightly less compact form than anticipated from the denaturing gel analysis. The protein in these fractions was confirmed to be Tbj1 by means of Western analysis with anti-Tbj1 antibody after acetone precipitation of selected fractions corresponding to the peak area (Figure 4.2.A; Western blot below chromatogram). The presence of a small quantity of Tbj1 aggregates that eluted at the void volume of the column (ca. 8 ml) was also confirmed by Western analysis (results not shown). Tbj1 was eluted after the BSA standard (66 kDa; 16 ml) and the ovalbumin standard (45 kDa; 18ml) but before the lysozyme standard (14.4 kDa; 22 ml). This indicates that Tbj1 eluted close to the low-end resolution limit of the column used. The result suggests that Tbj1 purifies as a compact monomer, which was anticipated as the analysis in Chapter 2 indicated the absence of a dimerisation domain in the protein.

The same result was obtained for Tcj1 (Figure 4.2.B), which was anticipated due to the similarity in terms of primary sequence and size between Tbj1 and Tcj1. It was interesting to note that the Tcj1 chromatogram also displayed a peak around the void volume of the column, but of a smaller magnitude than that of Tbj1 (Figure 4.2.A). While Western analysis confirmed the identity of Tcj1 in the fractions corresponding to the major peak (ca. 300 mAU), it was shown that the aggregates in this case were not Tcj1 aggregates, but could potentially be aggregates of a protein that co-purified with Tcj1 (data not shown). In spite of the addition of large quantities of ATP to the wash buffers in all Tcj1 purifications, a significant amount of endogenous *E. coli* DnaK was found to co-purify alongside this protein (Chapter 3; Figure 3.10), and it is proposed that

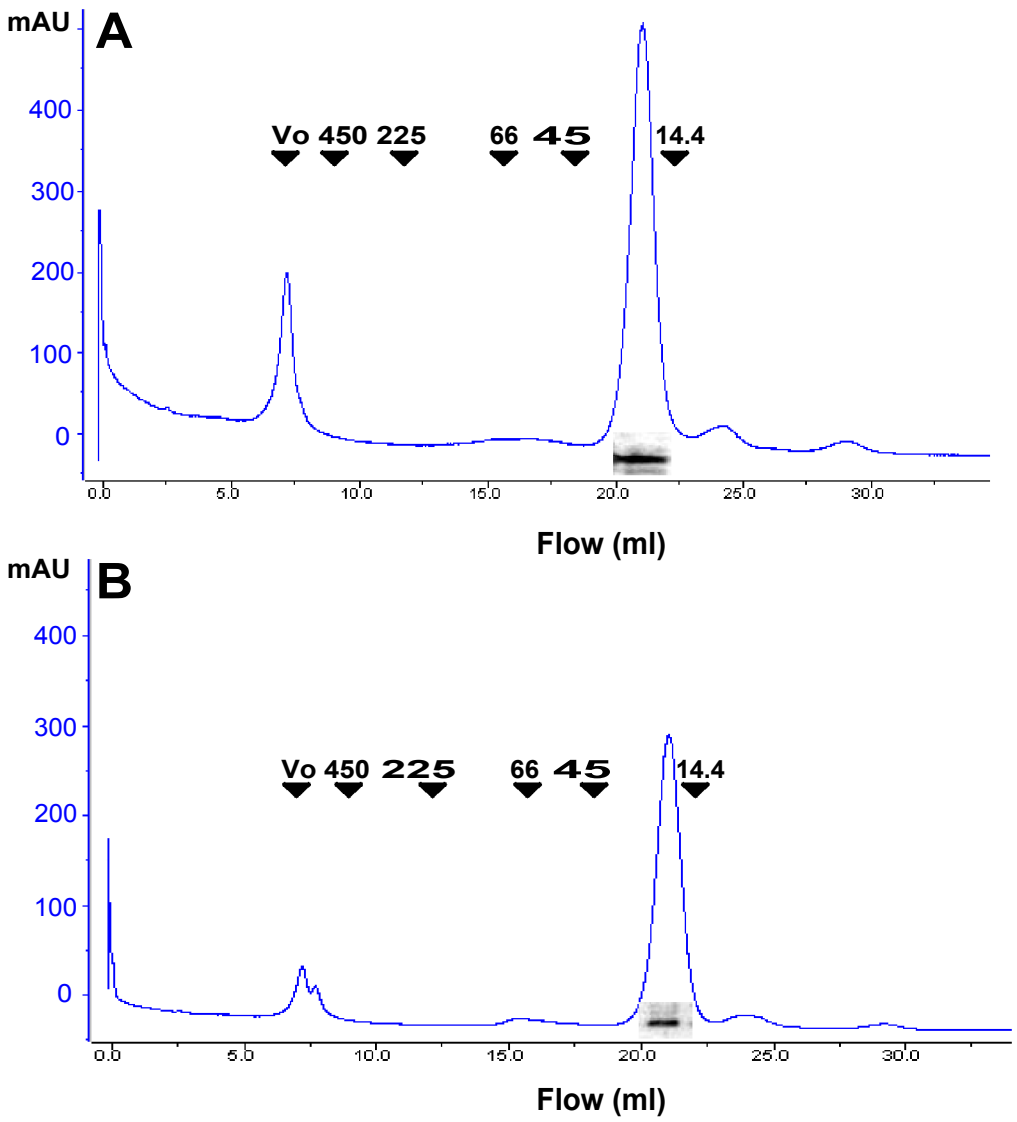


Figure 4.2. SEC-FPLC analysis of Tbj1 and Tcj1

The SEC-FPLC chromatograms of Tbj1 [A] and Tcj1 [B] are shown in blue. The concentration of both proteins prior to application to the SEC-FPLC system was approximately 0.5 mg / ml. The volumes at which the standards were eluted from the column are indicated by black inverted triangles and the molecular mass of each standard (kDa) is given above. Vo indicates the void volume of the column. The following standards were used: ferritin (450 kDa); catalase (225 kDa); BSA (66 kDa); ovalbumin (45 kDa) and lysozyme (14.4 kDa). The identity of the Tbj1 protein was confirmed by means of Western analysis; Tbj1 was detected in a total of 7 fractions with rabbit anti-Tbj1 antibody (1:5000) and donkey anti-rabbit (1:5000) secondary antibody. The Western blots are shown below the peak fraction from which the sample was derived.

the peak that corresponded to an elution volume of approximately 8 ml could potentially consist of aggregated endogenous DnaK or another higher order contaminant that was present in the purified Tcj1 sample. The co-purification of *E. coli* DnaK is a matter of significant concern for any studies involving Tcj1 chaperone function, as the potential Tcj1 – DnaK partnership could lead to invalid results in biochemical assays. This was considered as a potential interfering factor in all biochemical studies on this protein, although it would appear that previous studies on the protein did not focus on the contamination aspect (Edkins *et al.*, 2004). The effect of high levels of endogenous DnaK contamination in purified Tcj1 was taken into account throughout the biochemical assays performed in the present study.

4.3.3. The use of MsHsp70 as a control Hsp70 in the biochemical characterisation of Tbj1 and Tcj1

An Hsp70 protein from the Alfalfa plant (*Medicago sativa*) is commercially available in a highly pure and active form from the Alfa Biogene International company (Germany) (http://www.alfabiogene.de/HSP70 page1.pdf). MsHsp70 has an overall sequence similarity of 76 % with human Hsp70, and possesses an overall functionality that is considered to be highly similar to human Hsp70 with respect to its ability to bind and hydrolyse ATP as well as interactions between the ATPase domain and the substrate binding domain. Most importantly, MsHsp70 possesses the ability to interact with a number of co-chaperones and substrates identical to those employed by human Hsp70. MsHsp70 was chosen as a control Hsp70 protein for the partial biochemical characterisation of Tbj1 and Tcj1 due to its purity and high levels of activity. Due to its high levels of similarity to human Hsp70, MsHsp70 is of interest in studies that involve parasites that reside within the human host at some stage of their lifecycle. The use of a commercial Hsp70 was an added advantage in the present case in order to confirm whether TcHsp70 purified according to the method outlined in Chapter 3 was indeed functionally active.

Bovine Hsc70, which was used as a template for the homology modelling of TcHsp70, and MsHsp70 (Figure 4.3) has an identity of 62.1 % and 61.7 % with these two proteins. The accuracy of the homology models generated for the peptide binding domains of TcHsp70 and MsHsp70 were verified using Procheck software (Laskowski *et al*., 1993; Morris *et al*., 1992). The results obtained from the Procheck analysis served to indicate that the homology models generated were accurate with respect to overall backbone structure, but that the orientation of a number of individual atoms may be different from that found in the most energy minimised conformation. The models were rendered in orientations in which the residues of interest would be most clearly visible.

An analysis of the primary amino acid sequence of MsHsp70 was performed, and a comparison with the sequence of TcHsp70 was analysed (Appendix H; Figure H.2). MsHsp70 and TcHsp70 were found to have a sequence identity of 71 % over their full-length sequences. The ATPase domain of TbHsp70 from *T. brucei* possesses a 95 % identity to that of TcHsp70, and a 78 % identity to MsHsp70. The extreme C-terminal regions of the substrate binding domain differ significantly between the two proteins, indicating that this region could be the centre of specialised functions, or the ability to bind to different partner proteins. The peptide binding region of TbHsp70 was shown to possess a 59 % identity with that of MsHsp70 (data not shown), while the peptide binding regions of TbHsp70 and TcHsp70 were found to be highly similar, with an 82 % identity (data not shown). Most of the binding sites deemed important for Hsp70-Hsp40 interaction to take place are well conserved in both TbHsp70 and TcHsp70, with the notable exception of Met^{404} in DnaK which is Ala^{404} in *T. brucei*, *T. cruzi* and *M. sativa*, and Ala^{429} which is Tyr^{429} in the same.

Both MsHsp70 and TcHsp70 possess a fully conserved C-terminal EEVD region, which is not present in TbHsp70 (Appendix H; Figure H.2). The EEVD motif plays a regulatory role in Hsp40 / Hsp70 function and it is clear that both TcHsp70 and MsHsp70 will display more conventional behaviour towards potential partner Hsp40 proteins than the TbHsp70 protein. Several residues that are predicted to have an important function within DnaK are not conserved in either TbHsp70 or MsHsp70 (Suh *et al*., 1998; Chang

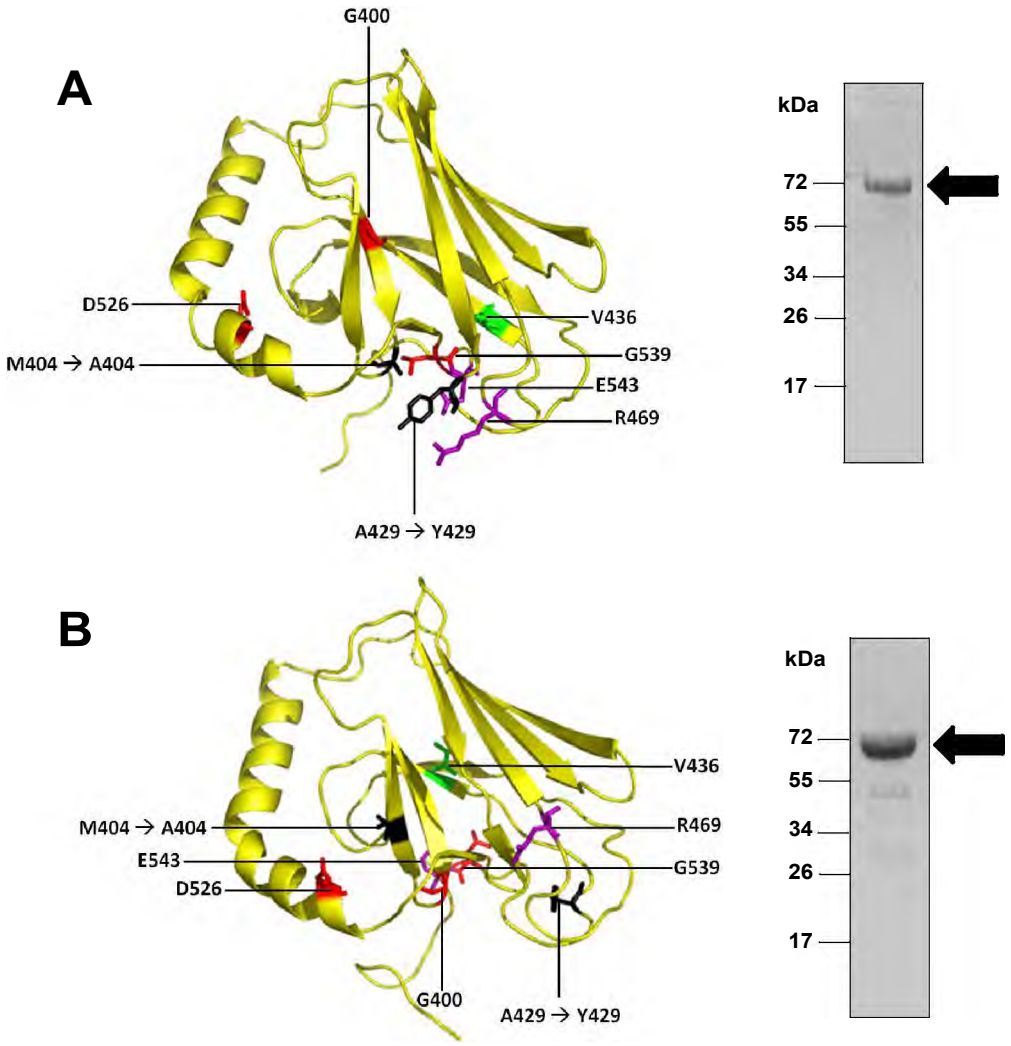


Figure 4.3. Comparison of the predicted TcHsp70 and MsHsp70 peptide binding domains

Comparison of [A] purified TcHsp70 (3.65 µg) with [B] MsHsp70 (2.5 µg) on a 12 % SDS-PAGE gel stained with Coomassie, with black arrows highlighting the two 70-kDa proteins. The homology models were generated using Modeller software, and visualised with Pymol (DeLano, 2002). The crystal structure of bovine Hsc70 (1YUW) was used as a template (Jiang et al., 2005). [A] – Model of the peptide binding domain of TcHsp70 and [B] – Model of the peptide binding domain of MsHsp70. Residues implicated to be important for function in the canonical DnaK chaperone and their corresponding residues as described in Chapter 2 in TcHsp70 / MsHsp70 are highlighted on the models: Black residues – residues important for interaction with Hsp40 proteins (Zhu et al., 1996); Red residues - residues that contact substrate in DnaK (Suh et al., 1998); Purple residues – residues implicated in stabilisation of the Hsp70 lid region (Ha et al., 1997; Chang et al., 2001)

et al., 2001; Zhu *et al*., 1996). The well conserved Gly^{400} residue proposed to be important for the interaction of DnaK with DnaJ (shown in black in Figure 4.5) is conserved in both TcHsp70 and MsHsp70. The Hsp70 arch residues Glu^{543} and Arg^{469} (shown in purple in Figure 4.5) are fully conserved in the peptide binding domain of TcHsp70, but not in that of MsHsp70, with Arg^{469} substituted for Val^{469}.

The MsHsp70 protein that was to be used as a control in further biochemical characterisation studies was visualised on a 12 % SDS-PAGE gel alongside purified TcHsp70 in order to compare the differences in purity and yield between the two proteins (Figure 4.3). The analysis revealed that TcHsp70 and MsHsp70 are of a similar size as anticipated from the literature, and are both resolved just below the 72 kDa molecular weight marker. Purified TcHsp70 appeared to possess very low levels of contaminants in the sample loaded onto the gel (10 µl), and despite the fact that the protein did not always deliver consistent yields in terms of quantity, the levels of contaminating proteins were low enough that the protein could be used in biochemical assays with confidence.

The result from SDS-PAGE analysis obtained for the MsHsp70 is similar to that shown in the product sheet from Alfa Biogene and indicated that the MsHsp70 protein sample used for biochemical assays was sufficiently pure (Figure 4.3.B).

4.3.4. TcHsp70 suppresses the aggregation of MDH in a dose-dependent manner

An investigation into the ability of TcHsp70 to suppress the aggregation of a model substrate was conducted prior to determining whether or not Tbj1 and Tcj1 can assist TcHsp70 in suppressing the aggregation of MDH. Upon addition to the heated sample buffer, MDH was found to aggregate rapidly (Figure 4.4; black trace). Six independent samples of MDH were assayed in this manner in order to ensure that the MDH standard displayed the same pattern of aggregation upon heating to 48°C prior to the addition of TcHsp70 (data not shown). In order to assess the accuracy of the results, a control

reaction was utilised in which BSA was substituted for the Hsp70 being analysed. This served to indicate whether or not any aggregation suppression observed upon the addition of TcHsp70 was in fact due to the addition of the Hsp70 to the system, and not merely the stabilisation of the MDH by another protein added to the system. BSA was able to lower the amount of aggregate formation to a moderate degree (17 % reduction in aggregation) but for the purposes of this study, values of at least 50 % aggregation suppression were considered a partially successful result, while values >80 % aggregation suppression were considered a successful result.

Different concentrations of TcHsp70 (ranging from 0.01μM to 0.36 μM) were added to MDH (0.72 μM) at a temperature of 48°C for 30 minutes. The progress of the aggregation reaction was monitored spectrophotometrically at 360 nm, with an increase in absorbance values indicating the progression of aggregate formation (Figure 4.4). The results (Figure 4.4) indicated that the addition of TcHsp70 to a heat denatured sample of MDH has a marked effect in terms of aggregation suppression, and that the aggregation suppression ability of TcHsp70 was concentration dependent. At a ratio of 2:1 (MDH:TcHsp70), the aggregation of MDH was suppressed by 89 % compared to MDH alone. The addition of lower concentrations of TcHsp70 to MDH resulted in a decrease in aggregation suppression efficiency (Figure 4.4). TcHsp70 was able to suppress aggregate formation with approximately half the efficiency when used at 0.18 μM versus 0.36 μM. This relationship was not observed when the concentration was lowered to levels beyond 0.09 μM. The results obtained for the lower concentrations of TcHsp70 were found to cluster on the graph with an average aggregation suppression of 56 %, suggesting that suppression of aggregation was no longer showing a linear proportion. An analysis of the initial rates of aggregation (Figure 4.4; black box inset) of all the reactions performed with TcHsp70 indicated that, the rate at which aggregation occurred was slowed down with respect to the rate of aggregation of MDH on its own (Figure 4.4; black trace) and the BSA control (data not shown). When no TcHsp70 was added to the sample, the MDH aggregated rapidly and reached a plateau of aggregation within 420 seconds. In contrast, the addition of TcHsp70

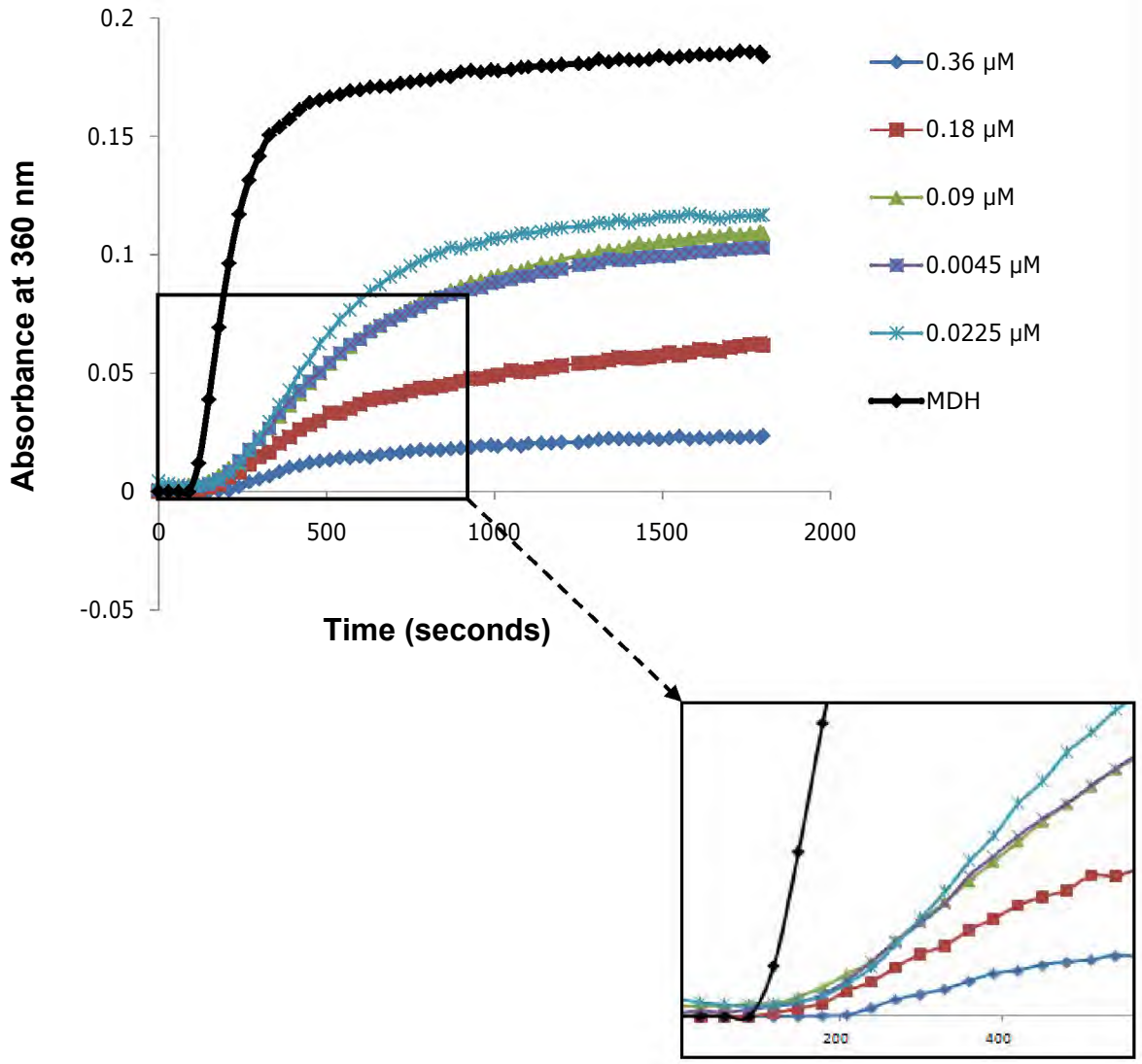


Figure 4.4. TcHsp70 suppresses the aggregation of MDH in a concentration-dependent manner

The ability of different concentrations of TcHsp70 to suppress the aggregation of a model substrate, MDH, was determined by measuring the absorbance at 360 nm. The graph shows average results of the rate of MDH aggregation (black line). The different concentrations of TcHsp70 added to MDH (0.72 µM) are indicated in colours matching those shown in the figure legend on the right of the graph. The black box and the insert to the figure shows an enlargement of the initial rate of aggregation for each reaction.

appears to slow down the rate of aggregate formation from the moment of addition to the solution. The ability of MsHsp70 to suppress the aggregation of MDH was assessed in the same manner as that of TcHsp70 (data not shown). Higher concentrations of MsHsp70 (0.36 µM) were able to suppress the aggregation of MDH by 77 %, which indicated that in the *in vitro* system used here, MsHsp70 was not as effective at aggregation suppression as TcHsp70. Interestingly, lower concentrations of MsHsp70 (0.18 µM and lower) resulted in levels of aggregation suppression that were comparable to that of the BSA control (24-36 % aggregation suppression) indicating that MsHsp70 is not effective in the *in vivo* aggregation suppression of MDH at low concentrations. Because effective suppression of MDH aggregation was observed in both TcHsp70 and MsHsp70 at a concentration of 0.36 µM to a concentration of 0.72 µM MDH, this ratio of Hsp70 to MDH was used in subsequent analyses to determine whether Tbj1 and Tcj1 were able to assist these proteins in aggregation suppression *in vitro*.

4.3.5. Tbj1 and Tcj1 are able to suppress the aggregation of MDH in partnership with TcHsp70 and MsHsp70

The ability of Tbj1 and Tcj1 to suppress the aggregation of MDH on their own or in partnership with TcHsp70 and MsHsp70 was investigated (Figure 4.5). The results shown are of experiments performed in triplicate with three independently purified batches of protein and have been calculated as a function of the percentage of the aggregation value of MDH derived from the absorbance reading at 360 nm at 1101 seconds. This time point was chosen for all calculations as it corresponded to a point at which the aggregation reactions had reached a plateau in terms of their absorbance and from which no further increases were seen. The value obtained for MDH aggregation was taken as 100 % aggregation, and all calculations were related to the MDH standard. The BSA standard was used as outlined previously in Section 4.3.4. A Type I Hsp40 protein from *T. cruzi*, Tcj2, was added to MDH at equimolar concentrations as a positive control reaction. Tcj2 was able to reduce total aggregate

formation to 43 % (57 % aggregation suppression) in the absence of an Hsp70 protein, indicating that it possesses independent chaperone properties (Figure 4.5). MDH (0.72 µM) treated with TcHsp70 (0.36 µM) resulted in 13 % aggregation, while the addition of MsHsp70 (0.36 µM) resulted in 25 % aggregation (Figure 4.5).

Tbj1 (0.72 µM) and Tcj1 (0.72 µM) were added to heated assay buffer in order to determine whether these proteins were aggregation-prone in isolation prior to assessing their ability to assist in the aggregation suppression of MDH. No increase in absorbance at 360 nm was observed in any of these reactions performed in triplicate (data not shown) indicating that neither Tbj1 nor Tcj1 are prone to self-aggregation. Due to the high levels of endogenous DnaK contamination in the Tcj1 purifications, the data obtained for this protein was recorded but not shown in the present analysis. The addition of 0.72 µM Tbj1 to MDH resulted in a rapid increase in aggregate formation to levels higher than that observed for MDH only (133 %) (Figure 4.5). This indicated that Tbj1 does not possess independent chaperone properties, which was anticipated for a Type III Hsp40 protein which does not possess a substrate-binding domain. This result was particularly significant in light of the fact that Tbj1 did not display any aggregate formation when heated in the absence of MDH (data not shown).

The same phenomenon was observed when Tcj1 was added to MDH in the absence of an Hsp70 protein, leading to an increase in aggregation to 139 % (data not shown). The addition of Tbj1 to TcHsp70 and MDH resulted in 10 % aggregation, while the addition of Tbj1 to MsHsp70 resulted in 7 % aggregation (Figure 4.5). Both these values indicate an enhancement in suppression of aggregation compared to the Hsp70 alone, especially in the case of MsHsp70 where the addition of Tbj1 reduced aggregation from 25 % to 7 % (Figure 4.5).

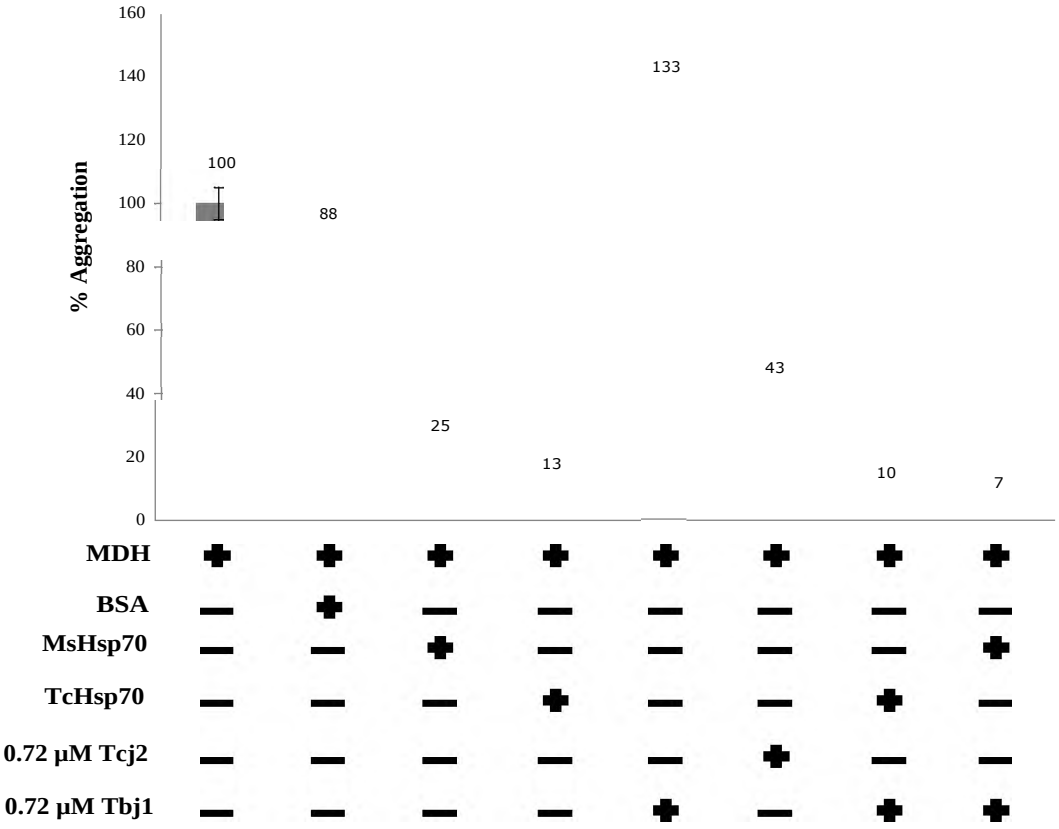


Figure 4.5. Aggregation suppression of MDH with TcHsp70 and MsHsp70 in partnership with Tbj1

Average results obtained for MDH aggregation suppression assays performed in triplicate with three independently purified batches of protein. All samples were assayed at 48°C over a 1101 second period. The '+' symbols indicate constituents present in a given reaction, while the '–' symbols indicate constituents that have been omitted from a given reaction. The actual values (% aggregation) are given above each bar of the bar graphs. Tcj2 refers to the Tcj2 positive control and error bars in the figure indicate the standard error obtained for each experiment.

These data suggest that Tbj1 (and Tcj1) are unable to suppress MDH aggregation alone, but are able to assist Hsp70-based suppression of aggregation.

4.3.6. Tbj1 and Tcj1 are unable to stimulate the ATPase activity of TcHsp70

The ability of Tbj1 to stimulate the ATPase activity of TcHsp70 was assessed using a modified version of the ammonium molybdate assay developed by Chifflet and colleagues (Chifflet *et al*., 1988; Ludewig, pers. comm.). A number of control reactions were utilised in order to determine the validity of the data obtained. ATP was assayed in the absence of Hsp40 and Hsp70 proteins and a value of 0.021 nmol Pi / min was obtained, which was subsequently deducted from all Pi values in the study in order to account of the effect of spontaneous ATP hydrolysis in the system where relevant. TcHsp70 was heat inactivated by boiling for 20 minutes in order to determine if the inactivated protein resulted in a lower ATPase activity than the purified protein. The ATPase activity obtained for boiled TcHsp70 was 0.09 nmol Pi / min / mg, which indicated that the TcHsp70 used for the study was at least partially active in that it possessed an ATPase activity of 0.56 nmol Pi / min / mg (Figure 4.6). TcHsp70 in the absence of ATP displayed no ATPase activity (Figure 4.6). Tbj1 was also assayed in the presence of ATP (0.03 nmol Pi / min / mg) and in the absence of ATP (0.02 nmol Pi / min / mg) (data not shown). The same control reaction performed with Tcj1 resulted in the opposite phenomenon; Tcj1 alone in the presence of ATP resulted in a calculated ATPase activity of 0.60 nmol Pi / min / mg which appeared to indicate that Tcj1 was stimulating the ATPase activity of endogenous co-purified DnaK. The removal of ATP from the reaction resulted in a decrease of ATPase activity to zero (data not shown). The addition of Tbj1 (0.4 μM) to TcHsp70 (0.4 μM) in the absence of ATP, resulted in an ATPase activity of 0.10 nmol / min / mg (Figure 4.6) which indicated the basal reading for this system, i.e. any values lower than 0.10 nmol / min / mg were not considered ATPase stimulation as they could be attributed to the presence of reagents in the system.

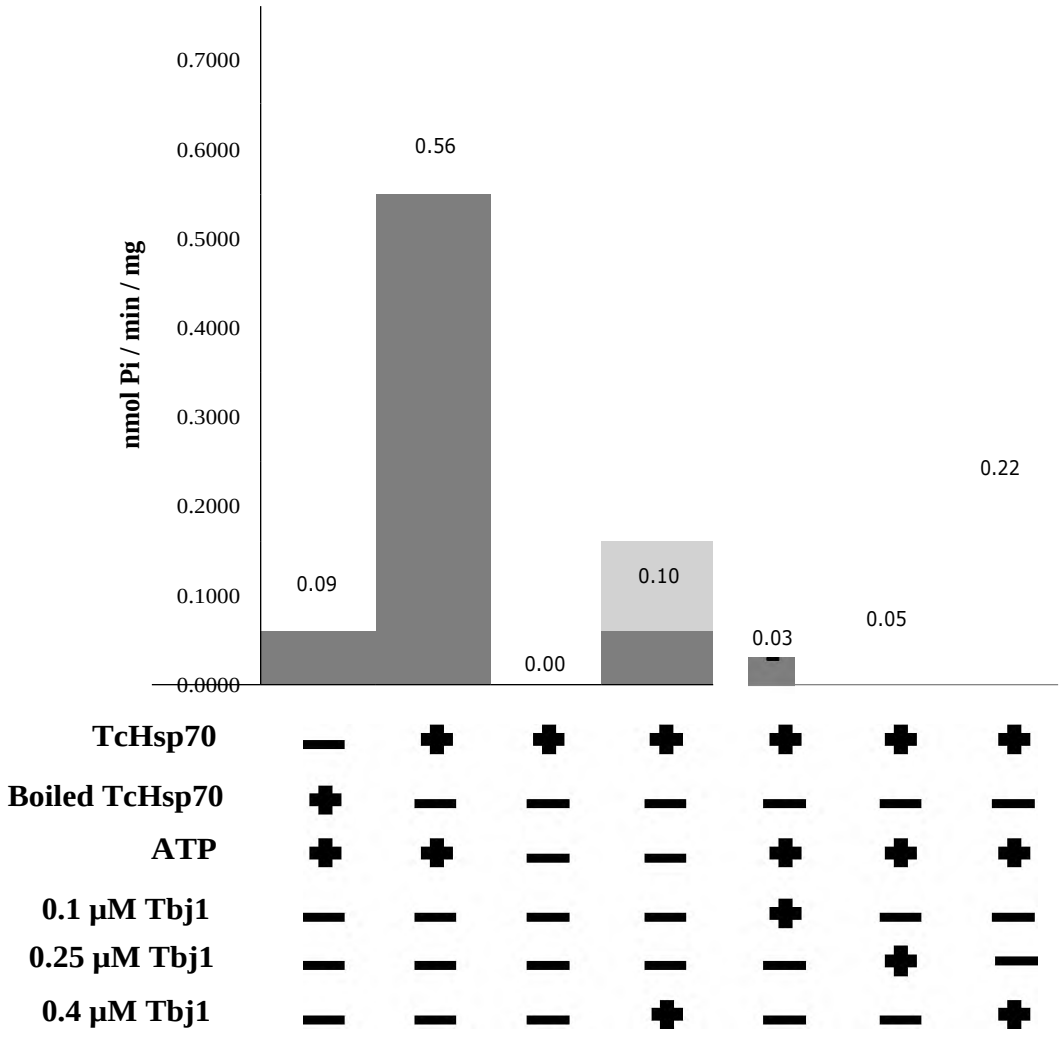


Figure 4.6. Assay of the effect of Tbj1 on the ATPase activity of TcHsp70

Each bar on the graph indicates a separate experiment. The symbols listed below each bar represent the constituents added to each reaction mixture, with '+' indicating the presence of a constituent and '-' its absence. The actual Pi values are given above the graphs and the lines denote standard errors in the experiment. The concentration of Hsp40 protein varied as indicated below the graph and TcHsp70 protein concentration was maintained at (0.4 µM) for the duration of the experiment, and samples were assayed after a period of 3 hours. The ATPase activity is given in nmol phosphate released (Pi) per minute per mg of TcHsp70 and added ATP has been corrected for in the figure. The presence of a constituent in the assay mixture is indicated by a (+) symbol below the figure, while the absence of a constituent is indicated by a (-) symbol.

Tbj1 did not have a stimulatory effect on the ATPase activity of TcHsp70 at equimolar concentrations (0.2205 nmol Pi / min / mg) but rather appeared to have an inhibitory

effect when compared to the basal ATPase activity of TcHsp70 (0.56 nmol Pi / min / mg) (Figure 4.6). In light of the results obtained for the ATPase stimulation of TcHsp70 in the presence of Tbj1 and Tcj1, an attempt was made to determine whether or not the effect may be enhanced by the use of different ratios of Tbj1 and Tcj1 with respect to a fixed concentration of TcHsp70 (0.4 µM). Previous studies have indicated that the concentration of Hsp40 added to a given Hsp70 may alter the ATPase stimulation to some extent (Edkins *et al*., 2004). Three different concentrations were used for Tbj1 and Tcj1 as a preliminary study, with the intention of investigating concentration ratios that offered promise for enhanced ATPase stimulation in further detail. Lower concentrations of Tbj1 appeared to increase the inhibitory effect of Tbj1 on the ATPase activity of TcHsp70, with the addition of 0.1 µM of Tbj1 resulting in an ATPase activity of 0.03 nmol Pi / min / mg and the addition of 0.25 µM of Tbj1 resulting in an activity of 0.05 nmol Pi / min / mg (Figure 4.6).

Tcj1 also did not have any marked stimulatory effect (0.19 nmol Pi / min / mg) over the basal ATPase activity obtained for TcHsp70 on its own (data not shown). This correlated with the finding of Edkins and colleagues (2004), which showed that Tcj1 was unable to stimulate the ATPase activity of TcHsp70 both in the presence and absence of a model substrate. Tcj1 displayed the same behaviour with respect to concentration-dependent inhibition of TcHsp70 ATPase activity as Tbj1 in spite of the endogenous DnaK present in the samples (data not shown).

4.3.7. Tbj1 and Tcj1 are unable to stimulate the ATPase activity of MsHsp70

The ability of Tbj1 and Tcj1 to stimulate the ATPase activity of MsHsp70 was assessed as described for TcHsp70 above, with identical control reactions. MsHsp70 possessed a very high basal ATPase activity (38.07 nmol Pi / min / mg) which was considerably reduced to 0.0 nmol Pi / min / mg upon boiling of the protein, when added ATP was

taken into account (Figure 4.7). The unboiled protein displayed some activity in the absence of ATP (0.22 nmol Pi / min / mg).

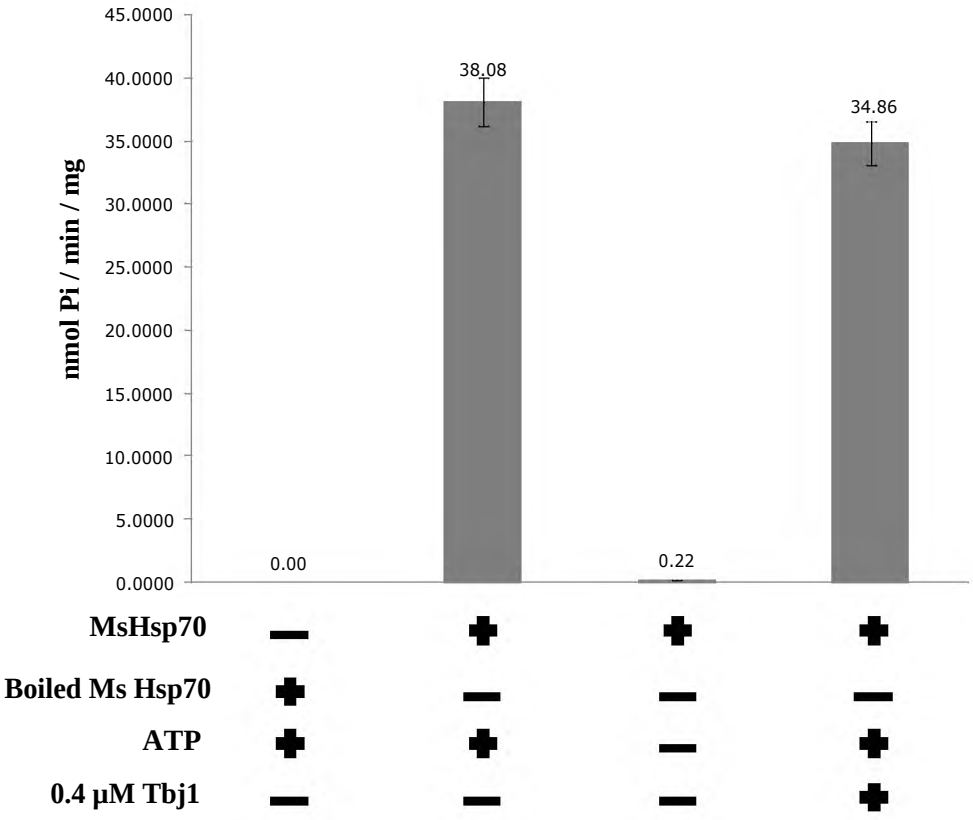


Figure 4.7. Assay of the effect of Tbj1 on the ATPase activity of MsHsp70

Each bar on the graph indicates a separate experiment. The symbols listed below each bar represent the constituents added to each reaction mixture, with '+' indicating the presence of a constituent and '-' its absence. The actual Pi values are given above the graphs and the lines denote standard errors in the experiment. The concentration of Hsp40 protein and Hsp70 protein was maintained at (0.4 µM) for the duration of the experiment, and samples were assayed after a period of 3 hours. The ATPase activity is given in nmol Phosphate released (Pi) per minute per mg of TcHsp70 and added ATP has been corrected for in all calculations. The presence of a constituent in the assay mixture is indicated by a (+) symbol below the figure, while the absence of a constituent is indicated by a (-) symbol.

The addition of Tbj1 to MsHsp70 at equimolar concentrations resulted in a net decrease of the ATPase activity of MsHsp70 (34.86 nmol / min / mg) (Figure 4.7) which was proportionately lower than the decrease observed when Tbj1 was added to TcHsp70 (Figure 4.6). However, Tbj1 was still shown to have a slight inhibitory effect on the ATPase activity of MsHsp70. This was also observed for Tcj1 which resulted in an ATPase activity of 32.20 nmol Pi / min / mg when added to MsHsp70 (data not shown).

4.4. DISCUSSION AND CONCLUSIONS

A preliminary *in vitro* biochemical analysis of Tcj1 and its previously uncharacterised *T. brucei* ortholog, Tbj1, was presented here. This is the first biochemical analysis performed on Tbj1 and will hopefully form a foundation for future studies on this protein. This work also attempts to provide more information on the biochemical behaviour of Type III Hsp40 proteins, which have not been well-characterised as a group due to the extent of their sequence diversity.

4.4.1. Tbj1 is unable to reverse the thermosensitivity of *E. coli* OD259 cells

The present chapter describes the first *in vivo* complementation study performed on Tbj1. Previous studies on trypanosomal Hsp40s in our group utilised a yeast *in vivo* complementation system for Hsp40s from the *T. cruzi* parasite (Edkins *et al*., 2004). Due to the lack of biochemical characterisation that has been performed on the *T. brucei* Hsp40 complement, it is unclear whether these proteins display behaviour more similar to the DnaJ proteins of bacteria or to the eukaryotic system that is exemplified by the yeast Hsp40 complement. A bacterial complementation system (*E. coli* OD259) was utilised for the purposes of the present work, to investigate whether Tbj1 is able to functionally replace *E. coli* DnaJ and CpbA.

The results indicated that Tbj1 was unable to functionally replace DnaJ and CpbA (Figure 4.1), suggesting that Tbj1 does not possess independent chaperone activity or to interact with DnaK as a co-chaperone. This result was not unanticipated due to the substitution of the Lys / Arg^{26} residue (Lys / Arg^{26} — Tyr^{26}) in the J-domain of Tbj1 (His^{26} in the sequence amplified during the course of this work), which has been predicted to be crucial for the interaction between the J-domain and a partner Hsp70 (Hennessy *et al*, 2005). The residues predicted to be important for the stabilization of Helices I and II (Tyr^{7}, Leu^{10}, Ile^{21}) are well-conserved within the Tbj1 J-domain and the HPDKN loop

region is fully conserved within this protein. The major areas of difference between the Tbj1 J-domain and the *E.coli* DnaJ J-domain lie in Helix IV, especially with regards to the QKRAA motif. Furthermore, Tbj1 appears to lack a canonical substrate binding domain and thus could be unable to recognize denatured *E. coli* proteins.

4.4.2. Tbj1 and Tcj1 purify in compact monomeric form

Certain Hsp40 proteins such as *E. coli* DnaJ, Djla and Sis1have previously been shown to exist as homodimers by means of X-ray scattering and size-exclusion chromatography experiments (Shi *et al*., 2005; Sha *et al*., 1999; Lakhal *et al*., 2008). The requirements of Type III Hsp40s with respect to dimerisation are not well characterised, largely due to the lack of available literature on Type III Hsp40s in general. The SEC-FPLC analysis performed in the present study revealed that both Tbj1 and Tcj1 purify as compact monomers with no dimer formation apparent (Figure 4.2).

4.4.3. Tbj1 and Tcj1 are unable to suppress the aggregation of a model substrate, but are able to assist Hsp70 proteins in aggregation suppression

A study of two potential partner Hsp70s for Tbj1 and Tcj1, MsHsp70 and TcHsp70 has revealed that TcHsp70 is able to successfully suppress the aggregation of a model substrate over a number of substoichiometric concentrations (Figure 4.4) while MsHsp70 displays a very poor ability to suppress aggregation (data not shown). The difference in the ability of Hsp70 proteins to successfully suppress aggregation in the absence of partner Hsp40 proteins has been documented in the literature. Boshoff and colleagues (2008) showed that *E. coli* DnaK is unable to prevent the thermal aggregation of a model substrate such as MDH, while *Agt* DnaK is able to prevent the aggregation of MDH in a concentration dependent manner (Boshoff *et al*., 2008).

Neither Tbj1 nor Tcj1 were able to suppress the aggregation of MDH when added to the thermally aggregated mixture (Figure 4.5 and data not shown). The addition of Tbj1, and more particularly Tcj1 to the thermally aggregating MDH substrate resulted in increased values of aggregation (133 % and 139 % respectively) which were significantly higher than the values obtained for thermally aggregated MDH on its own (Figure 4.4). This phenomenon has previously been observed during aggregation suppression assays using a Type II Hsp40 from *E. coli* and a model substrate (rhodanese) (Chae *et al*., 2004). Addition of CbpA alone to rhodanese resulted in an increase in aggregation of 20-30 %, while the addition of CbpA and DnaK to rhodanese resulted in aggregation suppression (Chae *et al*., 2004). The *E. coli* nucleotide exchange factor, GrpE, was not necessary in the above reaction, despite being necessary to effect the ATPase stimulation of DnaK by CbpA (Liberek et al., 1991).

This result corroborates findings in literature that indicate that Type III Hsp40s do not possess independent chaperone properties (Gur *et al*., 2004; Qiu *et al*., 2006). Tbj1 and Tcj1 displayed no increase in aggregate formation when incubated at 48°C in the absence of MDH (data not shown). The Type I Hsp40, Tcj2 (kindly donated by M.H. Ludewig) was added to MDH as a control to confirm the difference in aggregation suppression between the canonical Type I Hsp40s and the Type III Hsp40s. Tcj2 was able to reduce MDH aggregation by 57 %.

It would appear from this data that Tbj1 and Tcj1 are able to interact with MDH by means of a transient type of interaction and that these proteins form a complex that collectively aggregates. This suggests that both Tbj1 and Tcj1 are able to enhance aggregate formation in the absence of a suitable partner Hsp70 protein, as was observed for CbpA (Chae *et al*., 2004). It thus appears that while Tbj1 and Tcj1 are not prone to spontaneous aggregation at increased temperatures, they are capable of aggregating in conjunction with a protein that is aggregating, leading to increased levels of aggregation. This suggested that Tbj1 and Tcj1 may participate in processes involving protein and / or aggregation assembly.

While showing a complete inability to suppress the thermal aggregation of MDH when added to the protein without a partner Hsp70 protein, both Tbj1 and Tcj1 were able to suppress the thermal aggregation of MDH in partnership with MsHsp70 (Figure 4.5 and data not shown). The presence of endogenous DnaK in the purified Tcj1 samples led to a lack of confidence in the results obtained. Further studies to produce highly pure Tcj1 are needed prior to any further characterisation in this regard. Tbj1 showed a slight increase in the ability to suppress the thermal aggregation of MDH over that of TcHsp70 on its own (3 % difference) which could suggest that these two proteins are able to work together, and that Tbj1 is somehow able to enhance the ability of TcHsp70 to suppress aggregation. This could potentially occur by means of substrate stabilisation rather than direct interaction with TcHsp70.

The addition of Tbj1 (0.72 μM) to MsHsp70 (0.36 μM) resulted in a significant reduction of observable aggregate formation (7 % aggregation and 93 % suppression). This reduction of aggregate formation in the presence of Tbj1 and MsHsp70 was particularly interesting in light of the inefficiency of MsHsp70 alone with respect to aggregation suppression of MDH. It would appear that MsHsp70 is unable to function efficiently in the absence of a partner Hsp40 protein, which is interesting considering the high ATPase activity of the protein. It is important to note that while the above analysis is of interest from a purely biochemical perspective, the concentrations used are far higher than would occur in an *in vivo* system and that extensive further investigations would need to be conducted in order to determine the validity of the above observations with respect to an *in vivo* system and the Hsp40-Hsp70 partnerships in such systems.

4.4.4. Tbj3. and Tcj3. are unable to stimulate the ATPase activity of TcHsp70 and MsHsp70

ATPase stimulation studies revealed that TcHsp70 as purified during the course of the present investigation had a basal ATPase activity of 0.56 nmol Pi / min / mg, which is lower than the value reported in the literature for a pET-14b-TcHsp70 construct (40

nmol / min / mg; Edkins *et al*., 2004). It was unclear whether the basal ATPase activity for TcHsp70 reported in the present study was accurate or whether it was merely the artifactual remnants of the ATPase activity of an inactive Hsp70 protein. A comparison with the ATPase activity of other Hsp70 proteins in the literature indicates that most Hsp70s possess a low basal ATPase activity, as is evidenced by the need of these proteins for partner Hsp40s that are able to stimulate their activity. The heat shock cognate 70 (Hsc70) protein has an ATPase activity of 0.2 nmol / min / mg while *E. coli* DnaK has a reported ATPase activity of 2.5 nmol / min / mg (Tsai *et al*., 1996). It is known that the purification method employed (i.e. denaturing vs. native purification) can affect the ATPase activity displayed by Hsp70 proteins, as shown in the work of Shonhai and others (Shonhai, 2007; Matambo *et al*., 2004). The Hsp70 from the malaria parasite, *P. falciparum* displayed an ATPase activity of 14.6 nmol / min / mg in spite of the use of a denaturing purification method, indicating that if the protein is active in itself, the purification method will not remove all ATPase activity (Matambo *et al*., 2004).

In contrast to the ATPase activity determined for TcHsp70, MsHsp70 displayed a very high basal ATPase activity of 38.07 nmol / min / mg indicating that the protein was highly active. An attempt was made to determine whether Tbj1 or Tcj1 could stimulate the ATPase activity of either TcHsp70 or MsHsp70. Previous studies indicated that Tcj1 was unable to stimulate the ATPase activity of TcHsp70 (Edkins *et al*., 2004). This was confirmed in the present work, which showed that Tbj1 and Tcj1 were both unable to stimulate the ATPase activity of either TcHsp70 or MsHsp70 (data not shown). The addition of lower concentrations of Tbj1 to TcHsp70 resulted in increased apparent inhibition of the ATPase activity of TcHsp70 (Figure 4.6). In spite of the endogenous DnaK contamination in Tcj1, the same effect was observed in this protein (data not shown).

The potential effect of the His_6-tag on the ATPase activity of the proteins studied here has not been determined, but previous studies have indicated that the presence of a His_6-tag does not interfere with ATPase activity assays, and does not lead to an increase in observable ATPase activity (Minami *et al*., 1996). It was thus assumed for

the purposes of the present work that the effect of the His_6-tags of all the proteins assayed was negligible. Although maximal stimulation of the ATPase activity of Hsp70 proteins has been found to occur in the presence of a substrate and a suitable co-chaperone, the findings of Edkins and colleagues (2004) have indicated that the presence of a substrate did not alter the ability of Tcj1 to stimulate the ATPase activity of TcHsp70 (Laufen *et al*., 1999; Edkins *et al*., 2004).

Tbj1, while being an ortholog to Tcj1 and being successfully overexpressed during the course of the present study, leaves many questions as to its biochemical activities and potential functionality *in vivo*. The apparent ability of Tbj1 to assist in the suppression of thermal aggregation, but its inability to stimulate the ATPase activity of two different partner Hsp70 proteins could serve to indicate that this protein possesses a specialised function within the *T. brucei* parasite that differs from the well-conserved function of Type I Hsp40s. At this point it is unclear what this function is, as studies on Type III Hsp40s are not very numerous. Further experimental investigation and biochemical characterisation will hopefully reveal more about the *in vitro* abilities and *in vivo* functions of this protein.

4.4.5. Conclusions

Preliminary *in vitro* biochemical characterisation indicates that Tbj1 is a Type III Hsp40 that purifies as a compact monomer in a heterologous expression system and is able to assist in the aggregation suppression of a model substrate while remaining unable to stimulate the ATPase activity of two different Hsp70 proteins. These data suggest that Tbj1 possesses atypical chaperone properties, which could indicate a specialised function in the *T. brucei* parasite.

CHAPTER 5

Conclusions and future work

5.1. GENERAL DISCUSSION AND CONCLUSIONS

A thorough *in silico* analysis of the J-domains of the Type III Hsp40 complement of *T. brucei* as well as the Type III and Type IV Hsp40s of the TriTryps has been undertaken during the course of the present work. This was coupled with an in depth *in silico* analysis of the cytoplasmic Hsp70 proteins from *T. brucei* and *T. cruzi* in order to investigate potential *in vitro* partnerships.

The coding regions of two novel Type III TbHsp40s and one putative Type IV TbHsp40 have been successfully cloned in the pQE-1 vector. A novel Type III TbHsp40, Tbj1, has been successfully overproduced in a heterologous *E. coli* expression system and purified prior to preliminary *in vitro* and *in vivo* biochemical characterisation. This work aims to serve as a platform for further studies in the field of *T. brucei* molecular chaperones, particularly Hsp40s as well as provide more insight into the chaperone function of Type III Hsp40s.

5.1.1. Levels of J-domain conservation in Type III TbHsp40s

A total of 47 Type III TbHsp40s were identified in *T. brucei*, with a great diversity in terms of their molecular masses and primary amino acid sequences. All of these proteins were confirmed to be true Type III Hsp40s because of the absence of the Cys / Gly repeat region and the zinc finger region that are found in Type I Hsp40s (Cheetham and Caplan, 1998; Botha *et al.*, 2007). As the J-domain was the only defining Hsp40 domain that these proteins possessed (Cheetham and Caplan, 1998), this domain was studied more fully in order to determine the level of conservation of key residues with respect to those of other well-characterised Hsp40 proteins. A general consensus sequence for the Type III TbHsp40 J-domains from *T. brucei* was derived from a sequence alignment of all these J-domains (Chapter 2; Figure 2.4). Residues such as Tyr^{7} and Leu^{10}, which have previously been shown to be involved in the maintenance of the structural integrity of the J-domain (Hennessy, 2004) were found to be well-

conserved. Interestingly, more diversity in residues implicated in Hsp40-Hsp70 function was observed (Chapter 2; Figure 2.3 and 2.4). The Lys / Arg^{26} residue (*E. coli* numbering) has been shown to be crucial for Hsp40-Hsp70 interaction (Genevaux *et al*., 2002; Hennessy *et al*., 2004). This residue was shown to be conserved in several of the J-domains studied in *T. brucei*, but was shown to frequently be substituted with a His^{26} or Val^{26} residue (Chapter 2; Figure 2.3). This could have implications for the specificity of Hsp40-Hsp70 interactions in *T. brucei*, and could potentially serve to define discrete partnerships for these proteins.

The complete conservation of the crucial HPD motif, often followed by a Lys^{36} residue, suggested that the Type III Hsp40s in *T. brucei* would be able to stimulate the ATPase activity of a partner Hsp70 protein (Hennessy *et al*., 2005). While residues in Helix I and II of the Type III TbHsp40 J-domains were found to be well-conserved with respect to those of well-characterised Hsp40s, this conservation did not extend to the residues of Helix III and IV (Chapter 2; Figures 2.3 and 2.4). The KFK motif in Helix III was not conserved in most instances, with the substitution for an EFK / VFK motif being most common. The greater level of conservation of Lys^{48} over Lys^{46} has previously been shown in other J-domain analyses (Hennessy, 2004). The Helix IV region of the *T. brucei* J-domains studied was very poorly conserved with respect to those of canonical Hsp40 proteins, with the exception of the Arg^{63} residue that has previously been shown to be important for J-domain function (Hennessy *et al*., 2005).

The substitution of key residues in the J-domains of the Type III Hsp40s in *T. brucei* is proposed to play an important role in the specificity of these proteins in their interaction with Hsp70s. The importance of certain key residues in the consensus sequence for functionality of the J-domain will be investigated in future studies by means of rational mutagenesis coupled with *in vitro* and *in vivo* assay methods. The aim with this study would be to mutate residues that differ from those of canonical Hsp40s to the residues found in canonical Hsp40s and

observe the changes that result from such substitutions. This will give an enhanced understanding of the importance of these residues for Hsp40 function in the parasite and hopefully assist in elucidating more information on residues that are critical for the Hsp40-Hsp70 partnership with respect to Type III Hsp40 proteins.

5.1.2. The *T. brucei* Type IV Hsp40 complement

A total of 6 proteins with poorly conserved J-domains containing abrogated HPD motifs were identified in *T. brucei* (Chapter 2; Table 2.7). Type IV Hsp40s were initially studied *in silico* in the malaria-causing parasite, *P. falciparum*, and it is proposed that these proteins are unable to function as regular Hsp40 proteins, but preside over highly specialised functions in the cell. Future work will focus on two of these Hsp40s, Tbj31 (the J-domain of which possesses the greatest level of similarity to that of canonical Hsp40s) and Tbj47, which has previously been shown to be essential for *T. brucei* viability *in vivo* (Subramaniam *et al*., 2006). The expression of these proteins, and examination of their biochemical and molecular chaperone abilities with respect to aggregation suppression and ATPase stimulation of a partner Hsp70 will be investigated in order to gain an understanding of the nature of their function. This will be complemented with binding assays in order to determine whether or not the abrogated J-domains are able to successfully bind to Hsp70 proteins. As the study of Type IV Hsp40s is still a very new field, the possibility that some of these proteins may, in fact, possess no discernable Hsp40 activity in the classical sense cannot be overlooked at this stage.

5.1.3. The Hsp70 complement of *T. brucei* and *T. cruzi*

An investigation of the Hsp70 complement of *T. brucei* and *T. cruzi* revealed that the cytosolic members of this protein family were found to be more similar to Hsp70s from other eukaryotes (Chapter 2; Figure 2.8 and Table 2.9). It was

interesting to note that the EEVD motif was absent from all TbHsp70s, but were found in the *T. cruzi* Hsp70s. This could have major implications on the function and behaviour of TbHsp70 proteins in general as this motif is important for the interaction between Hsp40 and Hsp70 (Freeman *et al*., 1996; Michels *et al*., 1999). While the possibility of sequencing errors can never be ignored in any *in silico* investigation, the absence of an EEVD motif in the TbHsp70s could indicate that the TbHsp40 and TbHsp70 complements have evolved a unique mode of interaction that is specially adapted for the requirements of the parasite. This would be an attractive system for future study in terms of drug design, as highly specialised parasitic systems that are essential for parasite survival but differ markedly from that of the human host make for a good starting point as drug targets.

The Hsp70 complement of *T. brucei* was shown to occur in both the BSF and Procyclic stages of the parasite lifecycle, and in all organelles of the parasite, which was in contrast to the Hsp40 proteins that were generally only found to occur in either the BSF or Procyclic lifecycle stage. This suggests that the Hsp70 proteins in the parasite potentially interact with one group of Hsp40s during the Procyclic lifecycle stage, and a different set of Hsp40s during the BSF lifecycle stage. The possibility that a number of Hsp40s may be constitutively expressed in the parasite for the purpose of general cellular homeostasis across all lifecycle stages cannot be ignored. Further studies involving RNAi of the full TbHsp40 complement are necessary to assist in determining the role of these proteins.

While the TbHsp40 and TbHsp70 complement appears to be well-conserved with respect to canonical Hsp40 and Hsp70 proteins, the *in silico* data generated here suggests that certain specialised modes of action of these proteins may occur in the parasite. A further study of these interactions could potentially reveal whether or not the kinetoplastida have evolved specialised molecular chaperone systems that enable them to successfully survive the transition from an insect vector to a mammalian host.

5.1.4. The coding regions of three novel TbHsp40s were successfully cloned

The present work outlines the first known attempt to clone the coding region of Type III Hsp40 proteins and express these proteins in a heterologous *E. coli* system. Previous investigations have focussed on the Hsp40 complement of *T. cruzi* (Tibbetts *et al*., 1998; Salmon *et al*., 2001; Edkins *et al*., 2004) and frequently relied on the use of cDNA libraries constructed by others. As no cDNA library was available for *T. brucei* and because of the lack of introns in *T. brucei* genes (Huang and van der Ploeg, 1999) it was decided to amplify the coding regions of these proteins directly from *T. brucei* genomic DNA. This method has previously been used with great success in biochemical studies on Hsp70 proteins from *T. cruzi* (Olson *et al*., 1994). One major advantage in using the direct amplification from genomic DNA is that trypanosomal parasites often have unusually long leader sequences and poly-A tails (Tibbetts *et al*., 1998) which can complicate analysis of novel genes. For the purposes of the present study, the coding regions were not codon harmonised or optimised as a careful study of the favoured codons in *T. brucei* and a comparison with those in *E. coli* revealed that very few, if any, of the codons in the TbHsp40 coding regions of interest were predicted to prove problematic (Chanda *et al*., 2007; Horn, 2008).

The sizes of the coding regions amplified corresponded to the sizes predicted on the GeneDB database, and subsequent DNA sequencing confirmed the identity of the sequences, although minor strain differences were observed in the sequences of Tbj1 and Tbj47 (Chapter 3; Figure 3.5 and Appendix I). The coding regions of Tbj1, Tbj47 and Tbj51 were successfully cloned into the pQE-1 expression vector (Qiagen, U.S.A.). Tbj47 and Tbj51 were found to be unable to express in a heterologous *E. coli* system and it is proposed that the system used is unsuitable for the successful expression of these proteins. Future work will focus on alternative expression systems, particularly eukaryotic expression systems such as the yeast expression system and insect cell lines in order to

determine whether this could potentially facilitate the overproduction of Tbj47 and Tbj51 (Gräslun *et al*., 2008).

Tbj1 was successfully overexpressed and purified using a heterologous *E. coli* expression system (Chapter 3; Figure 3.8) although the protein was found to occur in the insoluble fraction. Future studies will focus on producing soluble Tbj1 in order to assess whether the performance of this protein *in vitro* assays is affected when a native purification method is adopted. This is crucial in order to determine whether the results of the study outlined here in terms of biochemical analysis present an accurate assessment of the activity of Tbj1, or whether certain results could be ascribed as artefacts due to the denaturing purification adopted. It is also important to note that, as this protein was expressed in a prokaryotic expression system, certain post-translational modifications such as glycosylation (Maier and Steverding, 2008) that it may possess in its endogenous source will be lacking in the form studied in this thesis. Hence, while Tbj1 was successfully overexpressed and purified here, it is important to consider that the behaviour of the protein *in vivo* could potentially differ from that produced *in vitro*, which makes future *in vivo* analyses such as RNAi and localisation studies important to confirm the findings presented here.

5.1.5. Tbj1 is unable to reverse the thermosensitivity of the *E. coli* OD259 strain

Full-length Tbj1 was shown to be unable to reverse the thermosensitivity of a temperature-sensitive strain of *E. coli* (*E. coli* OD259 (Chapter 4; Figure 4.1) indicating that is unable to functionally replace DnaJ and CbpA *in vivo* (Nicoll *et al*., 2006). Western analysis confirmed that this result was not due to a lack of expression of His_6-Tbj1 in these cells. The overall lack of sequence identity between Tbj1 and DnaJ indicated that this result is not to be unexpected, particularly with respect to the $Arg^{26} \rightarrow His^{26}$ substitution and the abrogation of the KFK and QKRAA (Suh *et al*., 1999) motifs in the J-domain of the protein

(Chapter 3; Figure 3.7). Mutations in the KFK and QKRAA motifs have previously been shown to diminish the ability of Hsp40 proteins to substitute for DnaJ and CbpA *in vitro* (Nicoll *et al*., 2007). The substitution of the basic Lys / Arg^{26} residue has previously been shown to compromise functionality of the J-domains investigated (Nicoll *et al*., 2007).

The complete inability of Tbj1 to functionally replace DnaJ and CbpA *in vivo* indicates that this protein could potentially preside over a significantly different functionality of these proteins. The significance of the use of the full-length protein in this case, however, must not be overlooked. Several studies involving complementation have involved the insertion of the J-domain of a protein of interest into the backbone of another Hsp40 that is well-known to functionally replace DnaJ and CbpA in *E. coli* OD259 cells (Nicoll *et al*., 2007). The impact that other residues in the Tbj1 sequence could have on its ability to interact with DnaK cannot be ignored. The expression of Tbj1 could also potentially be lethal to *E. coli* OD259 cells when they are placed under stressful conditions such as increased temperatures. The absence of complementation in this case must be attributed to the inability of Tbj1 to functionally replace DnaJ and CbpA by interacting with *E. coli* DnaK.

Future analysis will focus on domain swapping experiments, in which the J-domain of Tbj1 is inserted into the *Agt* DnaJ backbone and the *Agt* DnaJ J-domain is inserted into the Tbj1 backbone in order to determine which of these two chimeric proteins, if any, could functionally replace DnaJ and CbpA in *E. coli*. Another factor that will be investigated is whether or not the mutagenesis of unconserved amino acids in the J-domain of Tbj1 to residues that feature in canonical Hsp40 J-domains (e.g. the His / Tyr^{26} residue which is normally a Lys / Arg^{26} residue) could confer the ability to bind to DnaK on Tbj1. Rational mutagenesis studies on all key residues in the J-domain of Tbj1 will be attempted in order to determine whether or not these changes could enhance the function of this protein *in vivo*, both in a heterologous and a homologous (*T. brucei* cell

culture) system. This investigation will be coupled to the backbone study discussed above, as the role of other residues beyond the J-domain in Type III Hsp40 function has not been well-characterised to date. As a possibility exists that a different result could be obtained in a eukaryotic complementation system, the yeast complementation system used previously in our research unit with some success will also be investigated.

5.1.6. Tbj1 purifies in compact monomeric form and increases the aggregation of a thermally denatured substrate in the absence of Hsp70

SEC-FPLC analysis revealed that Tbj1 purifies as a compact monomer (Chapter 4; Figure 4.2) and the SDS-resistant dimer that was observed when samples containing Tbj1 were not boiled and resolved on SDS-PAGE gels was not detected during SEC-FPLC. While the disappearance of this dimeric structure could be attributed to increased salinity of the SEC-FPLC buffer with respect to previous runs in which 150 mM NaCl was added, it is proposed that in solution, Tbj1 occurs, and most probably functions, in monomeric form. This is in contrast to many other Hsp40 proteins, particularly Type I and Type II Hsp40s, which are known to dimerise in order to perform their *in vivo* function (Aron *et al*., 2005). While SEC-FPLC is a reliable technique for preliminary studies on purified proteins, additional techniques such as circular dichroism (CD) analysis (Borges *et al*., 2005; Rogers *et al*., 2008) to confirm the three dimensional structure of Tbj1 in solution are needed. Optimisation of a native purification method for the protein is also essential in order to assess whether or not any native structures may have been disrupted by the denaturing purification described here, although an *in silico* analysis of the C-terminal domain of Tbj1 has suggested that dimerisation of this protein *in vivo* is unlikely (Chapter 2 and Appendix H; Figure H.1).

Tbj1 was shown to be unable to suppress the thermal aggregation of MDH, and was found to dramatically increase the rate of MDH aggregation, in spite of the fact that the Tbj1 protein itself does not aggregate when exposed to elevated

temperatures (Chapter 4; Figure 4.8 and data not shown). Because Type III Hsp40s are predicted to not possess independent chaperone activity (Cheetham and Caplan, 1998), the inability of Tbj1 to suppress the aggregation of MDH in the absence of a partner Hsp70 protein was expected. The increase in aggregation of a denatured model substrate upon the addition of a molecular chaperone has previously been demonstrated, with the Type II Hsp40, CbpA (Chae *et al*., 2004). Tbj1 was, however, able to assist in aggregation suppression of MDH in the presence of both TcHsp70 and MsHsp70. While it is unlikely that Tbj1 binds to TcHsp70 or MsHsp70 due to its inability to stimulate the ATPase activity of these proteins, and due to a lack of conservation in J-domain residues proposed to be crucial for Hsp70 binding, it is possible that Tbj1 in this case acted to stabilise the aggregating MDH in a way that enhanced access to the aggregates by Hsp70. Tcj1 displayed the same behaviour during the course of this assay. It has previously been suggested that Hsp40s bind non-native substrates and target them towards Hsp70 proteins (Langer *et al*., 1992), although Type III Hsp40s have typically been proposed to possess highly specialised peptide binding domains that only recognise specific stubstrates (Cheetham and Caplan, 1998). The increased aggregation of MDH in the presence of Tbj1 and absence of an Hsp70 protein appears to suggest that some type of interaction between the two proteins does occur, which suggests that Tbj1 may possess the ability to interact with a broader range of substrates than would be expected for the Type III Hsp40 class in general. This could allude to some type of housekeeping function in the parasite, particularly as Tcj1 has been shown not to be upregulated in response to heat stress (Tibbetts *et al*., 1998), and it is proposed that Tbj1 behaves in the same way.

However, the ability of Type III Hsp40 proteins to assist Hsp70 proteins in the suppression of thermal aggregation of a model substrate has been confirmed in the literature. DjlA, a Type III Hsp40 protein from *E. coli*, was shown to be able to assist DnaK in the aggregation suppression of chemically denatured luciferase *in vitro* (Genevaux *et al*., 2001). Because the addition of Tbj1 (and Tcj1) to MDH

in the presence of TcHsp70 and MsHsp70 made a small, yet significant difference in the aggregation suppression ability of these proteins, which was consistently observed, future studies will aim to further characterise this by attempting to determine the nature of the interaction and subsequently create a proposed model for the mechanism of this reaction. SEC-FPLC, qualitative binding assays / affinity pull-down assays and quantitative binding assays such as surface plasmon resonance (SPR) could assist in elucidating whether or not Tbj1 is able to bind to a potential partner Hsp70 or selected substrate proteins. Because of the specificity of Hsp40-Hsp70 interactions, and acknowledging that this might have played a role in the outcome of the present investigation, a potential partner Hsp70 for Tbj1 predicted to reside in the *T. brucei* nucleus (Tb927.7.1030) will be biochemically characterised in partnership with Tbj1 in order to determine whether the ability to suppress the aggregation of MDH is improved with respect to the data obtained in the present work.

5.1.7. Tbj1 is unable to stimulate the ATPase activity of Hsp70

Tbj1 was unable to stimulate the ATPase activity of TcHsp70 and MsHsp70 (Chapter 4; Figure 4.6). This finding is supported by the literature in that Tcj1 was also found to be unable to stimulate the ATPase activity of TcHsp70 (Edkins *et al*., 2004). This study presents the first *in vitro* ATPase data for Tbj1, however, due to the high levels of sequence identity between the two proteins, it was anticipated that their behaviour *in vitro* with respect to a partner Hsp70 would be similar. The addition of Tbj1 (and Tcj1) to TcHsp70 and MsHsp70 resulted in apparent inhibition of the basal ATPase activity of the two proteins (Chapter 4; Figure 4.6 and data not shown). This was based on the fact that the ATPase activity of both TcHsp70 and MsHsp70 in the presence of Tbj1 was lower than the basal activity of these two proteins, indicating potential inhibition rather than no ATPase stimulation per se, which would have simply resulted in values equalling that of the basal ATPase rate. While previous studies have shown that the J-domain alone is sufficient for ATPase stimulation, other regions in Hsp40s,

notably the G / F rich region and the C-terminal domains have been shown to also be important (Wall *et al*., 2007). Studies in which the J-domain alone was added to an Hsp70 protein have indicated that inhibition, rather than stimulation of ATPase activity takes place (Michels *et al*., 1999). The inhibitory effect of some Hsp40 proteins on the activity of Hsp70 has been observed previously (Cheetham *et al*., 1995; Sheng *et al*., 1997). This observation does not apply to all Type III Hsp40s that only contain the J-domain and is apparent in the finding that DjIa from *E. coli* was shown to stimulate the ATPase activity of DnaK in a similar manner to that of *E. coli* DnaJ (Genevaux *et al*., 2001).

The highly reproducible nature of this phenomenon with respect to Tbj1 and TcHsp70 / MsHsp70 appears to indicate that Tbj1 does, in fact, interact in some way with the Hsp70 proteins. However, it is not apparent whether the Hsp70 proteins studied in the course of the present investigation are recognised as substrate proteins due to the presence of denatured aggregates in the purified protein sample as observed during SEC-FPLC. Once again, the integrity of the purified protein plays a key role in such *in vitro* investigations. The lack of conservation of key residues in the J-domain of Tbj1 that are proposed to be important for Hsp40-Hsp70 interaction does indicate that a real possibility exists that the binding of these proteins may not be efficient enough for ATPase stimulation; however, if any binding does occur, the conservation of the HPD tripeptide should, in theory allow for the stimulation of the ATPase activity of Hsp70. The lack of conservation of certain key residues alone could also potentially not abrogate binding of Tbj1 to a partner Hsp70 protein completely, as Hsp40-Hsp70 interaction occurs through other regions including the Helix II of the Hsp40 protein as a whole (Greene *et al*., 1998). Because the Helix II of Tbj1 was shown to be conserved with respect to that of canonical Hsp40s, the possibility of interaction with Hsp70 cannot be ruled out in the absence of future investigations involving SPR and CD studies on the proteins investigated here.

Because of the specificity of Hsp40-Hsp70 interactions, the possibility that Tbj1 cannot stimulate the ATPase activity of TcHsp70 and MsHsp70 simply due to the fact that it does not recognise them as partner proteins cannot be overlooked. As the EEVD motif on Hsp70 has previously been shown to be important for the interaction with partner Hsp40 proteins (Freeman *et al.*, 1995), and because none of the TbHsp70s possess this motif, it is possible that *T. brucei* has developed a specialised mode of interaction with Hsp70 proteins; this can only be assessed when the ATPase assay described in this work is performed on specific TbHsp70 proteins.

5.2. THE POTENTIAL ROLE OF Tbj1 *IN VIVO*

Tbj1 is a Type III Hsp40 that does not possess independent chaperone activity and potentially exerts an inhibitory function of the ATPase activity of Hsp70 proteins *in vitro*. It is potentially involved in regulatory systems in the *T. brucei* parasite.

The inability of Tbj1 to stimulate the ATPase activity of TcHsp70 and MsHsp70 *in vitro*, but its ability to assist these two proteins in the aggregation suppression of MDH suggests that a previously uncharacterised mode of action may have occurred here. While previous studies have shown that aggregation suppression *in vivo* does not require Hsp40 induced ATPase stimulation of Hsp70 to occur, it is unclear what the situation would be *in vitro* (Michels *et al.*, 1999). The fact that Type I and Type II Hsp40s are able to suppress aggregation on their own, indicates that aggregation suppression is an ATP-independent mechanism.

In the absence of further studies, a conclusive hypothesis cannot be rendered from this work. However, the preliminary data presented here suggests that Tbj1 is able to stabilise non-native substrates, and potentially assist them to an Hsp70 in an ATP-independent aggregation suppression reaction. This would appear to confirm the hypothesis on which the present work was built that proposed that

the Type III Hsp40s in *T. brucei* do not exhibit the same behaviour as canonical chaperones. Tbj1 could also potentially exert an inhibitory effect on the ATPase activity of Hsp70 proteins, possibly as part of a regulatory system in the *T. brucei* parasite. As Tcj1 has previously been shown to not be upregulated in response to heat shock (Tibbetts *et al*., 1998), it is possible that Tbj1 is also not upregulated in response to heat shock and plays a role in the general cellular homeostasis of the *T. brucei* parasite.

What role inhibition of Hsp70 function could play in *T. brucei* is not clear at present, however, the inhibition of Hsp70 has been linked to the development of a disease-stage and subsequent morbidity in numerous organisms (Michels *et al*., 1999). As *T. brucei* parasites are known to deliberately limit their numbers in the mammalian host by the generation of non-proliferative stumpy forms which involves complex changes in the parasite (Roditi and Liniger, 2002) and because Tbj1 has been shown to be expressed in the bloodstream form of the parasite as opposed to the procyclic form, it is possible that Tbj1, if it does indeed inhibit Hsp70, could function as part of the parasite system of limiting proliferation in order to enhance the length of survival of their mammalian hosts. The predicted localisation of Tbj1 to the nucleus of the parasite could support this hypothesis in that the nucleus is most often the location where alterations in the cell are initiated. Future work to investigate this could yield highly useful information in the fight against HAT.

The preliminary biochemical characterisation of Tbj1 *in vitro* will hopefully serve as a platform from which future studies of this protein can proceed. Further biochemical analysis *in vitro* is required in order to validate and expand on the data obtained during the course of the present investigation. This includes the additional analyses recommended at the end of each section in this chapter. Following from these studies, and informed by them, *in vivo* investigations will aim to examine the localisation of Tbj1 in the *T. brucei* parasite by means of confocal microscopy as well as the function of Tbj1 by means of RNAi analysis.

Tbj1, but of Type III Hsp40 interactions in Kinetoplastid parasites.

REFERENCES

Acharya, P., Kuman, R. and Tatu, U. (2007). Chaperoning a cellular upheaval in malaria: Heat shock proteins in *Plasmodium falciparum*. *Molecular and Biochemical Parasitology*, **153**, 85-94

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

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

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

Ansede, J.H., Anbazhagan, M., Brun, R., Easterbrook, J.D., Hall, J.E. and Boykin, D.W. (2004). *O*-alkoxyamidine prodrugs of furamidine: *in vitro* transport and microsomal metabolism as indicators of *in vivo* efficacy in a mouse model of *Trypanosoma brucei* rhodesiense infection. *Journal of Medicinal Chemistry*, **47**, 4335-4338

Arif, A.A., Gao, L., Davis, C.D. and Helm, D.S. (1999). Antibody response to heat shock proteins and histopathology in mice affected with *Trypanosoma cruzi* and maintained at elevated temperature. *Journal of Parasitology*, **85**, 1089-1099

Aron, R., Lopez, N., Walter, W., Craig, E.A. and Johnson, J. (2005). *In vivo* bipartite interaction between the Hsp40 Sis1 and Hsp70 in Saccharomyces cerivisiae. *Genetics*, **168**, 1873-1882

Asea, A. (2005). Stress proteins and initiation of the immune response: chaperokine activity of Hsp72. *Exercise Immunology Review*, **11**, 34-45

Atwood, 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**, 1818-1822

Balana-Fouce, R. And Reguera, R.M. (2007). RNA interference in Trypanosoma brucei: a high-throughput engine for functional genomics in trypanosomatids? *TRENDS in Parasitology*, **23 (8)**, 348-351

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. Journal of Biological *Chemistry*, **271**, 14840-14848

Baneyx, F. and Pulambo, J.L. (2003). Improving heterologous protein folding via molecular chaperones and foldase co-expression. *Methods in Molecular Biology*, **205**, 179-197

Bangs, J.D., Uyetake, L., Brickman, M.J., Balber, A.E. and Boothroyd, J.C. (1993). Molecular cloning and cellular localisation of a BiP homologue in *Trypanosoma brucei*. *Journal of Cell Science*, **105**, 1101-1113

Bardwell, J.C.A., Tilly, K., Craig, E., King, J., Zyclicz, M. and Georgopoulos, C. (1986). The nucleotide sequence of the *Escherichia coli* K12 dnaJ+ gene. A gene that encodes a heat shock protein. *Journal of Biological Chemistry*, **261 (4)**, 1782-1785

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. et al. (2005). What the genome sequence is revealing about trypanosome antigenic variation. *Biochemical Society Transactions*, **33**, 986-989

Barry, J.D. and McCulloch, R. (2001). Antigenic variation in trypanosomes: enhanced phenotypic variation in the eukaryotic parasite. *Advances in Parasitology*, **49**, 1-70

Barry, J.D. and Emery, D.L. (1984). Parasite development and host responses during the establishment of *Trypanosoma brucei* infection transmitted by tsetse fly. *Parasitology*, **88**, 67-84

Basha, E., Lee, G.J., Demeler, B. And Vierling, E. (2004). Chaperone activity of cytosolic small heat shock proteins from wheat. *European Journal of Biochemistry*, **271**, 1426-1436

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

Baxevanis, A.D. (2003). The Molecular Biology Database Collection: 2002 update. *Nucleic Acids Research*, **31**, 1-12

Bellstedt, D.U., Human, P.A., Rowland, G.F. and Van der Merwe, K.J. (1987). Acid-treated, naked bacteria as immune carriers for protein antigens. *Journal of Immunological Methods*, **98**, 249-255

Bendsten, J.D., Nielsen, H., von Heijne, G. and Brunak, S. (2004). Improved prediction of signal peptides: SignalP 3.0. *Journal of Molecular Biology*, **340**, 783-795

Bente, M., Harder, S., Wiesgigl, M., Heukeshoven, J., Gelhaus, C., Krause, E., Clos, J. And Brucchaus, I. (2003). Developmentally induced changes of the proteome in the protozoan parasite, *Leishmania donovani*. *Proteomics*, **3**, 1811-1829

Bentivoglio, M., Grassi-Zucconi, G., Olsson, T. and Kristensson, K. (1994). *Trypanosoma brucei* and the nervous system. *Trends in Neurosciences*, **17 (8)**, 325-329

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**, 84-94

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**, 416-422

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

Biteau, N., Bringaud, F., Gibson, W., Truc, P. and Baltz, T. (2000). Characterisation of Trypanozoon isolates using a repeated coding sequence and microsatellite markers. *Molecular and Biochemical Parasitology*, **105**, 185-201

Blatch, G.L. and Lässle, M. (1999). The tetratricopeptide repeat: a structural motif mediating protein-protein interactions. *Bioessays*, **21**, 932-939

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

Bock, J.H. and Langer, P.J. (1993). Sequence and genomic organisation of the hsp70 genes of *Leishmania amazoensis*. *Molecular and Biochemical Parasitology*, **62**, 187-197

Boguski, M.S. (1994). Bioinformatics. *Current Opinion in Genetics and Development*, **4**, 383-388

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

Boorstein, W.R., Ziegelhoffer, T. and Craig, E.A. (1994). Molecular evolution of the Hsp70 multigene family. *Journal of Molecular Evolution*, **38**, 1-17

Bouteille, B., Oukem, O., Bisser, S. and Dumas, M. (2003). Treatment perspectives for human African Trypanosomiasis. *Fundamental and Clinical Pharmacology*, **17**, 171-181

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. *Journal of Biological Chemistry*, **280 (14)**, 13671-13681

Bork, P., Sander, C., Valencia, A. and Bukau, B. (1992). A module of the DnaJ heat shock proteins found in malaria parasites. *Trends in Biochemical Sciences*, **17**,129

Boshoff, A., Stephens, L.L. and Blatch, G.L. (2008). The *Agrobacterium tumefaciens* DnaK: ATPase cycle, oligomeric state and chaperone properties. *The International Journal of Biochemistry and Cell Biology*, **40**, 804-812

Botha, M., Pesce, E-R. And Blatch, GL. (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

Boykin, D.W., Kumar, A., Hall, J.E., Bender, B.C. and Tidwell, R.R. (1996). Anti-Pneumocystis activity of bis-amidoximes and bis-*O*-alkylamidoxime prodrugs. *Bioorganic and Medicinal Chemistry Letters*, **6,** 3017-3020

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

Bridges, D.J., Pitt, A.R., Hanrahan, O., Brennan, K., Voorheis, H.P., Herzyk, P., de Koning, H.P. and Burchmore, R.J. (2008). Characterisation of the plasma membrane subproteome of bloodstream form *Trypanosoma brucei*. *Proteomics*, **8** (1), 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**, 153-154

Brodsky, J.L., Hamamoto, S., Feldheim, D. And Schekman, R. (1993). Reconstitution of protein translocation from solubilised yeast membrane reveals topologically distinct roles for BiP and cytosolic Hsc70. *Journal of cell biology*, **120**, 95-102

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

Brodsky, O. And Cronin, C.N. (2006). Economical parallel protein expression screening and scale-up in Escherichia coli. *Journal of Structural and Functional Genomics*, **7**, 101-108

Bruce, D. (1895). Preliminary report on the tsetse fly disease or nagana in Zululand Durban. Bennet and Davis.

Buchner, J. (1999). Hsp90 and Co. – a holding for folding. *Trends in Biochemical Sciences*, **24**, 136-141

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

Burchmore, R.J.S., Ogbunude, P.O.J., Engana, B. And Barrett, M.P. (2002). Chemotherapy of Human African Trypanosomiasis. *Current Pharmaceutical Design*, **8**, 257-267

Buscaglia, C.A. and Noia, J.M. (2003). *Trypanosoma cruzi* clonal diversity and the epidemiology of Chagas disease. *Microbes Infect*, **5**, 419-427

Campbell, K.S., Mullane, K.P., Aksoy, I.A., Stubdal, H., Zalvide, J., Pipas, J.M., Silver, P.A., Roberts, T.M., Schaffhausen, B.S. and DeCaprio, J.A. (1997). DnaJ / Hsp40 chaperone domain of SV40 large T antigen promotes efficient viral DNA replication. *Genes and Development*, **11**, 1098-1110

Carrello, A., Allan, R.K., Morgan, S.L., Owen, B.A.L., Mok, D., Ward, B.K., Minchin, R.F., Toft, D.O. and Ratajczak, T. (2004). Interaction of the Hsp90 cochaperone cyclophilin 40 with Hsc70. *Cell Stress and Chaperones*, **9 (2)**, 167-181

Carreira, M.A.C., 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**, 141-146

Carrington, M., Miller, N., Blum, M., Roditi, I., Wiley, D. And Turner, M. (1991). Variant specific glycoprotein of *Trypanosoma brucei* consists of two domains each having an independently conserved pattern of cysteine residues. *Journal of Molecular Biology*, **221**, 823-835

Cascardo, J.C., Buzeli, R.A., Almeida, R.S., Otoni, W.C. and Fontes, E.P. (2001). Differential expression of the soybean BiP gene family. *Plant Science*, **160**, 273-281

Chadli, A., Bonhouche, I., Sullivan, W., Stensgard, B., McMahon, N., Catelli, MG. And Toft, D.O. (2000). Dimerisation and N-terminal domain proximity underlie the function of the molecular chaperone heat shock protein 90. *Proceedings of the National Academy of Science U.S.A.*, **97**, 12524-12529

Chae, C., Sharma, S., Hoskins, J.R. and Wickner, S. (2004). CbpA, a DnaJ homolog, is a DnaK co-chaperone, and its activity is modulated by CbpM. *Journal of Biological Chemistry*, **279 (32)**, 33147-33153

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

Chambers, J.W., Fowler, M.L., Morris, M.T. and Morris, J.C. (2008). The anti-trypanosomal agent Lonidamine inhibits *Trypanosoma brucei* hexokinase 1. *Molecular and Biochemical Parasitology*, **158**, 202-207

Chanda, I., Pan, A., Saha, S.K. and Dutta, C. (2007). Comparative codon and amino acid composition analysis of TriTryps-conspicuous features of Leishmania major. *FEBS Letters*, **581 (30)**, 5751-5758

Chanez, A-L., Hehl, A.B. and Schneider, A. (2006). Ablation of the single dynamin of T.brucei blocks mitochondrial fission and endocytosis and leads to a precise cytokinesis arrest. *Journal of Cell Science*, **119 (14)**, 2968-2974

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

Cheetham, M.E., Anderton, B.H., Jackson, A.P. (1996). *Biochemical Journal*, **319**, 103-108

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

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

Chou, P.Y. and Fasman, G.D. (1978). Prediction of protein secondary structure of proteins form their amino acid sequences. *Advances in Enzymology*, **47**, 4-147

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

Clayton, C.E. (2002). Life without transcriptional control? From fly to man and back again. *EMBO Journal*, **21 (8)**, 1881-1888

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

Colasante, C., Ellis, M., Ruppert, T and Voncken, F. (2006). Comparative proteomics of glycosomes from bloodstream form and procyclic culture form *Trypanosoma brucei brucei*. *Proteomics*, **6 (11)**, 3275-3293

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

Cox, F.E.G. (2004). History of sleeping sickness (African trypanosomiasis). *Infectious Disease Clinics of North America*, **18**, 231-245

Cross, G.A.M. (1975). Identification, purification and properties of clone specific glycoprotein antigens constituting the surface coat of *Trypanosoma brucei*. *Parasitology*, **71**, 393-417

Cullinan, S.B. and Whitesell, L. (2006). Heat shock protein 90: A unique chemotherapeutic target. *Seminars in Oncology*, **33 (4)**, 457-465

Cyr, D.M., Lu, X. and Douglas, M.G. (1992). Regulation of Hsp70 function by a eukaryotic DnaJ homolog. *Journal of Biological Chemistry*, **267**, 20927-20931

Daniel, S., Bradley, G., Longshaw, V.M., Söti, C., Csermely, P. and Blatch, G.L. (2008). Nuclear translocation of the phosphoprotein Hop (Hsp70/Hsp90 organising protein) occurs under heat shock, and its proposed nuclear localisation signal is involved in Hsp90 binding. *Biochimica et Biophyica Acta*, **1783**, 1003-1014

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

Daugaard, M., Rohde, M. And Jaattela, M. (2007). The heat shock protein 70 family: Highly homologous proteins with overlapping and distinct functions. *FEBS Letters*, **581**, 3702-3710

Daugherty, D.L., Rozema, D., Hanson, P.E. and Gellman, S.H. (1998). Artificial Chaperone-assisted refolding of citrate synthase. *Journal of Biological Chemistry*, **273 (51)**, 33961-33971

DeLano, W. L. (2002). The PyMOL molecular graphics system. *DeLano Scientific, San Carlos, CA, U.S.A.* (http://www.pymol.org)

Delespaux, V. and de Koning, H.P. (2007). Drugs and drug resistance in African trypanosomiasis. *Drug Resistance Updates*, **10**, 30-50

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

Demand, J., Alberti, S., Patterson, C. And Hohfeld, J. (2001). Cooperation of a ubiquitin domain protein and an E3 ubiquitin ligase during chaperone / proteasome coupling. *Current Biology*, **11**, 1569-1577

Dias, J.C. et al., (2002). The impact of Chagas disease control in Latin America: a review. *Mem. Inst. Oswaldo. Cruz.*, **97**, 603-612

Donelson, J.E., Gardner, M.J. and El-Sayed, N.M. (1999). More surprises from Kinetoplasts. *Proceedings of the National Academy of Sciences U.S.A.*, **96**, 2579-2581

Dutra, W.O., Rocha, M.O.C. and Teixeira, M.M. (2005). The clinical immunology of Chagas disease. *TRENDS in Parasitology*, **21 (12)**, 581-587

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

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. *International Journal of Biochemistry and Cell Biology*, **36**, 1585-1598

Engman, D.M., Dragon, E.A. and Donelson, J.E. (1990). *Journal of Immunology*, **144**, 3987-3991

Engelman, D.M., Steitz, T.A. and Goldman, A. (1986). Identifying nonpolar transbilayer helices in amino acid sequences of membrane proteins. *Annual Review of Biophysics and Chemistry*, **15**, 321-351

Eramnian, D., Eswar, N., Shen, M-Y., and Sali, A. (2008). How well can the accuracy of comparative protein structure models be predicted ? *Protein Science*, **17**, 1881-1893

El-Sayed, N.M., Myler, P.J., Blandin, G., Berriman, M., Crabtree, J., Aggarawal, G., Caler, E., Renauld, H., Worthey, E.A., Hertz-Fowler, C., Ghedin, E., Peacock, C., Bartholomeu, D.C., Haas, B.J., Tran, A-N., Wortman, J.R., Alsmark, U.C.M., Angiuoli, S., Anupama, A., Badger, J., Bringaud, F., Cadag, E., Carlton, J.M., Cerqueira, G.C., Creasy, T., Delcher, A.L., Djikeng, A., Embley, T.M., Hauser, C., Ivens, A.C., Kummerfeld, S.K., Pereira-Leal, J.B., Nilsson, D., Peterson, J., Salzberg, S.L., Shallom, J., Silva, J.C., Sundaram, J., Westernberger, S., White, O., Melville, S.E., Donelson, J.E., Andersson, B., Stuart, K.D. and Hall, N. (2005). Comparative genomics of Trypanosomatid parasitic protozoa. *Science*, **309**, 404-409

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

Feldman, D.E. and Frydman, J. (2000). Protein folding *in vivo*: The importance of molecular chaperones. *Current Opinion in Structural Biology*, **10**, 26-33

Ferrante, A. And Allison, A.C. (1983). Alternative pathway activation of complement by African trypanosomes lacking a glycoprotein coat. *Parasite Immunology*, **5**, 491-498

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

Finkelstein, D.B. and Strausberg, S. (1983). Identification and expression of a cloned yeast heat shock gene. *Journal of Biological Chemistry*, **258**, 1908-1913

Floridi, A. And Lehninger, A.L. (1983). Action of the antitumor and antispermatogenic agent Ionidamine on electron transport in Ehrlich ascites tumor mitochondria. *Archives of Biochemistry and Biophysics*, **226**, 73-83

Folgueira, C. And Requena, J.M. (2007). A postgenomic view of the heat shock proteins in kinetoplastids. *FEMS Microbiology Reviews*, **31 (4)**, 359-377

Foucher, A.L., McIntosh, A., Douce, G., Wastling, J., Tait, A. and Turner, C.M.R. (2006). A proteomic analysis of arsenical drug resistance in Trypanosoma brucei. *Proteomics*, **6 (9)**, 2726-2732

Forde, R.M. (1902). Some clinical notes on a European patient in whose blood a trypanosome was observed. *Journal of Tropical Medicine*, **5**, 261-263

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

Freeman, B.C., Myers, M.P., Schumacher, R. and Morimoto, R.I. (1995). *EMBO Journal*, **14**, 2281-2292

Garimella, R., Liu, X., Qiao, W., Liang, X., Zuiderweg, E.R.P., Riley, M.I. et al. (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, S. C., Buchberger, A., Laufen, T., Mayer, M. P., Schröder, H., Valencia, A., and Bukau, B. (1998). Mutations in the DnaK chaperone affecting interaction with the DnaJ co-chaperone. *Biochemistry*. **95**, 15229-15234

Gässler, C.S., Wiederkehr, T., Brehmer, D., Bukau, B. And Mayer, M.P. (2001). Bag-1M accelerates nucleotide release for human Hsc70 and Hsp70 and can act concentration-dependent as positive and negative cofactor. *Journal of Biological Chemistry*, **276**, 32538-32544

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

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

Genevaux, P., Wawrzynow, 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 colanic acid capsule requires DjlA-DnaK interaction. *Journal of Biological Chemistry*, **276**, 7906-7912

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

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

Girard, M., Poupon, V., Blondeau, F. and McPherson, P.S. (2005). The DnaJ domain protein, RME-8 functions in endosomal trafficking. *Journal of Biological Chemistry*, **280**, 40135-40143

Glass, D.J., Polvere, R.I. and van der Ploeg, L.H.T. (1986). Conserved sequences and transcription of the hsp70 gene family in *Trypanosoma brucei*. *Molecular and Cellular Biology*, **6**, 4657-4666

Geerts, S., Holmes, P.H., Diall, O. and Eisler, M.C. (2001). African bovine trypanosomiasis: the problem of drug resistance. *Trends in Parasitology*, **17**, 25-28

Goebl, M. and Yanagida, M. (1991). The TPR snap helix: a novel protein repeat motif from mitosis to transcription. *Trends in Biochemical Sciences*, **16**, 173-177

Goldschmidt, H., Sheiner, L, Bütikofer, P., Roditi, I., Uliel, S., Günzel, M., Engster, M. and Michaeli, S. (2008). Role of protein translocation pathways across the endoplasmic reticulum in *Trypanosoma brucei*. *Journal of Biological Chemistry*, **283**, 32085-32098

Goloubinoff, P. and De Los Rios, P. (2007). The mechanism of Hsp70 chaperones: (entropic) pulling the models together. *TRENDS in Biochemical Sciences*, **32 (8)**, 372-380

Gräslund, S., Nordlun, P., Weigelt, J., Bray, J., Gilead, O., Knapp, S., Opperman, U., Arrowsmith, C., Hui, R., Minga, J., dhe Paganona, S., Park, H-W., Savchenko, A., Yee, A., Edwards, A., Vicentelli, R., Cambillau, C., Kim, R., Kim, S-H., Rao, Z., Shi, Y. Et al. (2008). Protein production and purification. *Nature Methods*, **5 (2)**, 135-146

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 Sciences, U.S.A.*, **95**, 6108-6113

Greener, T., Zhao, X., Nojima, H., Eisenberg, E. and Greene, L.E. (2000). Role of cyclin G-associated kinase in uncoating clathrin-coated vesicles from non-neuronal cells. *Journal of Biological Chemistry*, **275**,1365-1370

Griffith, F.L. (1898). The Petrie Papyri: Hieratic Papyri from Kahun and Gurob (Principally of the Middle Kingdom). (Text and Plates) London: Bernard Quaritch.

Gupta, R.S. and Singh, B. (1994). Phylogenetic analysis of 70 kD heat shock protein sequences suggests a chimeric origin for the eukaryotic cell nucleus. *Current Biology*, **4**, 1104-1114

Gur, E., Biran, D., Shechter, N., Genevaux, P. Georgopoulos, C. and Ron, E.Z. (2004). The *Escherichia coli* DjlA and CbpA proteins can substitute for DnaJ in DnaK-mediated protein disaggregation. *Journal of Bacteriology*, **186 (21)**, 7236-7242

Haddow, J.D., Haines, L.R., Gooding, R.H., Olafson, R.W. and Pearson, T.W. (2005). Identification of midgut proteins that are differentially expressed in trypanosome-susceptible and normal tsetse flies (*Glossina mortisans mortisans*). *Insect Biochemistry and Molecular Biology*, **35**, 425-433

Hageman, J. and Kampinga, H.H. (2009). Computational analysis of the human HSPH / HSPA / DNAJ family and cloning of a human HSPH / HSPA / DNAJ expression library. *Cell Stress and Chaperones*, **14**, 1-21

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

Hartl, F.U. (1996). Molecular chaperones in cellular protein folding. *Nature*, **381**, 571-579

Häusler, T. and Clayton, C. (1996). Post-transcriptional control of hsp70 mRNA in *Trypanosoma brucei*. *Molecular and Biochemical Parasitology*, **76**, 57-71

Hendrick, J.P. and Hartl, F-U. (1993). Molecular chaperone functions of heat-shock proteins. *Annual Review of Biochemistry*, **62**, 349-384

Hennessy, F. (2004). Characterisation of the J-domain amino acid residues important for the interaction of DnaJ-like proteins with Hsp70 chaperones. Phd Thesis, Rhodes University, Grahamstown, Republic of South Africa.

Hennessy, F., Cheetham, M.E., Dirr, H.W. and Blatch, G.L. (2000a). Analysis of the levels of conservation of the J domain among the various types of DnaJ-like proteins. *Cell Stress Chaperones*, **5 (4)**, 347-358

Hennessy, F., Nicoll, W.S., Zimmerman, R., Cheetham, M.E. and Blatch, G.L. (2000b). Not all J-domains are created equal: Implications for the specificity of Hsp40-Hsp70 interactions. *Protein Science*, **14**, 1697-179

Hennessy, F., Boshoff, A. and Blatch, G.L. (2005). 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

Hertz-Fowler, C., 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 resourse for prokaryotic and eukaryotic organisms. *Nucleic Acids Research*, **32**, 339-343

Hide, G., Cattand, P., LeRay, D., Barry, J.D. and Tait, A. (1990). The identification of *Trypanosoma brucei* subspecies using repetitive DNA sequences. *Molecular Biochemistry and Parasitology*, **39**, 213-235

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

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

Hockney, R.C. (1994). Recent developments in heterologous protein production in *Escherichia coli*. *TRENDS in Biotechnology*, **12 (11)**, 456-463

Höfeld, J., Minami, Y. and Hartl, F-U. (1995). Hip, a novel cochaperone involved in the eukaryotic Hsc/Hsp40 reaction cycle. *Cell*, **83**, 589-598

Holt, S.E., Aisner, D.L., Baur, J., Tesmer, V.M., Dy, M., Oulette, M., Trager, J.B., Morin, G.B., Toft, D.O., Shay, J.W., Wright, W.E. and White, M.A. (1999). Functional requirement of p23 and Hsp90 in telomerase complex. *Genes and Development*, **13,** 817-826

Horn, D. (2008). Codon usage suggests that translational selection has a major impact on protein expression in Trypanosomatids. *BMC Genomics*, **9**, 2

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**, 585-587

Holzer, A.M., Martiniuk, F. and Levis, W.R. (2007). Heat-shock proteins as drugs: Potential applications in cancer, infections and autoimmune and atopic diseases. *Journal of Drugs in Dermatology*, **6 (4)**, 393-399

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

Huang, L., Jacob, R.J., Pegg, S.C-H., Baldwin, M.A., Wang, C.C., Burlingame, A.L. and Babbitt, P.C. (2001). Functional assignment of the 20 S proteasome from Trypanosoma brucei using mass spectrometry and new bioinformatics approaches. *Journal of Biological Chemistry*, **30**, 28327-28339

Huang, J. and van der Ploeg, L.H.T. (1991). Maturation of polycistronic pre-mRNA in *Trypanosoma brucei*: Analysis of *trans* splicing and poly (A) addition at nascent RNA transcripts from the *hsp70* locus. *Molecular and Cellular Biology*, **11**, 3180-3190

Hunt, C. And Morimoto, R.I. (1985). Conserved features of eukaryotic hsp70 genes revealed by comparison with the nucleotide sequence of human hsp70. *Proceedings of the National Academy of Science, U.S.A.*, **82**, 6455-6459

Hutchinson, O.C., Picozzi, K., Jones, N.G., Mott, H., Sharma, R., Welburn, S.C. and Carrington, M. (2007). Variant Surface Glycoprotein gene repertoires in *Trypanosoma brucei* have diverged to become strain-specific. *BMC Genomics*, **8**, 234-24

Ivens, A.C. et al. (2005). The genome of the Kinetoplastid parasite, *Leishmania Major*. *Science*, **309**, 436

Jakob, U., Lilie, H., Meyer, I. And Buchner, J. (1995). Transient interaction of Hsp90 with early unfolding intermediates of citrate synthase – implications of heat shock *in vivo*. *Journal of Biological Chemistry*, **270**, 7288-7294

Jameson, B.A. and Wolf, H. (1988). The antigenic index: a novel algorithm for predicting antigenic determinants. *CABIOS*, **4**.

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

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

Johnson, B.D., Chadli, A., Felts, S.J., Bonhouche, I., Catelli, M.G. and Toft, D.O. (2000). Hsp90 chaperone activity requires the full-length protein and interaction among its multiple domains. *Journal of Biological Chemistry*, **275**, 32499-32507

Jones, A., Faldas, A., Foucher, A., Hunt, E., Tait, A., Wastling, J.M. and Turner, C.M. (2006). Visualisation and analysis of proteomic data from the procyclic form of *Trypanosoma brucei*. *Proteomics*, **6 (1)**, 259-267

Jubete, Y., Maurizi, M.R. and Gottesman, S. (1997). Role of the heat shock protein DnaJ in the Lon-dependent degradation of naturally unstable proteins. *Journal of Biological Chemistry*, **271**, 30798-30803

Kabani, M., Beckerich, J-M. and Brodsky, J.L. (2003). The yeast Sis1p and Fes1p proteins define a new family of nucleotide exchange factors. *Curr. Genom.*, **4**, 263-273

Karzai, A. W., and McMacken, R. (1996). A bipartite signaling mechanism involved in DnaJ-mediated activation of the *Escherichia coli* DnaK protein. *J. Biol. Chem.* **271,** 11236–11246

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

Kelley, W.L. (1999). Molecular chaperones: How J-domains turn on Hsp70s. *Current Biology*, **9**, R305-308

Kelley, L.A. and Sternberg, M.J.E. (2009). Protein structure prediction on the web: a case study using the Phyre server. *Nature Protocols*, **4**, 363-371

Kennedy, P.G.E. (2006). Diagnostic and neuropathogenesis issues in human African trypanosomiasis. *International Journal for Parasitology*, **36**, 505-512

King, J. And Laemmli, U.K. (1971). Polypeptides of the tail fibres of bacteriophage T4. *Journal of Molecular Biology*, **62 (3)**, 465-477

Kissinger, J.C. (2006). A tale of three genomes: the kinetoplastids have arrived. *Trends in Parasitology*, **22 (6)**, 240-243

Kleine, F.K. (1909). Weitere wissenschaftliche Beobachtungen über die Entwicklung von Trypanosomen in Glossinene. *Deutsche medizinische Wochenschrift*, **35**, 924-925

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

Kristjanson, P.M., Swallow, B.M., Rowlands, G.J. et al. (1999). 'Measuring the costs of African animal trypanosomiasis, the potential benefits of control and returns to research.' *Agricultural Systems*, **59**, 79-98

Kunji, E.R.S., Harding, M., Butler, J.G. and Akamine, P. (2008). Determination of the molecular mass and dimensions of membrane proteins by size exclusion chromatography. *Methods*, **46**, 62-72

Lakhal, F., Bury-Mone, S., Nomane, Y., Le Goi, N., Paillard, C. And Jacq, A. (2008). Djla, a membrane-anchored DnaJ-like protein, is required for cytotoxicity of clam pathogen Vibrio tapetis to hemocytes. *Applied and Environmental Microbiology*, **74 (18)**, 5750-5758

Lamb, J.R., Tugendreich, S. and Hieter, P. (1995). Tetratrico peptide repeat interactions: to TPR or not to TPR ? *Trends in Biochemical Sciences*, **20**, 257-259

Lambrecht, F.L. (1985). Trypanosomes and hominid evolution. *Bioscience*, **35**, 640-646

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

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

Lässle, M., Blatch, G. L., Kundra, V., Takatori, T., Zetter, B. R. (1997). Stress-inducible, Murine Protein mSTI1. Characterisation of binding domains for heat shock proteins and *in vitro* phosphorylation by different kinases. *Journal of Biological Chemistry,* **272**, 1876-1884

Lathe, G.H. and Ruthven, C.R. (1955). The separation of substances and estimation of their relative molecular sizes by the use of columns of starch in water. *Biochemical Journal*, **62 (4)**, 665-674

Laufen, T., Zuber, U., Buchberger, A. and Bukau, B. (1998). DnaJ proteins in proteins. In: Molecular Chaperones in the Life Cycle of proteins. Edited by A.L. Fink and Y. Goto. New York: Dekker, p 241-274

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

Lee, M.G. and van der Ploeg, L.H. (1990). Transcription of the heat shock 70 locus in *Trypanosoma brucei*. *Molecular and Biochemical Parasitology*, **41**, 221-231

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

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

Lian, H-Y., Zhang, H., Zhang, Z-R., Loovers, H.M., Jones, G.W., Rowling, P.J.E., Itzhaki, L.S., Zhou, J-M. and Perrett, S. (2007). Hsp40 interacts directly with the native state of the yeast prion protein Ure2 and inhibits formations of amyloid-like fibrils. *Journal of Biological Chemistry*, **282**, 11931-11940

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

Lill, R., Dowhan, W. And Wickner, W. (1990). The ATPase activity of SecA is regulated by acidic phospholipids, SecY, and the leader and mature domains of precursor proteins. *Cell*, **60**, 271-280

Lindquist, S. And Craig, E.A. (1988). The heat shock proteins. *Annual Review of Genetics*, **22**, 631-677

Liniger, M., Bodenmüller, K., Pays, E., Gallati, S., and Roditi, I. (2001). Overlapping sense and antisense transcription units in *Trypanosoma brucei*. *Molecular Microbiology*, **40**, 869-878

Lipman, N.S., Jackson, L.R., Trudel, L.J. and Weis-Garcia, F. (2005). Monoclonal versus polyclonal antibodies: distinguishing characteristics, applications and information resources. *ILAR Journal*, **46**, 258-268

Longshaw, V.M., Chapple, J.P., Balda, M.S., Cheetham, M.E. and Blatch, G.L. (2004). Nuclear translocation of the Hsp70 / Hsp90 organising protein mSTI1 is regulated by cell cycle kinases. *Journal of Cell Science*, **117**, 701-710

Ludewig, M.H. (2005). Preliminary bioinformatic analysis of the Hsp40 and Hsp70 proteins in the *T.brucei* genome. Unpublished data

Ludewig, M.H. (2009). Personal Communication.

Luscher, A., de Koning, H.P. and Maser, P. (2007). Chemotherapeutic strategies against *Trypanosoma brucei*: drug targets vs. drug targeting. *Current Pharmaceutical Design*, **13 (6)**, 555-567

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

Luu, V-D., Brems, S., Hoheisel, J.D., Burchmore, R., Guilbride, D.L. and Clayton, C. (2006). Functional analysis of *Trypanosoma brucei* PUF1. *Molecular and Biochemical Parasitology*, **150**, 340-349

Macedo, A.M. et al., (2004). *Trypanosoma cruzi*: genetic structure of populations and relevance of genetic variability in the pathogenesis of Chagas disease. *Mem. Inst. Oswaldo Cruz*, **99**, 1-12

Maier, A. and Sterverding, D. (2008). Expression and purification of non-glycosylated Trypanosoma brucei transferring receptor in insect cells. *Experimental Parasitology*, **120** (2), 205-207

Maloney, A. And Workman, P. (2002). Hsp90 as a new therapeutic target for cancer therapy: the story unfolds. *Expert Opinion on Biological Therapy*, **2**, 3-4

Masocha, W., Rottenberg, M.E. and Kristensson, K. (2007). Migration of African Trypanosomes across the blood-brain barrier. *Physiology and Behaviour*, **92**, 110-114

Matambo, T.S., Odunuga, O.O., Boshoff, A. and Blatch, G.L. (2004). Overproduction, purification and characterisation of the *Plasmodium falciparum* heat shock protein 70. *Protein Expression and Purification*, **33**, 214-222

Matthews, K.R. (2005). The developmental cell biology of *Trypanosoma brucei*. *Journal of Cell Science*, **118**, 283-290

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

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

McLaughlin, S.H., Smith, H.W. and Jackson, S.E. (2002). Stimulation of the weak ATPase activity of human Hsp90 by a client protein. *Journal of Molecular Biology*, **315**, 787-798

McCulloch, R. (2004). Antigenic variation in African trypanosomes: monitoring progress. *Trends in Parasitology*, **20**, 117-121

McKean, P.G. (2003). Coordination of cell cycle and cytokinesis in *Trypanosoma brucei*. *Current Opinon in Microbiology*, **6**, 600-607

Melville, M.W., Tan, S.L., Wambach, M., Song, J., Morimoto, R.I. and Katze, M.G. (1999). The cellular inhibitor of the PKR protein kinase, P58(IPK), is an influenza virus-activated co-chaperone that modulates heat shock protein 70 activity. *Journal of Biological Chemistry*, **274**, 3797-3803

Melville, M.W., Katze, M.G. and Tan, S.-L. (2000). P58^{IPK}, a novel cochaperone containing tetratricopeptide repeats and a J-domain with oncogenic potential. *Cellular and Molecular Life Sciences*, **57**, 311-322

Mhlanga, J.D.M. (1994). Antigenic variation in *Trypanosoma brucei*, a relationship with poly ADP-ribose polymerase. Thesis: University of Sussex, UK

Mhlanga, J.D.M., Bentivoglio, M. and Kristensson, K. (1997). Neurobiology of Cerebral Malaria and African Sleeping Sickness. *Brain Research Bulletin*, **44** (5), 579-589

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. *Journal of Biological Chemistry*, **271 (32)**, 19617-19624

Moncayo, A. (2003). Chagas disease: current epidemiological trends after the interruption of vectorial and transfusional transmission in the Southern Cone countries. *Mem. Inst. Oswaldo Cruz*, **98**, 577-591

Moore, J.H. (2007). Bioinformatics. *Journal of Cellular Physiology*, **213**, 365-369

Mount, D. (1985). Computer analysis of sequence, structure and function of biological macromolecules. *BioTechniques*, **3**, 102-112

Morris, S.A., Tanowitz, H.B., Wittner, M. and Bilezikian, J.P. (1990). Pathophysiological insights into the cardiomyopathy of Chagas disease. *Circulation*, **82**, 1900-1909

Morris A L, MacArthur M W, Hutchinson E G & Thornton J M (1992). Stereochemical quality of protein structure coordinates. *Proteins*, **12**, 345-364.

Mosser, D.D., Caron, A.W., Bourget, L., Meriin, A.B., Sherman, M.Y., Morimoto, R.I. and Massie, B. (2000). The chaperone function of hsp70 is required for protection against stress-induced apoptosis. *Molecular and Cellular Biology*, **20 (19)**, 7146-7159

Mulenga, C., Mhlanga, J.D.M., Kristensson, K. and Robertson, B. (2001). *Trypanosoma brucei brucei* crosses the blood-brain barrier while tight junction proteins are preserved in a rat chronic disease model. *Neuropathology and Applied Neurobiology*, **27**, 77-85

Myler, P., Audelman, L., DeVos, T. et al. (1999). *Leishmania major* Friedlin chromosome I has an unusual distribution of protein-coding genes. *Proceedings of the National Academy of Sciences U.S.A.*, **96**, 2902-2906

Nathan, D.F., Vos, M.H. and Lindquist, S. (1997). *In vivo* functions of the Saccharomyces cerivisae Hsp90 chaperone. *Proceedings of the National Academy of Science U.S.A.*, **94,** 12949-12956

Neckers, L. and Neckers, K. (2005). Heat-shock protein 90 inhibitors as novel cancer chemotherapeutics – an update. *Expert Opinion In Emerging Drugs*, **10**, 137-149

Nicolet, C.M. and Craig, E.A. (1989). Isolation and characterisation of STI1, a stress-inducible gene from *Saccharomyces cerivisiae*. *Molecular and Cellular Biology*, **9**, 3638-3646

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

Nicoll, W.S., Botha, M., McNamara, C., Schlange, M., Pesce, E-R., Boshoff, A., Ludewig, M.H., Zimmerman, 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. *International Journal of Biochemistry and Cell Biology*, **39**, 736-751

Nyalwidhe, J., Maier, U-G. and Lingelbach, K. (2003). Intracellular parasitism: cell biological adaptations of parasitic protozoa to a life inside cells. *Zoology*, **106**, 341-348

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

O'Brien, M.C., Flaherty, K.M. and McKay, D.B. (1996). Lysine71 of the chaperone protein Hsc70 is essential for ATP hydrolysis. *Journal of Biological Chemistry*, **271**, 15874-15878

Odunuga, O.O., Hornby, J.A., Bies, C., Zimmerman, R., Pugh, D.J. and Blatch, G.L. (2003). Tetratricopeptide repeat motif-mediated Hsc70-mSTI1 interaction: Molecular characterisation of the criticial contacts for successful binding and specificity. *Journal of Biological Chemistry*, **278**, 6896-6904

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

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*. *Journal of Biological Chemistry*, **269**, 3868-3874

Paba, J., Santana, J.M., Teixera, A.R.L., Fontes, W., Sousa, M.V and Ricart, C.A.O. (2004). Proteomic analysis of the human pathogen, *Trypanosoma cruzi*. *Proteomics*, **4**, 1052-1059

Paggi, M.G., Faniculli, M., Perrotti, N. et al. (1988). The role of mitochondrial hexokinase in neoplastic phenotype and its sensitivity to Lonidamine. *Annals of the New York Academy of Sciences*, **551**, 358-360

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

Panigrahi, A.K., Zikova, A., Dalley, R.A., Acestor, N., Ogata, Y., Anupama, A., Myler, P.J. and Stuart, K.D. (2008). Mitochondrial complexes in *Trypanosoma brucei*: a novel complex and a unique oxidoreductase complex. *Molecular and Cellular Proteomics*, **7 (3)**, 534-545

Pays, E., Lips, S., Nolan, D., Vanhamme, L. And Pérez-Morga, D. (2001). The *VSG* expression sites of *Trypanosoma brucei*: multipurpose tools for the adaptation of the parasite to mammalian hosts. *Molecular and Biochemical Parasitology*, **114**, 1-16

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

Pearl, L.H. and Prodromou, C. (2000). Structure and *in vivo* function of Hsp90. *Current Opinion in Structural Biology*, **10**, 46-51

Pearl, L.H. and Prodromou, C. (2006). Structure and mechanism of the hsp90 molecular chaperone machinery. *Annual Reviews of Biochemistry*, **75**, 271-294

Pedersen, J., Lauritzen, C., Madsen, M.T. and Dahl, S.W. (1999). Removal of N-terminal polyhistidine tags from recombinant proteins using engineered aminopeptidases. *Protein Expression and Purification*, **15**, 389

Pepin, J. and Milord, F. (1991). African trypanosomiasis and drug-induced encephalopathy: risk factors and pathogenesis. *Transactions of the Royal Society of Tropical Medicine and Hygiene,* **85**, 222-224

Pepin, J. and Milord, F. (1994). The treatment of human African trypanosomiasis. *Advances in Parasitology*, **33**, 1-47

Phillips, J.J., Yao, Z-P., Zhang, W., McLaughlin, S., Laue, E.D., Robinson, C.V. and Jackson, S.E. (2007). Conformational dynamics of the molecular chaperone Hsp90 in complexes with a co-chaperone and anticancer drugs. *Journal of Molecular Biology*, **372 (5)**, 1189-1203

Picard, D. (2002). Heat shock protein 90, a chaperone for folding and regulation. *Cellular and Molecular Life Sciences*, **59**, 1640-1648

Pierpaoli, E.V., Sandmeier, E., Schönfeld, H.J. and Christen, P. (1998). Control of the DnaK chaperone cycle by substoichiometric concentrations of the co-chaperones DnaJ and GrpE. *Journal of Biological Chemistry*, **273**, 6643-6649

Popp, S., Packschies, L., Radzwill, N., Vogel, K.P., Steinhoff, H.J., Reinstein, J. (2005). Structural dynamics of the DnaK-peptide complex. *Journal of Molecular Biology*, **347**, 1039-1052

Powers, M.V. and Workman, P. (2006). Targeting of multiple signalling pathways by heat shock protein 90 molecular chaperone inhibitors. Conference paper. Endocrine-Related Cancer, 13, S125–S135.

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

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

Qiao, X., Chuang, B-F., Jin, Y., Muranjan, M., Hung, C-H., Lee, P-H and Lee, M.G-S. (2006). Sorting signals required for trafficking of the cysteine-rich acidic repetitive transmembrane protein in *Trypanosoma brucei*. *Eukaryotic Cell*, **5 (8)**, 1229-1242

Qiu, X.-B., Shao, Y.-M, Miao, S. and Wang, L. (2006). The diversity of the DnaJ/Hsp40 familiy, the crucial partners for Hsp70 chaperones. Cellular and Molecular Life Sciences, **63**, 2560-2570

Requena, J., Jiminez-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 1393-1406

Requena, J.M., Lopez, M.C., Jiminez-Ruiz, A., de la Torre, J.C. and Alonso, C. (1988). A head-to-tail tandem organisation of hsp70 genes in *Trypanosoma cruzi*. *Nucleic Acids Research*, **16**, 1393-1406

Rocha, M.O., Ribeiro, A.L. and Teixeira, M.M. (2003). Clinical management of chronic Chagas cardiomyopathy. *Front. Biosci.*, **8**, e44-e54

Rodgers, L., Gamez, A., Riek, R. and Ghosh, P. (2008). The Type III Secretion chaperone SycE promotes a localised disorder-to-order transition in the natively unfolded effector YopE. *Journal of Biological Chemistry*, **283,** 20857-20863

Roditi, I. And Liniger, M. (2002). Dressed for success: the surface coats of insect-borne protozoan parasites. *Trends in Microbiology*, **10**,128-134

Roe, S.M., Prodroumo, C., O'Brien, R., Ladbury, J.E., Piper, P.W. and Pearl, I.H. (1999). *Journal of Medical Chemistry*, **42,** 260-266

Rost, B. (1999). Twilight zone of protein sequence alignments. *Protein Engineering*, **12**, 85-94

Rout, M.P. and Field, M.C. (2001). Isolation and characterisation of subnuclear compartments from *Trypanosoma brucei. Journal of Biological Chemistry*, **276 (41)**, 38261 – 38271

Rüdiger, S., Buchberger, A. and Bukau, B. (1997). Interaction of Hsp70 chaperones with substrates. *Nature Structure Biology*, **4 (5)**, 342-349

Russell, R., Karzai, A.W., Mehl, A.F. and McMacken, R. (1999). DnaJ dramatically stimulates ATP hydrolysis by DnaK: insight into targeting of Hsp70 proteins to polypeptide substrates. *Biochemistry*, **38**, 4165-4176

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. Journal of Biological Chemistry*, **276 (47)**, 43970-43979

Sanchez, E.R. (1990). Hsp56: a novel heat shock protein associated with untransformed steroid receptor complexes. *Journal of Biological Chemistry*, **265**, 22067-22070

Santoro, M.G. (2000). Heat shock factors and the control of the stress response. *Biochemical Pharmacology*, **59**, 55-63

Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) Molecular Cloning: a Laboratory Manual. Volumes I, II and III. Cold Spring Harbour Laboratory Press.

Sargeant, T. J., Marti, M., Caler, E., Carlton, J. M., Simpson, K., Speed, T. P., and Cowman, A. F. (2006). Lineage-specific expansion of proteins exported to erythrocytes in malaria parasites**.** *Genome Biol*. **7**: R12

Sayers, E.W., Barrett, T., Benson, D.A., Bryant, S.H., Canese, K., Chetvermin, 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. (2009). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research*, **37**, 5-15

Sha, B., Lee, S. And Cyr, D.M. (1999). The crystal structure of the peptide-binding fragment from the yeast Hsp40 protein Sis1. *Structure Fold Des*, **8**, 799-807

Sheng, Q, Denis, D., Ratnofsky, M., Roberts, T.M., DeCaprio, J. and Schaffhausen, B. (1997). *Journal of Virology*, **71**, 9410-9416

Shi, Y-Y., Hong, X-G. and Wang, C-C. (2005). The C-terminal (331-376) sequence of *Escherichia coli* DnaJ is essential for dimerisation and chaperone activity. *Journal of Biological Chemistry*, **280 (24)**, 22761-22768

Scheibel, T., Weikl, T. and Buchner, J. (1998). Two chaperone sites in Hsp90 differing in substrate specificity and ATP dependence. *Proceedings of the National Academy of Sciences, U.S.A.*, **95 (4)**, 1495-1499

Schlieker, C., Bukau, B. and Mogk, A. (2002). Prevention and reversion of protein aggregation by molecular chaperones in the *E.coli* cytosol: implications for their applicability in biotechnology. *Journal of Biotechnology*, **96**, 13-21

Schlendstedt, G., Harris, S., Risse, B., Lill, R. and Silver, P. (1995). A yeast DnaJ homologue, Scj1p, can function in the endoplasmic reticulum via a conserved domain that specifies interactions with Hsp70s. *Journal of Cell Biology*, **129**, 979-988

Schneider, A. (2001). Unique aspects of mitochondrial biogenesis in trypanosomatids. *International Journal of Parasitology*, **31**, 1403-1415

Schneider, A., Bursac, D. and Lithgow, T. (2007). The direct route: a simplified pathway for protein import into the mitochondrion of trypanosomes. *TRENDS in Cell Biology*, **18 (1)**, 12-18

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

Sedbrook, J.C., Chen, R. and Masson, P.H. (1999). ARG1 (Altered Response to Gravity) encodes a DnaJ-like protein that potentially interacts with the cytoskeleton. *Proceedings of the National Academy of Science, U.S.A.*, **96**, 1140-1155

Shiau, A.K., Harris, S.F., Southworth, D.R. and Agard, D.A. (2006). Structural analysis of *E.coli* hsp0 reveals dramatic nucleotide-dependent conformational arrangements. *Cell*, **127**, 329-340

Schulte, T.W., An, W.G. and Neckers, L.M. (1997). Geldanamycin-induced destabilization of Raf-1 involves the proteasome. *Biochemical and Biophysical Research Communications*, **239**, 655-659

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

Shonhai, A. (2007). Molecular characterisation of the chaperone properties of *Plasmodium Falciparum Heat* shock protein 70. Phd thesis, Rhodes University, Grahamstown, Republic of South Africa

Shonhai, A., Botha, M., De Beer, T.A.P., Boshoff, A. and Blatch, G.L. (2008). Structure-function study of a *Plasmodium falciparum* Hsp70 using three dimensional modelling and *in vitro* analyses. *Protein and Peptide Letters*, **15**, 1117-1125

Shonhai, A., Boshoff, A. And Blatch, G.L. (2007). The structural and functional diversity of Hsp70 proteins from *Plasmodium falciparum*. *Protein Science*, **16**, 1803-1818

Sibley, L.D. and Andrews, N.W. (2000). Cell invasion by un-palatable parasites. *Traffic*, **2**, 100-106

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

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

Smejkal, R.M., Wolff, R. and Olenich, J.G. (1988). *Experimental Parasitology*, **65**, 1-9

Smith, D.F. (1993). Dynamics of heat shock protein 90-progesterone receptor binding and the disactivation loop model for steroid receptor complexes. *Molecular Endocrinology*, **7**, 1418-1429

Smith, D.F. (2004). Tetratricopeptide repeat cochaperones in steroid receptor complexes. *Cell Stress and Chaperones*, **09 (2)**, 109-121

Smith, D.F., Whitesell, L. and Katsanis, E. (1998). Molecular Chaperones: Biology and Prospects for Pharmacological Intervention. *Pharmacological Reviews*, **50**, 493-510

Sogin, M.L., Gunderson, J.H., Elwood, H.J., Alonso, R.A. and Peattie, D.A. (1989). Phylogenetic meaning of the kingdom concept: an unusual ribosomal RNA from *Giardia lamblia*. *Science*, **243**, 75-77

Sõti, C., Nagy, E., Giricz, Z., Vigh, L., Csermely, P. and Ferdinandy, P. (2005). Heat shock proteins as emerging therapeutic targets. *British Journal of Pharmacology*, **146**, 769-780

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. *Journal of Biological Chemistry*, **274**, 30534-30539

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

Stevens, J.R., Noyes, H.A., Dover, G.A. and Gibson, W.C. (1999). The ancient and divergent origins of the human pathogenic trypanosomes, *Trypanosoma brucei* and *T.cruzi*. *Parasitology*, **118**, 107-116

Steverding, D. (2008). The history of African trypanosomiasis. *Parasites and Vectors*, **12** (1), 3

Szabo, A., Korszun, R., Hartl, F.U. and Flanagan, J. (1996). A zinc finger-like domain of the molecular chaperone DnaJ is involved in binding to denatured protein substrates. *EMBO Journal*, **15**, 408-417

Szyperski, T., Pellechia, M., Wall, D., Georgopoulus, C., and Wuthrich, K. (1994) *Proc. Natl. Acad. Sci. U. S. A.* **91,** 11343–11347

Tavaria, M., Kola, I. and Anderson, R.L. (1997). The hsp70 genes of mice and men. In: Guidebook to Molecular Chaperones and Protein-Folding Catalysts. Ed. Gething, M-J. Oxford University Press, Oxford, 49-52

Terada, K. and Mori, M. (2000). Human DnaJ homologs dj2 and dj3, and bag-1 are positive co-chaperones of Hsc70. *Journal of Biological Chemistry*, **275**, 24728-24734

Tibbetts, R.S., Kim, I.Y., Olson, C.L., Barthel, L.M., Sullivan, M.A., Windquist, A.G., Miller, S.D. and Engman, D.M. (1994). Molecular cloning and characterization of the 78-kilodalton glucose-regulated protein of *Trypanosoma cruzi*. *Infection and Immunity*, **62**, 2499-2507

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

Toutain, C.M., Clarke, D.J., Leeds, J.A., Kuhn, J., Beckwith, J., Holland, I.B. and Jacq, A. (2003). The transmembrane domain of the DnaJ-like protein Djla is a dimerisation domain. *Molecular Genetics and Genomics*, **268**, 761-770

Tsai, J. And Douglas, M.D. (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**, 9347-9354

Tunac, J. (1989). A new high-aeration capacity shake-flask system. *J. Ferm. Bioeng.* **68**, 15–159

Ueguchi, C., Kakeda, M., Yamada, H. and Mizuno, T. (1994). An analogue of the DnaJ molecular chaperone in *Escherichia coli*. *Proceedings of the National Academy of Sciences U.S.A.*, **91**, 1054-1058

Vago, A.R., Andrade, L.O., Leite, A.A., d'Avila, R.D., Macedo, A.M., Adad, S.J., Tostes, S., Moreira, M.C., Filho, G.B. and Pena, S.D. (2000). Genetic characterisation of *Trypanosoma cruzi* directly from tissues of patients with chronic Chagas disease: differential distribution of genetic types into diverse organs. *Am. J. Pathol.*, **156**, 1805-1809

Van der Ploeg, L.H.T., Giannini, S.H. and Cantor, C.R. (1985). Heat shock genes: regulatory role for differentiation in parasitic protozoa. *Science*, **228 (4706)**, 1443-1446

Van Deursen, F.J., Thornton, D.J. and Matthews, K.R. (2003). A reproducible protocol for analysis of the proteome of Trypanosoma brucei by 2-dimensional gel electrophoresis. *Molecular and Biochemical Parasitology*, **128**, 107-110

Vera, A., Gonzalez-Montalban, N., Aris, A. and Villaverde, A. (2007). The conformational quality of insoluble recombinant proteins is enhanced at low growth temperatures. *Biotechnology and Bioengineering*, **96**, 1101-1106

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**, 189-201

Vickerman, K. (1997). Landmarks in trypanosome research. In: Hide, G., Mottram, J.C., Coombs, G.H. and Holmes, P.H. (Eds.) Trypanosomiasis and Leishmaniasis. CAB International, Oxford, pp. 1-37

Volk, W.A., Benjamin, D.C., Kadner, R.J. and Parsons, J.T. (1986). Essentials of Medical Microbiology 3rd Ed. Lippencott Company U.S.A. pp. 756-759

Wall, D., Zylicz, M. and Georgopoulos, C. (1994). *Journal of Biological Chemistry*, **269**, 5446-5451

Walsh, P., Bursac, D., Law, Y.C., Cyr, D. and Lithgow, T. (2004). The J-protein family: modulating protein assembly, disassembly and translocation. *European Molecular Biology Organisation Reports,***23** (**5)**, 567–571.

Wegele, H., Haslbeck, J., Reinstein, J. and Buchner, J. (2003). Sti1 is a novel activator of the Ssa proteins. *Journal of Biological Chemistry*, **278**, 25970-25976

Wegele, H., Wandinger, S.K., Schmid, A.B., Reinstein, J. and Buchner, J. (2006). Substrate transfer from the chaperone Hsp70 to Hsp90. *Journal of Molecular Biology*, 356, 802-811

Weickert, M.J., Doherty, D.H., Best, E.A. and Olins, P.O. (1996). Optimization of heterologous protein production in *Escherichia coli*. *Current Opinion in Biotechnology*, **7 (5)**, 494-499

Welch, W.J. and Feramisco, J.R. (1982). Purification of the major mammalian heat shock proteins. *Journal of Biological Chemistry*, **24**, 14949-14959

Wendel, S., Brener, Z., Camargo, M.E. and Rassi, A. (1992). Chagas disease – American trypanosomiasis: the impact on transfusion and clinical medicine. In: International Society Blood Transfusion – ISBT Brazil, 92, São Paulo: International Society of Blood Transfusion, p 13-29

Weich, H., Buchner, J., Zimmerman, M., Zimmerman, R. and Jakob, U. (1993). Hsc70, immunoglobulin heavy chain binding protein, and Hsp90 differ in their ability to stimulate transport of precursor proteins into mammalian microsomes. *Journal of Biological Chemistry*, **268**, 7414-7421

Werner-Washburne, M. and Craig, E.A. (1989). Expression of members of the Saccharomyces cerevisiae hsp70 multigene family. *Genome*, **31**, 684-689

Werner-Washburne, M., Stone, D.E. and Craig, E.A. (1987). Complex interactions among members of an essential subfamily of hsp70 genes in *Saccharomyces cerevisiae*. *Molecular and Cellular Biology*, **7**, 2568-2577

Williams, B.I. (1996). African Trypanosomiasis. In the Wellcome Trust Illustrated History of Tropical diseases. Edited by Cox, F.E.G. London: The Wellcome Trust, 178-191

World Heath Organisation (2002). Control of Chagas disease. *WHO Technical Reports Series*, **905**, 109

Workman, P. (2004). Combinatorial attack on multistep oncogenesis by inhibiting the Hsp90 molecular chaperone. *Cancer Letters*, **206**, 149-157

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 a novel dimerisation motif for Hsp40. *Journal of Molecular Biology*, **346**, 1005-1011

Young JC, Schneider C, Hartl FU. (1997). *In vitro* evidence that hsp90 contains two independent chaperone sites. *FEBS Letters,* **418(1-2)**, 139–143.

Zhao, R.M., Davey, M., Hsu, Y.C., Kaplanek, P., Tong, A. And Parsons, A.B. et al. (2005). Navigating the chaperone network: an integrative map of physical and genetic interactions mediated by the Hsp90 chaperone. *Cell*, **120**, 715-727

Ziegelbauer, K. and Overath, P. (1993). Organisation of two invariant surface glycoproteins in the surface coat of *Trypanosoma brucei*. *Infection and Immunity*, **61**, 4540-4545

APPENDICES

Appendix A: General Experimental procedures

A.1. Preparation of competent cells

E. coli glycerol stocks archived at -80°C were streaked onto 2 x YT agar plates (1.6 % tryptone, 1.0 % yeast, 1.5 % agar, 0.5 % NaCl) with no antibiotic (*E.coli* JM109, DH5α, XL1Blue, BL21), choramphenicol (*E.coli* BL21 pLysS) or kanamycin (*E.coli* OD259) and incubated overnight. A single colony from each plate was selected and used to inoculate a 5 ml 2 x YT broth culture and grown overnight with shaking at the temperature favoured by the strain of competent cells being prepared. Four flasks containing 100 ml 2 x YT broth were inoculated with different volumes of the overnight culture (1.5, 1.0, 0.7 and 0.3 ml respectively) and the cultures were grown at 37°C with shaking at 180 rpm until the $_{A600}$ of the culture with the 1.5 ml inoculum reached a value of 0.6 – 0.8. The cultures were harvested at 6 000 *g* for 20 minutes in a JA-14 rotor (Beckman, U.S.A.) prior to re-suspension of the pellet in 4 ml of cold RF1 solution (15 % w/v glycerol, 100 mM KCl, 50 mM $MnCl_2$, 30 mM CH_3COOK, 10 mM $CaCl_2$; pH 5.8) and incubation on ice for 20 minutes. After the incubation period these were centrifuged at 5 000 rpm for 20 minutes in a JA-14 rotor (Beckman, U.S.A.) and the pellets obtained were pooled and re-suspended in 2 ml RF2 solution (15 % glycerol w/v, 10 mM MOPS, 10 mM KCl, 75 mM $CaCl_2$; pH 6.8). The cells were divided into 200 µl aliquots and stored at -80°C until required.

A.2. Transformation of competent cells

An aliquot of plasmid DNA (50-100 ng; 1.5 µl) was added to 100 µl of competent cells that were previously thawed on ice and incubated on ice for 30 minutes. The cells were heat-shocked at 42°C for 45 seconds and immediately placed on ice for 5 minutes. A volume of ice cold 2 x YT broth (900 µl) was added to the cells prior to incubation at 37°C for one hour. The cells were pelleted by centrifugation at 12 000g for 2 minutes

and re-suspended in 200 µl of the supernatant broth prior to plating on 2 x YT plates containing appropriate antibiotics.

A.3. Small scale preparation of plasmid DNA

A - High quality plasmid DNA for DNA sequencing was isolated using the Qiaprep kit (Qiagen, U.S.A) which is based on a modified alkaline lysis method of plasmid isolation (Birnboim and Doly, 1979). A plasmid of interest was transformed into *E.coli* JM109 cells and a single colony was selected and used to inoculate a 5 ml 2 x YT broth containing 100 µg / ml ampicillin. The inoculated culture was grown overnight with shaking at 37°C. After 16 hours growth, the entire bacterial culture was pelleted by centrifugation at 12 000 *g* for 2 minutes. The pellet obtained was re-suspended in buffer P1 (250 µl), prior to the addition of buffer P2 (250 µl) and buffer N3 (350 µl). The sample was centrifuged at 12 000 g for 10 minutes before the supernatant was applied to a Qiaprep spin column in a 2 ml collection tube and centrifuged at 12 000 *g* for 60 seconds. The column was washed with 500 µl buffer PB prior to washing with 750 µl buffer PE. After each washing step, the column was centrifuged at 12 000 *g* for 60 seconds and the washing solution was discarded. Upon completion of the washing steps, the column was centrifuged for an additional 60 seconds to remove any residual ethanol prior to elution of the bound plasmid DNA after a 1 minute incubation period in buffer EB and centrifugation at 12 000 g for 60 seconds.

B - In cases where less pure plasmid DNA preparations were required, as was the case during large-scale screening experiments, plasmid DNA was isolated using the Alkaline lysis method (Ausubel *et al.*, 1991; Sambrook *et al.*, 1989). An overnight culture was prepared from a single transformant as outlined above and pelleted by centrifugation at 12 000 g for 2 minutes. The pellet was resuspended in 100 µl Solution 1 (50 mM glucose, 25 mM Tris-HCl, 10 mM EDTA; pH 8.0) prior to the addition of 5 µl RNAse A (10 mg / ml) and vortexing until the pellet was fully resuspended. The mixture was incubated at room temperature for 5 minutes. Solution II (150 µl; 0.2 M NaOH, 1 % w / v SDS; prepared fresh) was added prior to incubating on ice for 2 minutes. After the

incubation period, 250 µl ice cold Solution III (3 M potassium acetate; pH 5.0) was added prior to another incubation period on ice for 5 minutes. The sample was centrifuged at 12 000 g for 10 minutes and the supernatant obtained was transferred to a sterile eppendorf tube to which 450 µl isopropanol had been added. The supernatant-isopropanol mixture was incubated at room temperature for 2 minutes and centrifuged at 4°C for 30 minutes to yield a white pellet. The supernatant obtained was discarded and the pellet resuspended in 400 µl TE buffer (10 mM Tris-HCl, 1 mM EDTA; pH 8.0), 40 µl sodium acetate (3 M sodium acetate) and 800 µl ice cold ethanol (100 %) before being incubated at – 20°C for 30 – 60 minutes. The plasmid DNA was collected after a further 30 minute centrifugation step at 12 000 g (4°C). The pellet obtained was washed with 200 µl ethanol (70 %), pelleted by centrifugation and resuspended in 25 µl TE buffer (10 mM Tris-HCl, 1 mM EDTA; pH 8.0).

A.4. Agarose gels

Agarose (0.8 %) was added to 60 ml of 0.5 x TBE buffer (45 mM Tris base, 45 mM boric acid, 1 mM EDTA, pH 8.0) and heated in a microwave oven for approximately 90 seconds until fully dissolved. The molten agarose was cooled to 55 - 60°C prior to the addition of ethidium bromide to a final concentration of 0.5 µg / ml. After cooling further, the agarose gel was poured into a casting mould with a spacer comb and left to set for 20 minutes. Prior to loading the DNA samples in loading buffer (6x loading buffer; 0.25 % bromophenol blue, 30 % glycerol), the spacer comb was removed and the gel placed into an electrophoresis tank filled with 0.5 x TBE (45 mM Tris base, 45 mM boric acid, 1 mM EDTA, pH 8.0) buffer. The samples were resolved at 110 V with the DNA moving towards the anode. The electrophoresis run was allowed to proceed until the gel front had travelled $^{3}/_{4}$ of the length of the gel. For smaller DNA fragments, a 2 % agarose gel was used. Upon completion of the electrophoresis, agarose gels were visualized using the Chemidoc XRS imaging system (Bio-Rad, U.S.A.).

A.5. DNA restriction digestion protocol

Plasmid DNA (8 µl containing 100-300 ng) was added to 10x New England Biolabs (NEB; U.S.A.) restriction buffer appropriate to the given restriction enzyme (4 µl), BSA (2 µl) and sterile deionised water (24 µl). The digestion reaction was initiated by the addition of 0.5 U of restriction enzyme and allowed to incubate at 37°C for 1 – 3 hours. The reaction was terminated by the addition of DNA loading buffer (0.25 % w / v bromophenol blue, 50 % glycerol v /v).

Table A.1. Restriction enzymes and buffers used for diagnostic restriction digests of *T.brucei* and *T.cruzi* heat shock proteins

Vector name	Enzyme 1	Enzyme 2	NEB Buffer no.	BSA
pQE1-Tbj1	KpnI	PmlI	1	Yes
pQE1-Tbj47	HindIII	PvuII	2	No
pQE1-Tbj51	KpnI	HindIII	2	No
pET23b-Tcj1	XhoI	EcoRI	3	Yes

A.6. Polymerase chain reaction (PCR)

PCR reactions were set up containing *T. brucei* genomic DNA (100 ng; 1.5 µl), forward primer (4 µM), reverse primer (4 µM), 10 x PCR buffer (5 µl; 50 mM KCl; 10 mM Tris-HCl; 1.5 mM $MgCl_2$), dNTP mix (1 µl), Expand High Fidelity DNA polymerase (1 U) and autoclaved distilled water (35.7 µl). The PCR parameters are outlined in Figure 1.1.

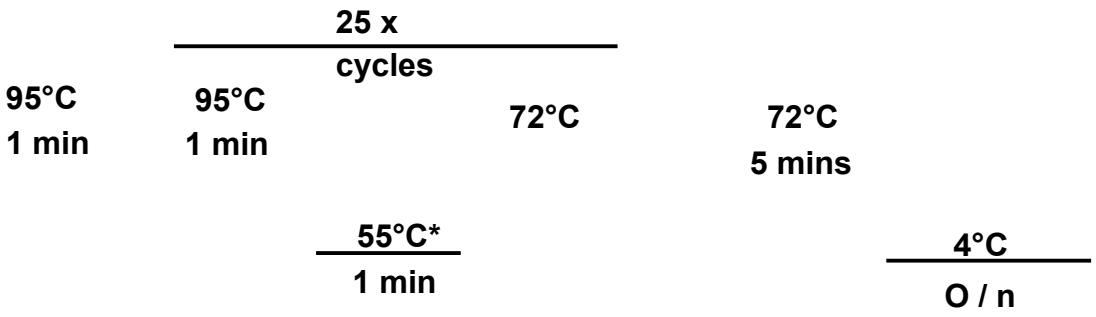


Figure A.1. Parameters for PCR amplification of TbHsp40s from *T.brucei* genomic DNA.

These parameters were used for all PCR amplifications from *T.brucei* genomic DNA. The temperature (*) varied according to that of the primer set for a given coding region.

PCR products were analysed on 0.8 – 1.0 % gels as that were prepared as outlined in section A.4.

A.7. Ligation of PCR fragments into pGEM-T Easy®

PCR products (ca. 100 ng) were added to 2 x rapid ligation buffer (60 mM Tris-HCl, pH 7.8, 20 mM MgCl2, 20 mM DTT, 2 mM ATP, 10 % polyethylene glycol), pGEM-T Easy® vector (50 ng) and T4 DNA ligase (0.3 U). The ligation mixture was incubated overnight at 4°C prior to transformation and screening as outlined above in A.2. and A.3.

A.8. Gel purification of DNA bands using the DNA gel extraction kit (Fermentas, U.S.A.)

DNA fragments of interest were gel purified using the DNA gel extraction kit (Fermentas, U.S.A.). pGEM-T Easy® constructs containing the insert of interest as well as digested pQE-1 plasmids (Qiagen, U.S.A.) were resolved on 1 % agarose gels

prepared in 0.5 x TBE buffer containing no ethidium bromide. Subsequent to agarose gel electrophoresis, the gel was stained in a solution of ethidium bromide (1 mg / ml) for approximately 20 minutes. The stained gels were subjected to long-wave U.V. (365 nm) for as short a period as possible and the relevant bands were excised and added to DNA binding solution (6 M sodium iodide; 3 x gel volume) and 500 µl TBE conversion buffer prior to heating at 55°C for 5-10 mins until the gel slices were completely melted. Silica powder suspension (5 µl) was added and vortexed. The mixture was subsequently incubated at 55 °C for 5 minutes and vortexed every 2 minutes to maintain an even suspension of silica beads. After the 5 minute incubation period, the samples were centrifuged at 12 500 g for 30 seconds in order to sediment the silica powder suspension bound to the free DNA in solution. The pellet was washed three times with ice cold wash buffer (concentrated wash buffer containing Tris-HCl, NaCl and EDTA in 95 % ethanol) and resedimented each time by centrifuging at 12 500 g for 20 seconds. After the final wash step, the pellet was air dried for 15 minutes to remove any residual ethanol and resuspended in 25 µl TE buffer. The beads were pelleted and discarded. Gel purified DNA was always used fresh in ligation reactions.

A.9. Ligation of DNA fragments into the pQE-1 plasmid vector (Qiagen, U.S.A.)

Gel purified plasmid DNA as well as insert DNA in appropriate concentrations (1:4) were added to 10 x ligation buffer (660 mM Tris-HCl, 50 mM $MgCl_2$, 50 mM dithiothreitol (DTT), 10 mM ATP, pH 7.5), 1 U of T4 DNA ligase and sterile deionised water to make the reaction up to 20 µl. Ligation reactions were allowed to proceed for 1 hour at room temperature or overnight at 4°C prior to transformation and screening of plasmid DNA.

A.10. DNA sequencing

Plasmids of interest were isolated using the Qiagen miniprep kit (Qiagen, U.S.A.) as outlined above in A.3.A. Plasmid DNA was eluted in autoclaved distilled water instead

of in TE buffer. A cycle sequencing reaction was prepared using the BigDye V3.1 Terminator cycle sequencing kit (Applied Bioystems, USA) and containing DNA template (500 ng, typically 2 – 5 µl), primer (3.2 pm), Big Dye Buffer (2 µl), Big Dye mix (4 µl) and autoclaved distilled water to make up to 20 µl. The following thermal cycling protocol was adhered to over 25 cycles: 96 °C for 10 seconds, 50 °C for 5 seconds and 60 °C for 4 minutes. The extension products obtained were purified from unincorporated dye terminators using the the Zymo kit according to the kit manufacturers instructions. After clean-up, the samples were eluted in 8 µl autoclaved distilled water and air dried overnight at room temperature. DNA sequencing was performed on a 3100 ABI Genetic analyser sequencing machine by a qualified technician at the Rhodes University DNA sequencing facility (Rhodes University, Grahamstown; South Africa). Sequence information obtained was analysed using the AlignX and Contig Express components of the Vector NTI® molecular analysis suite (Informax, U.S.A.).

A.11. Primers used for the DNA sequencing of pQE-1 plasmids

Coding region amplified	Primer name	Primer sequence (5' -3')	Tm (°C)
pQE-1 forward of MCS	pQE-1For	CCC GAA AAG TGC CAC C	54.0
pQE-1 reverse of MCS	pQE-1Rev	TTA GCT CCT GAA AAT CTC G	49.2

A.12. Induction studies

Induction studies were performed on all proteins prior to purification in order to determine whether adequate protein expression levels were obtained. The plasmids of interest were transformed in *E.coli* XL1 Blue or BL21 competent cells and a single colony was used to inoculate a 25 ml 2 x YT broth starter culture containing 100 µg / ml ampicillin, which was grown overnight with shaking at 37°C for 16 hours. The overnight starter culture was used to inoculate a 225 ml 2 x YT broth culture containing

100 µg / ml ampicillin which was grown with shaking until it reached an $_{A600}$ of between 0.6 – 0.8 prior to initiation of induction with the addition of 1 mM IPTG. Samples (1 ml) were taken at hourly intervals for 6 hours post-induction as well a pre-induction and 16 hour post induction sample. The absorbance was measured for each sample, and ice cold PBS (137 mM NaCl, 2.7 mM KCl, 10 mM NaH_2PO_4, 2 mM KH_2PO_4; pH 7.4) and SDS-PAGE loading buffer were added according to the formula:

$[A_{600} / 0.5] / 150 * 10$ = volume PBS in µl

[volume PBS in µl] / 6 = volume of SDS-PAGE loading buffer in µl

This was done to ensure that an equal quantity of protein was loaded per lane so that induction would be clear from SDS-PAGE analysis. Samples were stored at -20°C until analysed using SDS-PAGE and Western blotting.

A.13. SDS-Page analysis of proteins

Protein samples were analyzed on acrylamide gels prepared according to a modified protocol based on the standard method of Laemmli (1970). The acrylamide gels were resolved using the Bio-Rad minigel system (Bio-Rad, U.S.A.). Acrylamide resolving gels (12 %) were prepared (0.375 M Tris-HCl, pH 8.8, 0.1 % SDS, 12 % acrylamide) and allowed to set prior to the addition of stacking gels (4 %; 0.125 M Tris-HCl, pH 6.8, 0.1 % SDS, 4 % acrylamide). The polymerization reactions were initiated by the addition of 0.05 % ammonium persulphate (APS) and 0.005 % N,N,N',N'-tetramethylethylenediamine (TEMED). Protein samples were prepared by adding 1 part SDS-loading buffer (0.0625 M Tris-HCl, pH 6.8, 10 % glycerol, 2 % SDS, 5 % β-mercaptoethanol and 0.05 % bromophenol blue) to 6 parts of the protein sample and boiled for 20 minutes. After loading the samples into the stacking gel, the gels were resolved in SDS-PAGE running buffer (25 mM Tris-HCl, 192 mM glycine and 1 % SDS)

for 1 hour at 200 V. Gels were stained using Coomassie stain (40 % methanol, 0.7 % acetic acid, 0.075 % Coomassie brilliant blue R250) for 1 hour – overnight and destained with Coomassie destain solution (40 % methanol, 0.7% acetic acid) for 3 hours with shaking at 40 rpm.

A.14. Quantification of purified protein with the Bradford assay

A protein standard curve was prepared from a 1 mg / ml BSA with standards ranging from 0 – 120 µg / ml. The BSA standards (100 µl) were added to 200 µl Bradford reagent (Sigma, U.S.A.) and allowed to develop for 5 minutes prior to reading the absorbance at 595 nm. Protein samples were diluted 1:10 and assayed using the same method.

A.15. Charging of Ni-sepharose beads

Sepharose slurry (2 ml; GE Healthcare, U.K.) was pipetted into a 15 ml falcon tube and centrifuged at 500 g for 3 mins in order to sediment the sepharose beads prior to removal of the supernatant. Distilled water (5 gel volumes) was used to resuspend the beads and removed by centrifugation at 500 g for 3 minutes. The washed beads were resuspended in 0.5 column volume of 0.1 M $NiSO_4$ and unbound nickel was removed by centrifugation at 500 g for 3 minutes. The charged column was washed three times with distilled water (5 column volumes each time) and resedimented by centrifuging at 500 g for 3 minutes. After the final wash step, the charged nickel-sepharose column was resuspended in 1 column volume of denaturing lysis buffer (8M Urea, 300 mM NaCl, 100 mM Tris-HCl, 10 mM Imidazole; pH 8.0) and stored at 4°C until required.

A.16. Protein purification

All proteins were purified using a standard denaturing protocol utilizing native wash and elution steps. A single colony that was previously positively identified by means of

plasmid isolation and a diagnostic digest was inoculated into a 2 x YT overnight starter culture (30 ml) containing 0.1 mg / ml ampicillin and grown with shaking (180 rpm) at 37°C for 12 -16 hours. A 225 ml flask of 2 x YT broth was inoculated with the overnight starter culture and induced with 1 mM IPTG when the $_{A600}$ had reached 0.6 – 0.8. The culture was induced for a period of 5 hours. After 5 hours, the induced cultures were centrifuged at 6 000 g for 20 minutes in a JA – 14 rotor (Beckman Coulter, U.S.A.) at 4°C and the pellets were resuspended in 10 ml denaturing lysis buffer (8M Urea, 300 mM NaCl, 100 mM Tris-HCl, 10 mM Imidazole; pH 8.0) with 100 µl of a 100 mM stock of PMSF added. Subsequent to thorough resuspension, the lysate was aliquotted into microcentrifuge tubes and stored at -20°C overnight. Subsequent to thawing at room temperature, the protein lysates were centrifuged for 20 minutes at 13 000 rpm in a refrigerated microcentrifuge (Eppendorf; Germany) at 4°C and the cleared lysates were added to a 1 ml charged nickel Sepharose (GE Healthcare, U.K.) column that was previously charged as outlined in A.15. The lysate was allowed to bind the nickel column for a period of 4 – 6 hours on ice with gentle shaking. The nickel bead-bound lysate was subsequently collected by means of centrifugation at 500 g for 3 minutes. Between 3 – 5 washes of 2-3 column volumes each was performed with native wash buffers with different imidazole concentrations, depending on the levels of expression of the individual protein. The Tbj1-lysate bound to the nickel-sepharose beads was purified by means of 4 wash steps of 2-3 column volumes each with native wash buffer 1 (100 mM Tris-HCl, 300 mM NaCl, 100 mM Imidazole; pH 8.0). ATP (0.6 mM) was added to each wash in order to remove any residual *E.coli* DnaK contamination. When the target protein of interest was sufficiently pure (single band on an SDS-PAGE gel) and enough contaminants had been washed from the column, the protein was eluted with native elution buffer (100 mM Tris-HCl, 300 mM NaCl, 1 M Imidazole; pH 8.0) and stored at 4°C prior to analysis by SDS-PAGE and Western blotting in order to confirm protein identity and integrity. Samples for SDS-PAGE analysis were collected throughout the purification process and analysed until the purification was sufficiently optimised and the process considered sufficiently reproducible and robust to render this unnecessary.

A.3.7. Complementation assays

Plasmids encoding full length proteins were transformed into temperature-sensitive *E.coli* OD259 cells. Plasmid DNA (100 ng) was added to 100 µl OD259 competent cells and incubated on ice for 30 minutes prior to heat shock at 37°C for 45 seconds. The cells were cooled on ice for 5 minutes prior to the addition of ice cold 2 x YT broth and incubation at 30°C for 1 hour. After the incubation period, the cells were collected by centrifuging at 12 500 g for 2 minutes and resuspended in 200 µ 2x YT broth before plating out onto 2 x YT plates containing ampicillin (100 µg / ml) to select for the pQE-1 plasmids and kanamycin (50 µg / ml) to select for OD259 cells that retain the plasmid rendering them sensitive to temperature.

A single colony from each plate was screened using the miniprep method outlined in section A.3.B. and A.5. to ensure that the right constructs were being used for the complementation assay. After confirmation of plasmid identity was successfully completed, a single colony from each plate was used to inoculate a 5 ml 2 x YT overnight culture containing ampicillin (100 µg / ml) kanamycin (50 µg / ml), which was allowed to grow at 30°C with shaking. The overnight cultures (0.5 ml) were used to inoculate 50 ml 2 x YT broth cultures containing ampicillin (100 µg / ml) and kanamycin (50 µg / ml), which was allowed to grow at 30°C with shaking for between 3 – 5 hours until the $_{OD600}$ reached 2.0 before aliquots were taken and diluted to an OD of 0.2. Serial dilutions ranging from 10^{0}-10^{-10} were made from the diluted culture and 2 µl of each dilution were spotted onto plates containing ampicillin (100 µg / ml), kanamycin (50 µg / ml) and IPTG (50 µM). Identical samples and controls were plated onto 5 duplicate plates that were incubated at 30°C, 37°C, 38.5°C, 40°C and 42°C overnight. OD259 cells in the same dilution range were plated onto 2 x YT plates containing kanamycin only (50 µg / ml) and incubated at the temperatures described above. After an overnight incubation period, the plates were scanned and results were documented.

A.18. MDH assays

Proteins of interest were purified as described in section A.16. and dialysed against ice cold autoclaved buffer (50 mM Tris-HCl; 200 mM NaCl; pH 8.0) for 4 hours with stirring at 4°C in order to remove residual imidazole. Proteins were quantified using the Bradford assay prior to use and were stored at 4°C for a maximum of 7 days during the course of this experiment. The MDH standard reaction was performed by adding MDH (0.72 µM) to MDH assay buffer (50 mM Tris-HCl; 100 mM NaCl, pH 7.4) and heating to 48°C in a Heλios spectrophotometer. Once the mixture had reached temperature, a rate scan over a 30 minute period was performed at an absorbance of 360 nm in order to monitor the aggregation process. Proteins of interest were carefully quantified and added to MDH and / or potential partner proteins in different ratios, and each time the standard assay procedure above was followed. The data obtained was analysed using a Microsoft Excel spreadsheet. The aggregation experiment was performed in triplicate with three separate batches of each purified protein, i.e. protein from three independent purifications.

A.19. ATPase assays

Proteins of interest were purified as described in section A.16. and dialysed against ice cold autoclaved buffer for 4 hours with stirring at 4°C in order to remove residual imidazole. Proteins were quantified using the Bradford assay and the molar concentration determined using the formula [(mg protein / size of protein in Da) = mM concentration] prior to use and were stored at 4°C for a maximum of 7 days during the course of this experiment. Samples were prepared by adding 0.4 µM of the Hsp70 protein being tested to an equimolar concentration of Hsp40 protein, 10 mM HEPES buffer (pH 7.4), 10 mM $MgCl_2$, 20 mM KCl and 0.5 mM DTT. Samples were incubated at 37°C for 5 minutes prior to the addition of 600 µM ATP to initiate the ATP hydrolysis reaction. From this point onwards, 50 µl samples were taken every 30 minutes over a 3 hour period, and added to 50 µl SDS (10 %) in a 96-well flat bottomed microtitre plate to terminate the reaction. Upon completion of sample collection, colour development was

initiated by the addition of 50 µl of ammonium molybdate (1 % in 1 N HCl), 50 µl ascorbic acid (6 % in phosphate free water), 125 µl sodium citrate / acetic acid (2 % each) in order. The 96-well plates were incubated at 37°C for 30 minutes to allow for full colour development and read at an absorbance of 850 nm. Values obtained were calculated using a phosphate standard curve prepared from KH_2PO_4 in tandem with the ATPase assay reaction.

Appendix B: Sample standard curves

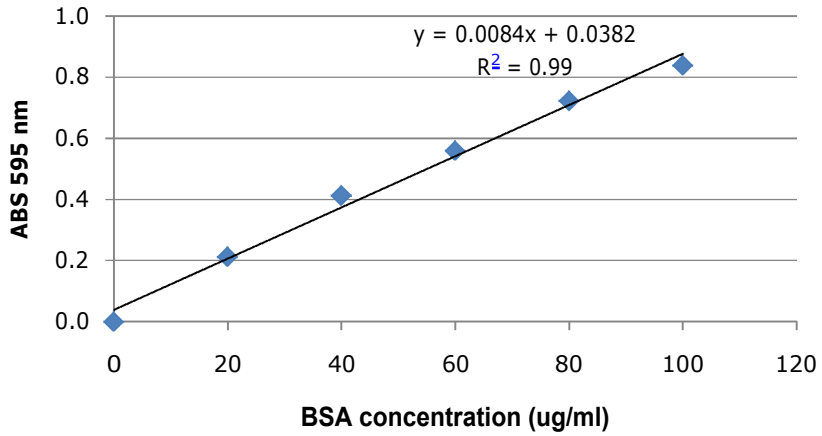


Figure B.1. Sample BSA standard curve

The plot shows average absorbance values that were obtained when a stock solution of BSA (1 mg / ml) is diluted to form standards ranging from 0 – 120 µg / ml and assayed using the Bradford method at 595 nm. The standard curve was used in protein quantification for biochemical characterization.

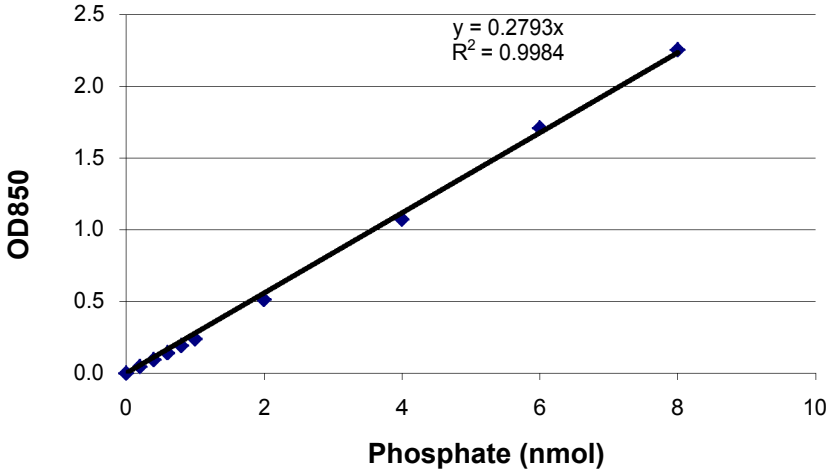


Figure B.2. Sample phosphate standard curve

The plot shows average absorbance values that were obtained when a stock solution of 0.5 nmol of KH_2PO_4 was assayed using the Ammonium Molybdate method at an absorbance of 850 nm. The standard curve was utilized in order to determine that ATPase activity of Tb / TcHsp70s by measuring the rate of phosphate release.

Appendix C: Primers to amplify TbHsp40s from *T.brucei* genomic DNA

Coding region amplified	Primer name	Primer sequence (5' -3')	Tm (°C)
Tbj1	Tbj1_fp_KpnI	GGT ACC ATG GGG TCA GAT GTG	58.5
	Tbj1_rp_PmlI	CAC GTG CTA TCT ACT TCG GTC ATA G	60.3
Tbj47	Tbj47_fp_PvuII	GCA GCT ATG CGA GGC ATT AC	56.4
	Tbj47_rp_HindIII	AAG CTT TTA CGT ACC CGG C	55.5
Tbj51	Tbj51_fp_KpnI	GGT ACC AAA ATG GAG ACG ACA TGC AAG	59.8
	Tbj51_rp_HindIII	AAG CTT TCA CCA AAA TCC ACT ACT GCC	59.6

Appendix D: Genotypes of *E. coli* strains used

E.coli strains	Genotype
OD259 / WKG190	MC4100 *ara*D139 Δ*ara*714 Δ*cbpA*::*kan dna*J::Tn*10*-42
XL1-Blue	*rec*A1 *end*A1 *gyr*A96 *thi*-1 *hsd*R17 *supE*44 *rel*A1 *lac* [F' *pro*AB *lacIq*ZΔ*M*15 *Tn*10 (Tetr)]
BL21(DE3)	*hsdS* gal [λ*c*I*ts* 857 *ind*1 Sam7 *nin*5 *lac*UV5-T7 *gene 1*]
JM109	e14–(McrA–) *recA1 endA1 gyrA96 thi-1 hsdR17* (*r*K – *m*K(+) *supE44 relA1* Δ(*lac-proAB*) [F´ *traD36 proAB lacI*qZΔ*M15*
DH5α	F- endA1 glnV44 thi-1 relA1 gyrA96 deoR nupG lacZdeltaM15 hsdR17

Appendix E: Plasmids used

Plasmid	Function	Other
pQE-1	Plasmid containing a T5 promoter; allows for cloning in-frame with a His_6-tag which is removable	Qiagen (U.S.A.)
pET-23b	Plasmid containing T7 promoter, 6 x His tag	Novagen (Germany)
pGEM-T Easy PCR cloning		Promega (U.S.A.)

All plasmids contained the β-lactamase gene (encoding Ampicillin resistance) and were selected for by including 0.1 mg / ml in cultures.

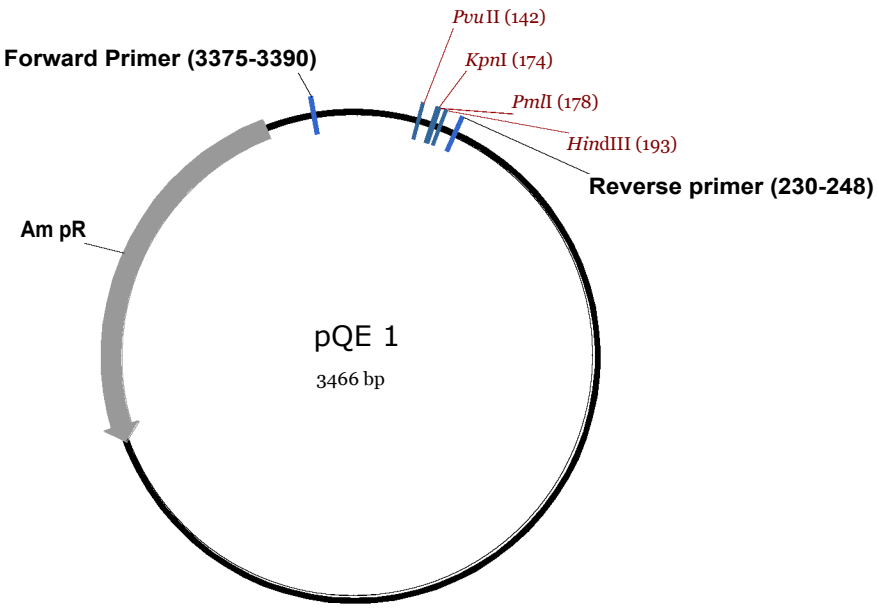


Figure E.1. Plasmid map of the pQE-1 expression vector

The β-lactamase gene (AmpR) is shown in grey, while relevant restriction sites in the multiple cloning site (MCS) are shown in black.

Appendix F: Special reagents and chemicals

Name of Chemical / Reagent	Vendor / supplier
Acrylamide	Sigma (U.S.A.)
Agarose	Whitehead Scientific (South Africa)
Acetic acid (glacial)	Merck (Germany)
Ammonium molybdate	Merck (Germany)
Ammonium persulphate	Merck (Germany)
Ampicillin	Roche (Germany)
Ascorbic acid	Sigma (Germany)
Adenosine triphosphate (ATP)	Roche (Germany)
Chemiluminiscence Western blotting kit	Amersham (U.S.A.)
β-mercaptoethanol	Merck (Germany)
Bovine serum albumin (BSA)	Roche (Germany)
Bromophenol blue	Sigma (U.S.A.)
Calcium chloride	Merck (Germany)
Coomassie brilliant blue R250	Amersham (U.S.A.)
Chloramphenicol	Roche (Germany)
Dithiothreitol (DTT)	Roche (Germany)
dNTP mix for PCR	Roche (Germany)
Ethidium bromide	Sigma (U.S.A.)
Glycerol	Merck (Germany)
Glycine	Sigma (U.S.A.)
N'-2-ethanesulfonic acid (HEPES)	Sigma (U.S.A.)
Hybond C-Extra	Amersham (U.S.A).
BioRad membrane for Western transfer	Bio-Rad (U.S.A.)
Imidazole	Sigma (U.S.A.)
Isopropyl-1-thio-D-galacopyranoside (IPTG)	pEQLabs (Germany)
Potassium chloride	Merck (Germany)
Potassium dihydrogen phosphate	Merck (Germany)
Lambda DNA	Promega (U.S.A.)
Lysozyme	Roche (Germany)
Methanol	Merck (Germany)
Malate dehydrogenase (MDH)	Roche (Germany)
Magnesium chloride	Merck (U.S.A.)
Monoclonal mouse anti-His antibody	Amersham (U.S.A.)
Ni-chelating sepharsoe	Amersham (U.S.A.)
Phenylmethylsulfonyl fluoride (PMSF)	Sigma (U.S.A.)
Polyethylene glycol (PEG) 20 000	Merck (Germany)
Ponceau S	Amersham (U.S.A.)
Restriction enzymes	New England Biolabs (U.S.A.)
Sodium chloride	Sigma (U.S.A.)
Sodium dodecyl sulphate (SDS)	Sigma (U.S.A.)
Sodium hydroxide	Merck (Germany)

SnakeskinTM pleated dialysis tubing	Pierce (U.S.A.)
Tetramethylethylenediamine (TEMED)	Sigma (U.S.A.)
Tris base	Merck (Germany)
Tryptone	Merck (Germany)
Tween-20	Merck (Germany)
Urea	Merck (Germany)
Yeast	Merck (Germany)

Kits used

Qiagen Miniprep kit	Qiagen (U.S.A.)
Fermentas gel purification kit	Fermentas (U.S.A.)

Appendix G: Uniform resource locators for bioinformatics of the Tritryp Hsp40 and Hsp70 complement

Website	URL	Use
BLAST	http://www.ncbi.nlm.gov/BLAST	Identification of similar sequences
ClustalW	http://www2.ebi.ac.uk/clustalw/	Multiple sequence alignments
Expasy	http://www.expasy.ch/tools/	Protein annotations
Prosite	http://au.expasy.org/prosite/	Protein annotations
SWISS-MODEL	http://swissmodel.expasy.org/SWISS-MODEL.html	Homology modeling of proteins
GeneDB	http://www.genedb.org/	Retrieval of *T. brucei* sequences
WOLF PSORT	http://wolfpsort.org/	Localisation prediction

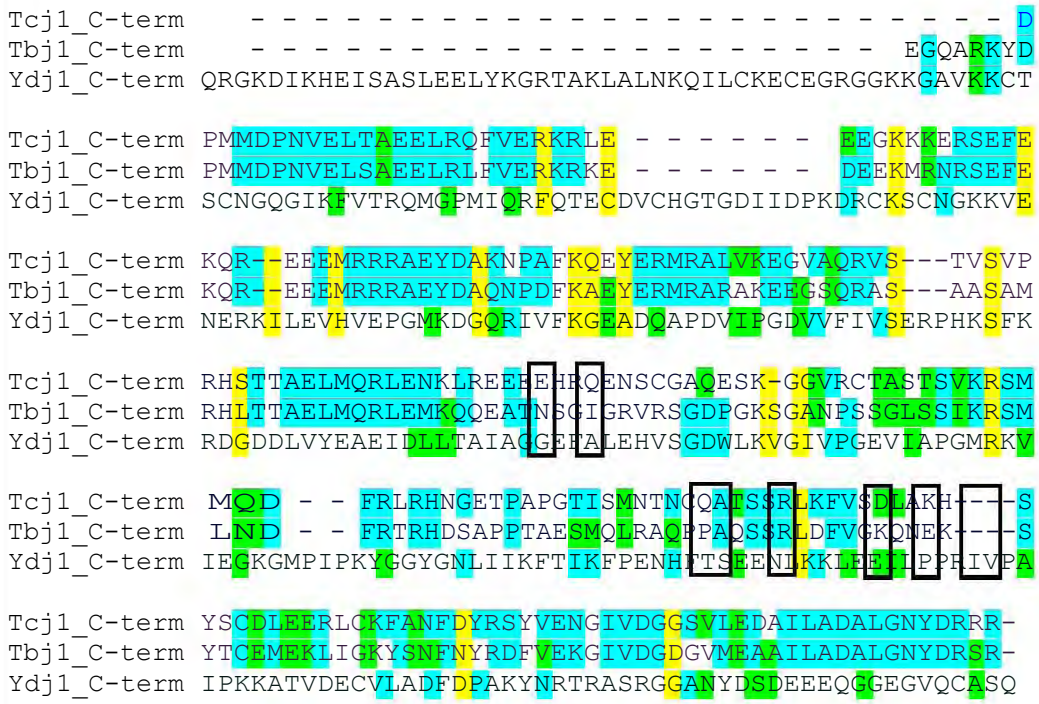


Figure H.1. Aligment of the C-terminal domains of Tbj1 and Tcj1 with that of Ydj1 from yeast

The sequence alignment of the C-terminal domains of Tbj1, Tcj1 and Ydj1 is shown. Amino acid residues that have been implicated in dimerisation of Ydj1 and their equivalent residues are highlighted with black boxes. These residues include L274, I278, F335, P336, F340, L346, L349, I352 and L353. [Similar residues] – green; [Identical residues] – yellow background with red writing; [Conservative changes] – turquoise and [Non-similar residues] – black writing on a while background. The alignment was generated using the Align X component of the Vector NTI® suite.

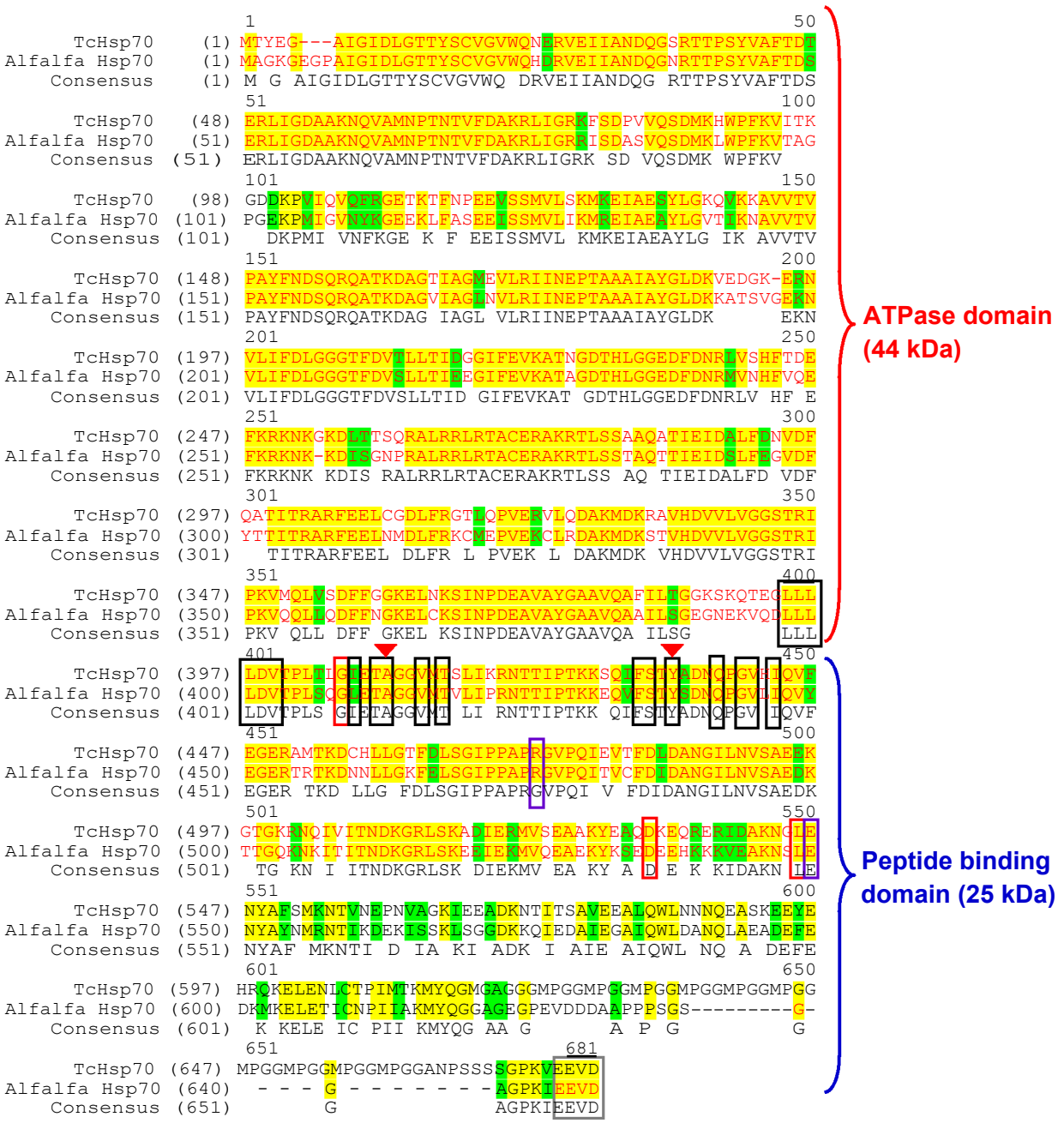


Figure H.2. Sequence alignment of TcHsp70 and MsHsp70

TcHsp70 and MsHsp70 were aligned. The linker region is outlined in black. The two proteins have an overall identity of 71 % over their full length sequence. The alignment was generated using the AlignX component of the Vector NTITM suite. Key features that are significant to Hsp70 function are highlighted: Black boxes – residues in contact with substrate in DnaK (Zhu et al., 1996); Red arrows - residues M404 and A429 comprising the arch; Red boxes – residues (G400, D526, G539) that are important for the DnaK – DnaJ interaction (Suh et al., 1998); Purple boxes – residues implicated in stabilisation of the Hsp70 lid region (Ha et al., 1997; Chang et al., 2001); the grey box shows the highly conserved EEVD region.

Appendix I: Sequences of Tbj1, Tbj47 and Tbj51 from *T. brucei* genomic DNA

The alignments below show the sequences obtained by amplification from *T. brucei* genomic DNA during the course of this study (labeled with the appropriate plasmid name) compared with the protein sequence obtained from the GeneDB database (www.genedb.org) after translation using Vector NTI software (Informax, U.S.A.).

```
 pQE1Tbj1     (1) MKHHHHHHQLMGSDVFELIGNTALYEVLGVPRTATDAEIRRAYHKLAVIY
     Tbj1     (1) ----------MGSDVFELIGNTALYEVLGVPRTATDAEIRRAYYKLAVVY
Consensus     (1)           MGSDVFELIGNTALYEVLGVPRTATDAEIRRAYHKLAVIY

 pQE1Tbj1    (51) HPDKNPEGVEVFKEVSFAHSILSDPTQREMYDNQRLRTHIEGRARKYDPM
     Tbj1    (41) HPDKNPEGVEVFKEVSFAHSILSDPTQREMYDNQRLRTHIEGQARKYDPM
Consensus    (51) HPDKNPEGVEVFKEVSFAHSILSDPTQREMYDNQRLRTHIEG ARKYDPM

 pQE1Tbj1   (101) MDPNVELSAEEL-LFVERKR-EDEEKMRNRSEFEKQREE--EMRAEYDAQ
     Tbj1    (91) MDPNVELSAEELRLFVERKRKEDEEKMRNRSEFEKQREEEMRRAEYDAQ
Consensus   (101) MDPNVELSAEEL LFVERKR EDEEKMRNRSEFEKQREE    RAEYDAQ

 pQE1Tbj1   (147) NPDFRPSTSECVLQRKRGLSVRLLRQPCAIHLTTAELMQRLEMKQQEATN
     Tbj1   (141) NPDFKAEYERMRARAKEEGSQRASAASAMRHLTTAELMQRLEMKQQEATN
Consensus   (151) NPDFK          K  S  R        HLTTAELMQRLEMKQQEATN

 pQE1Tbj1   (197) SGIGRVRSGDPGKSGANPSSGLSSIKRSMLNDFRTRHDSAPPTAESMQLR
     Tbj1   (191) SGIGRVRSGDPGKSGANPSSGLSSIKRSMLNDFRTRHDSAPPTAESMQLR
Consensus   (201) SGIGRVRSGDPGKSGANPSSGLSSIKRSMLNDFRTRHDSAPPTAESMQLR

 pQE1Tbj1   (247) AQPPAQSSRLDFVGKQNEKSYTCEMEKLIGKYSNFNYRDFVEKGIVDGDG
     Tbj1   (241) AQPPAQSSRLDFVGKQNEKSYTCEMEKLIGKYSNFNYRDFVEKGIVDGDG
Consensus   (251) AQPPAQSSRLDFVGKQNEKSYTCEMEKLIGKYSNFNYRDFVEKGIVDGDG

 pQE1Tbj1   (297) VMEAAILADALGNYDRSR
     Tbj1   (291) VMEAAILADALGNYDRSR
Consensus   (301) VMEAAILADALGNYDRSR
```

Figure I.1. Sequence alignment of the translated protein sequence of Tbj1 amplified from *T. brucei* genomic DNA with the sequence of Tbj1 from GeneDB

Alignment of the protein sequence from Tbj1 as obtained from the GeneDB database, and the protein sequence translated from the DNA sequence obtained from Tbj1 that was amplified from *T. brucei* TREU genomic DNA in the present study that was inserted in the pQE-1 expression vector. The alignment was performed using the Align X component of the Vector NTITM suite.

```
pQE1-Tbj47     (1) MKHHHHHHQLMRGITLALAPPSLLFVPKIQRRCFNVIQRGNDREVDSLFA
     Tbj47     (1)           MRGITLALAPPSLLFVPKIQRRCFNVIQRGNDREVDSLFA
 Consensus     (1)           MRGITLALAPPSLLFVPKIQRRCFNVIQRGNDREVDSLFA

pQE1-Tbj47    (51) LLGFAGDNEAHRIRRTRAELRQGFMREAMKLKDPQNDKSDAAKLEKLREA
     Tbj47    (41) LLGFAGDNEAHRIRRTRAELRQGFMREAMKLKDPQNDKSDAAKLEKLREA
 Consensus    (51) LLGFAGDNEAHRIRRTRAELRQGFMREAMKLKDPQNDKSDAAKLEKLREA

pQE1-Tbj47   (101) YQLLSNDRFRVQYAAHHYASPDASLHLLVDGGQVAANFNPEHQSFNFVDH
     Tbj47    (91) YQLLSNDRFRVQYAAHHYASPDASLHLLVDGGQVAANFNPEHQSFNFVDH
 Consensus  (101) YQLLSNDRFRVQYAAHHYASPDASLHLLVDGGQVAANFNPEHQSFNFVDH

pQE1-Tbj47   (151) AISRADMSPSSRSSSDKQRSFSDFTGQYNSVIGNTGCSTDARPYNAPEAR
     Tbj47   (141) AISRAAMSPSSRSSSDKQRSFSDFTGQYNSVIGNTGCSTDARPYNAPEAR
 Consensus  (151) AISRA MSPSSRSSSDKQRSFSDFTGQYNSVIGNTGCSTDARPYNAPEAR

pQE1-Tbj47   (201) AAINGASINFMLRISFDESVLGCTKTAVYEKNVSCQRCSGNGRMVLKRPR
     Tbj47   (191) AAINGAGINFMLRISFDESVLGCTKTAVYEKNVSCQRCSGNGRMVLKRPR
 Consensus  (201) AAINGA INFMLRISFDESVLGCTKTAVYEKNVSCQRCSGNGRMVLKRPR

pQE1-Tbj47   (251) KCPQCRGRGSTHLPSATYHIERLCTYRNGDGVTPPPKCSGCRGAGVVPGH
     Tbj47   (241) KCPQCRGRGSTHLPSATYHIERSCTYCNGDGVTPPPKCSGCRGAGVVPGH
 Consensus  (251) KCPQCRGRGSTHLPSATYHIER CTY NGDGVTPPPKCSGCRGAGVVPGH

pQE1-Tbj47   (301) TVQVPVDIRPGTTNMTACRLRGMGHDGVRGGVAGDLIVTVLVQEHRVFHR
     Tbj47   (291) TVQVPVDIRPGTTNMTACRLRGMGHDGVRGGVAGDLIVTVLVQEHRVFHR
 Consensus  (301) TVQVPVDIRPGTTNMTACRLRGMGHDGVRGGVAGDLIVTVLVQEHRVFHR

pQE1-Tbj47   (351) DGLDLHMVLPITLSTALLGGMVSVPLLHGPFCTRVPPCVRNGQQIRLSGR
     Tbj47   (341) DGLDLHMVLPITLSTALLGGMVSVPLLHGPFCTRVPPCVRNGQQIRLSGR
 Consensus  (351) DGLDLHMVLPITLSTALLGGMVSVPLLHGPFCTRVPPCVRNGQQIRLSGR

pQE1-Tbj47   (401) GVTLGGSGVLTNAEEGIDTDSSASKQEQQQRGDLYIHLLVVIPKGEELTG
     Tbj47   (391) GVTLDGSGVLTNAEEGIDTDSSASKQEQQQRGDLYIHLLVVIPKGEELTG
 Consensus  (401) GVTL GSGVLTNAEEGIDTDSSASKQEQQQRGDLYIHLLVVIPKGEELTG

pQE1-Tbj47   (451) AQRSALEQFVVEQDGNGAEGVDDITPTALKRRFRHWLPGT
     Tbj47   (441) AQRSALEQFVVEQDGNGAEGVDDITPTALKRRFRHWLPGT
 Consensus  (451) AQRSALEQFVVEQDGNGAEGVDDITPTALKRRFRHWLPGT
```

Figure I.2. Sequence alignment of the translated protein sequence of Tbj47 amplified from *T. brucei* genomic DNA with the sequence of Tbj1 from GeneDB.

Alignment of the protein sequence from Tbj47 as obtained from the GeneDB database, and the protein sequence translated from the DNA sequence obtained from Tbj47 that was amplified from *T. brucei* TREU genomic DNA in the present study that was inserted in the pQE-1 expression vector. The alignment was performed using the Align X component of the Vector NTITM suite.

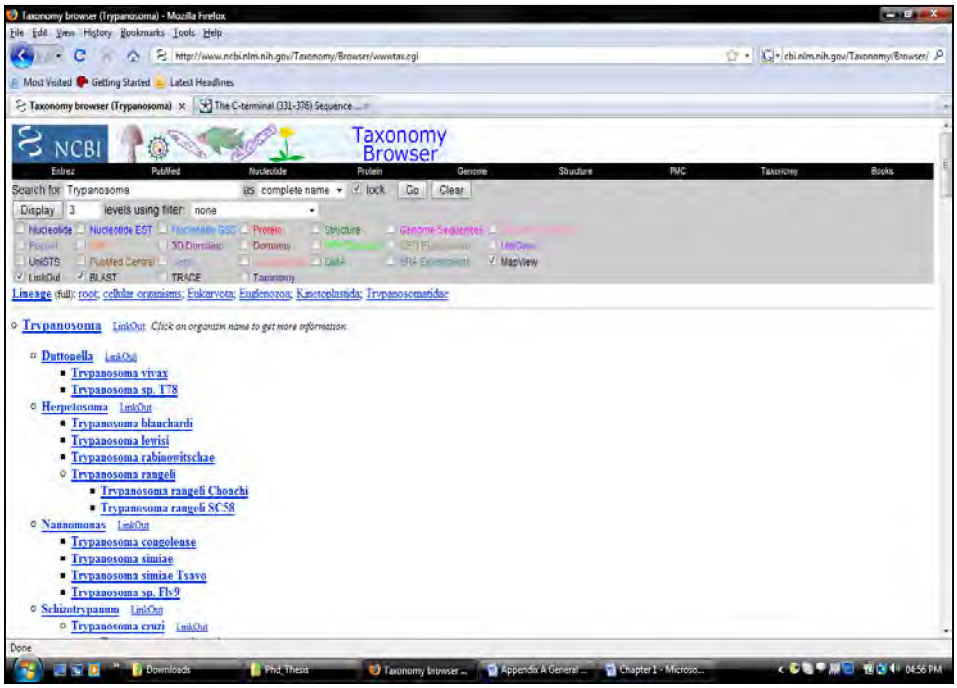
```
pQE1-Tbj51     (1) MKHHHHHHQLMETTCKEDVPQVRFPSPTSSKSMTTSPQKLSILGDITAEP
     Tbj51     (1)           METTCKEDVPQVRFPSPTSSKSMTTSPQKLSILGDITAEP

pQE1-Tbj51    (51) LSARRSFNDVSKVREPDGRWCRKFLSTYFRCELCQNVVTDPVQILPNVLL
     Tbj51    (41) LSARRSFNDVSKVREPDGRWCRKFLSTYFRCELCQNVVTDPVQILPNVLL

pQE1-Tbj51   (101) VCRRCALSRRVPSKDLQELPVSLTRAFEDLYEGQREVQLPSASRMSEIGR
     Tbj51    (91) VCRRCALSRRVPSKDLQELPVSLTRAFEDLYEGQREVQLPSASRMSEIGR

pQE1-Tbj51   (151) ASTTSDEAICKRRMVRVKRPDAVQEGSDHSQTFPLAASINGTLNASGRSS
     Tbj51   (141) ASTTSDEAICKRRMVRVKRPDAVQEGSDHSQTFPLAASINGTLNASGRSS

pQE1-Tbj51   (201) VPLLLNKALCELENKEYCMRVEIESAEASGALELKKSYNSDMRKITARSC
     Tbj51   (191) VPLLLNKALCELENKEYCMRVEIESAEASGALELKKSYNSDMRKITARSC

pQE1-Tbj51   (251) GTSKTLKTDADQKYEQAEYTLALELYTKAIELQPRDRLTRLTALYGNRSS
     Tbj51   (241) GTSKTLKTDADQKYEQAEYTLALELYTKAIELQPRDRLTRLTALYGNRSS

pQE1-Tbj51   (301) AYFMAMRYAECIADCMKVVELDPNNVKLFARAAKAAAIMGDLTAAVSHME
     Tbj51   (291) AYFMAMRYAECIADCMKVVELDPNNVKLFARAAKAAAIMGDLTAAVSHME

pQE1-Tbj51   (351) SIPEERVTPNIISEREKYKNGLDTYKRAESSFGKSDSDDAWQMLVAQFSD
     Tbj51   (341) SIPEERVTPNIISEREKYKNGLDTYKRAESSFGKSDSDDAWQMLVAQFSD

pQE1-Tbj51   (401) TIFFRIRYAESLQNQKRFLKAVEVLDVVPQERRTPKLLYIMAACLFMCGF
     Tbj51   (391) TIFFRIRYAESLQNQKRFLKAVEVLDVVPQERRTPKLLYIMAACLFMCGF

pQE1-Tbj51   (451) EHFDKARTCLEDVQQLDENCAQLLKVLNIVDEGKQKGNQYFQQKKFVAAM
     Tbj51   (441) EHFDKARTCLEDVQQLDENCAQLLKVLNIVDEGKQKGNQYFQQKKFVAAM

pQE1-Tbj51   (501) EHYTTAIGAAVNNNQILRILYCNRAASYKEVGKYREAIEDCTRTIQLDPA
     Tbj51   (491) EHYTTAIGAAVNNNQILRILYCNRAASYKEVGKYREAIEDCTRTIQLDPA

pQE1-Tbj51   (551) FSKAYARRARCHQALSDFASAIRDFKAAIKYDPNDQELPRELRSCEQSMA
     Tbj51   (541) FSKAYARRARCHQALSDFASAIRDFKAAIKYDPNDQELPRELRSCEQSMA

pQE1-Tbj51   (601) KEGERERDYYYVLGVSRNATEREIKARYRELSLRWHPDKCMSLPEEERVV
     Tbj51   (591) KEGERERDYYYVLGVSRNATEREIKARYRELSLRWHPDKCMSLPEEERVV

pQE1-Tbj51   (651) AERKFKIIVEAHTTLIDAVKRRDYDLKMEKERLTRSGGFGGFNGYSSETF
     Tbj51   (641) AERKFKIIVEAHTTLIDAVKRRDYDLKMEKERLTRSGGFGGFNGYSSETF

pQE1-Tbj51   (701) RGHSNRFRQGSSGFW
     Tbj51   (691) RGHSNRFRQGSSGFW
```

Figure I.3. Sequence alignment of the translated protein sequence of Tbj51 amplified from *T. brucei* genomic DNA with the sequence of Tbj1 from GeneDB

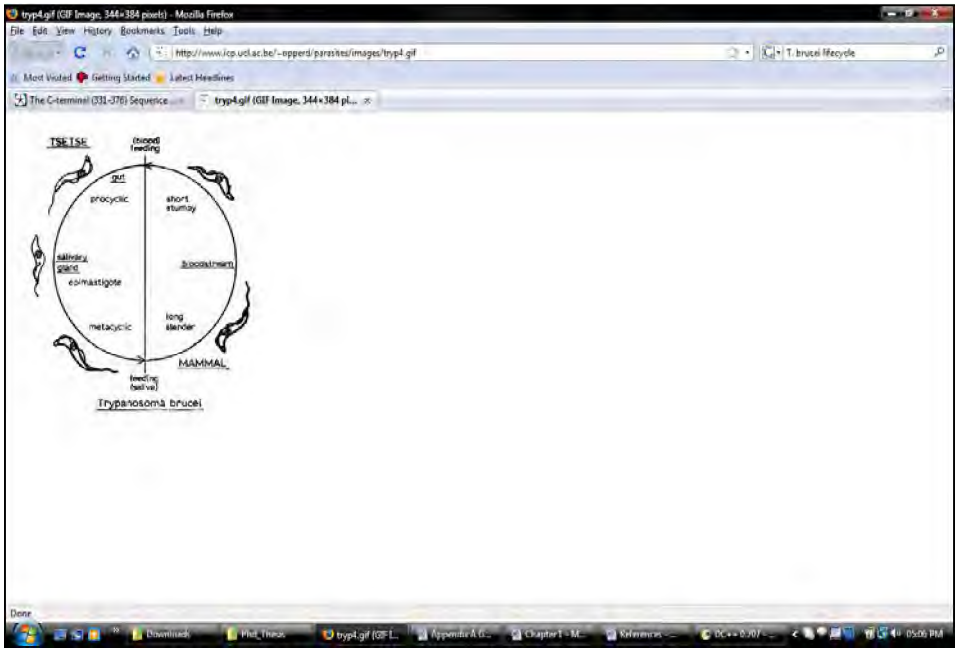
Alignment of the protein sequence from Tbj51 as obtained from the GeneDB database, and the protein sequence translated from the DNA sequence obtained from Tbj51 that was amplified from *T. brucei* TREU genomic DNA in the present study that was inserted in the pQE-1 expression vector. The alignment was performed using the Align X component of the Vector NTITM suite.

Appendix J: Full copies of internet references used

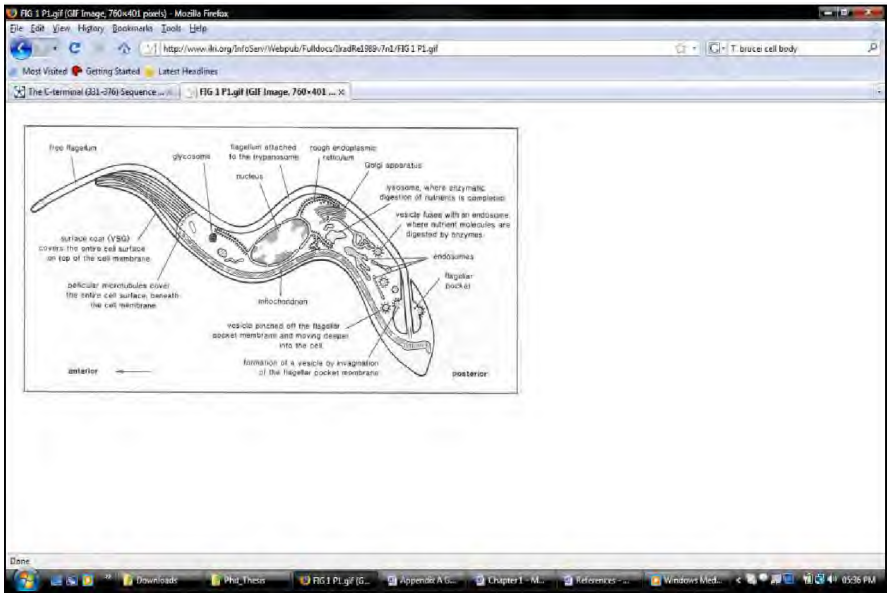
Internet 1: http://www.ncbi.nlm.nih.gov/Taxonomy/Browser/



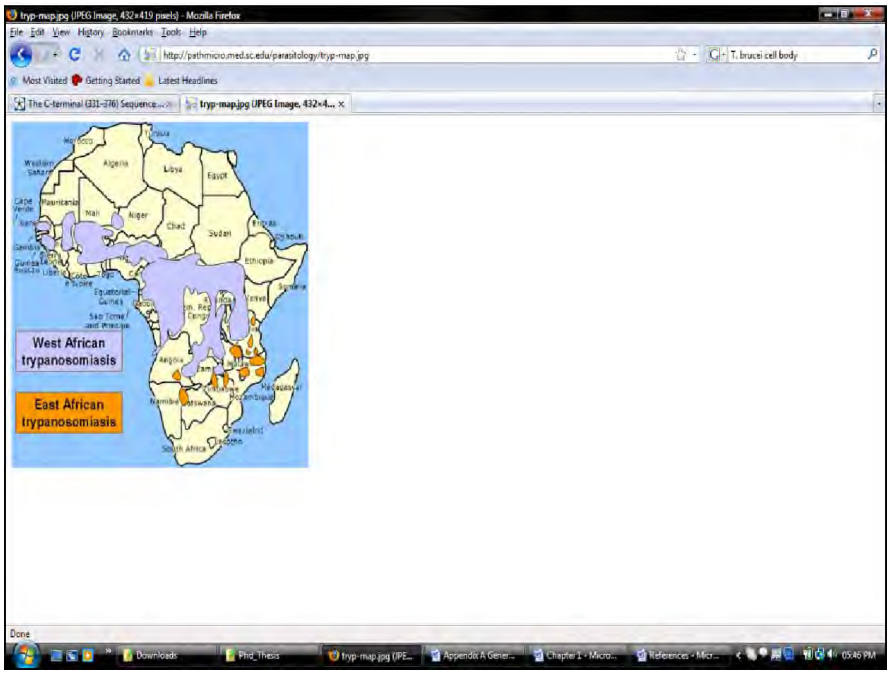
Internet 2: http://www.icp.ucl.ac.be/~opperd/parasites/images/tryp4.gif



Internet 3: http://www.ilri.org/InfoServ/Webpub/Fulldocs/IlradRe1989v7n1/FIG%201%20P1.gif



Internet 4: http://pathmicro.med.sc.edu/parasitology/tryp-map.jpg



Appendix K: Summary of Procheck results for homology modelling

The table below shows the results obtained when the homology models generated during the course of the present work were analysed by Procheck software for Windows NT.

Table K.1. Results of Procheck analysis performed on homology models

Model	Planar gps	M/C bond lengths	M / C bond angles
Tbj31	100 %	100 %	84 %
TcHsp70	100 %	89.1 %	75.1 %
Tbj51	100 %	92.5 %	78.8 %
Tbj1	100 %	89.8 %	76.3 %
Tcj1	100 %	92.5 %	76.9 %
TbHsp70	100 %	88.7 %	76.4 %
MsHsp70	100 %	85.9 %	70.8 %