A step forward in defining Hsp90s as potential drug targets for human parasitic diseases

A mini-thesis submitted in partial fulfilment of the requirement for the degree of

MASTER OF SCIENCE OF RHODES UNIVERSITY

by

Coursework/Thesis

in

Bioinformatics and Computational Molecular Biology in the Department of Biochemistry, Microbiology and Biotechnology Faculty of Science

by

Ngonidzashe Faya

December 2013

ABSTRACT

Parasitic diseases remain a health burden affecting more than 500 million people worldwide with malaria having the highest mortality rate. The parasites can be transferred to the human bodies either through the mouth by ingestion of contaminated food and water or through the skin by bug bites or direct contact to environments harbouring them. Epidemiological control seems to be impossible since there is failure to control the insect vectors as well as practice of hygiene. Therefore, this has led to the development of a number of vaccines, chemotherapy and disease control programs. However, parasites have increasingly developed resistance to traditionally used anti-parasitic drugs and due to that fact; there is need for alternative medication for parasitic diseases. Heat shock protein 90 (Hsp90) facilitates the folding of proteins in all living cells and their role is more important to parasites because of their environmental changes, from vector to host. Hsp90s play a major role; therefore this justifies the need for a deeper analysis of the parasitic Hsp90s. Recent studies have revealed that, the *Plasmodium sp.* Hsp90 has an extended linker region which increases the protein's affinity for ATP and its inhibitors. Therefore we hypothesize that there are also significant features in other parasitic Hsp90s which would lead to Hsp90 being defined as potential drug targets. In the present study an attempt was made to gain more insight into the differences in primary structure of human and parasitic Hsp90s. The sequences were retrieved from the NCBI database and analysis was done in three groups basing on the localization of the Hsp90. The physicochemical properties were calculated and in every group, the protozoan Hsp90s showed significant differences when compared to the human orthologs. Multiple sequence alignments (MSA) showed that endoplasmic reticulum Hsp90s have an extended region in the middle domain indicating their ability to bind to a unique subset of client proteins. Sequence identities between the human and parasites showed that the protozoan Hsp90s are less related to the human Hsp90s as compared to the other parasites. Likewise, motif analysis showed the trypanosomatids and apicomplexan groups have their own unique set of motifs and they were grouped together in the phylogenetic analysis. Phylogenetic analysis also showed that, the protozoan Hsp90s forms their own clades in each group while the helminths did not form in endoplasmic reticulum group. In this study, we concluded that, Hsp90 can be a potential drug target for the protozoan species and more specifically those from the apicomplexan and trypanosomatids groups.

DECLARATION

I, **Ngonidzashe Faya**, declare that this thesis submitted to Rhodes University is a master piece of my own and has not been submitted elsewhere for a degree.

Signature……

Date……

ACKNOWLEDGEMENTS

To begin with, I thank the LORD Almighty for taking me this far.

My greatest gratitude goes to my supervisor, Prof. Özlem Tastan Bishop for tirelessly working with me to make this project a success. The commitment and motivation throughout will never go unappreciated. Thank you!!!

I also want to thank Rhodes University and financial aid office for the opportunity to study and the provision of a scholarship (Rhodes University Prestigious Scholarship).

I also want to thank my colleagues at Rhodes University; Research Unit in Bioinformatics (RUBi), for the support, motivation, assistance and friendship.

I also want to express my heartfelt gratitude to my family and close friends for the love, support, and motivation and above all, believing in me.

Last but not least, I want to specifically thank the following people who sacrificed their times in helping me to be able to complete the project successfully:

- Dr Adrienne Edkins
- Thomas Musyoka Mutemi
- Rowan Hatherley
- Daphine Mundondo

DEDICATION

This thesis is dedicated to the loving memory of my dad,

Jonathan Faya

Table of Contents

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

3D – Three Dimension

BLAST - Basic Local Alignment Tool

BLASTP - Basic Local Alignment Tool for Protein sequences

ER - Endoplasmic Reticulum

GA - Geldanamycin

GRP – Glucose Regulated Protein

HSP90 - Heat Shock Protein 90

MAFFT - Multiple Sequence Alignment (based on) Fast Fourier Transform

MAST - Motif Alignment & Search Tool

MEGA - Molecular Evolutionary Genetic Analysis

MUSCLE - MUltiple Sequence Comparison by Log – Expectation

PDB - Protein Data Bank

PROMALS - PROfile Multiple Alignment with Local Structure

TRAP1 – Tumor necrosis factor Receptor-Associated Protein 1

SYMBOLS USED

α - Alpha

β - Beta

LIST OF WEB SERVERS AND WEB-BASED APPLICATIONS

BLAST - www.ncbi.nlm.nih.gov/BLAST/

InterProScan - http://www.ebi.ac.uk/Tools/pfa/iprscan/

MAFFT - http://mafft.cbrc.jp/alignment/server/index.html

MEME - http://meme.nbcr.net/meme/cgi-bin/meme.cgi

MUSCLE - http://www.ebi.ac.uk/Tools/msa/muscle/

PROMALS3D - http://prodata.swmed.edu/promals3d/promals3d.php

Scan PROSITE - http://prosite.expasy.org/scanprosite/

SIAS - http://imed.med.ucm.es/Tools/sias.html

CHAPTER ONE

1. Introduction

1.1. Parasites and Parasitic Diseases

Parasites are organisms that get their food from the host they live in or on, either at the expense of that host organism or not. Three main classes of parasites have been found to cause diseases in humans and these are protozoa, ectoparasites and helminths. Protozoa are one-celled organisms that multiply in humans as their survival method and at the same time causing infections. They are transferred to humans through an arthropod vector and basing on their locomotion method, the protozoans are sub-divided into four groups (Table 1.1) (Honigberg et al. 1964). Helminths multicellular organisms are large to be seen through a naked eye and there are three main groups which are the flatworms, thorny-headed worms and roundworms. The ectoparasites are also multicellular that attach or burrow into the skin and feed on the host's blood for survival.

Table 1.1:The classification of human disease-causing protozoa basing on their mode of locomotion.

Protozoan group	**Locomotion**	**Example**
Sporozoa	Adult stage is not motile	*Plasmodium* sp.
Sarcodina	amoeba	*Entamoeba* sp.
Ciliophora	flagellates	*Balantidium* sp.
Mastigophora	ciliates	*Leishmania* sp.

Parasitic diseases have caused a tremendous health burden in Africa and malaria is on top with the most deaths globally (Fig 1.1). Malaria takes approximately 90% share of deaths caused by parasitic diseases each year (WHO, 2008). The rest of the diseases share the remaining 10% and some like filariasis cause a small number of deaths per year which when averaged produce a very small number that is approximately equal to zero (Date et al. 2007a). Vector strategy is the most used by human parasites and Table 1.2 shows a summary of the parasite, disease it cause and its vector. The *Leishmania sp.* that is responsible for causing leishmaniasis, a chronic disease that affects millions world-wide, is used as an example to illustrate how the vector-host system works.

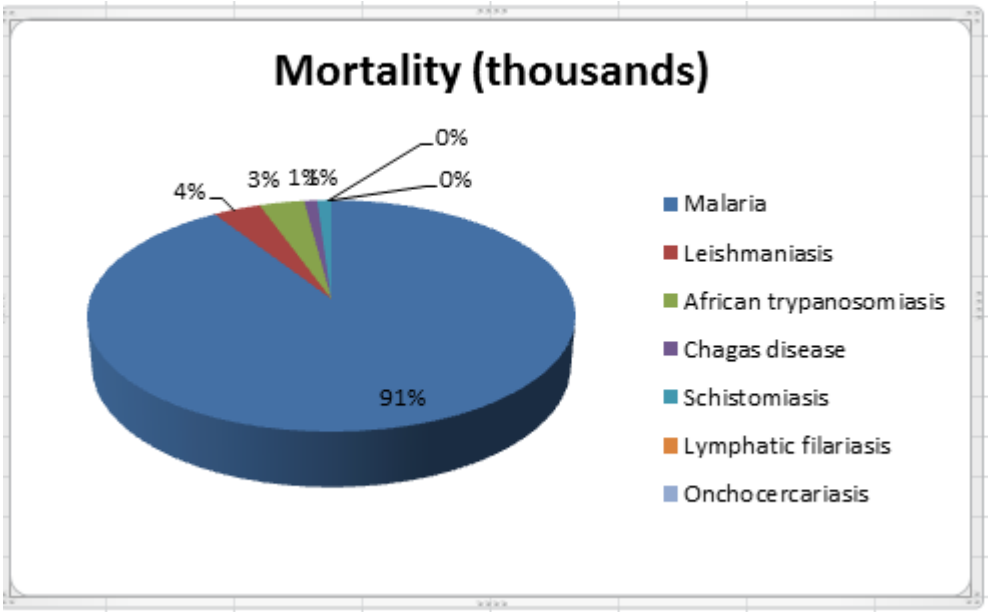


Figure 1.1: A diagrammatical representation of the deaths caused by parasitic diseases globally per year. (Date et al. 2007b) and modified accordingly. 91 % of deaths caused by parasitic disease are due to malaria while lymphatic filariasis and onchocercariasis can be contained and are not a threat to human life as they have 0 percent mortality rate.

Sandflies are the vectors for the protozoa (Fig 1.2) and the protozoa exist as promastigotes in the gut of the female sand-fly. The parasites are transmitted to the mammalian tissues during feeding where they undergo transition stages to amastigote forms. These transitions occur in the macrophages and multiplication occurs by binary fission (Wirth et al. 1986).

Treatments of parasitic diseases in general, rely solely on chemotherapy as there are challenges in availability of drugs and resistance to the existing drugs. Therefore, there is a need for alternative and cheaper anti-parasitic drugs. Parasites encounter stress impacts due to physiological changes when they are transferred from the vector to the host. Proteins tend to denature in situations like these, therefore molecular chaperones tend to play a major role in the survival of the parasite by maintaining the 3D structure of its proteins. Because of their important role, molecular chaperones are worthy to be analyzed more as drug targets for parasitic diseases.

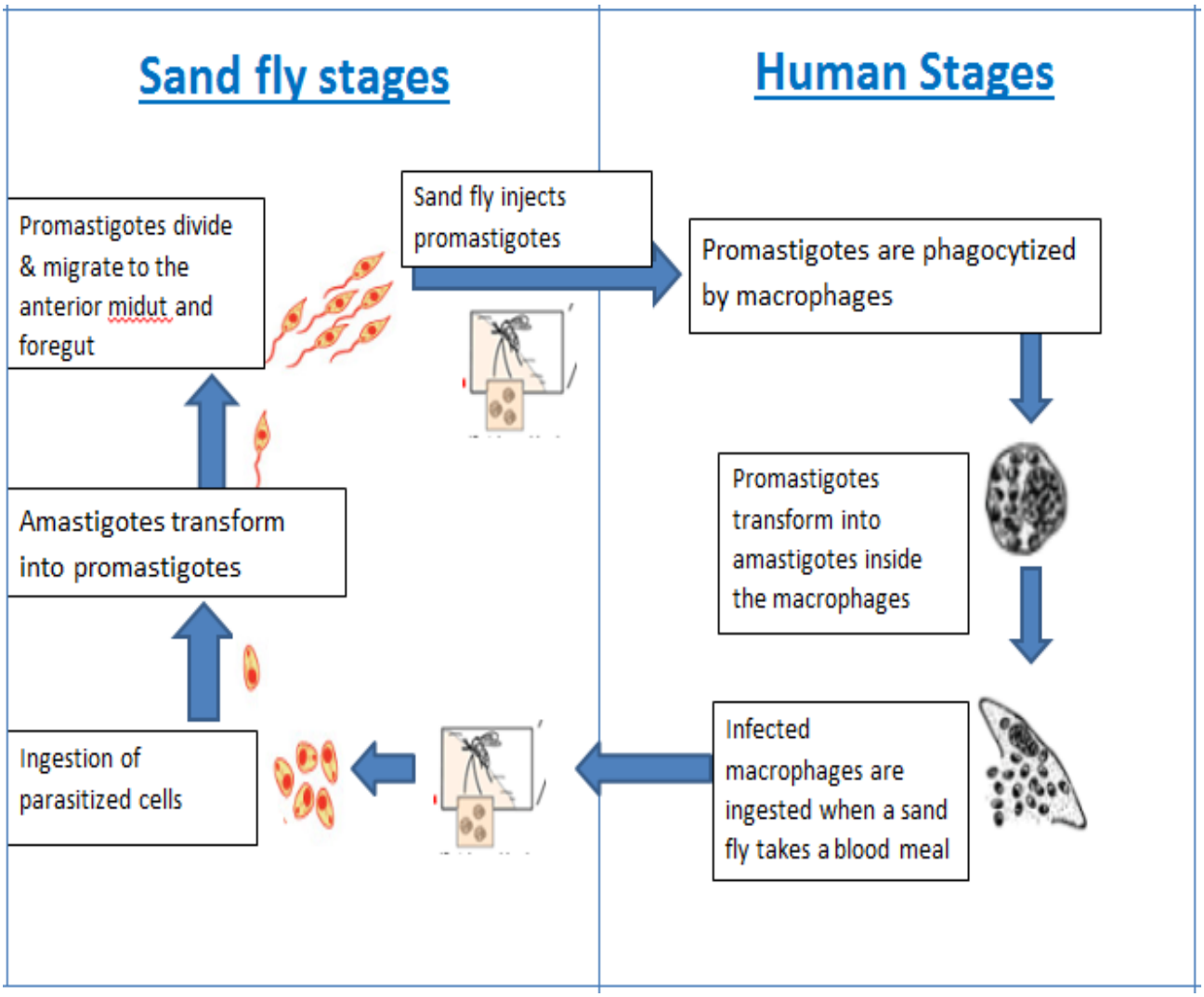


Figure 1.2: Diagrammatical representation of the *Leishmania* life cycle showing both the human and the sand fly stages

Table 1.2.A summary of the human parasitic diseases and vectors responsible for submitting the parasite

Type of parasite	**Disease**	**Parasite**	**Vector**
Protozoa	African sleeping sickness	*T. brucei* and *T. gambiense*	Tsetse fly
	Malaria	*Plasmodium sp.*	Mosquito
	Chagas disease	*T. cruzi*	Kissing bugs
	Leishmaniasis	*Leishmania sp.*	Sand fly
	Babesiosis	*Babesia sp.*	Tick
Helminths	Schistosomiasis	*Schistosoma sp.*	Fresh water snails
	Roundworm-lymphatic filariasis	*B. malayi*	Arthropods
	Filariasis	*Loa loa*	Horse fly
	Elephantiasis	*W. bancrofti*	Mosquito
Ectoparasites	Pediculosis	*P. humanus*	Head-to-head contact
	Scabies	*S. scabiei*	Skin-to-skin contact e.g. sexual activity

1.2. Molecular Chaperones

Molecular chaperones are a set of ubiquitously conserved proteins that are responsible for the folding or assembly of other macromolecular structures but are not included in these structures when the process of folding or assembly is complete. Their major function is to turn newly synthesized polypeptides into functional structures and re-fold denatured proteins back to a functional state (Hartl et al. 1996). In general, chaperones do not give the polypeptides their 3D shape but they help them to efficiently find their correct structure which is encoded from the amino acid sequence (Dobson & Karplus 1999). The majority of molecular chaperones do not behave as true catalysts as they do not increase the rate of folding. They are involved in preventing incorrect interactions of sticky protein folding intermediates and making sure that the

correct orientation is achieved. Therefore in this way they increase the yield of folded protein but not the rate (Hartl et al. 1996).

Chaperones are essential to the cells throughout their entire life-time but they play a vital role when they increase their levels when the cell suffers stress (Sõti et al. 2005). They are also involved in aetiology of diseases (Welch and Brown, 1996). Molecular chaperones are still classified by their molecular weights and the major families are Hsp60, Hsp70, Hsp90, Hsp100 and small Hsps which have a molecular weight between 12 and 43 kDa (Laudanski & Wyczechowska 2006). Hsp70 and Hsp90 are responsible for protein holding while Hsp60 and small Hsps focus more on folding hence given the name chaperonins. The Hsp70 family performs more chaperone functions than other chaperones since they are involved in stabilizing unfolded newly synthesized proteins prior to their assembly into multi-molecular complexes; they are also involved in rearrangement of protein oligomers and the resolution of protein aggregates (Becker & Craig 1994). Hsp60s differ structurally from the Hsp70s but they share functional features in the sense that Hsp60s are also involved in the folding of newly synthesized polypeptides. The Hsp90s differ from the rest of the Hsp family both structurally and functionally. Hsp90 is involved in regulating the function of the folded proteins by binding to them in an ATP dependent mechanism which is also shared by Hsp70 and Hsp60 (Welch and Brown, 1996).

1.2.1. Hsp90 family

At least three different Hsp90 isoforms have been identified in a large number of eukaryotes with the cytosolic member being the most abundant (Table 1.3). The presence of TRAP1 and Grp94 in a cell depends on the complexity of the organism. Yeast cells do not have any homolog in the ER but cytosolic Hsp90 and TRAP1 are present. The cytosolic Hsp90 and the Grp94 are believed to have resulted from gene duplication that has occurred in the early stages of eukaryotic evolution because of their identities that are approximately 50% (Gupta, 1995). TRAP1 differs from the cytosolic Hsp90 only in the N- and C-termini but it resembles the HtpG protein (Song *et al.* 1995), both structurally and in size. Other than TRAP1 and Grp94, the human cell contains

two isoforms; Hsp90α and Hsp90β in the cytosol, making the cells to have four homologs. These two cytosolic proteins are a result of second gene duplication and are approximately 76% identical (Krone and Sass, 1994). Hsp90β is slightly larger than Hsp90α which makes it to be sometimes denoted as Hsp86 while Hsp90α is denoted as Hsp84.

Table 1.3: Members of the Hsp90 family with their locations in a cell

Name	**Localization**	**References**
Hsp90α/β	Cytosol	(Prohászka et al. 1998)
Grp94/96	Endoplasmic Reticulum	(Prohászka et al. 1998)
Trap1/Hsp75	Mitochondria	(Felts 2000)
HtpG	Bacterial cytosol	(Stechmann & Cavalier-Smith 2003)

1.2.2. Hsp90 structure

The Hsp90 chaperone exists as a homodimer and each monomer is made up of three domains (Fig 1.3); the N-terminal ATP binding domain, C-terminal dimerization domain and the middle domain which is responsible for client binding. These domains are highly conserved and depending with species the Hsp90 contains two highly charged regions; the 1st linker joining the C-terminal and middle domains and 2nd linker joining the C-terminal and the middle domain (Louvion et al. 1996). The two charged regions probably play a role in the binding properties of the chaperone but they disappear in the TRAP1/Hsp75 structure. To support the role of the charged domains, studies have shown that the cytosolic Hsp90 has a higher affinity for positively charged or hydrophobic proteins when compared to TRAP1 (Prohászka et al. 1998). The chaperone is hydrophobic, and under stress conditions, its hydrophobicity increases (Yamamoto et al. 1991).

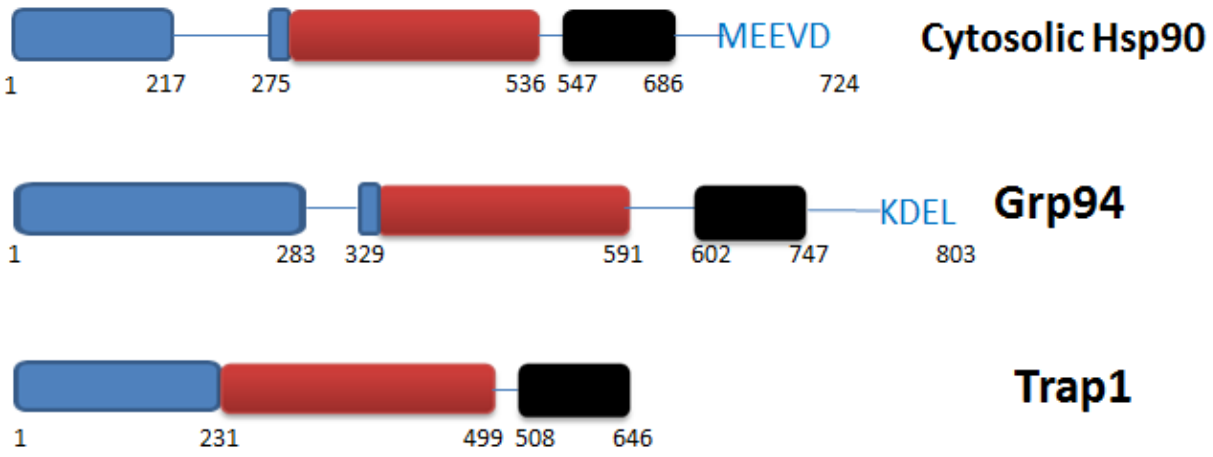


Figure 1.3: Domain architecture for human Hsp90 family members, the blue colour represents the N-terminal while the middle domain is in red and the C-terminal is black. Adapted from (Krukenberg et al. 2011), and modified accordingly.

The N-terminal domain contains the binding site for ATP and has shown structural similarity with DNA gyrase B hence is grouped under the GHKL superfamily with a novel ATP binding fold (Dutta et al. 2000). The N-terminal has a stretch of around sixty conserved amino acids (Nagamune et al. 1997)that are responsible for ATP binding (Young et al. 1997). Structural conformations in the N-terminal domain that may affect ATP binding or hydrolysis stop the function of the protein both *in vivo* and *in vitro* (Grenert 1999). The charged linker region is involved in the association of the chaperone and the client proteins such as steroid receptors (Dao-Phan et al. 1997) and CK-II (a protein kinase) (Miyata and Yahara, 1995). Genetic studies have revealed that the charged region is not essential for the survival of the chaperone (Louvion et al. 1996) but increases the affinity for ATP and its inhibitors (Pallavi et al. 2010). The highly conserved middle domain plays two major roles that are binding of substrates (Park et al. 2011; Meyer et al. 2003) and activating the ATP hydrolysis by binding to co-chaperones p23 and Aha1 (Pearl et al. 2006; Ali et al. 2006; Meyer et al. 2004). The C-terminal is the dimerization domain and is also suggested to play a role in substrate binding (Hagn et al. 2011).

1.2.2.1. Cytosolic Hsp90

The cytosolic Hsp90 are characterized by the C-terminal motif MEEVD that plays a major role in co-chaperone binding (Meng et al. 1996). This binding is made possible by the presence of

tetratricopeptide repeat (TPR) domains in co-chaperones that bind the Hsp90 at its MEEVD motif. The cytosolic Hsp90 is structurally flexible, since under apo conditions, its structure is in a "V" shaped conformation(Shiau et al. 2006) (Fig 1.4) while in pH dependent manner, alternations between open and closed conformations are observed (Krukenberg et al. 2011). The closed state structure of the cytosolic Hsp90 has been observed to show high similarity to the ER homolog, Gpr94 (Dollins et al. 2007). Based on the homologs that have been structurally examined; i.e. Hsp90α, Grp96, TRAP1 and HtpG, the apo state conformational tends to be universal (Krukenberg et al. 2009) and this shows that the flexibility of the protein is important when it comes to function. The way the chaperones interact with client proteins is mainly determined by the flexibility of the chaperone.

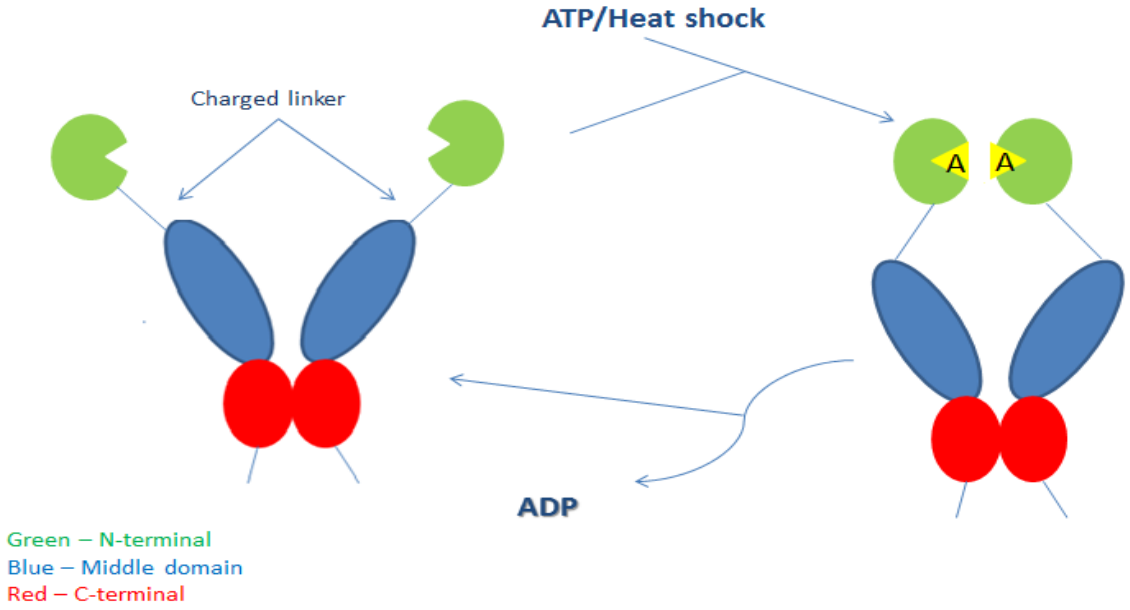


Figure 1.4: The Hsp90 structure in an open and closed conformation. Under stress condition, ATP binds to the N-terminal domain of the open conformation to form a closed conformation.

1.2.2.2. ER Hsp90

Grp94 also forms dimers and there is a distinct structural difference between the cytosolic Hsp90 and Grp94 in their N- and C- termini where approximately the first fifty amino acids (fig 1.5) are distinct and from this the conformational change in cytosolic Hsp90 performed by the first twenty four amino acids in dimerization of the N-terminal (Richter et al. 2006) is likely to be different in Grp94 (Fig 1.5). The charged linker is short in Grp94 and is E/D rich followed by a

poly-K stretch as illustrated by Fig 1.6. The aligned sequences of the cytosolic Hsp90 and Grp94 indicate that the WDWE Trp zipper motif in strand 9 is missing in the cytosolic Hsp90 (Fig 1.6).

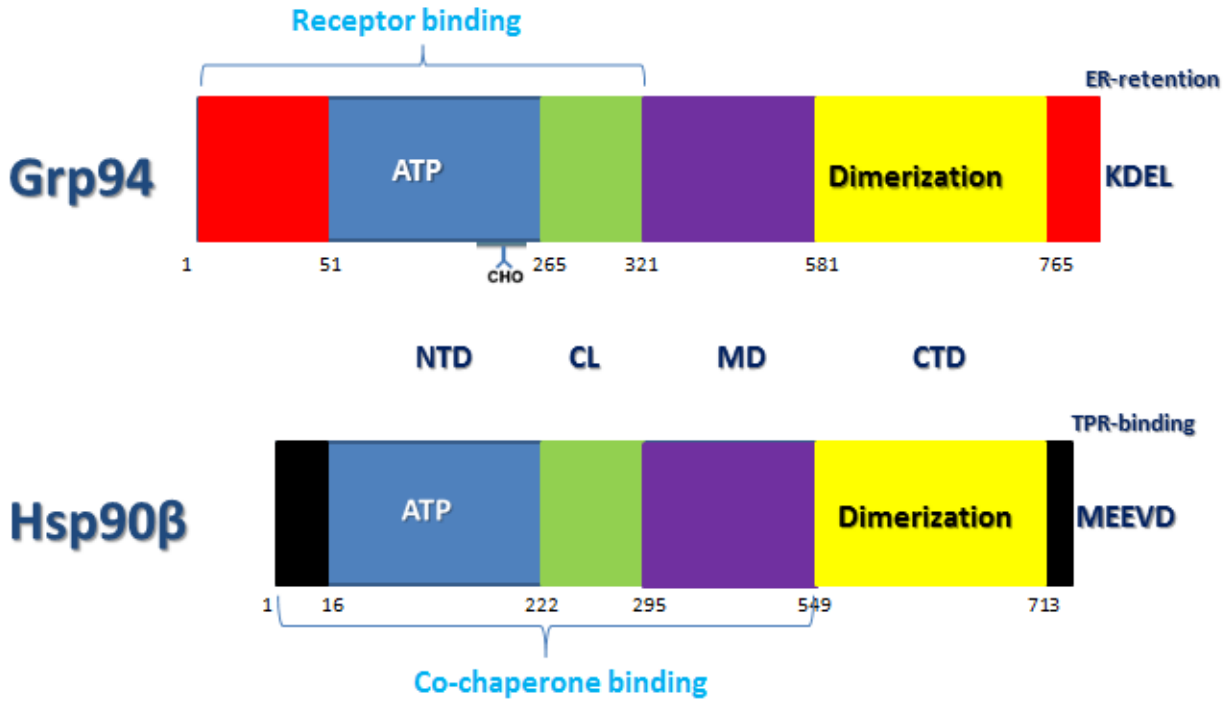


Figure 1.5: Diagrammatical representation of the structural features of human Hsp90β and Grp94. Blue, N-terminal domain (NTD), Green, Charged linker (CL), purple, middle domain (MD), yellow, C-terminal domain (CTD), red and black, sites with distinct differences. KDEL is the ER retention ligand on the Grp94 while MEEVD binds TRP contain proteins to the Hsp90. Adapted from Marzec et al. 2011 and modified accordingly.

The compounds that bind competitively to the ATP binding site for cytosolic Hsp90 have similar affinity for Grp94 (Shahinas et al. 2010) and so far all the compounds that inhibit the site for Hsp90 have been found to have similar effects in the Grp94. However, there is one compound that binds specifically to Grp94 but does not bind to the Hsp90 site. This is an adenine analogue, NECA (Soldano et al. 2003). NECA's specificity is caused by a modified lid on the binding site that is formed by an insertion of about five amino acids between helix 4 and helix 5 in the Grp94.

Figure 1.6: The aligned sequences of the cytosolic Hsp90 and the Grp94 showing strand 8 and strand 9. The blue colour shows similar residues, red, basic residues, purple shows Asn and Gln. Green shows aromatic residues, orange shows aliphatic residues, brown shows proline and black shows small side chain residues. The Grp94 has Leu substituted by Pro at position 300 and is highlighted in yellow. The Trp zipper motif within strand 9 is highlighted in red.

The C-terminal of the Grp94 provides a homo-dimerization interface just like the cytosolic Hsp90 and is made possible by its interaction with the middle domain (Yamada et al. 2002). The C-terminal ends with an approximately fifty five residue acidic region, which is longer by about twenty residues when compared to the cytosolic Hsp90's terminus and is more acidic in Grp94s. The C-terminal contains KDEL sequence as the terminus instead of the MEEVD sequence of the cytosolic Hsp90 (Fig 1.6) and the KDEL terminus specifies ER retention. Instead of the KDEL peptide binding to co-chaperones like the MEEVD, it is a binding site for an intracellular receptor responsible for retrieving the proteins that have escaped the ER. The Grp94 is a glycoprotein unlike cytosolic Hsp90 that is glycosylated at Asn-196 (Qu et al. 1994) where the oligosaccharide with 2 N-acetyl-glucosamine residues and 8 mannose binds. The availability of the sugars determines the presence of the attached oligosaccharide (Wearsch and Nicchitta, 1996).

1.2.2.3. Mitochondrial Hsp90

TRAP1 has been found to be structurally similar to yeast Hsp90 and like other Hsp90s, it forms a tight homodimer (Leskovar et al. 2008). Phylogenetic analysis indicates that the protein evolved from the common ancestor as other Hsp90, as will be shown in the next section. However, a

study on protein features suggests that TRAP1 evolved earlier than the other Hsp90s (Chen et al. 2005). The cytosolic and ER Hsp90s contain a C-terminus conserved functional motif (MEEVD and KDEL respectively). Sequence analysis revealed that this motif is lacking in the TRAP1 C-terminus (Leskovar et al. 2008).

1.2.3. Hsp90 function

Most of the studies regarding the Hsp90 chaperone function have been done on the cytosolic Hsp90s and the requirement for the chaperone's function is still restricted to a number of client proteins that are unstable and those that require help to properly fold. Based on the studies so far, the Hsp90s bind to proteins that are partially folded or in a nearly folded state (Street et al. 2011). The cytosolic Hsp90 in eukaryotes is associated with co-chaperones that aid in client recruitment and folding and regulate conformation dynamics of the chaperone. Co-chaperones associations differ with the organism as well as the client (Johnson & Brown 2009) e.g. the *in vitro* reconstitution of the progesterone receptor require Hsp70, Hsp40, Hop and p23 (Kosano 1998) while the kinase Chk1 does not require p23 but Cdc37 instead (Arlander et al. 2006).

Cytosolic and ER Hsp90s have been observed to have a conserved cooperation with Hsp70s during client folding (Wegele et al. 2006) and Hsp70s are also ATP dependent chaperones that are highly conserved throughout all species (Mayer & Bukau 2005). Hsp70s helps with folding of unfolded or partially unfolded polypeptides through an ATP-regulated cycle and also functions with Hsp40s that modulates their ATPase activity. Most of the proteins folded by the Hsp70 system are directed to the Hsp90 for completion and the reasons for this requirement are not yet understood. Hsp70 is a very important co-chaperone for Hsp90 client folding since the mutations of Hsp40, an Hsp70 co-factor, in *S. cerevisiae* is known to disrupt the activity of Hsp90 client system (Kimura et al. 1995; Mayer & Bukau 2005). Besides Hsp70, at least eight other co-chaperones are involved in helping the Hsp90 in the complete folding process of the steroid receptor. Fig 1.7 shows the steps likely to happen in this process of complex association where Hsp70 binds the target protein together with its co-chaperone Hip. Hop a very important co-chaperone that links Hsp70 and its co-chaperone to Hsp90 which also comes with its co-chaperone such as p23, Cyclosporin A-binding immunophilin (Cyp) 40 and FKBP52. When p23 binds to the Hsp90-Hsp70-Hop complex, Hsp70, Hip and Hop dissociate from the complex allowing completion of folding (Dittmar et al. 1997). The target protein is then released and is

already activated for its function. The folding pathway of kinases is not that clear but it has been elucidated that CDC37 recognizes the kinase client proteins(Stepanova et al. 1996).

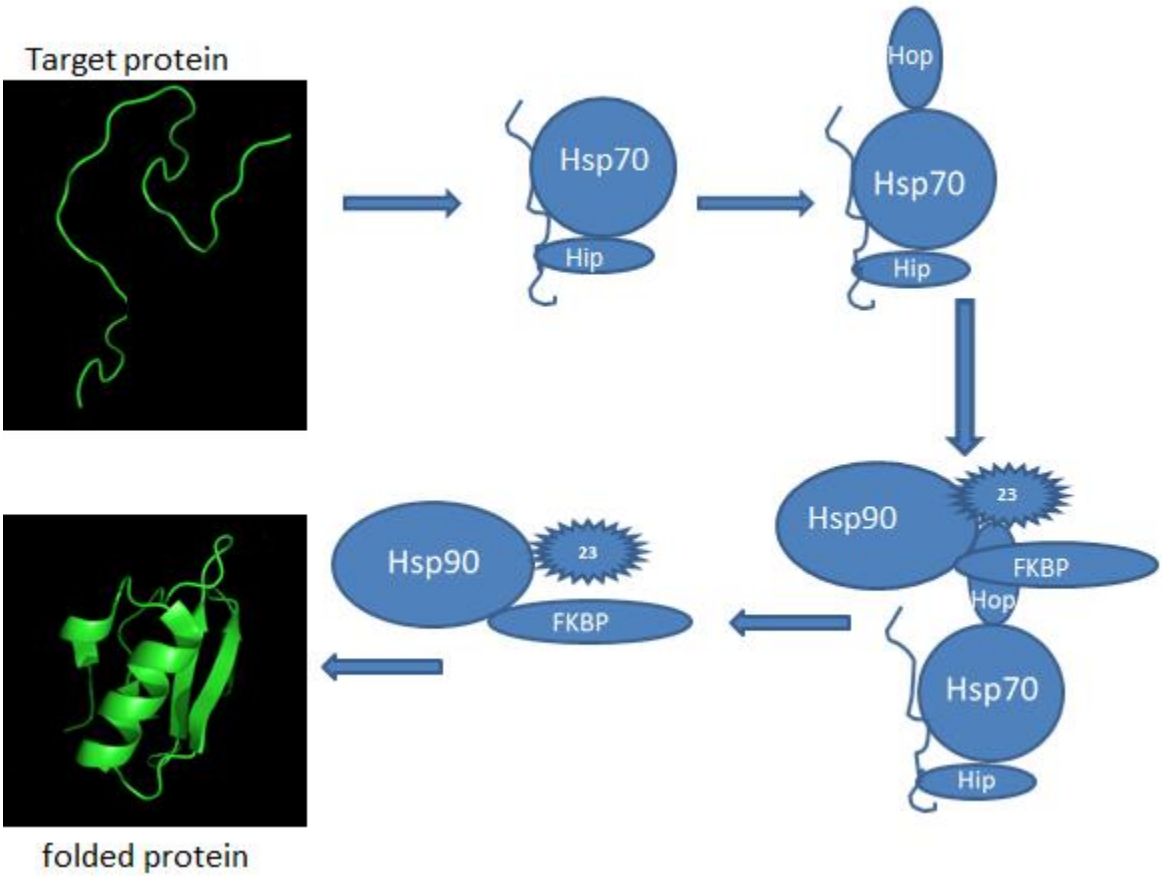


Figure 1.7: The folding process of the steroid receptor by the chaperone complex. Adapted from (Prohászka et al. 1998) and modified accordingly.

The folding of newly synthesized glucocorticoid receptors showed that Hsp90 associate with Hsp40, Hsp70 and other co-chaperones (Smith et al. 2008) to form a multi-subunit complex. The unfolded receptor binds to both Hsp70 and Hsp40 and this polypeptide is transferred to the Hsp90 by another co-chaperone called Hop which through its TPR domains binds both Hsp90 and Hsp70. The intermediate complex is formed when both Hop and the client are bound to the Hsp90 and Hsp90 requires ATP to further process the client. The binding of ATP is aided by other co-chaperones which contain a TPR domain such as Cyp40 and another co-chaperone binding to the dimerized amino-termini (Richter et al. 2006) Each client has its specific co-chaperones has been observed with kinases where Cdc37 binding is exclusive to kinases (Caplan

et al. 2007)TRAP1 however, shows functional differences as it does not bind to co-chaperones, p23 and Hop (Chen et al. 2005).

The fact that the isoforms of the cytosolic Hsp90 share high levels of similarity does not mean that they have identical functions. The Hsp90α and Hsp90β are 85% identical but distinct functions have been identified. Hsp90α has been identified to be secreted extracellularly (Tsutsumi et al. 2008) and is correlated with tumor invasiveness while Hsp90β has specific role in the anti-apoptotic functions of CpG-B oligodeoxynucleotide and Bcl2 (Cohen-Saidon et al. 2006). The cytosolic isoforms form mostly homodimers and there are slight differences in the C-terminal dimerization domain which explains why the Hsp90β is less stable than α-homodimers. They tend to form similar chaperone complex during folding process for example in nuclear hormone receptor complex (Mendel & Ortí 1988) and filamentous actin complex (Terasawa et al. 2005).

1.2.4. Evolutionary relationship

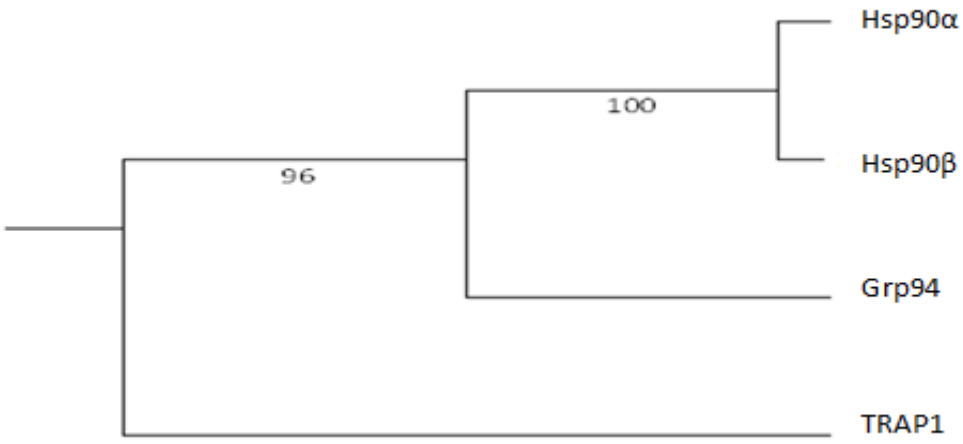


Figure 1.8: A simple phylogenetic analysis of the human Hsp90 family members. The values at the nodes are bootstrapping values used for validation and the sequences were retrieved from the NCBI database; Hsp90α (NP_005339.3), Hsp90β (NP_001258900.1), Grp94 (AAK74072.1) and TRAP1 (NP_001258978.1).

Phylogenetic studies have shown that there is a relationship that exists between the members of the Hsp90 family (Stechmann & Cavalier-Smith 2003) i.e. they share a common ancestor where Grp96 and cytosolic Hsp90 share the last common ancestor. The branching of the chaperones in the phylogenetic tree as shown in figure 1.8 perhaps indicates different functions. Evolution of the Hsp90 family can be explained well by the gene duplications that could have occurred a long time back. Basing on the branching patterns of figure 1.8; the cytosolic sequences together with Grp96 form paralogs with Hsp75 indicating that the mitochondria evolved earlier than the other organelles.

1.3. Parasitic Hsp90s

Parasites often alternate between two different hosts during their life cycle and the hosts always present different environments for the parasite and for it to survive, adaptation measures should be taken (Royet al. 2012). Temperature, pH, ionic strength and the host's immune system always present a challenge to the parasite and the ability of the Hsp90 to adapt to these changes makes the parasite to depend on it to adapt and proliferate. These Hsp90s are mostly expressed on the surface of the parasites for immune recognition therefore overexpression of Hsp90 is a defense mechanism for many parasites e.g. *Leishmania* sp.(Streit et al. 1996) and an antigen in infections by other parasites such as *Schistosoma mansoni* (Johnson et al. 1989).

Malaria parasites always take advantage of the Hsp90 machinery to adapt in the environmental changes of the human host as there are elevated temperatures when it is transferred from the vector to the human host. Malaria patients develop symptoms such as elevated body temperatures that will lead to the activation of heat shock response and as well the *Plasmodium* sp.(Banumathy et al. 2003). The Hsp90 of *Giardia lamblia* is the only example so far that is coded by two different genes annotated as HspN since it codes for the N-terminal of Hsp90 and HspC for coding the C-terminal (Kamikawa et al. 2011) and they are spliced together in trans splicing (Roy et al. 2012). There are many other protein coding genes in between them on the chromosome and the reason for this split is probably a survival method against inhibitors.

There are some unique features in the protozoan primary structures of Hsp90 on the ATP binding domain i.e. a homologous substitution of human K98 to R clusters (Pallavi et al. 2010). The affinity for ATP or inhibitors for this modified site has not yet been examined systematically. Another difference is extended linker region between the N-terminal and the middle domain.

The biochemical characteristics between the protozoan parasites and humans are different also in the sense that the activity of ATP hydrolysis is 6 times higher in protozoa as has been shown by *P. falciparum* (Pallavi et al. 2010).

1.4. Hsp90 as a Drug Target

The importance of Hsp90s to cellular processes as well as disease states makes the protein to be a potential drug target (Krukenberg et al. 2011). Research has shown its implication as an anti-cancer drug target (Jhaveri et al. 2012). Hsp90 plays a major role in the survival of parasites during their life cycles (Table 1.4) and inhibition of their Hsp90 reduces their chance of surviving after host invasion, therefore Hsp90 could be considered as a potential drug target for parasitic diseases. The presence of an extended linker region in the protozoan Hsp90 gained attention in regulating conformational dynamics and plays a part in the affinity for inhibitors such as GA (Roy et al. 2012). Studies have proved that the charged linker region increase the Hsp90's affinity for ATP and its inhibitors (Pallavi et al. 2010) and by this the *P. falciparum* Hsp90 was inhibited and the human Hsp90 unaffected.

Table 1.4: The biological aspects of a selected protozoan Hsp90 are represented and the significance of Hsp90 in the protozoa together with effects of inhibition revealed also. Adapted from Roy et al. and modified accordingly.

Species	**Life cycle**	**Role of Hsp90**	**Possible inhibition**	**Reference**
P. falciparum	Ring to trophozite	Regulation of ring to trophozoite transition	GA and 17-AAG inhibits growth at the ring stage	Kamikawa et al. 2011
L. donovani	Promastigote to Amastigote	Regulates developmental transition from promastigote to amastigote	GA causes apoptosis to the parasite	Wiesgigl et al, 2001
T. evansi		Regulates cell-cycle growth	17 AAG causes cell death	Kamikawa et al. 2011
T. gondii	Bradyzoite to tachyzoite	Controls invasion, cell growth and differentiation	GA ceases infection by inhibiting transition from bradyzoite to tachyzoite	Peroval et al. 2006

1.5. Future Studies and Potential Knowledge Gap

Despite the studies that have been done so far on the Hsp90, there is still need for other discoveries to solve current problems such as personalized medicine and drug resistance. Personalized medicine and continuous studies of gene mutations will improve pharmaceutical drug development. This is because some individuals develop side effects after taking drugs such as aspirin therefore studying the genetic variations of every individual will help in determining how they will respond to the drugs. However, the Hsp90 is a highly conserved protein in all species therefore inhibition of parasitic Hsp90 using inhibitors like geldanamycin (GA) and its derivatives lacks specificity. Therefore, this could lead to the human Hsp90 being inhibited as well, causing unwanted effects on the patient. Due to lack of specificity, there is a need to find significant differences in the primary and tertiary structure between the human and parasitic chaperone and utilize on targeting those sites. Parasite are well known for developing resistance to vaccines therefore the idea of combining Hsp90 inhibitors and other drugs should be considered as an option.

1.6. Problem Statement

Parasitic drugs that are in current use are slowly becoming ineffective due to resistance by the parasite therefore other approaches that are affordable have to be taken to stop parasitic infections. The Hsp90s have been found to play a major role in all existing cells and without these proteins; the cell is prone to any environmental change. The Hsp90 is ATP dependent therefore inhibiting the ATP binding site will cause detrimental effects on the cell. The Hsp90s have been targeted as cancer drugs and inhibitors can selectively attack the chaperones from the cancer cells without affecting the normal cells (Calderwood et al. 2006). So likewise, the inhibitors of such kind can also be used for targeting heat shock proteins from parasitic organisms causing fatal consequences to the parasite. GA has already been proposed to have some antimalarial activity on the *P. falciparum* parasite as it interferes with the function of the parasite's Hsp90 (Banumathy et al. 2003). Pallavi et al. were able to inhibit the *Plasmodium* and *Trypanosoma* Hsp90s using low concentrations of the inhibitor (Pallavi et al. 2010). Therefore if GA is able to interfere with the parasite's Hsp90, it means the parasitic Hsp90s have some

unique features which can also be found on other parasitic organisms and can also suffer the action of inhibitors on them.

1.7. Hypothesis

We hypothesize that there is a significant difference between the human and parasitic orthologs physicochemical properties in the primary structure of the protein hence Hsp90 may be a selective drug target for human parasitic diseases.

1.8. Aim

In this study, we were aiming to have a clear picture of the relationship between the human and parasitic Hsp90s in terms of the physicochemical properties, motifs, evolution and sequence level analysis.

1.9. Specific Objectives

- Retrieval of the human Hsp90 proteins and their parasitic orthologs from protein databases by reverse BLAST
- Primary structure analysis through physicochemical properties, motifs and MSA.
- Phylogenetic tree analysis

CHAPTER TWO

2. Sequence Retrieval and Protein Properties Analysis

Four different Hsp90 homologs have been identified to be expressed by *H. sapiens* and they are localized in the cytosol, ER and mitochondria. The Hsp90 proteins are highly conserved and their 3 domains are conserved throughout all kingdoms. In this chapter, parasitic orthologs were retrieved from the NCBI database and the physicochemical properties calculated. The properties for human sequences were compared against the parasitic orthologs. Additionally, statistical tools were implemented to visualize and analyse the properties as well as to see the relationships that exists between different properties.

2.1. Introduction

Hsp90sare highly conserved molecular chaperones, found throughout all kingdoms playing a major role in the folding and regulating of client proteins (Hartl et al. 2009). The features that make the client proteins dependent on Hsp90 are still unclear but the whole process is ATP dependent and physicochemical properties determine client specificity(Young et al. 1997). As been explained in Chapter 1, there are 3 homologs found in parasitic organisms: cytosolic Hsp90, TRAP1 and GRP94 but humans have two genes that encode cytoplasmic Hsp90 isoforms (Gupta et al. 1995) thereby making them to be the only organism with 4 homologs in this study.

Physicochemical properties are known to influence the protein complex to attain its native and stable conformation successfully (Banerjee et al. 2010) and hydrophobicity has been found to have the major effect on protein folding (Dill 1990). Since the Hsp90s are highly conserved throughout all kingdoms, it is therefore important to analyze if the protein properties are conserved as well. This is to the best of our knowledge the first extensive and comparative analysis of the physicochemical properties of Hsp90s between human and parasites. In this study, various physicochemical properties that are known to have an influence on structure and function were analysed: hydrophobicity, aromaticity, molecular weight (Mr), isoelectric point (pI), aliphatic index, instability index and grand average hydropathy (GRAVY). Hydrophobic residues are embedded in the core of a protein molecule while the hydrophilic residues that

interact with water are located in the surface region (Keskin et al. 2008). The R-group in an amino acid determines whether a protein is hydrophobic or hydrophilic.

Proteins tend to degrade rapidly when exposed to high temperatures, extreme pH and naturally from premature termination. However, proteins have different half-lives (Rechsteiner, 1987) meaning that they have different features within them that elicit proteolysis. Sequence specific properties have been suggested to have an influence (Rogers et al. 1986). Therefore, instability index is an estimate measure of the stability of a protein in a test tube using sequence information. A value less than 40 indicates that the protein is stable (Guruprasad et al. 1990). Isoelectric point (pI) is the pH value at which the net charge of a protein is 0 but the surface is covered with a charge. Proteins are stable at pI where a pI value of 7 indicates neutrality of a protein. pI value less than 7 indicate that the protein is acidic, while greater than 7 indicates that the protein is basic(Sarma et al. 2012).

Tryptophan, tyrosine, phenylalanine are amino acid residues that are aromatic and contribute the aromaticity of a protein. Aromaticity is the measure of the relative frequency of aromatic amino acids in a protein (Lobry and Gautier, 1994). Basically, relative frequency is another term for proportion, dividing the number of aromatic residues by the total number of residues in a sequence. When exposed to UV light proteins and peptides fluoresce at different levels due to the differences in the number of the aromatic residues. Most of the emissions in a protein results from tryptophan and little from phenylalanine and tyrosine. Aromaticity is calculated using the formula:

$$Aromaticity = \sum_{i=1}^{20} \delta i * fi$$

Where δ_i is 1 when the amino acid in an aromatic and is 0 when otherwise, f_i is the relative frequency of the amino acid of kind i in a protein (Lobry and Gautier, 1994).

The GRAVY score is a measure of the overall hydropathicity of a protein where the positive scores indicate non-interaction with water and negative values indicate interactions with water. The formula for calculating the property is as follows:

$$GRAVY = \sum_{i=1}^{20} \alpha i * fi$$

where f_i is the relative frequency of amino acid of kind i in a protein and α_i is the hydropathy index of that amino acid (Kyte et al. 1982). The GRAVY score can be used to predict the type of a protein basing on the fact that trans-membrane proteins have a higher score than the globular proteins. Valine, isoleucine, leucine and alanine are amino acid residues containing hydrophobic, aliphatic side chains. Due to their hydrophobic properties, these residues are located inside most protein molecules (Keskin et al. 2008).

Aliphatic index is defined as an estimate measure of the relative volume occupied by these side chains and the general formula to calculate aliphatic index is as follows:

$$Aliphatic\ Index = X(Ala) + a * X(Val) + b * (X(Ile) + X(Leu))$$

where X(Ala), X(Val), X(Ile) and X(Leu) are mole percentages of the amino acids and the coefficients a and b are the relative volumes of the side chains; a=2.9 for Val and b=3.9 for Leu/Ile (Ikai 1980). Aliphatic index measurements are used to check for stability of globular proteins in increasing temperatures where the value is regarded as a positive factor for the increase of thermostability (Ikai 1980). A high value indicates that a protein is stable in varying temperatures.

In this Chapter, sequences were retrieved accordingly from the biological databases and grouped according to their localization. The physicochemical properties of each group were calculated. The properties studied were those mentioned above where the human Hsp90s were compared against the vector and parasitic Hsp90s in the same group. We observed that the physicochemical properties are generally conserved. However, the *Plasmodium sp.* had variations in Mr and GRAVY in all the groups. We also observed that the range of values for aromaticity and GRAVY in human and protozoan Hsp90s were quite different. Furthermore, we did some statistical analysis to try and understand if location of the Hsp90 has influences on the properties. In all this we observed that the physicochemical properties are generally conserved and the environment has an influence in overall properties of a protein.

2.2. Methods

Four human Hsp90 proteins were retrieved from the NCBI database (Table 2.1) and used to retrieve their homologs in parasitic disease causing species and the vectors responsible for transmission (Table 2.2). The parasitic sequences were compared to their human orthologs as well as the vectors for different physicochemical properties and the methods and tools used are as described below.

2.2.1. Sequence retrieval

The human Hsp90 sequences were retrieved from the NCBI-Entrez database and these sequences were then divided into three groups depending on their localization in the cell resulting in Hsp90α and Hsp90β being grouped into the cytosolic Hsp90 (group A) while Grp94 was grouped into the ER Hsp90 (group B) and the Hsp75, in the mitochondrion Hsp90 (group C). The sequences were then used as queries to obtain members of the same group from different parasites using BLASTP tool from NCBI. The BLAST tool default parameters of BLOSUM-62 scoring matrix, word-size of 3, and gap existence of 11 and extension cost of 1 were used. Reverse BLAST was done in a two way process where the protein from the human was used to search the parasitic protein and then the parasitic protein ideally returns the human protein used to get it as the best hit.

2.2.2. Physicochemical property analysis

SignalP v4.0 (Petersen & Nielsen 2011)which uses both neural network and hidden Markov model methods was used to predict the signal peptide in the protein sequences. The default settings were used. The Biopython module was used in writing python scripts (see Appendix 1, Script A1.1) to calculate the normalized protein properties which are Mr, aromaticity, pI, instability index, aliphatic index and GRAVY.

2.2.3. Statistical analysis

Statistical calculations were performed using R. An R-script was used to plot the boxplots, scatter plots and calculate the Pearson correlation coefficients (see Appendix 1, Script A1.2). The Kruskal-Wallis test was used to assess the statistical significance ($p<=0.05$) of the differences between the physicochemical properties measured in the analysis of Hsp90 from different groups (see Appendix 1, Script A1.3).

2.3. Results

2.3.1. Sequence retrieval

2.3.1.1. Human Sequences

Two cytosolic homologs; Hsp90α and Hsp90β, were retrieved. Hsp90α has 2 isoforms and Hsp90β has 3 isoforms as shown in Table 2.1. The ER Hsp90 has no isoforms while the mitochondrion Hsp90 has 2 isoforms. The isoforms were aligned using the Mafft program to see the differences and to decide which sequence to use as a query for retrieving parasitic and vector sequences.

Table 2.1 Human Hsp90 sequences retrieved from the NCBI database. The accession number of the respective isoform is also presented.

Location		Isoform: Accession number		
Cytosol	Hsp90α	1: NP_001017963.2	2: NP_005339.3	
	Hsp90β	A: NP_031381.2	B: NP_001258900.1	C: NP_001258901.1
ER		NP_003290.1		
Mitochondria		1: NP_057376.2)	2: NP_001258978.1	

All the isoforms in different locations differ only in the N-terminal site (Fig 2.1; Appendix 2, Fig A2.1) indicating that the N-terminal of the Hsp90 has some variations. The isoform 1 of Hsp90α has got an extra 120 amino acid residues at the start of the N-terminal. Both sequences were used as queries to retrieve the other sequences and all the hits that were produced by isoform 1 were lacking at least 123 amino acid residues at the start of the N-terminal, therefore making isoform 2 as the best choice for use in further analysis of the proteins. The Hsp90β isoforms have got the same amino acid residues in the first 78 positions where after that, there is a long and short insertion in isoform a, and c respectively. The isoform "a" was used for further analysis since when comparing the same BLAST hits produced by the isoforms, isoform a's hit had a higher identity. Isoform 1 of Hsp75 also produced a hit with a higher identity when compared to isoform 2 therefore it was also used for further analysis.

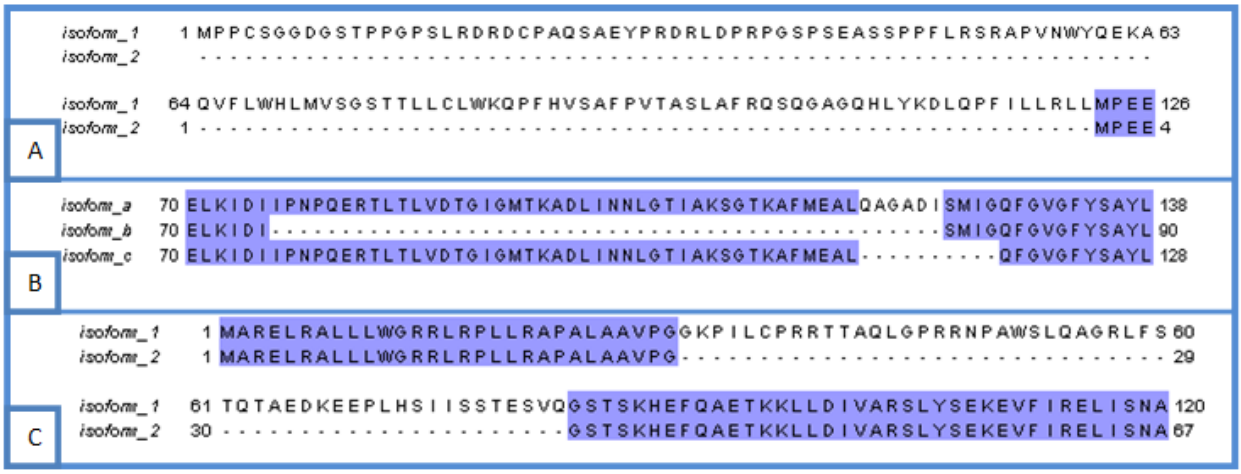


Figure 2.1 Pairwise alignments using the MAFFT online program showing differences between human isoform sequences. "A" represents the 2 Hsp90α isoforms where isoform 1 has an insertion in the N- terminal. "B" represents the 3 Hsp90β. "C" represents the 2 Hsp75 isoforms with isoform 1 having an insertion in the N-terminal.

2.3.1.2. Parasites and vector sequences

Sequences from 31 different parasitic species were retrieved from the NCBI database; includes 22 protozoans, 8 helminths and 1 ectoparasite. Only one organism from the ectoparasites, *Pediculus humanus corporis*, had sequences for the Hsp90 in the databases. 20 sequences from different vectors were also retrieved (Table 2.2.). Using the reverse BLAST method, we confidently selected the sequences as true orthologs. This is because all the sequences, when used as queries ideally returned the human sequence that was used to find them as the best hit. An exception was only found in the group C sequences where they will return human TRAP1 sequence as the best hit. But either way, TRAP1 and Hsp75 are both mitochondrion proteins therefore this gave us the confidence to use these sequences. The parasites that are transferred by these vectors are also shown. The group A proteins possess an –MEEVD motif at the end of their C-terminal site as has been observed by Meng *et al.* while the group B proteins contain the –KDEL motif which specifies signal retention. The group C proteins did not have a common motif at the end of the C-terminal domain which could be an indicator of different properties of the protein. *Giardia sp*. lacks the mitochondrion Hsp90 because the organism lacks the organelle.

Table 2.2.: Summary of the orthologs retrieved from the NCBI database. The table shows the species name accession number of its sequences from the cytosol, ER and mitochondrion. In parenthesis is the abbreviation of the

species name and the localization where _A: cytosol, _B: ER and _C: mitochondrion. The colored boundaries show different groups; blue: protozoa, red: helminths, green: ectoparasite and black: vectors.

. Species	Location					
	Cytosol		ER		Mitochondria	
Parasites	Accession Number	Short name	Accession Number	Short name	Accession Number	Short name
Babesia bovis	XP_001611554.1	(**BBO_A**)	XP_001610762.1	(**BBO_B**)	XP_001611009.1	(**BBO_C**)
Babesia equi	AFZ79107.1	(**BEQ_A**)	EKX74076.1	(**BEQ_B**)	XP_004833240.1	(**BEQ_C**)
Giardia intestinalis	BAJ33526.1	(**GIN_A**)	EES98386.1	(**GIN_B**)	---	---
Leishmania braziliensis	XP_001567804.1	(**LBR_A**)	XP_001566424.1	(**LBR_B**)	XP_001568035.1	(**LBR_C**)
Leishmania infantum	XP_003392730.1	(**LIN_A**)	AAF67727.1	(**LIN_B**)	XP_003392758.1	(**LIN_C**)
Leishmania major	XP_001685762.1	(**LMA_A**)	XP_003722150.1	(**LMA_B**)	XP_001686000.1	(**LMA_C**)
Leishmania mexicana	XP_003878279.1	(**LME_A**)	XP_003872486.1	(**LME_B**)	XP_003878495.1	(**LME_C**)
Plasmodium falciparum	AAA66179.1	(**PFA_A**)	XP_001350620.1	(**PFA_B**)	XP_001348591.1	**(PFA_C)**
Plasmodium vivax	XP_001613451.1	(**PVI_A**)	XP_001617311.1	(**PVI_B**)	---	---
Plasmodium berghei	---	---	XP_676606.1	**(PBE_B)**	XP_677759.1	**(PBE_C)**
Plasmodium yoelii yoelii	---	---	XP_725468.1	**(PYY_B)**	XP_725649.1	**(PYY_C)**
Naegleria gruberi	XP_002682619.1	**(NGR_A)**	---	---	XP_002674011.1	**(NGR_C)**
Toxoplasma gondii	XP_002368278.1	(**TGO_A**)	EEE29549.1	(**TGO_B**)	XP_002370104.1	(**TGO_C**)
Trypanosoma brucei	A44983	(**TBR_A**)	XP_843950.1	(**TBR_B**)	CBH17173.1	(**TBR_C**)
Trypanosoma cruzi	EKF32565.1	(**TCR_A**)	EKF98429.1	(**TCR_B**)	XP_820924.1	(**TCR_C**)
Acanthamoeba castellanii	XP_004367478.1	**(ACA_A)**	XP_004339231.1	**(ACA_B)**	XP_004344035.1	**(ACA_C)**
Babesia microti	CCF75932.1	**(BMI_A)**	CCF75129.1	**(BMI_B)**	CCF74928.1	**(BMI_C)**
Cryptosporidium hominis	XP_665730.1	**(CHO_A)**	XP_668238.1	**(CHO_B)**	---	---
Cryptosporidium muris	XP_002142400.1	**(CMU_A)**	XP_002141043.1	**(CMU_B)**	---	---
Cryptosporidium parvum	XP_626924.1	**(CPA_A)**	XP_628530.1	**(CPA_B)**	---	---
Blastocystis hominis	CBK21615.2	**(BHO_A)**	CBK22622.2	**(BHO_B)**	CBK24472.2	**(BHO_C)**
Entamoeba histolytica	XP_653162.1	**(EHI_A)**	XP_649964.1	**(EHI_B)**	---	---
Brugia malayi	XP_001901767.1	**(BMA_A)**	XP_001899398.1	**(BMA_B)**	XP_001895498.1	**(BMA_C)**
Loa loa	XP_003135662.1	**(LLO_A)**	XP_003140246.1	**(LLO_B)**	XP_003141395.1	**(LLO_C)**
Toxocara cati	ACO55135.1	**(TCA_A)**	---	---	---	---
Trichinella spiralis	XP_003374556.1	**(TSP_A)**	XP_003379158.1	**(TSP_B)**	---	---
Wuchereria bancrofti	EJW88125.1	(**WBA_A**)	EJW87335.1	(**WBA_B**)	EJW83191.1	(**WBA_C**)
Schistosoma japonicum	AAW27659.1	(**SJA_A**)	AAW25122.1	(**SJA_B**)	CAX73028.1	(**SJA_C**)
Schistosoma mansoni	XP_002578418.1	(**SMA_A**)	AAF66929.1	(**SMA_B**)	XP_002577510.1	(**SMA_C**)

Clonorchis sinensis	GAA47176.1	**(CSI_A)**	---	---	GAA47789.1	**(CSI_C)**
Pediculus humanus corporis	XP_002432348.1	**(PHC_A)**	XP_002428463.1	**(PHC_B)**	XP_002425720.1	**(PHC_C)**
Vectors (Parasite)	Accession Number	Short name	Accession Number	Short name	Accession Number	Short name
Aedes aegypti (Plasmodium)	XP_001649752.1	**(AAE_A)**	XP_001662951.1	**(AAE_B)**	XP_001654758.1	**(AAE_C)**
Anopheles albimanus (Plasmodium)	AAB05638.1	**(AAL_A)**	---	---	---	---
Anopheles darling (Plasmodium)	EFR27233.1	**(ADA_A)**	ERF27675.1	**(ADA_C)**	EFR22087.1	**(ADA_C)**
Anopheles gambiae (Plasmodium)	XP_308800.3	**(AGA_A)**	XP_321706.5	**(AGA_B)**	XP_311405.3	**(AGA_C)**
Culex quinquefasciatus (Brugia and *Loa loa)*	XP_001865484.1	**(CQU_A)**	XP_001844128.1	**(CQU_B)**	XP_001861262.1	**(CQU_C)**
Felis catus (Toxoplasma)	XP_003988041.1	**(FCA_A)**	XP_003989234.1	**(FCA_B)**	XP_003999038.1	**(FCA_C)**
Ixodes scapularis (Babesia)	XP_002413149.1	**(ISC_A)**	XP_002413149.1	**(ISC_B)**	XP_002403553.1	**(ISC_C)**
Lucilia cuprina (Entamoeba)	AEF38377.1	**(LCU_A)**	---	---	---	---

2.3.2. Protein properties

2.3.2.1. Signal peptides

All the group A and group C Hsp90s lacked the peptide signal since the proteins are synthesized and released in the compartment they perform their function while all the proteins in group B had the signal peptide present except Hsp90s from 5 species (Appendix 3, Table A3.1). The most probable answer for the lack of signal peptides in these species could be that the sequences deposited to the database had their signal peptides removed already. The group B proteins are synthesized in the cytosol and perform their functions in the ER so there is need for the signal peptide to direct the proteins to their final destination. The peptide signals had their length in the range of 16-30 amino acid residues long with the human sequence having a length of 20 amino acid residues. Multiple sequence alignment (MSA) showed that the Hsp90s signal peptides are rich in branched-chain amino acids (BCAA) (Fig 2.2). BCAA are amino acids that have an aliphatic side-chain; valine, leucine and isoleucine, and they are conserved in protozoan Hsp90s.

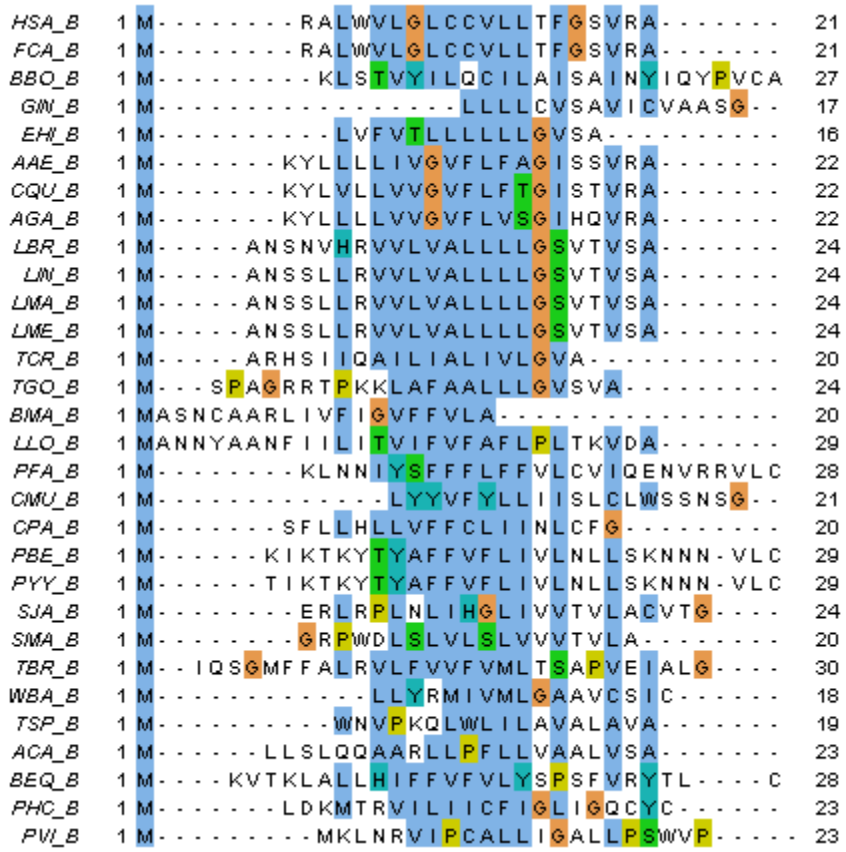


Figure 2.2.MSA for the signal peptides of group B Hsp90s. ClustalX coloring in Jalview is used.

2.3.2.2. Physicochemical properties

Table 2.3: Summary of the means and standard deviations of all the properties in each group.

Property	Group A		Group B		Group C	
	Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*
Mr	82.3kDa	1.856k	91.6kDa	4.534k	80.2kDa	7.872k
pI	4.99	0.078	5.09	0.318	6.27	0.525
Aliphatic index	81.01	3.599	84.59	4.205	86.13	4.344
Hydrophobicity	-0.455	0.052	-0.385	0.086	-0.267	0.078
Aromaticity	0.079	0.003	0.080	0.005	0.084	0.006
Instability index	42.13	4.573	38.60	3.792	42.08	4.548
GRAVY	-0.626	0.058	-0.556	0.09	-0.411	0.099

The mean and SD values for physicochemical properties in each group were calculated (Table 2.3). Mr of Hsp90s is generally high in group B, followed by A and lastly C. The large SD value of 7.872k shows that group C Hsp90s has quite different Mr values. The pI values do not deviate much from the mean as can be seen from the small SD values. Group A and group B have pI's that are closely related while group C pI's are uniquely high. The aliphatic indices are high in all the groups and the values increases from A to B and to C. The small SD values of hydrophobicity in all the groups show that the values for each Hsp90 are closely related. In all the groups, hydrophobicity value is negative. Aromaticity level is almost the same in all groups as can be seen from small mean differences and the SD is small. Instability indices are above 40 in group A and C indicating their short half-lives while in group B, is below 40. All the Hsp90s interact with water as the GRAVY values are negative.

Group A

The Mr ranges from 78 to 86 kDa (Table 2.4) with the *Plasmodium sp.* having quite large proteins of around 86 kDa as can be seen from the outliers in the boxplot (Fig 2.3) as well as the protruding peaks in the bar graphs (Fig 2.4). Boxplots are used to show the shape of the distribution with the middle line showing the central value (median). The first and third quartiles are the edges of the box. Outliers are points that are plotted individually only if they are 1.5 times smaller or bigger than the first and third quartiles respectively. *N. gruberi* has a uniquely low Mr weight of 78kDa while the values for other Hsp90s in this group are quite conserved as supported by a small standard deviation. The aromaticity values ranges from 0.07 to 0.08 with a very small standard deviations. This indicates conservation of the property; however the human Hsp90s have got the lowest aromaticity values. Instability index ranges from 35 to 51 with a mean value of 42. Basing on this value and the bar graphs, Hsp90s in this group have a conserved instability index property and are generally unstable in a test tube. The hydrophobicity values ranges from -0.41 to -0.55 with a very small standard deviation of 0.05. The negative value shows that the Hsp90s are hydrophilic in nature and the hydrophobicity property is conserved.

Table 2.4 Physicochemical properties for group A Hsp90s. The background colors indicates different sections in group A. Grey-Human and host, blue – protozoan , green – helminths and white – ectoparasite.

Name	Molecular weight	Aromaticity	Instability Index	pI	GRAVY	Aliphatic Index
HSA_AA	84656	0.074	41.94	4.94	-0.7503	79.37
HSA_AB	83261	0.075	42.15	4.97	-0.6785	81.33
BBO_A	82509	0.083	40.94	5	-0.6646	79.54
BEQ_A	82483	0.08	42.67	4.99	-0.6108	82.81
GIN_A	80754	0.078	36.94	4.99	-0.5845	83.78
LBR_A	80659	0.078	51.23	5.08	-0.6467	74.35
LIN_A	80547	0.079	51.58	5.08	-0.6559	73.53
LMA_A	80403	0.079	51.03	5.1	-0.6531	73.39
LME_A	80569	0.08	51.74	5.04	-0.6321	74.25
PFA_A	86421	0.078	40.19	4.91	-0.7368	79.36
PVI_A	86338	0.078	37.23	5.07	-0.7142	79.4
TGO_A	81930	0.082	43.98	4.96	-0.6264	81.5
TBR_A	80726	0.078	48.8	5.16	-0.6326	76.27
TCR_A	80727	0.08	49.22	5.07	-0.6239	76.16
ACA_A	82635	0.084	40.65	5.12	-0.6296	82.77
BMI_A	82352	0.081	41.32	5.04	-0.6654	80.08
BHO_A	80653	0.078	45.25	4.81	-0.5907	82.75
CHO_A	80875	0.08	44.25	4.87	-0.5551	87.14
CMU_A	81098	0.079	40.81	4.83	-0.5247	87.36
CPA_A	82350	0.082	44.23	4.93	-0.5368	87.86
EHI_A	82989	0.078	45.14	4.97	-0.6338	83.33
NGR_A	78358	0.079	40.27	5.12	-0.5181	86.89
BMA_A	80854	0.084	39.35	4.95	-0.6482	79.77
WBA_A	82601	0.081	40.9	4.99	-0.6367	80.63
SJA_A	82215	0.081	45.78	4.98	-0.5872	78.92
SMA_A	82299	0.077	45.13	4.95	-0.5426	83.75
LLO_A	80497	0.081	40.74	4.93	-0.6444	79.94
TCA_A	83270	0.082	38.82	5.02	-0.6714	80.1
CSI_A	81816	0.08	41.22	5.02	-0.5669	82.89
PHC_A	83438	0.076	37.8	4.94	-0.6468	82.66
CQU_A	82031	0.079	37.57	4.9	-0.5672	83.39
ISC_A	84275	0.075	39.04	4.95	-0.6937	80.3
LCU_A	81687	0.077	35.51	4.93	-0.6117	81.16
FCA_A	84771	0.074	42.85	4.93	-0.7542	79.26
AAE_A	81531	0.077	39.12	4.94	-0.5924	83.44
AAL_A	82150	0.076	35.45	4.94	-0.6024	82.08
ADA_A	82603	0.077	36.22	4.94	-0.6014	81.6
AGA_A	82083	0.076	38.39	4.95	-0.6158	82.33

The pI's of group A Hsp90s ranges from 4.9 to 5.1 indicating that they are acidic in nature. The very small standard deviation indicates that the property is generally conserved. However, the *T. brucei* Hsp90 has the highest peak of 5.16 while *B. hominis* has the smallest peak of 4.81 pI value. GRAVY is important in understanding how a protein interacts with water and all the GRAVY results had a negative value indicating that the proteins interact well with water agreeing with the hydrophobicity values which showed that all the Hsp90s are hydrophilic. The aliphatic index values were very high, ranging from 78 to 87 showing that, the Hsp90s in this group are thermostable. However, *Leishmania sp.* and *Trypanosoma sp.* had the lowest aliphatic index values with the boxplot indicating the earlier as outliers. In this group, aromaticity was the only property that showed a distinct difference when the human Hsp90 was compared to the parasitic Hsp90s.

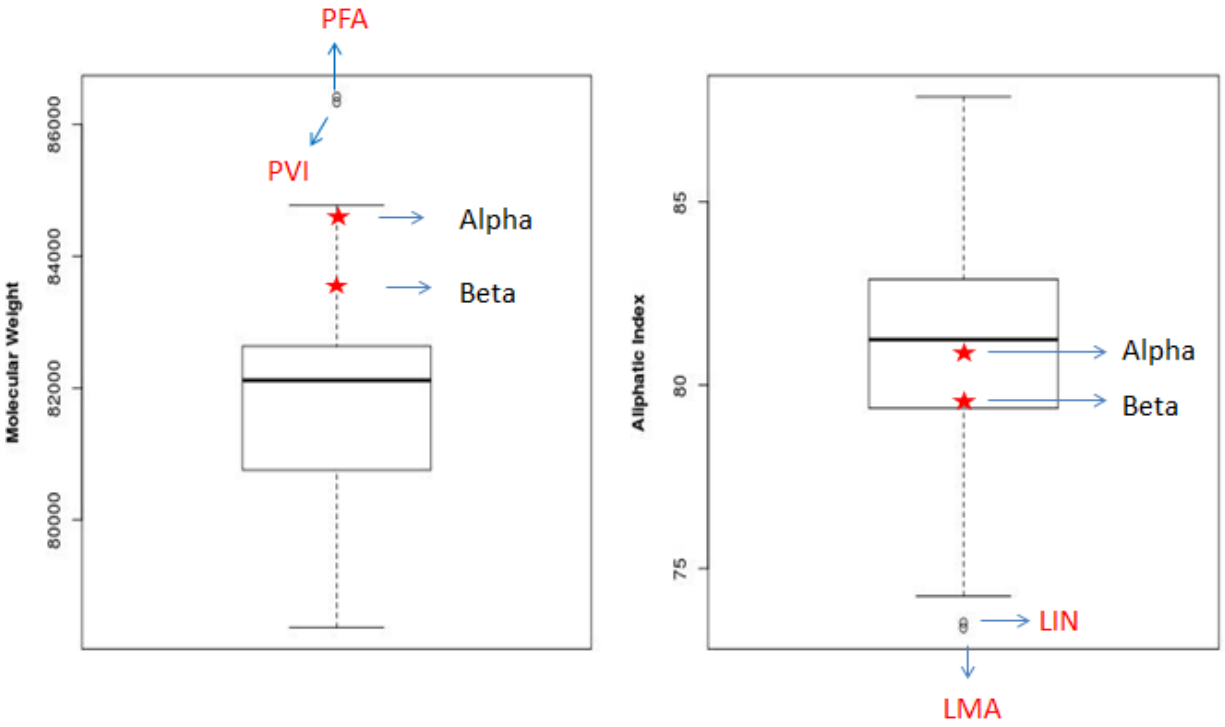


Figure 2.3.Boxplots for Mr and aliphatic index showing the outliers. The stars represent the position of the human Hsp90s.

Figure 2.4.Bar-graph showing the physicochemical properties of group A Hsp90s

Group B

The Mr ranges from 86-93 kDa (Table 2.5) with *E. histolytica* Hsp90 being the outlier (Fig 2.4 and Fig 2.5) having the least value as compared to the rest of the proteins in the group. Like group A properties, the aromaticity, pI, hydrophobicity and GRAVY are conserved as supported by the low standard deviations of 0.005, 0.318, 0.086 and 0.09 respectively. *B. hominis* has got a uniquely low GRAVY value as observed from the protruding peak (Fig 2.5) and its pI is the smallest in the group indicating that it is more acidic. Likewise, *Plasmodium sp.* and *T. brucei* have high pI values and the boxplot showed them as outliers (Fig 2.6). Unlike the group A and C Hsp90s, the mean for instability index is 38.60 which is less than the threshold of 40, meaning that the group B proteins are stable when placed in a test tube. The *B. hominis* Hsp90 has a the highest instability index value of 50 but the aliphatic index is generally high showing that the Hsp90s from the ER can survive a wide range of temperature changes.

Table 2.5 Protein properties for ER group (B). The background colors indicates different sections in group A. Grey- Human and host proteins, blue – protozoan sequences, green – helminths sequences and white – ectoparasites sequences.

Name	Mr	Aromaticity	Instability index	Hydrophobicity	pI	GRAVY	Aliphatic index
HSA_B	92465	0.07970	40.41	-0.51681	4.76	-0.7127	77.68
BBO_B	90845	0.08302	38.45	-0.35849	5.02	-0.5252	87.18
BEQ_B	97070	0.08028	42.77	-0.40968	4.87	-0.5962	88.83
LBR_B	88953	0.08895	34.65	-0.32147	4.98	-0.4798	83.71
LIN_B	86648	0.08560	34.73	-0.28534	4.90	-0.4320	86.21
LMA_B	86610	0.08690	33.30	-0.27626	4.98	-0.4184	86.47
LME_B	86814	0.08560	34.39	-0.28275	4.98	-0.4335	86.82
PFA_B	95014	0.08648	36.89	-0.40682	5.28	-0.5931	87.59
PVI_B	93690	0.08108	34.87	-0.39558	5.63	-0.5775	89.21
PBE_B	93078	0.08798	33.19	-0.33953	5.59	-0.5188	92.49
PYY_B	93302	0.08663	34.50	-0.37624	5.45	-0.5642	90.57
TGO_B	96789	0.07438	37.54	-0.49233	4.97	-0.6607	79.67
TBR_B	87763	0.08021	38.71	-0.35317	5.88	-0.5107	83.18
TCR_B	86787	0.07480	35.22	-0.41339	5.35	-0.5804	84.80
ACA_B	89734	0.07268	36.17	-0.34837	4.66	-0.4887	88.16
BHO_B	93980	0.08138	50.54	-0.62639	4.45	-0.8300	74.40
CHO_B	93545	0.09345	44.28	-0.27184	5.09	-0.4416	84.32
CMU_B	93720	0.07813	39.26	-0.28846	4.83	-0.4566	86.80
CPA_B	89189	0.09403	43.63	-0.27065	4.99	-0.4384	84.82
EHI_B	81197	0.07876	41.34	-0.27567	5.00	-0.4271	94.02
WBA_B	90969	0.08756	42.67	-0.36929	5.18	-0.5398	84.09
SJA_B	90725	0.07779	38.12	-0.35383	5.08	-0.5509	84.24
SMA_B	90497	0.07915	36.76	-0.32412	5.14	-0.5211	85.31
LLO_B	90528	0.08535	40.52	-0.44204	5.14	-0.6280	80.45
BMA_B	90906	0.08745	42.06	-0.40684	5.13	-0.5894	82.13
PHC_B	88930	0.07584	40.06	-0.35347	5.01	-0.5036	87.60
CQU_B	91041	0.08060	36.81	-0.44332	4.89	-0.6117	81.64
ISC_B	90281	0.07605	37.52	-0.48669	4.96	-0.6546	81.08
FCA_B	92494	0.07960	40.72	-0.52114	4.78	-0.7189	77.10
TSP_B	92227	0.07836	43.52	-0.47264	5.11	-0.6562	82.70
AAE_B	91125	0.08050	39.35	-0.42390	4.81	-0.5889	82.77
AGA_B	91343	0.07875	37.47	-0.43250	4.85	-0.5964	82.38

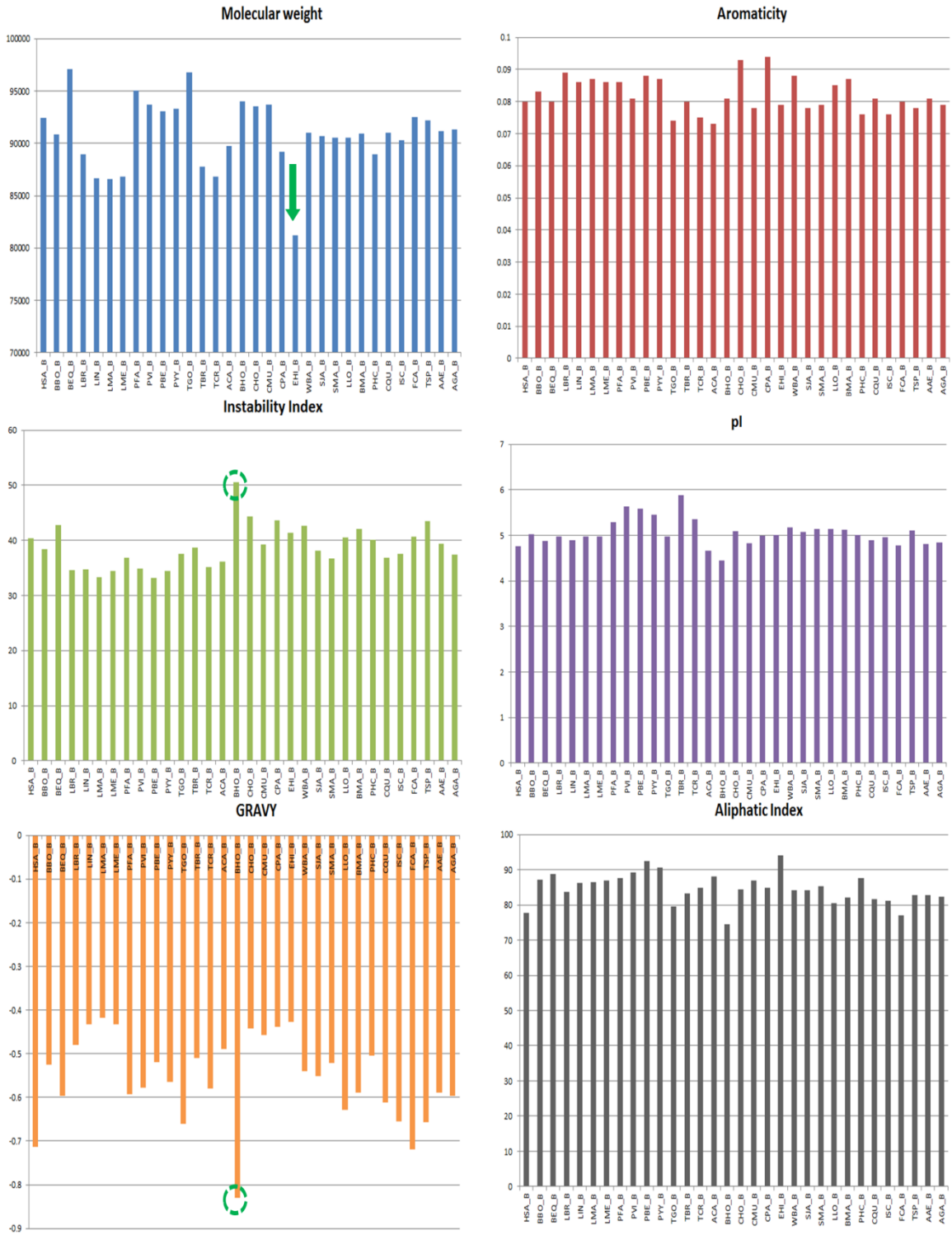


Figure 2.5.Bar-graph showing the physicochemical properties of group B Hsp90s

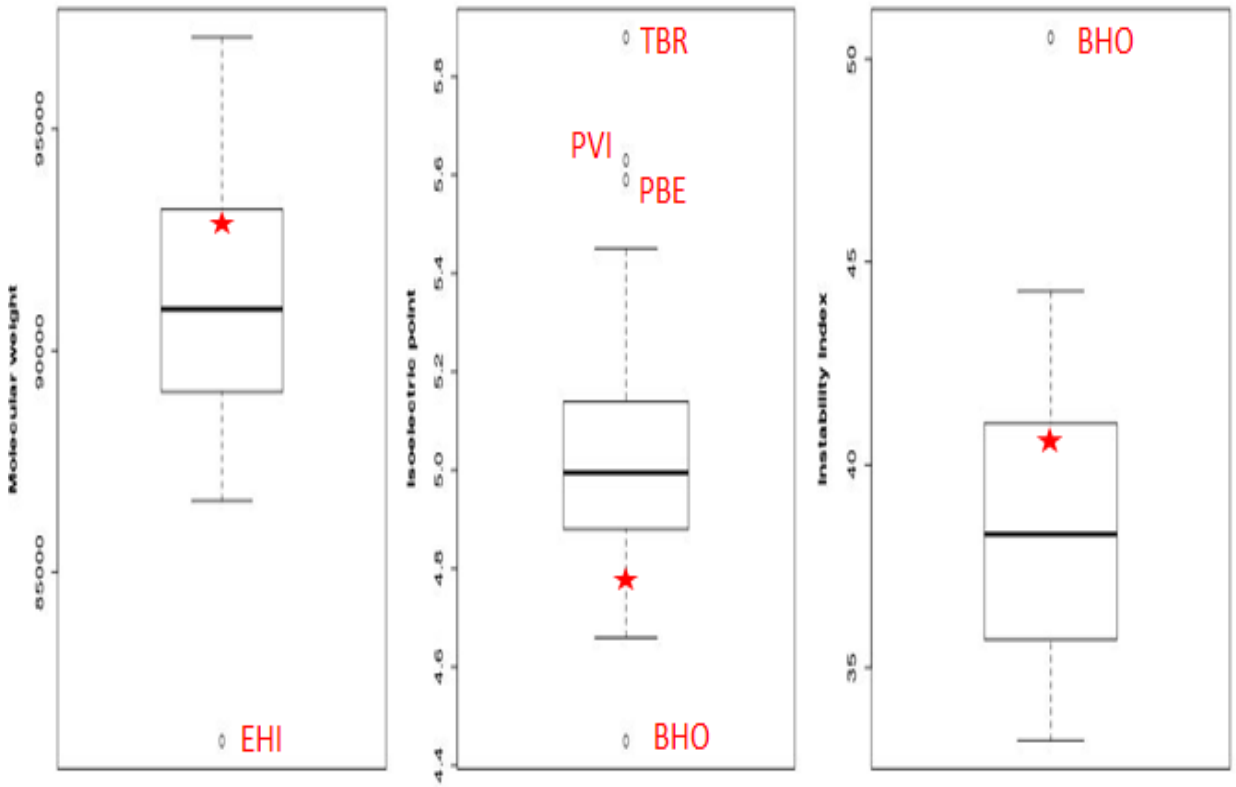


Figure 2.6.Boxplot showing the distribution of the Mr, pI and instability index values for the group B Hsp90s. The red stars represent the human Hsp90s.

Group C

The molecular weight ranges from 71-83kDa (Table 2.6) with Hsp90s of *P. falciparum*, *B. hominis* and *P. yoelli yoelli* as outliers (Fig 2.4) and having high molecular weight peaks (Fig 2.7). Like the other groups, the aromaticity, pI, hydrophobicity and GRAVY properties are conserved as supported by very small standard deviations of 0.006, 0.525, 0.078 and 0.099 respectively. *P. falciparum* has the lowest GRAVY value as shown in the boxplot (Fig 2.8) implying it has the highest interaction with water. The protozoan Hsp90s generally interact with water better than any other group as the GRAVY values are low and conserved. Basing on the pI values, the Hsp90 proteins are acidic as shown by the values which are less than 7 however the human Hsp90 has got a unique value of 7.69. This value makes it the only Hsp90 which is basic nature. The aliphatic index values are generally high with the *B. microtti*'s Hsp90 having a unique aromaticity peak with a value of 0.1. The mean of the instability index indicate that the Hsp90 proteins in this group are generally unstable when placed in a test tube. The *P. falciparum* in group C presents some unique properties as they are closely related to group B Hsp90s.

Table 2.6: Protein properties for group C proteins. The background colors indicates different sections in group A. Grey-Human and host proteins, blue – protozoan sequences, green – helminths sequences and white – ectoparasites sequences.

Name	Mr	Aromaticity	Instability index	Hydrophobicity	pI	GRAVY	Aliphatic index
HSA_C	74264	0.08141	42.20	-0.22888	7.69	-0.3137	93.07
BBO_C	70988	0.09791	34.36	-0.25361	5.89	-0.4034	83.85
BEQ_C	80383	0.09244	33.25	-0.22689	6.06	-0.3674	84.66
LBR_C	72213	0.08675	48.10	-0.32019	6.55	-0.4740	82.71
LIN_C	72010	0.08517	47.75	-0.31230	6.96	-0.4629	83.49
LMA_C	72040	0.08517	48.08	-0.31230	6.76	-0.4629	83.49
LME_C	72109	0.08517	47.97	-0.32492	6.60	-0.4757	83.33
PFA_C	106995	0.08846	45.41	-0.47573	5.38	-0.6859	82.83
PYY_C	96736	0.08451	42.30	-0.39789	5.50	-0.5779	82.86
TGO_C	87927	0.08217	42.32	-0.27307	5.46	-0.4014	77.36
TBR_C	84253	0.07171	40.33	-0.25896	5.83	-0.4137	80.88
TCR_C	83907	0.07467	46.55	-0.28267	6.15	-0.4320	80.48
ACA_C	81290	0.07913	43.32	-0.28104	6.15	-0.4233	78.46
BMI_C	79048	0.09957	39.05	-0.17316	5.84	-0.2981	90.00
NGR_C	80279	0.08757	40.90	-0.28955	6.39	-0.4530	85.37
BHO_C	93165	0.08475	39.16	-0.18281	6.29	-0.2926	85.69
WBA_C	76286	0.08148	40.10	-0.13630	6.42	-0.2603	92.74
SJA_C	80300	0.08807	45.73	-0.27699	6.26	-0.4206	88.75
SMA_C	80752	0.08381	42.63	-0.29261	5.91	-0.4494	89.83
BMA_C	76350	0.08309	40.19	-0.14837	6.42	-0.2653	93.04
CSI_C	86140	0.07301	45.89	-0.20078	5.87	-0.3199	92.69
LLO_C	76004	0.08296	40.02	-0.12000	6.57	-0.2415	92.73
PHC_C	78517	0.08708	36.03	-0.24819	6.95	-0.4109	90.41
CQU_C	80361	0.08085	35.35	-0.33475	6.46	-0.4901	86.57
ISC_C	78486	0.08000	49.87	-0.16857	6.63	-0.2734	86.26
FCA_C	77564	0.08076	42.32	-0.29075	6.38	-0.4043	86.39
AAE_C	79778	0.08108	36.99	-0.29445	6.66	-0.4313	86.84
ADA_C	80523	0.07854	41.28	-0.28191	6.71	-0.4070	88.50
AGA_C	74017	0.08116	36.93	-0.33997	5.87	-0.4764	85.88

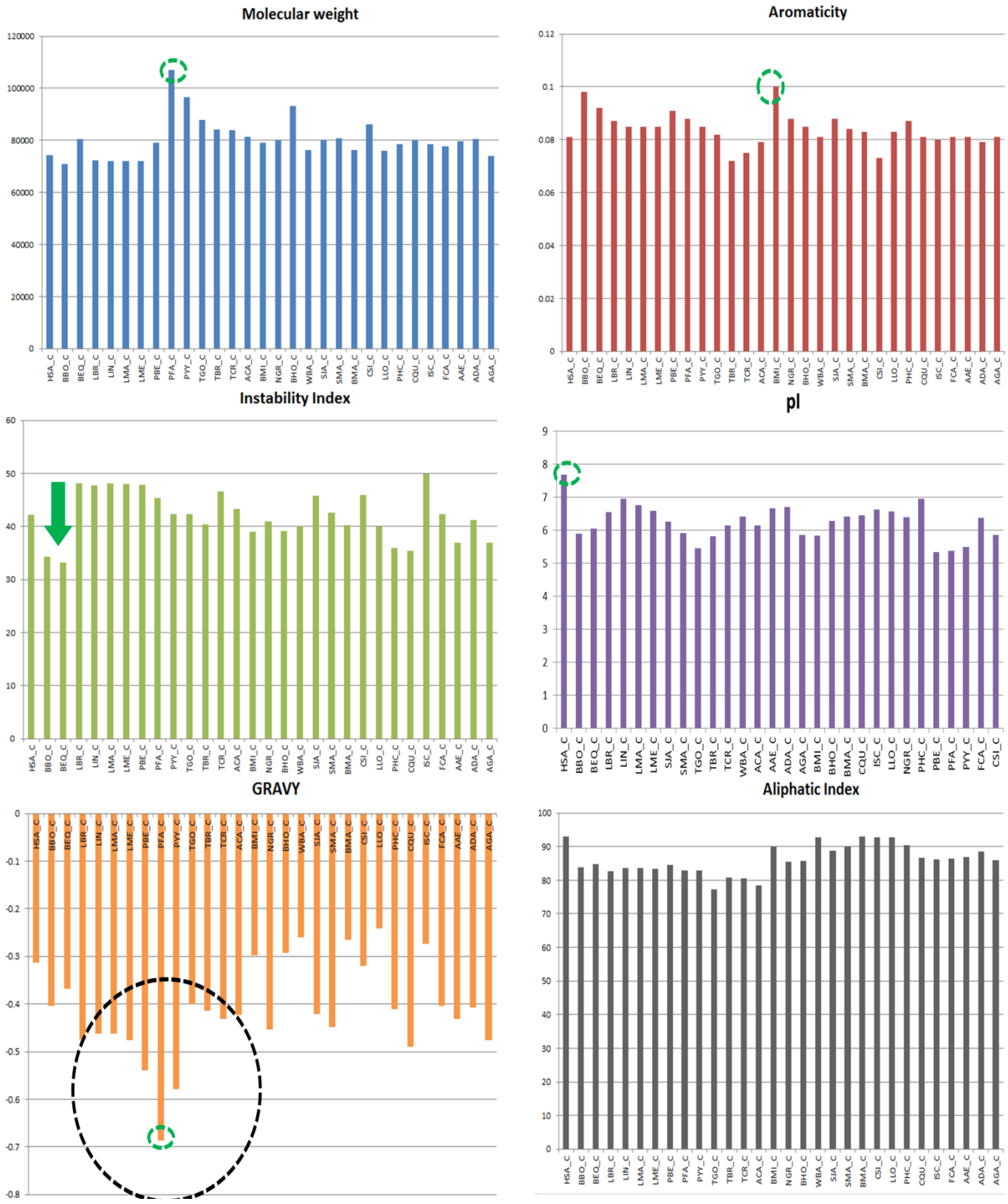


Figure 2.7.Bar-graph showing the physicochemical properties of group C Hsp90s

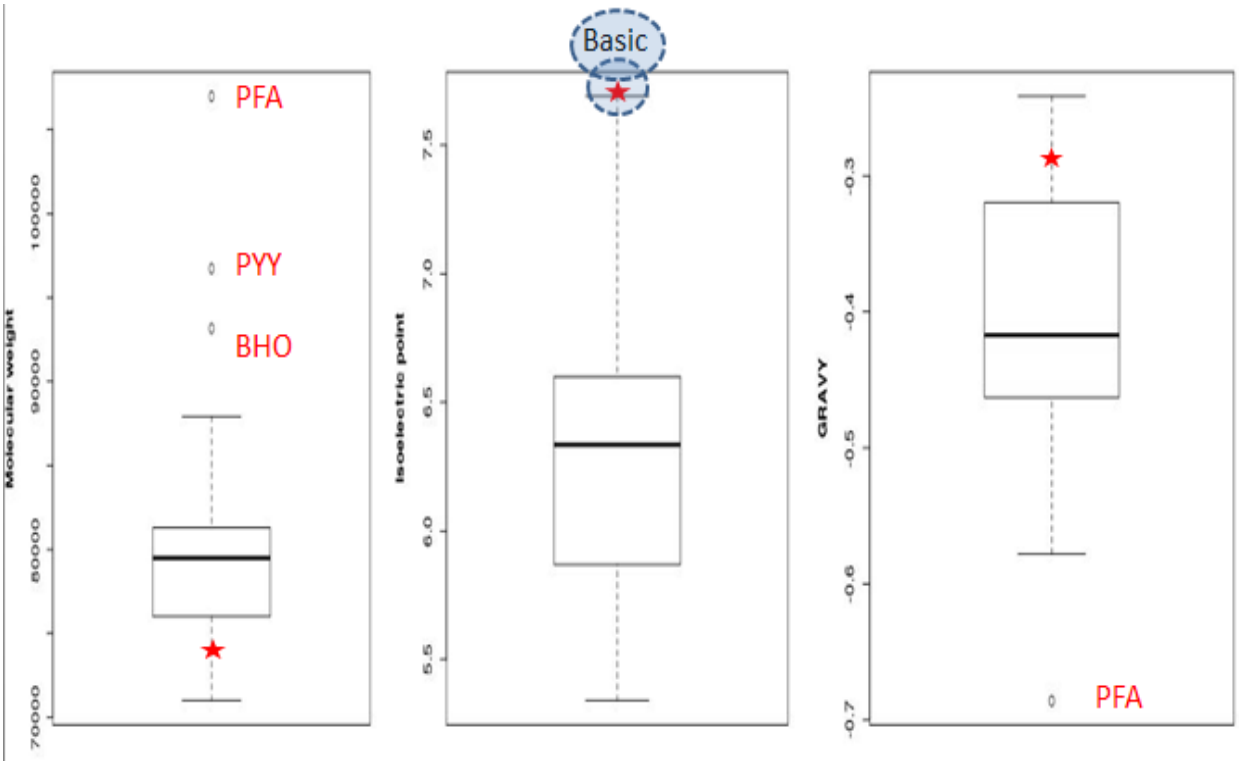


Figure 2.8.Box plots showing the distribution of Mr, pI and GRAVY values for group C Hsp90s. The stars represent the human Hsp90.

2.3.3. Statistical analysis

2.3.3.1. Correlation studies

Correlation studies are helpful in identifying the influence of a certain factor against another factor, in this study a physicochemical property vs. another property. The threshold for strong correlations was put at 0.65. Names of the Hsp90s were used as control since we knew that there is no relationship with the properties. Mr was removed from these studies because it was the only property that was not normalized therefore results from this could be by chance.

In group A, the strongest correlation was observed between hydrophobicity and GRAVY (Fig 2.9) followed by hydrophobicity and aliphatic index with values of 0.99 and 0.65 respectively. This strong relationship was expected as the hydrophobicity values (Table 2.3) were all negative indicating that Hsp90s are hydrophilic. Aliphatic index showed slight relationships with GRAVY, pI and instability index with values of 0.59, 0.53 and 0.55 respectively. The scatter plot clearly supports the correlation of hydrophobicity and GRAVY only in group A.

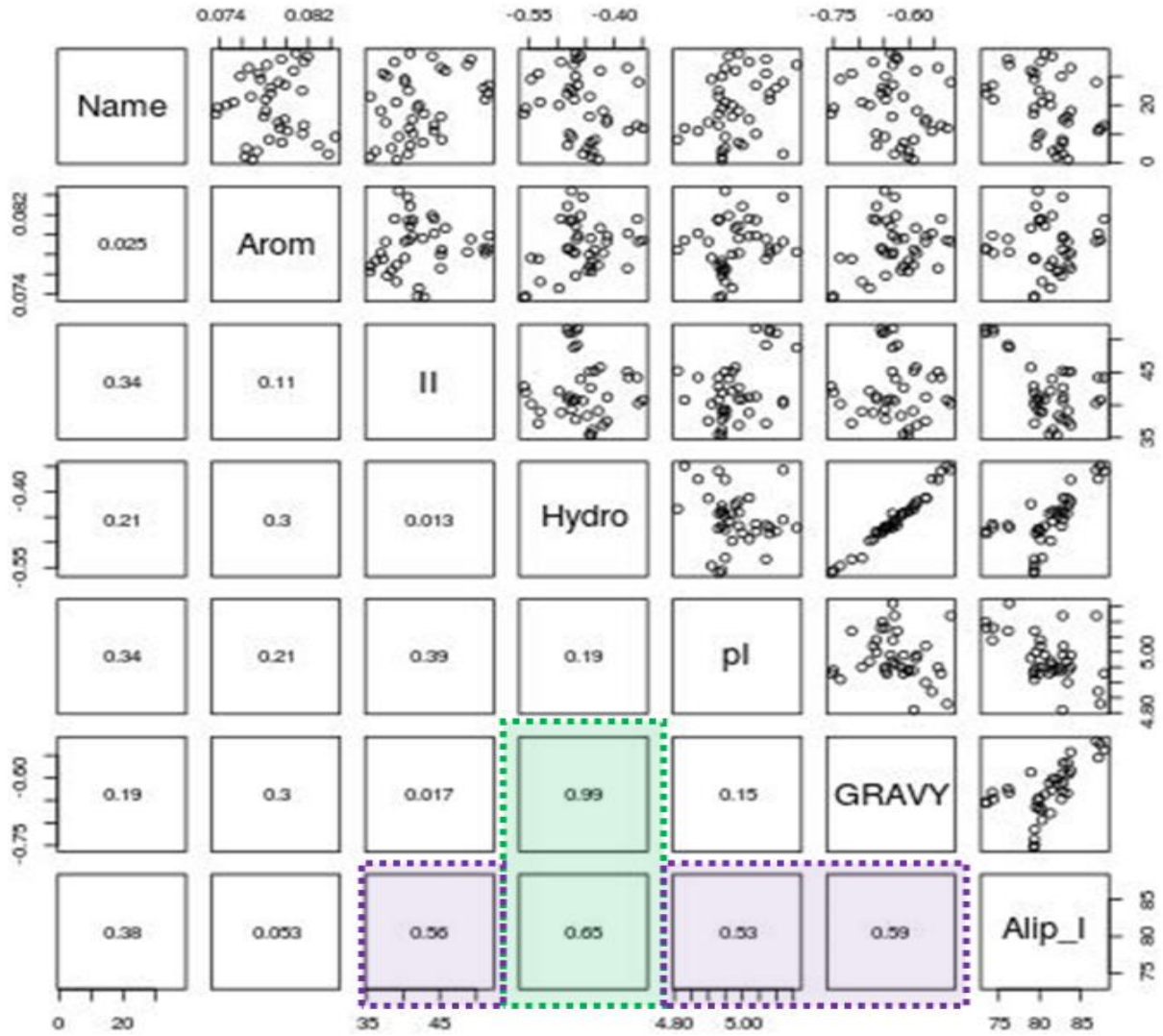


Figure 2.9.Scatter plot for group A Hsp90s with Pearson correlation coefficients. The purple indicating slight relations between properties while the green highlighting shows strong correlations.

In group B, hydrophobicity and GRAVY showed the strongest correlation with a value of 0.99 (Fig 2.10). The scatter plot could clearly support this correlation. Aliphatic index showed to have correlations with GRAVY and hydrophobicity with values of 0.71 and 0.72 respectively. The scatter plots also supported these relationships. The other properties, aromaticity, pI and instability index did not have any correlations with other properties.

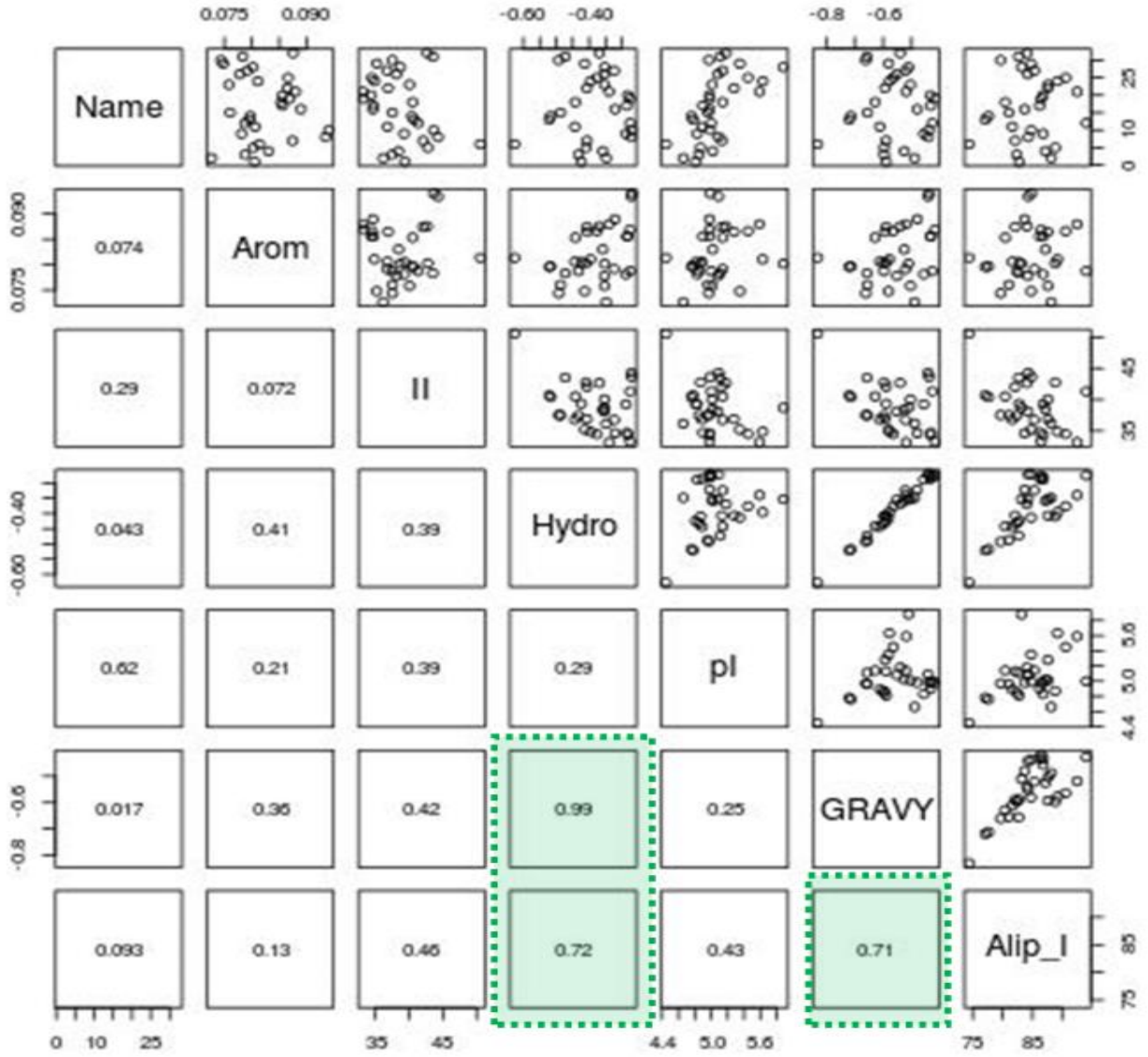


Figure 2.10.Scatter plot for group B proteins with Pearson correlation coefficients. The green highlighting shows properties that do have a correlation.

Group C Hsp90s, like other groups showed a strong correlation between hydrophobicity and GRAVY with a value of 0.99 that supported by linearized scatter plot. The correlation of aliphatic index with GRAVY and hydrophobicity (Fig 2.11) was not as strong as observed in group B Hsp90s. However, even though the Person coefficient values showed slight relations, the scatter plots did not clearly support the correlations.

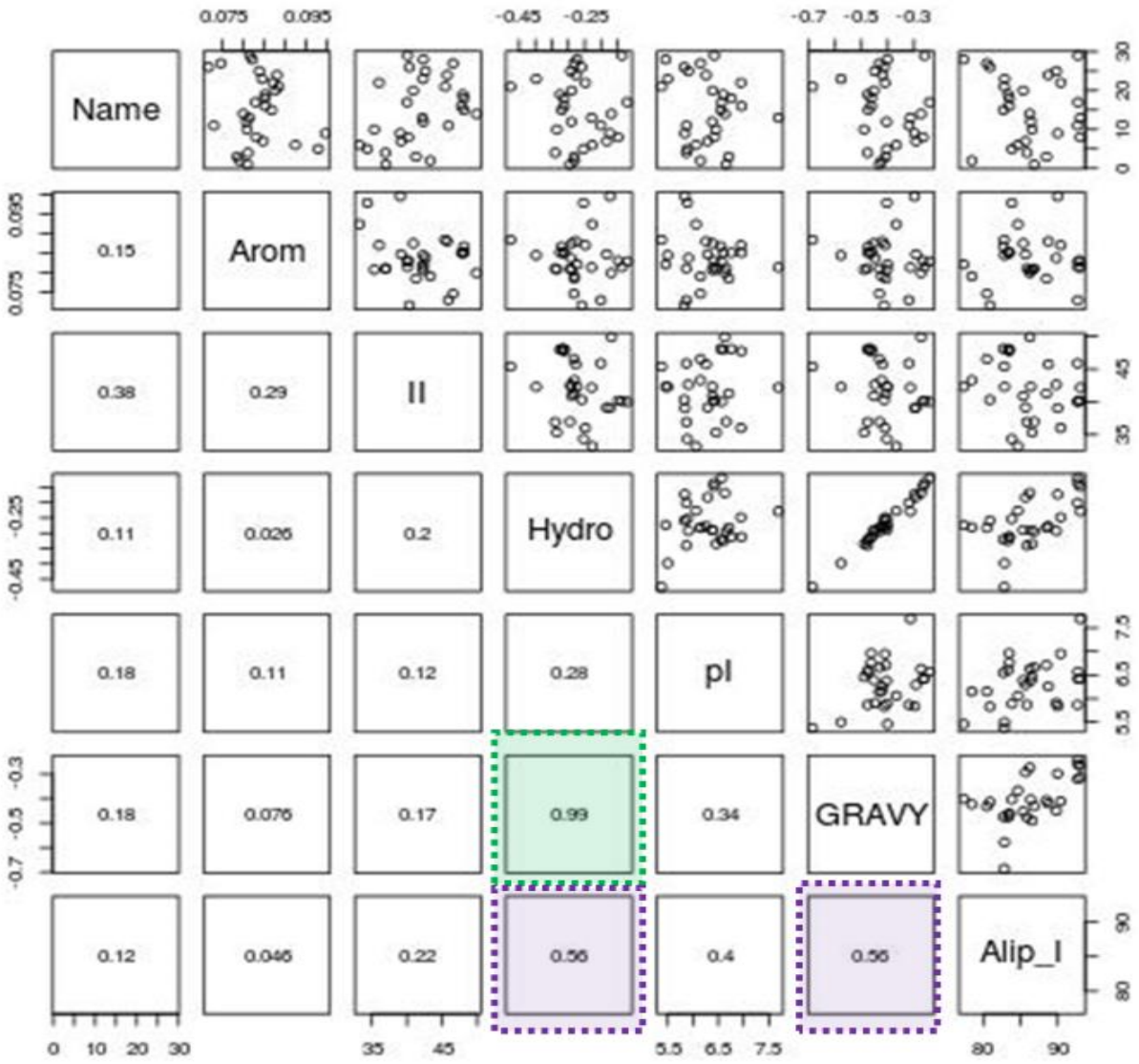


Figure 2.11.Scatter plot for group A Hsp90s with Pearson correlation coefficients. The purple indicating slight relations between properties while the green highlighting shows strong correlations.

2.3.3.2. Kruskal-Wallis test

The data produced from calculating the physicochemical properties in different groups is non-parametric and in this study we wanted to test whether samples originate from same distribution. Therefore, the Kruskal_Wallis test was used since it can be used to compare more than two populations using their medians. This method returns a *p*-value for a test of the null hypothesis that the data in each categorical group comes from the same distribution while the alternative

hypothesis simply means not all the samples come from the same distribution. The hypotheses are as follows:

H_0: same distribution

H_1: different distribution

If the *p*-value is less than α, which is set at 0.05, then reject the H_0 hypothesis (Spurrier 2003). The α-value is set at 0.05 because we are measuring at 95% significance level.

Table 2.7.The results of the *p*-values for Kruskal-Wallis test.

Physicochemical property		***p*-value**
Aromaticity		$1.168e^{-4}$
Instability index		$2.501e^{-3}$
Hydrophobicity		$3.706e^{-12}$
pI		$9.768e^{-14}$
GRAVY		$4.029e^{-12}$
Aliphatic index		$4.774e^{-06}$

All the *p*-values produced were less than 0.005 therefore we had evidence to reject the null hypothesis that the three groups had the same distribution for each property.

2.4. Discussion

In this Chapter, the Hsp90 homologs of each organism were placed in appropriate groups for an extensive analysis i.e. A: cytosolic, B: ER and C: mitochondria. The isoforms of the human Hsp90 all differ in the N-terminal site implying that the residues in the insertions could slightly cause some differences when they participate in the dimerization of the N-terminal domains during ATP binding (Richter et al. 2006). Other than the insertions, all the residues in the sequences are identical, indicating that molecules that bind to the nucleotide binding pocket are the same and have all the same affinity. Likewise, the client proteins and co-chaperones that are involved with the Hsp90 are also the same in all the isoforms.

The entire group A and C Hsp90s lacked a signal peptide because the Hsp90s are translated in their localization. The signal peptide target proteins to specific organelles (Patron and Waller 2007) therefore the presence of the signal peptides in group B Hsp90s mean they are transferred to the ER. The suggested general architecture of signal peptides is: the first 5 residues are mostly basic residues followed by rich region of hydrophobic residues (Von Heijne1990). The MSA of these signal peptides agreed with this suggestion as Lysine and Arginine residues were present in almost all the sequences and also the mid-region was observed to be conserved with hydrophobic residues. Another feature of peptide signals is the presence of polar residues near the cleavage site (Von Heijne 1990), of which our result was observed to have the same trend. However, a variation was observed in the *Plasmodium sp.* where there was a set of polar residues before the stretch of hydrophobic residues. This was also observed in a study on Hsp70s (Hatherley et al. 2013).

The overall analysis showed that the organelle in which the Hsp90 is found tends to play a role in determining the physicochemical properties of a protein. Orij et al, found that, the different organelles maintain a constant pH which is used to define the processes associated in that organelle (Orij et al. 2009). The ionization states of weak acids and weak bases including all the proteins are affected if pH levels changes. The average pI values were between 4.9 and 6.7 in all groups meaning they are weakly acidic in nature. Therefore, Hsp90 proteins have a negative charge in nature which supports what Csermely et al observed that; Hsp90 proteins tend to bind more to positively charged client proteins (Csermely et al. 1998). Group B Hsp90s have the highest Mr followed by group A and lastly group C. The small standard deviation clearly showed that group A Mr is conserved while there is a wide range of values in group B and group C. Hydrophobicity plays a major role in the folding of proteins. Cytosol Hsp90s showed to be more hydrophilic followed ER and lastly the mitochondrion Hsp90s. In this case, the environment is most likely to have a major influence as can be supported by the statistical *p*-value which is less than 0.05. The cytosol generally has higher water content than the organelles which explain why group A Hsp90s are the most hydrophilic. Aromatic residues are known to play a major role in the structural stabilization of proteins (Anderson et al. 1993) and studies have revealed that aromatic clusters and aromatic pairs increase the thermal stability of a protein (Serrano et

al.1991). Therefore, since the structure of Hsp90s is conserved, this explains why the aromaticity values are conserved as well.

GRAVY has been found to have a correlation with hydrophobicity (Barnejee et al. 2010); and in this study the same correlation was observed. It was also observed that the property is also highly conserved with very small standard deviations and all the proteins interact with water meaning they are hydrophilic. The TRAP1proteins have slightly higher GRAVY values ranging from -0.2 to -0.4 while group A and B proteins have values that are less than -0.6. This is probably due to the fact that the TRAP1 proteins are slightly acidic with the pI values close to 7, which is neutral. Water interact strongly with charged molecules therefore it interacts better with group A and group B proteins that a more acidic rather than the weakly acidic group C proteins. Guruprasad et al suggested that the primary determinants of the stability of a protein come from within the primary structure so calculating the instability index will give a rough idea of its half-life (Guruprasad et al. 1990). So, the instability indices show that only group B proteins are stable with a longer half-life. This information is useful when analyzing the proteins in vivo.

Group A

The molecular weights of the human and vector Hsp90s are quite high ranging from 82 to 84kDa while the range for parasitic Hsp90s is 82 and below. Mr is directly influenced by sequence length and this shows the parasitic Hsp90s are shorter as compared to the human sequences. This will be better analyzed in the next Chapter, through multiple sequence alignments and motif analysis. However, the *Plasmodium sp.* has larger Mr than the human hosts and their vectors. Kumar et al suggested that the extra residues of the *Plasmodium sp.* are found mainly in the charged linker region where they are involved in client binding and/or modulating ATP binding (Kumar et al. 2007). Therefore, the longer sequences of the human Hsp90s means that they have extra subsets of client proteins that binds to them.

The aromaticity for the human sequences and the vector sequences was observed to be slightly lower than the parasitic values. As has been suggested by Anderson et al, that aromatic residues are involved in maintaining the structure of a protein (Anderson et al. 1993),the parasites indicate to want to maintain their structure more because of high of environmental changes. The tertiary

structure of a protein is very important for its function and it should be maintained. The parasites interchange from the vector to the host and are exposed to various factors such as extreme pH and temperatures that will disrupt the 3D structure therefore parasitic Hsp90 take advantage of this feature to survive in all the environments. Instability indices are generally high in this group, well above 40. The trypanosomatids are observed to have the highest values with the rest of the group having a conserved property. Hoare et al observed that the life cycle of the trypanosomatids is short and they sometimes need only one host to complete it (Hoare et al. 1966) which explains why their Hsp90s have shorter half-lives.

The pI in this group is highly conserved as supported by the small standard deviation. The human sequences and the host sequences have pI values that are almost the same and the protozoan species have values that are slightly higher. The value shows that they are all weak acids but the values of the human host Hsp90s shows that the proteins bind strongly when compared to the protozoans Hsp90s. However, the *P. falciparum* and *Cryptosporidium sp.* Hsp90s have got pIs that are smaller than the human indicating that they bind to the client proteins more strongly. GRAVY value is used to estimate the interaction of a protein with water. Therefore, the negative value indicates the interaction with water and the lower the value the better the interaction (Kyte et al. 1982). The human Hsp90s were observed to have the lowest GRAVY values indicating their better interaction with water. The *Cryptosporidium sp.* had the lowest values and the rest of the parasitic Hsp90s had GRAVY values that were quite conserved. The reason for the human Hsp90s to interact more with water is likely to be influenced by the environment. The distribution of hydrophobicity and GRAVY is the same as they explain the same thing about water interaction. Aliphatic index is regarded as a positive factor for the increase of thermal stability of globular proteins (Ikai 1980). This property is quite conserved among human and vector sequences with the parasitic Hsp90s having remarkably high values. These high values indicate, the Hsp90s are stable over a wide range of temperatures. However, trypanosomatids i.e. *Leishmania sp.* and *Trypanosoma sp.* have uniquely low aliphatic indices when compared to their host. Lundkvist et al discovered that fever is not part of the symptoms of full blown sleeping sickness (Lundkvist & Bentivoglio 2004) which may explain the reason for uniquely low aliphatic index values.

Group B

The human, helminths, ectoparasite and vector Hsp90s showed a conserved molecular weight property of around 91kDa. This property was not conserved in the protozoans with the apicomplexan i.e. *Plasmosium sp, T. gondii, Cryptosporidium sp.* and *B. equii*, having uniquely larger Mr and the trypanosomatids having low Mr. The property is conserved per kind of protozoa. Aromaticity was highly conserved throughout, due to the fact that organelles tend to maintain certain environmental conditions themselves, therefore the parasitic Hsp90 does not have to be that unique for survival. The human, helminths, ectoparasite and vector Hsp90s have a conserved instability index property. The protozoans Hsp90s have quite low instability indices well below 40 except for the *Cryptosporidium sp.* The fact that the protozoans requires two organisms for them to complete their life cycles, and that, they always translocate to ER implies that they require a longer half-life if they are to survive which is supported by their low instability indices.

The pI property is highly conserved as all the values calculated were all around 5 with a very small standard deviation. As been suggested by Csermely et al, that the Hsp90s client proteins are positively charged, the pIs that are below 7 also supports that. Like in group A, the GRAVY values for the human and vector Hsp90s are very low indicating their better interaction with water. The *Leishmania sp.* and *Cryptosporidium sp.* had values that showed the lowest interaction with water. However, *B. hominis* has a uniquely low value that shows that it has the highest interaction with water when compared to other parasites and its hosts. The aliphatic indices in this group are highly conserved throughout the species. However, the protozoans Hsp90s have slightly higher aliphatic indices indicating that they are more stable over their host and vectors.

Group C

The Mr property is highly conserved around 80kDa with the apicomplexan organisms like other groups having slightly higher values. The conservation of the aromaticity property is like in grouping A where only the protozoan Hsp90s are observed to have some slight diversity. Like the group B Hsp90s, the organelle does most of the homeostasis. The instability property is quite conserved among the human, trypanosomatids and helminths Hsp90s. The apicomplexan showed some diversity as the Hsp90s from *Plasmodium sp.* were uniquely high while from *Babesia sp.*

was uniquely low. The best explanation for the shorter half-life of *Plasmodium sp.* Hsp90s could be that, once the Hsp90 helps with folding of proteins, they are shuttled to the host's erythrocytes where the host's Hsp90 takes over in the case of denaturation and refolding (Banumathy et al. 2003) and also about 2% of the genome is responsible for the production of Hsp90s implying they are expressed in abundance (Acharya et al. 2007).

The pI property in this group was observed to have slight differences when compared to each other. Human TRAP1 is the only protein that has an isoelectric point greater than 7, which indicates that it is positively charged in nature and is weakly basic. Therefore, due to this value, it is more likely that the human TRAP1 has got a different subset of client proteins as compared to others. The trypanosomatids have got uniquely high pI values indicating they are weakly acidic with the least binding strength as compared to other parasitic Hsp90s. Unlike other groups, the human Hsp90 had a uniquely GRAVY value that indicates its less interaction with water. The protozoan Hsp90s showed to interact better with water, with the *Plasmodium sp.* showing to interact more with water. Interestingly, the protozoan GRAVY and pI values tend to correlate, with species having low pIs having also low GRAVY values meaning water interacts more with molecules that have low pIs. The human and vector Hsp90s have got higher aliphatic indices than the parasites even though the property is generally conserved in the group.

2.5. Conclusion

In this Chapter, we observed that the human Hsp90s have isoforms that differ only in the N-terminal region due to insertions and deletions. However, the ER homolog does not have isoforms. Generally, the Hsp90's physicochemical properties are conserved as been shown by small standard deviations. This observation explains why the structures of the Hsp90s are conserved since the physicochemical properties are known to have influence in the folding of proteins. The apicomplexan Hsp90s have been observed to have Mr that is larger than human sequences which probably means they have a set of unique client proteins. The aromaticity property is quite conserved. However, due to the life cycle of the protozoans that require a host and a vector to be complete, they have a higher risk of denaturation. So, their aromaticity values are high in order to maintain their 3D structures. Only the human TRAP1 had a unique

isoelectric point that is different from all the Hsp90s in all the groups. Due to this nature, the mitochondrion Hsp90 can be promising targets. There is a strong relationship between hydrophobicity and GRAVY properties of the Hsp90s in all groups. A correlation between hydrophobicity and aliphatic index was also observed which shows hydrophobicity is mainly influenced by GRAVY and aliphatic index. The statistical analysis also showed that the distributions of the properties in each group are different meaning the environment also plays a major role in the overall properties of a protein.

CHAPTER THREE

3. Multiple Sequence Alignments (MSA), Motif and Phylogenetic analysis

The Hsp90s are essential for the survival of a cell and their functions are conserved across all kingdoms. Proteins that have a conserved function have features that are also conserved in their primary and tertiary structures. In this chapter, a comparison of the Hsp90 sequences is done through multiple sequence alignment and motif analysis. The motif results are further analysed through a python program that displays graphical outputs. The evolutionary distances between the human, vector and the parasitic sequences were determined by phylogenetic tree calculations.

3.1. Introduction

Hsp90s play a major role in the folding of newly synthesized proteins as well as regulation of gene expression and signal transduction processes (Walter & Buchner 2002). Generally, they support important cellular events. Parasites such as *Plasmodium* and *Leishmania* utilize the ability of Hsp90 to affect important cellular events throughout their survival (Banumathy et al. 2003) such as triggering transitions during their life cycles (Pallavi et al. 2010). For the reason that Hsp90 plays major role in development and growth, it has been implicated as a therapeutic target in a number of diseases. Successfully, it has been targeted as a cancer drug (Trepel 2010) and Hsp90-specific drugs have been tested in clinical trials (Hubbard et al. 2012). Not much has been done for parasitic infections in targeting the Hsp90 as potential drug target. Inhibition of Hsp90 function using GA has been observed to affect growth of the *Plasmodium* parasite in human erythrocytes *in vitro* (Banumathy et al. 2003) and *in vivo* (Pallavi et al. 2010). Pallavi and others also showed that inhibiting *T. evansi*'s Hsp90 would cure trypanosomiasis in mice. Therefore in this study a large scale analysis at the primary structure level was conducted to analyse other parasitic organisms.

Although, the function of Hsp90s is highly conserved throughout all kingdoms, there are significant differences between the parasitic and human Hsp90 at the primary structure level. These differences can be analysed through MSA and motif analysis. Phylogenetic analysis can also be used to find the sequences that are closely related.

MSA is a very important tool in sequence analysis as information about structure and function of a protein can be obtained. There are various automated programs for MSA but they cannot be fully relied on since different output can be obtained from the same set of input. Therefore, hand adjustment of the outputs might be necessary to increase the biological accuracy of the alignments. CLUSTALW (Thompson et al. 1994) was one of the first alignment programs to be created and is the most used program for MSA. However, according to Edgar et al. error rates are higher as compared to the modern programs (Edgar & Batzoglou 2006). This is because of the algorithm used, that does global alignment yet in most alignments the sequences are of different lengths. Mafft (Katoh et al. 2005), Muscle (Edgar 2004) and Promals3D (Pei et al. 2008) were used in this study for MSA. For high throughput, Mafft and Muscle have been praised for producing alignments quickly (Edgar 2004). Motif analysis is also important in analysing proteins at primary structure as well as to identify the function of unknown proteins. For this study, motif analysis was used for comparison of human sequences against parasitic sequences. Proteins in the same group or clade under evolutionary analysis share similar number of conserved motifs (Saha et al. 2013) therefore motif information can also be used to find closely related sequences. Various software programs are able to display motif information such as Scan Prosite software, InterPro Scan (Quevillon et al. 2005) and MEME (Bailey et al. 2009). MEME was preferred for this study as it searched for statistically significant motifs from MSA information resulting in new motifs being identified. On the other hand, ScanProsite and InterProScan produced fewer motifs and that have already been discovered. This is because they analyse sequences by searching for similar hits in the database. Therefore, in a bid to distinguish human and parasitic proteins, a method that identifies new motifs will be preferred hence MEME was used for further analysis

Phylogenetic studies on Hsp90 across all families have been carried out before (Fast et al. 2002). It has been found that the group A and group B Hsp90 constitute paralogous families which diverged a long time ago during evolution of eukaryotic cells (Chen et al. 2006). The group C are believed to have evolved separately from a common ancestor (Chen et al. 2006). In this study, phylogenetic studies are only restricted to the parasites that cause diseases to humans and the analysis are performed to observe the relations between sequences. The method group the data in such a way that, the objects placed in a clade are more related to each other than objects placed

in another. Proteins are naturally grouped into families and these families contain sequences that are conserved as evolution takes place.

3.2. Methodology

The sequences were analysed at primary structure level in the three groups (A, B and C) as indicated in Chapter 2 through MSA and motif analysis. The results obtained were then combined to see the suitability of the parasitic Hsp90 as potential drug targets. The evolutionary relationship analysis was also undertaken to see which parasitic proteins are more related to the human protein. The tools and the methods are discussed below.

3.2.1. Structure retrieval

All the structures used in this Chapter were retrieved from the PDB repository. PyMol (DeLano et al. 2002) program was used for structure visualization.

3.2.2. MSA

The alignment programs produce different results depending on the differences in size of the input sequences and also the type of algorithm implemented. In this study MAFFT, MUSCLE and PROMALS3D were used for each group of sequences. The G-INS-i strategy (Katoh et al. 2005) was implemented for MAFFT while the default settings were applied for the other programs. Jalview v2.7 software (Waterhouse et al. 2009) was used to view the alignments and the alignment with the best possible alignment was then taken for further analysis. Manual edit of the alignments was also done to improve the quality of the alignment. SIAS webserver was used to calculate the sequence identities. A python script (see Appendix 4, Script A4.1) was written to extract interesting regions in the alignment for further analysis.

3.2.3. Motif analysis

MEME v4.9.0 was used to get the motif information from each group of sequences. A python script was written for viewing the motif information in an easily understood manner. The script (see Appendix 4, Script A4.2) requires the MEME and MAST text files as inputs. A comma separated value (csv) file, heat map and bar graphs are the outputs of the script.

3.2.4. Phylogenetic tree analysis

Mega v5.05 software was the choice for constructing phylogenetic trees. A model test was done at complete and partial deletion to find the best evolutionary model that fits with the data. The top 3 models from each model test run were then used to construct trees and the best tree out of

the six trees was then used to further compare the human Hsp90 against its parasitic orthologs. The bootstrap of 500 replications was used to estimate the level of confidence.

3.3. Results

3.3.1. MSA

The alignment results from each group were split into four categories that are N-terminal domain, charged linker region, middle domain and C-terminal domain. The regions were analysed separately and the sequence identities (Fig 3.4) were calculated using a python script that uses the alignment file as the input and compare the human Hsp90 against all other sequences.

Group A Hsp90

Promals3D aligned the first residues (methionine) and the EEVD motif perfectly therefore the alignment result was used for further analysing group A sequences. (Appendix 5, Fig A5.1). The structure with a PDB ID, 2CG9, was the input structure for the alignment. The N-terminal domain is the ATP binding site and is conserved throughout the sequences with very few gaps/insertions in the alignment. Interestingly, *B. malayi* Hsp90 has a deletion in the conserved section of the N-terminal domain (Fig 3.1) but the residues important for ATP binding are conserved. This region has been mapped to the crystal structure (1US7) of the human N-terminal domain structure (Fig 3.2) and a missing α-helix secondary structure was observed. This region is located at the edge of the active site.

The charged linker region shows diversity (Fig 3.3) as indicated by a number of gaps/insertion in the alignment. The region is rich in acidic residues, Asp (D) and Glu (E) residues which also support the pI result in Chapter 2 that the group A Hsp90 are acidic in nature. The *Plasmodium sp*. sequences have an insertion in the linker region as has been observed by Pallavi and others (Pallavi et al.2010). Also, the *Cryptosporidium* species have short extensions of 5 residues rich in Asp residues. The middle domain as seen from appendix 1 is highly conserved with no unique features from sets of sequences. The C-terminal is also conserved and is characterized by the motif ME[DE]VD.

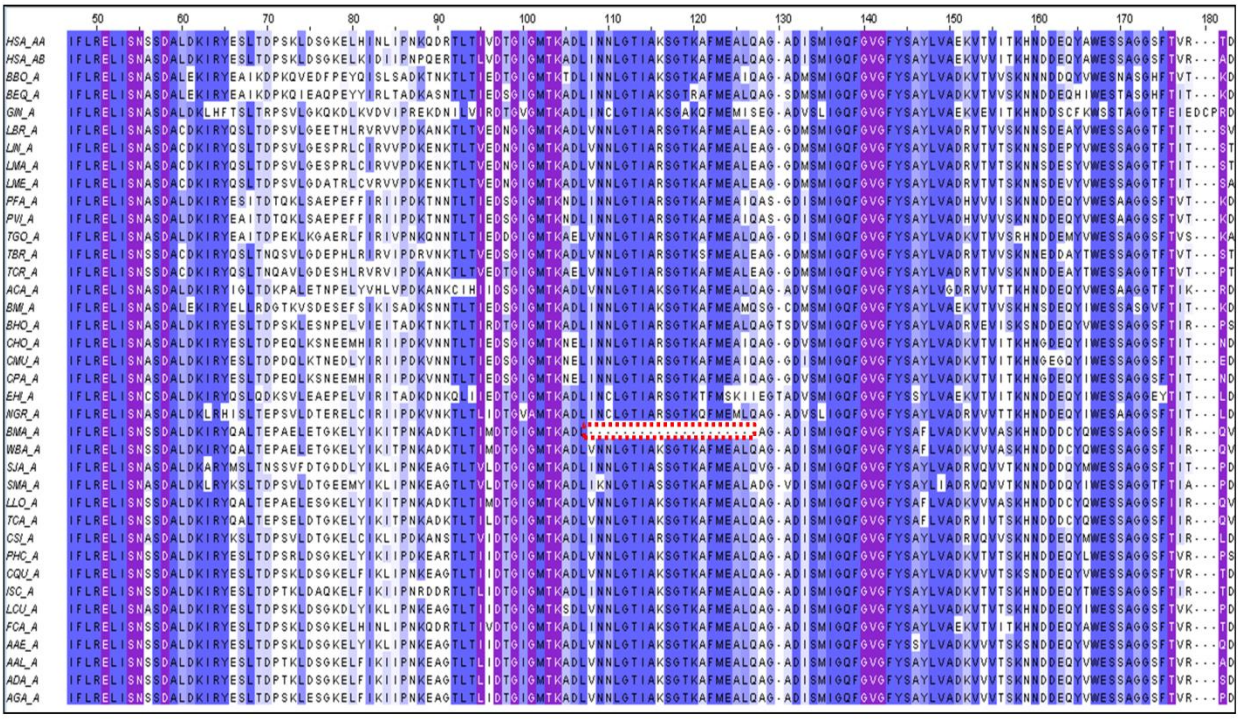


Figure 3.1.MSA of the N-terminal region of group A Hsp90s indicating a major deletion in the *B. malayi* Hsp90 sequence (red boundary). In purple are the residues involved in the ATP binding and in blue are the conserved regions.

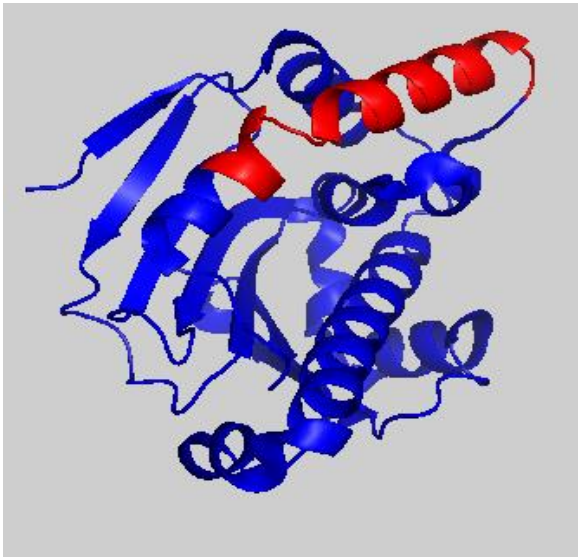


Figure 3.2.The N-terminal domain of the human cytosolic Hsp90 (1US7). In red is the missing region of *B. malayi* N-terminal domain.

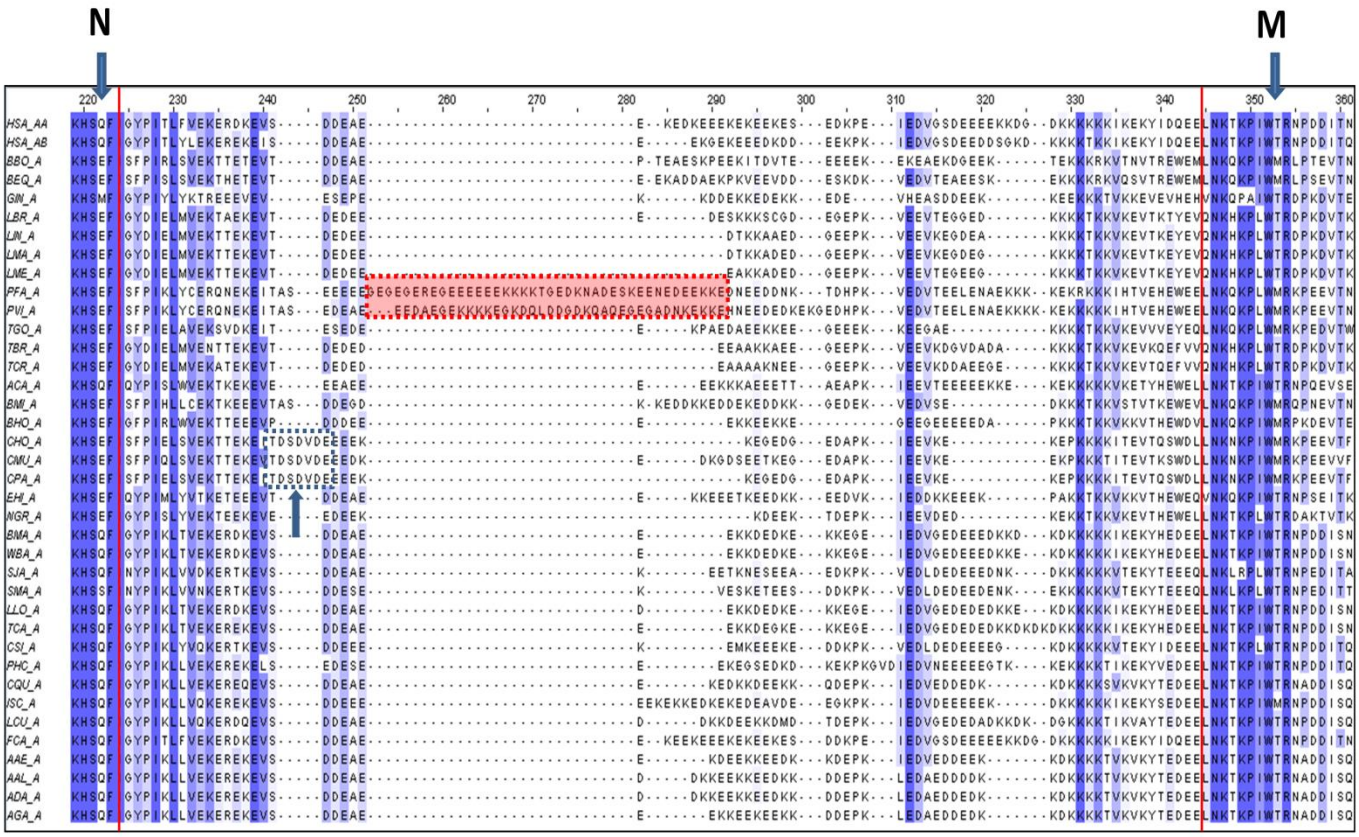


Figure 3.3.MSA for group A Hsp90 charged linker region. In blue are residues conserved throughout the sequences and in red background is the *Plasmodium sp.* insertion. The blue dotted background represents the conserved region in *Cryptosporidium sp.* Purple lines show the end of the N-terminal domain (N) and the beginning of the middle domain (M).

Group B Hsp90

There is no full structure for the ER Hsp90 in PDB therefore MUSCLE and MAFFT programs were used. The alignment result from MAFFT had more biological meaning therefore it was the best choice for further analysis. The N-terminal domain shows some diversity in approximately the first 50 residues and the ATP binding site is highly conserved throughout all sequences (Appendix 5, Fig A5.2). The group B Hsp90s contain signal peptides as discussed in Chapter 2 therefore, this could be the reason for the diversity observed. Unlike the group A charged linker region alignment, the charged linker region has no unique gaps/insertion however, the *Babesia* and *Trypanosoma* species have long deletions in the region.

Figure 3.4.MSA for the middle domain of group B Hsp90. The blue highlighting indicates residues conserved throughout the sequences. In A, red boundary indicates regions only common to the human and vector sequences, blue boundary: regions unique to the *Leishmania sp*. and yellow: regions unique to the plasmodium species. In B, red boundary shows region dissimilar to the non-protozoan sequences.

The middle domain of the group B shows variation (Fig 3.4A).The human and other vectors sequences have their own set of conserved residues while the *Leishmania sp.* and *Plasmodium sp.* have unique sets each. The group B protozoan species also have an extended middle domain with at least 20 amino acid residues longer (Fig 3.4B). The insertion is conserved in all protozoans there a structural architecture was constructed (Fig 3.5). The middle domain is involved in co-chaperone and client binding therefore this result shows that the protozoan could have their own set of client proteins unique to their host's and vector's client proteins. The C-

terminal domain is also conserved with the terminal KDEL motif conserved as well (Appendix 5, Fig A5.2). However, the *Babesia* species Hsp90 have some insertion making their C-terminal domain longer than any other species.

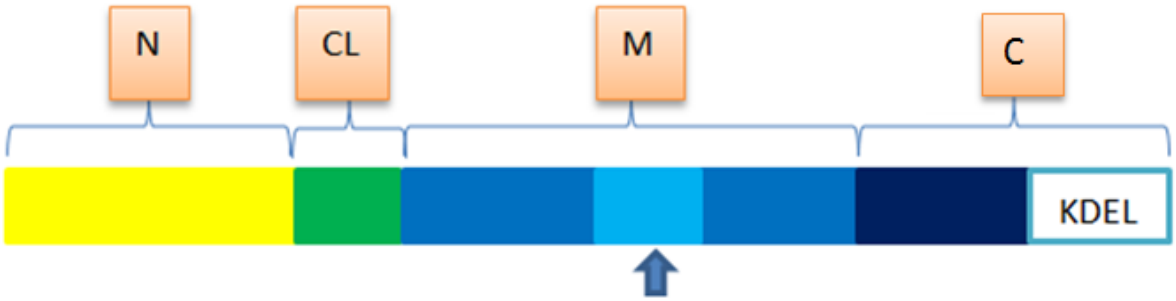


Figure 3.5.A schematic representation of the protozoan Hsp90 indicating its unique middle domain. The arrow shows the insertion. N: N-terminal domain, CL: charged linker, M: middle domain and C: C-terminal domain.

Group C Hsp90

As in group B, the MAFFT alignment result was used for further analysing sequences from this group. The overall result shows conservation in the group C Hsp90s (Appendix 5, Fig A5.3). The *Plasmodium* species have an extended N-terminal domain with an insertion of about 30 amino acid residues. The charged linker region does not exist in the TRAP1 sequences so the N-terminal domain is joined directly to the middle domain. The middle domain is highly conserved throughout the sequences but the *Plasmodium*, *Leishmania* and *Trypanosoma* species have insertions (Fig 3.6). There is no conserved motif at the end of the C-terminal but conserved motifs according to types of organisms are observed. The human and host sequences have L[ED][RK]H motif while the *Leishmania sp.* and *Trypanosoma sp.* have a unique P[TS]ADK. The helminths have a common SILTP motif (Fig 3.6).

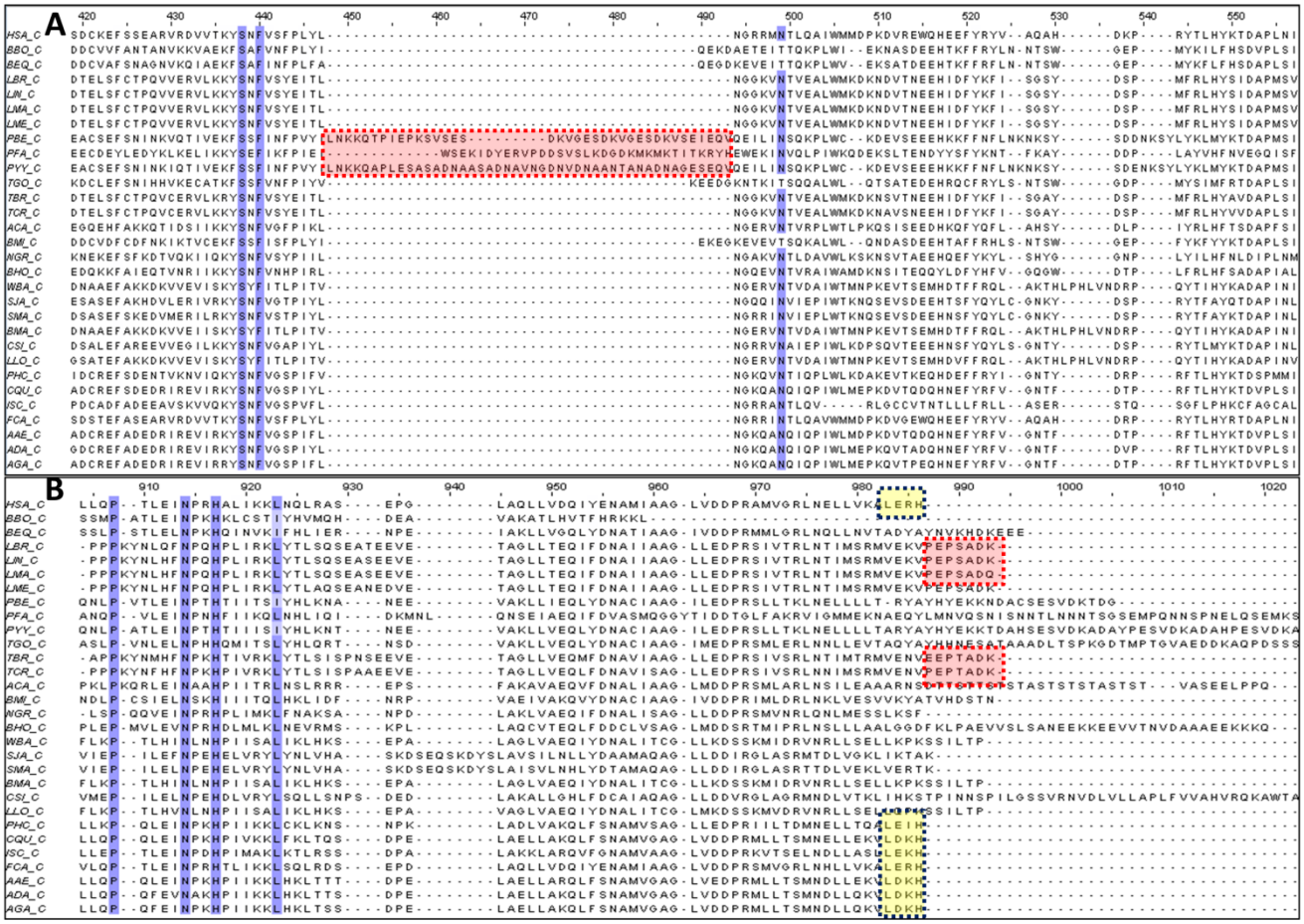


Figure 3.6.MSA for the middle and C-terminal domains of group C Hsp90. A (Middle domain): red boundary is the region only unique to the *plasmodium sp*. Blue line separates the two domains. B (C-terminal): yellow background represents host and the vector sequences; red boundary represents the trypanosomatids Hsp90s.

Sequence Identities

In group A sequences (Fig 3.7), the protozoan species had the lowest sequence identities (56-69%) against human Hsp90s, followed by the helminths group (70-79%) and lastly the ectoparasite group (81%). The protozoan species are unicellular organisms while the rest are multicellular. The ectoparasite had a sequence identity that is even higher than most of the vector sequences indicating that it is closely related to human Hsp90. The distribution of sequence identities in group B is the same as of group A but the values are slightly lower. The protozoan had sequence identity values ranging from 39-55% while the helminths and ectoparasite had value from 60-69% and 68.9 respectively. *Acanthamoeba* Hsp90 has a uniquely high sequence

identity of 55% as compared to other protozoan species in the group. Like other groups, the group C ectoparasite had the highest sequence identity (66.1%) when compared to other parasites. The helminths have higher values (56-64%) when compared to the protozoan Hsp90 (34-63%). The *Plasmodium* species Hsp90 have uniquely low sequence identities (below 50%).

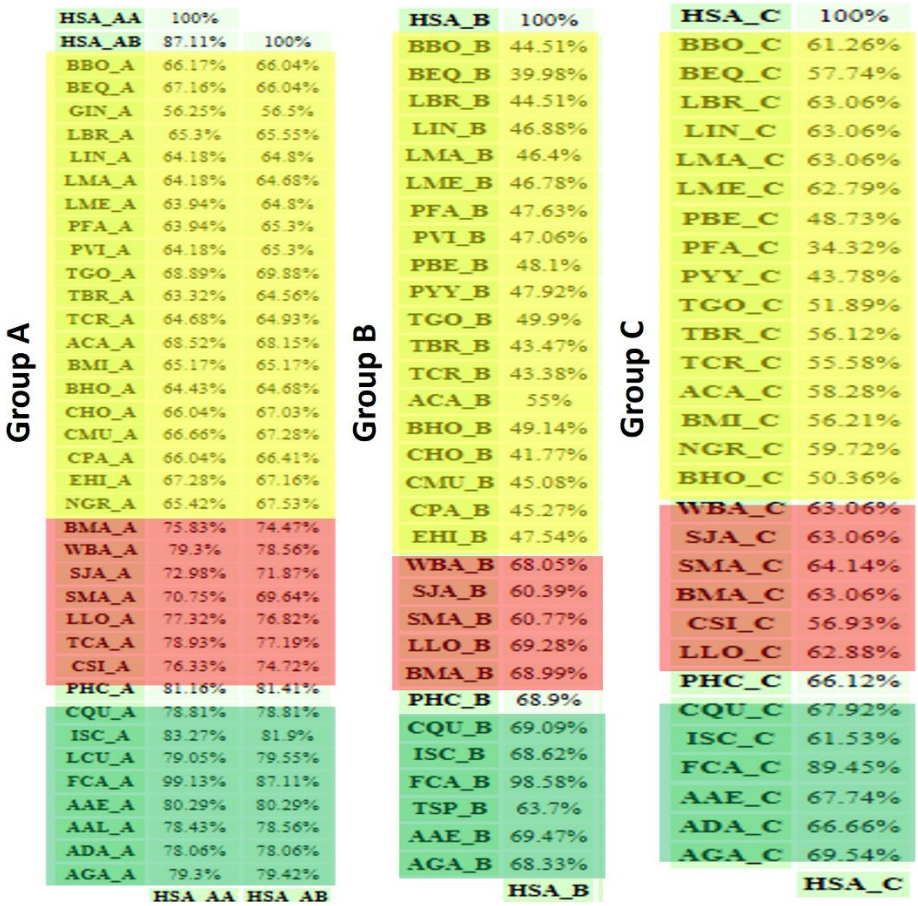


Figure 3.7.Sequence identity comparison of the Hsp90 of human against other sequences in the study. Yellow background: protozoan Hsp90, red background: helminths Hsp90, white background: ectoparasite Hsp90 and green background: vector Hsp90.

3.3.2. Motif analysis

Analysing motifs is very important in understanding the residues involved in maintaining the structure and function. Using the MEME default settings the program exhausted all the possible motifs in the sequence. The MEME output file was not informative for detailed analysis, therefore a python script was written to visualize the data in various ways (Appendix 4, Script A4.2). The script read overlapping motifs from the MAST output and removes them in the final output.

Group A

MEME output shows that there are 93 motifs in this group with about 18 motifs highly conserved throughout all sequences (Fig 3.8). The heat map represents motifs according to their conservation. More attention was paid to the motifs that are missing from the human Hsp90. Motif 21was observed to be uniquely missing from the protozoan and helminths Hsp90s. This motif is located in the highly conserved N-terminal domain. Motifs 30-93 are unique to 2 or 3 organisms. Length plays an important role in the number of motifs found per sequence (Appendix 6, Fig A6.1) as the longer sequences of *Plasmodium sp.* and *Brugia sp.* have more motifs. The number of motifs per sequence in group A ranges from 21-28.

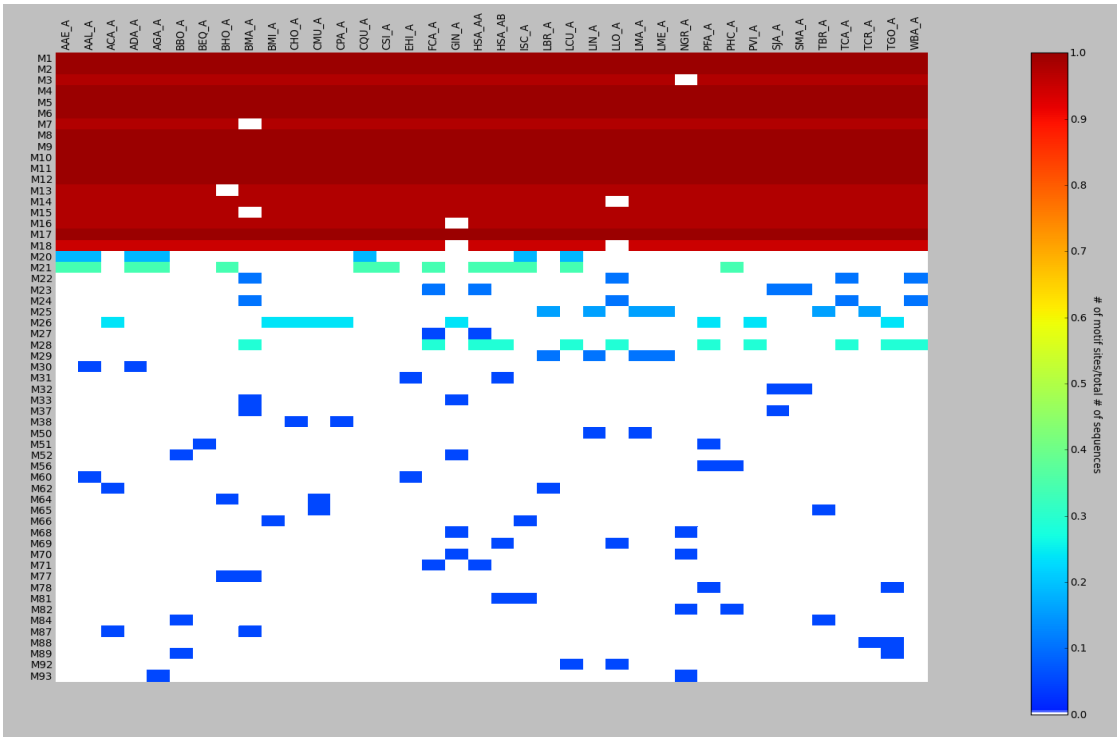
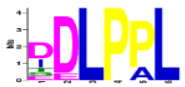


Figure 3.8.Heat map summarizing motif information for group A Hsp90. A: white regions show sequences that lack a motif and the level of conservation increases from blue to red.

Motifs that were used for further analysis are shown in Table 3.1. Based on the regular expression, the motifs that are only unique to the parasites are conserved e.g. motifs 24 and 26 for helminths and protozoan Hsp90 respectively. The unique motifs are located in the variable regions with the *Leishmania sp*. having a unique motif (28) rich in acidic residues. Motifs in this group are observed to be unique mostly in the middle domain and C-terminal domain. The N-

terminal domain motifs are highly conserved with almost all species having the same motif in this region. The helminths Hsp90 has a unique motif of about 12 residues found at the beginning of the N-terminal domain.

Table 3.1 A summary of unique motifs found in group A Hsp90s. The regular expressions are extracted from the MEME html file.

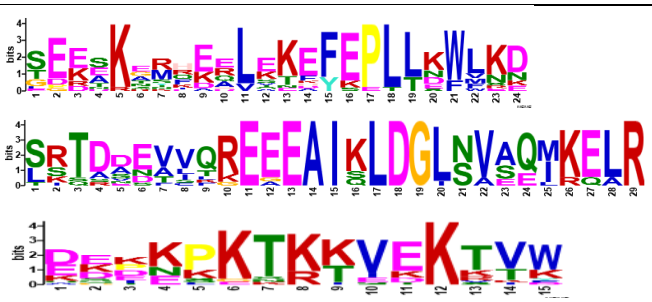
Motif(s)	**Species**	**Positions**	**Regular expression**
21	Host, vector and ectoparasite	174-182	
24	Helminths	1 - 7	
25	*Leishmania* and *Trypanosoma*	163 - 170	
26, 28	Protozoan	158 – 166, 275 – 283	 AND 

Group B

Motif diversity was observed in group B sequences (Fig 3.9A). Only 15 motifs out of the 100 motifs obtained were conserved throughout the species. The *Leishmania sp.*, *Plasmodium sp.* and *Cryptosporidium sp.* have motifs that are unique to themselves (Table 3.2). Interestingly, other protozoans from apicomplexan and trypanosomatids groups lack these motifs, indicating that they are only unique at genus level. These entire motif patterns are lacking in the human Hsp90. Motifs 17, 19 and 23 are absent in all the protozoan Hsp90 sequences (Appendix 6, Fig A6.2). *Leishmania sp.* and *Plasmodium sp.* have the highest number of motifs unique to themselves with 5 (Appendix 6, Fig A6.3) and 4 (Appendix 6, Fig A6.4) respectively. Their regular expressions show that, the motifs are conserved and some are 29 residues long. These long motifs are significant for the function of the Hsp90 as they are located in the conserved regions of the protein. The number of motifs per sequence range from 21-33 (Fig 3.9B) with the human sequence having as many as 29 motifs. Like in group A, N-terminal domain motifs are highly

conserved with slight variations in the first 50 residues. Diversity of motifs is observed in the middle domain and C-terminal domain.

Table 3.2: Motifs present to specific species. The regular expressions are extracted from the MEME html file.

Motif(s)	Species	Positions	Regular expression
17, 19, 23	All species except protozoan	42 - 71 601 – 625 664 - 679	
18, 33	*Leishmania* and *Trypanosoma*	607 - 557	
24, 26, 28, 31, 33	*Leishmania*	1 – 30 576 – 605 441 - 462 730 – 759 82 - 93	(see Appendix 6, Fig A6.3)
27	*Cryptosporidium*	714 - 764	
25, 29, 36, 37	*Plasmodium*	1 – 30 613 – 642 679 – 694 765 - 784	(see Appendix 6, Fig A6.4)

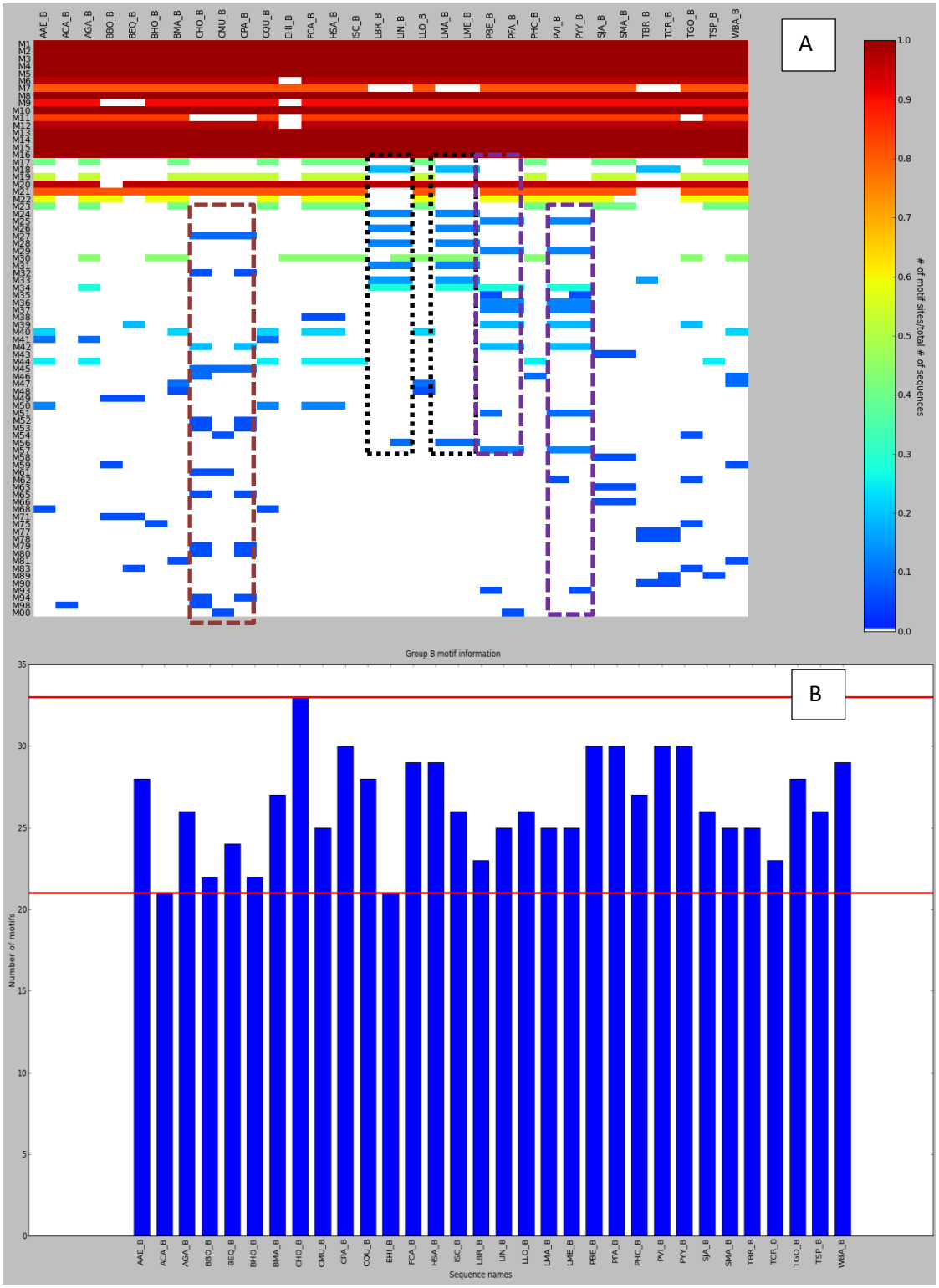


Figure 3.9.A summary of the motif information for group B Hsp90. A: white regions show sequence lacking a motif. Maroon boundary represents motif unique to *Cryptosporidium sp.*, black boundary represents motifs unique to *Leishmania sp.*, purple boundary: *Plasmodium sp.* B: the red lines show the minimum and highest bars.

Group C

As in the other groups, about 15 motifs are conserved throughout all the species. The group C sequences have the highest diversity with motif number per sequence ranging from 21-34 (Fig 3.10B). Motif 18, 24-29 are unique to protozoan Hsp90 (Fig 3.10A) with motif 28 located in the N-terminal domain and the rest in the middle domain. The *Plasmodium* species has a large number of unique motifs followed by the *Trypanosoma sp.* The *Leishmania* Hsp90 is highly conserved to the members of its genus as seen from the heat map that they have the same motifs and the bar graph showing equal number of 22 motifs. As in group B, *Leishmania* and *Trypanosoma* species have motifs that are only unique to them as they are in the trypanosomatid group. The human Hsp90 lacked unique motifs while parasitic Hsp90 had unique sections showing that there are regions important for the function and survival of these organism's (parasites) proteins.

Table 3.3. Motifs unique to certain species in group C Hsp90. The regular expressions were extracted from the MEME html file.

Motif(s)	**Species**	**Positions**	**Regular expression**
16	All species except protozoan	561 – 582	
18, 28, 35	*Leishmania* and *Trypanosoma*	462 – 491 50 – 58 586 – 592	(see Appendix 6, Fig A6.6)
21,27, 41	*Brugia*, *Wuchereria* and *Loa*	1 – 41 514 - 528 282 – 288	

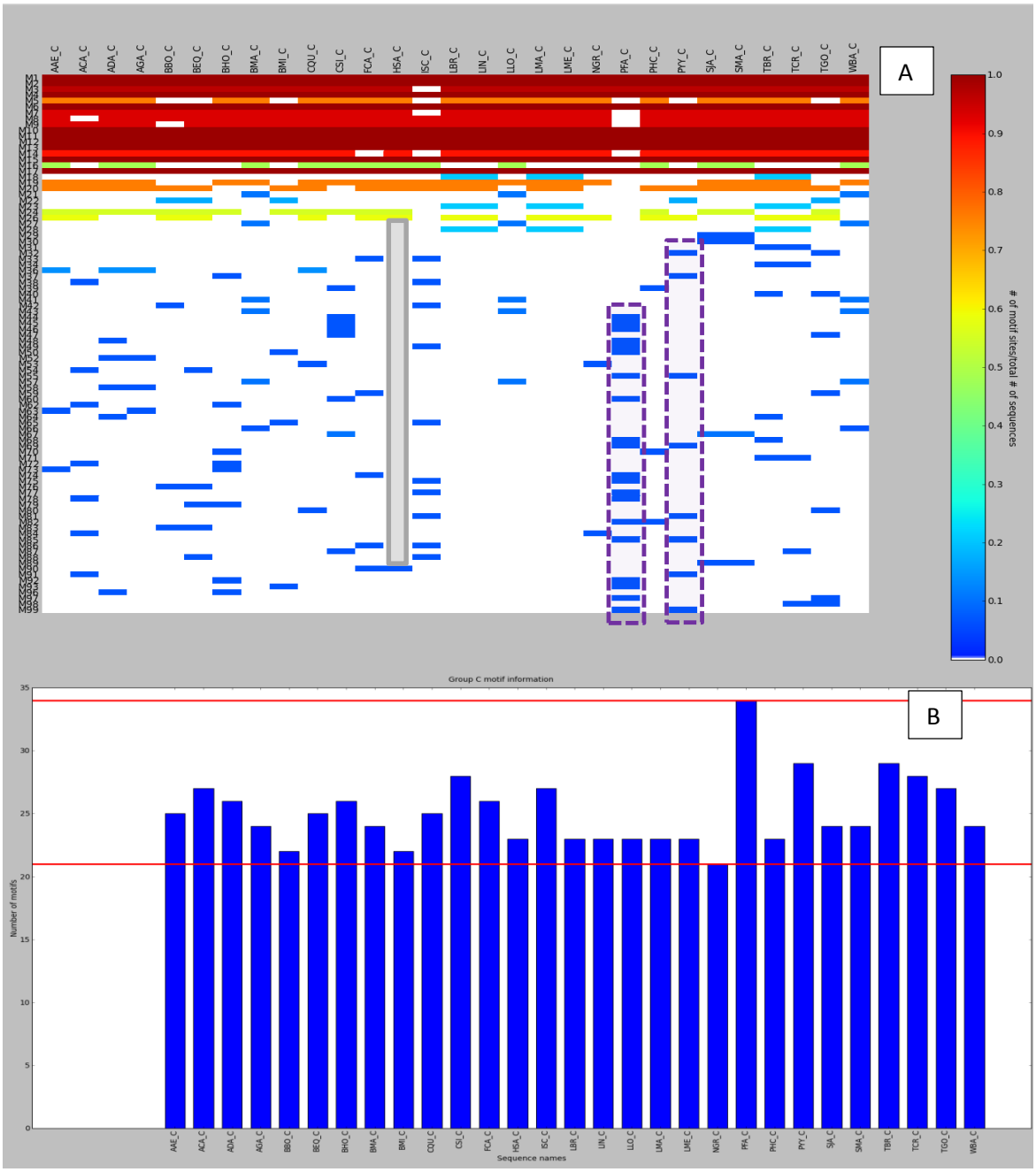


Figure 3.10.Group C Hsp90 motif information represented in a heat map (A) and bar graph (B). In purple boundary are the motifs unique to the *Plasmodium* species. The grey bar represents the human Hsp90 region absent of motifs.

3.3.3. Phylogenetic analysis

The evolutionary history was inferred by using the maximum likelihood method based on the General Reverse Transcriptase + Freq. model (Dimmic et al. 2002). The best evolutionary

models calculated using Mega5 at 95% deletion (Appendix 7, Table A1.1) and 100% deletion (Appendix 7, A7.2). Using the top 3 models of each calculation, 6 trees were constructed. The branching in all the 6 trees was the same and the tree with the highest bootstrap values (Fig 3.11) was then used for further analysis.

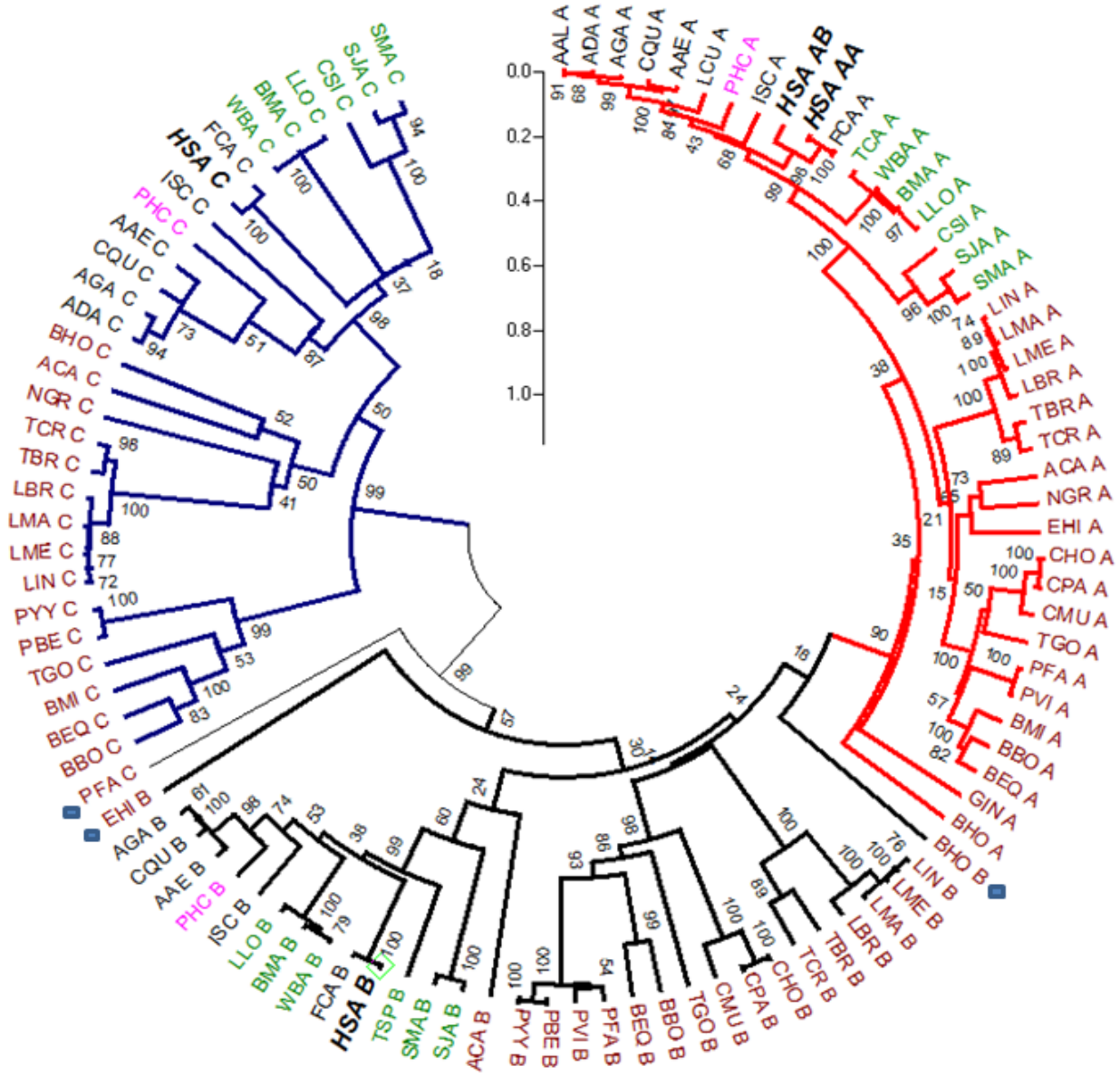


Figure 3.11.Molecular phylogenetic analysis by maximum likelihood method; the three colored branches represent red: group A proteins, black: group B proteins, blue: group C proteins. The names are colored according to the type with black representing host and vectors sequences, green: helminths and maroon: protozoan proteins. The blue squares indicate proteins that do not branch with other proteins of the same type of parasite.

Three distinct branches for the three groups were clearly observed with the group C sequences forming an out branch (Fig 3.11). The branches are supported by the high bootstraps values of 99%.

Group A

The human Hsp90 branches together with the vector sequences to form their own cluster. The bootstrap value of 99% is very high to support the branching. The ectoparasite also branches in the host's clade showing its close relation with the human Hsp90 as has been observed by its high sequence identity. The helminths Hsp90 form their own clades with very high bootstrap values of 99 and 100%. In the group A, the protozoan species branches out to form outgroup, even though the bootstrap value (38%) at the branching is very low. The protozoan branches in such a way that species from the same genus are branched together with very high bootstraps. The *Giardia* species is located at the base of the group.

Group B

The protozoan species branches out, to form a separate group but the branching is supported by low bootstrap value of 30%. As in group A phylogeny, the species form the same genus form clades indicating their close relation. The *B. hominis* (BHO_B) Hsp90 overlaps to be positioned at the basal of group A Hsp90 where it is closely related with its cytosolic homolog. The bootstrap value is low (18%) to confidently support the branching. Likewise the *A. castellani* (ACA_B) Hsp90 is not placed in the same branching with the other protozoan species. This was expected as was observed from a uniquely high sequence identity when compared to the human sequence. The helminths species were not branched together but *B. malayi* (BMA_B) and *W. bancrofti* (WBA_B) were grouped together with the human Hsp90 showing their close relation. This was well supported by the bootstrap of 99%. The *E. histolytica* (EHI_B) Hsp90 is also placed far from other protozoan species near the group C sequences. The bootstrap value of 57% can support that the EHI_B protein is not closely related to the other Hsp90.

Group C

Uniquely the group C *P. falciparum* (PFA_C) Hsp90 do not fall in the same branches with the Hsp90 from other organisms of the same type. The bootstrap value of 99% clearly supports that

PFA_C protein is closely related to the group B Hsp90. Also, the very low sequence identity that was obtained previously supports the idea. The protozoan branches out to form their group while the helminths are branch out together with the human Hsp90. The bootstrap value of 98% percent supports the fact that the group C helminths are closely related to the human Hsp90. The species from the same genus are observed to form their own clades.

3.4. Discussion

The Hsp90 primary structure is divided into four conserved regions that are an amino-terminal nucleotide binding domain, charged linker region, middle domain and the carboxyl-terminal dimerization domain (Prodromou et al. 1997). In this study, a closer analysis of these regions was undertaken.

Group A Hsp90

The N-terminal domain from different organisms is highly conserved (Fig 3.1).However, *B. malayi* Hsp90 has a deletion in this region that affects 2 α helices, but the residues that bind to ATP are still conserved. Valli and others in their studies highlighted that sensitivity to ATP differs according to species (Valli et al. 2010). This gap might have an effect in the sensitivity to ATP and its inhibitors such as GA and its homologs. The group A *Plasmodium sp.* have an extended linker region. Studies have revealed that the charged linker region plays a crucial role in the sensitivity of ATP (Hainzl et al. 2009) which explains why the Hsp90 from this organism was able to be selectively targeted using low concentrations of the inhibitor *in vivo* (Pallavi et al. 2010). In Pallavi's studies, under same conditions, the Hsp90 from *T. evansi* was also successfully targeted. Interestingly, the organism's Hsp90 lacks the extended charged linker but the inhibitor in low concentrations could still bind which showed that, increased sensitivity is not only influenced by the charged linker region but elsewhere in the protein. The middle domain was observed to be highly conserved with very few gaps/insertions in the alignment. The C-terminal is also conserved with a distinct insertion in the host and vector sequences. The C-terminal domain is also suggested to contribute to substrate binding (Hagn et al. 2011).

Therefore, the insertion unique to these sequences indicates that there are additional substrates that only bind to them.

The insertions found in the MSA were also observed to be motifs unique to those sequences. Due to the high conservation of the Hsp90 sequences, unique motifs were found in groups e.g. the protozoans had their own motifs (Table 3.1). However, *Leishmania* and *Trypanosoma* shared motifs unique to themselves indicating their close relation. This was also supported by a 100% bootstrap at the branch that separates the two genera.

Group B Hsp90

The N-terminal domain has an extended variable region at the beginning of the sequence. This is most likely due to the fact that the group B Hsp90 have signal peptides (as seen from Chapter 2), therefore these residues have no need for conservation as they will be chopped off in the functional protein. Diversity in the group B sequences is mainly found in the middle domain (Fig 3.3). The middle domain plays a major role in the activation of ATP hydrolysis in the N-terminal domain (Pearl et al. 2006) and binding of co-chaperones (Ali et al. 2006). The middle domain is believed to be involved in substrate binding (Park et al. 2011; Street et al. 2011). The middle domain of the protozoan Hsp90 is extended by approximately 45 residues suggesting that a lot more substrates bind to the domain. Also the *Leishmania and Plasmodium* species have a unique set of residues in the middle domain.

The motif analysis in group B showed that there is some variation among sequences even though the Hsp90 still maintains its function. Once again, the protozoan species had unique motifs with *Leishmania* and *Trypanosoma* having similar motifs. The Hsp90 from the two groups are closely related as this can be supported by the high valued bootstrap of 100%. Basing on the regular expression from the MEME html file, the residues have high bits values, meaning the motifs are highly conserved. The level of conservation indicates importance in function and in this case, the motifs might function in binding specificity.

Group C Hsp90

Parasitic group C proteins were more closely related to the human TRAP1 as they had sequence identities well above 50% as compared to the 40's of the group B sequences. The N-terminal

domain shows conservation only in the ATP binding site. The middle domain, like in other groups is highly conserved with the *Plasmodium sp.* having a region of about 40 residues unique to themselves. Unlike the C-termini of the group A and B sequences that have a MEEVD (Ballinger et al. 1999) and KDEL motifs respectively (Munro et al. 1987), group C do not have a unique motif. The human and the vector sequences contain an LEKH motif while the *Leishmania* and *Trypanosoma* Hsp90s contain an [ST]ADK motif.

Like the group A Hsp90, the group C motifs are conserved with at least 15 motifs present in almost all the species. The motifs are arranged in such a way that each genus has got its own unique motifs with the *Plasmodium sp.* having the highest number of motifs per sequence. The *Leishmania* and *Trypanosoma* Hsp90s share similarity in some motifs as has been observed in other groups.

Phylogenetic Studies

Hsp90s have got a high degree of conservation therefore evolutionary studies can be studied with confidence in a bid to distinguish the proteins using evolutionary distances. The group A and group B Hsp90 form their own branch while the group C Hsp90 have their own branch with very high bootstrap values of 99% each. Basically, the phylogeny analysis showed that the sequences arose from a common ancestor despite the sequence divergence among Hsp90 from different groups. As been suggested by Gupta (Gupta, 1995), the Hsp90 from group A and group B constitute paralogous gene families that was caused by a gene duplication during cell evolution a long time ago, our results also supported this theory (Fig 3.11). In all the groups, the protozoan species were observed to be at the base of the clades or forming their own clusters. This is because the protozoan species have been accepted to be the origin of eukaryotic diversity (Chen et al. 2006) and the Animalia kingdom can be traced back to the protist origin (Cavalier-Smith 2004). Our phylogeny clearly indicates that the protozoans Hsp90 are distantly related to the human Hsp90 while the ectoparasite and helminths are closely related. The unique branching of group C *P. falciparum* is most likely to be caused by the length of the sequence. The length of the sequence is in the same range with the Hsp90s of group B. Sequence identities (Appendix 7, Fig A7.1) clearly shows that the sequence is distantly related from the group B human Hsp90. Therefore, the length of the sequence has a major influence on the final phylogenetic tree output.

Phylogenies that have been done before placed *Giardia sp.* at the base of the clade (Arisue et al. 2005) and our phylogeny also clearly show that *G. intestanilis* is at the basal of the eukaryotic. From the cluster analysis, parasitic species from the same genus were observed to form their own clusters which mean they share high percentages of similarity and factors that affect the function of the other will definitely affect the other. The Hsp90 of the ectoparasite shows little differences with the human sequence as observed from the high sequence identities and absence of unique motifs. The use of the Hsp90 from the ectoparasite as a drug target will be difficult as there were no significant differences found. The major reason that could have led to the absence of major difference is the insufficient availability of the sequences in databases.

3.5. Conclusion

In this Chapter, we observed that the Hsp90 N-terminal domain has variations at the beginning of the sequences and the residues involved in binding ATP are conserved. The protozoans Hsp90s have higher sensitivity to ATP and studies previously suggested that the linker region is the causative in *Plasmodium sp.* However in this study, we observed that the other protozoan Hsp90s did not have an extended linker region therefore we suggested that sensitivity is also increased by residues elsewhere in the protein. Due to the insertions and unique motifs in the middle domain of the protozoan Hsp90s, we suggest that, this is another reason for increased sensitivity to ATP. Motif analysis showed that, despite the conservation of the Hsp90 across all kingdoms, there are still regions only unique to specific organisms. Phylogenetic analysis clearly indicated that the protozoan Hsp90s form their own clusters in every group. This result shows there are closely related and the functions and features might be conserved. In the tree, the human Hsp90 was placed at a distant from the protozoans in all the groups. Based on the phylogenetic tree and motif analysis, the *Leishmania* and *Trypanosoma* Hsp90are closely related. Finally, the Hsp90 from *Plasmodium*, *Leishmania* and *Trypanosoma* species showed distinct differences when compared to their human orthologs and we could confidently suggest them to be studied further in a bid to define them as potential drug targets for parasitic infections. The helminths and ectoparasites Hsp90s were difficult to define as we had insufficient results to support that. However, due to time constraints we were not able to model structures and map the motifs to analyse structural differences.

CHAPTER FOUR

4. Conclusions

This study paves a way in understanding the major differences in the physicochemical properties and primary structure of the human and parasitic Hsp90s. The approaches were conducted in a large dataset but similar kind of approach can be also implemented in smaller and even complex larger datasets of similar or different protein families. The physicochemical properties are important to know as they have an influence on the overall 3D structure of a protein thereby affecting the function. In a bid to define a drug target for human diseases, the target protein should be either be absent from the human body or should have significant differences when compared to the human protein else hazardous effects would be experienced.

In this study, over 80 Hsp90 sequences from disease causing parasites were retrieved (Table 2.2) and comparative analysis was carried out by calculating the physicochemical properties (Table 2.3, Table 2.4, Table 2.5 and Table 2.6). Basing on the standard deviations that are small, the properties are highly conserved throughout all the species. However, aliphatic index had higher standard deviation values and when individual species were compared (Appendix 1), the protozoan Hsp90s had uniquely slight high values. Aliphatic index can be used as a measure for thermal stability of a protein (Ikai 1980), therefore the protozoan Hsp90s are generally stable over a wide range of temperatures. This is so because of environmental changes that occurs during the life cycle of the parasite. Hydrophobicity is very useful in determining the structure of a protein and from the analysis we observed that this property is highly conserved (Appendix 1) with very small standard deviations (Table 2.3). Likewise, aromaticity, which is a property important for structural stability is also highly conserved. Therefore these results imply that, even the structures of the Hsp90s are highly conserved, so regions that differ in the tertiary structure will not have much influence on the function of the protein. The *Plasmodium sp.* Hsp90 showed property deviations in all the groups which could probably explain why it was the first to be successfully targeted

(Pallavi et al. 2010). Generally, the properties are highly conserved in groups i.e. apicomplexan and trypanosomatids showed conservation among their Hsp90s. Statistical analysis showed that distribution of the Hsp90 physicochemical properties is influenced by the environment.

MSA showed that the Hsp90s are generally conserved with sequence identities well above 40%. Studies revealed that the Hsp90s of *Plasmodium* and *Trypanosoma* have a higher affinity for ATP and its inhibitors (Pallavi et al. 2010) and the reason was because of an extended charged linker region. In our analysis, we observed that the extended region is absent in other protozoans which led us to the conclusion that affinity for ATP is also influenced by other regions elsewhere in the protein. In this study we also discovered that the ER protozoans have a unique extension in the middle domain which led us to design a new architecture of their structure (Fig 3.5). Motif analysis revealed that, the protozoans have more unique and conserved motifs. Trypanosomatids and apicomplexan Hsp90s had the highest number of unique and conserved motifs. Phylogenetic analysis revealed that the Hsp90s of protozoans are distantly related to the human while the helminths and ectoprasite are closely related to the human. With this information, we concluded that the Hsp90 from protozoans could be a way forward in designing anti-parasitic drugs. We did not have enough evidence to support the helminths and ectoparasite Hsp90s as potential drug targets for parasitic diseases.

Future work on defining parasitic Hsp90s as potential drug targets would involve structural modelling. Since there are no full structures for parasitic Hsp90s in the databases, homology modelling would be ideal, using the available bacterial and yeast structures. In this study we observed that the protozoan Hsp90s have a number of unique motif patterns. Therefore, mapping these unique patterns on the modelled structures will give information about their locations in structures. The comparisons can be done on the modelled human and parasite Hsp90 structures to observe the differences in motif positioning.

5. References

Acharya, P., Kumar, R. & Tatu, U., 2007. Chaperoning a cellular upheaval in malaria: heat shock proteins in Plasmodium falciparum. *Molecular and biochemical parasitology*, 153(2), pp.85–94.

Ali, A., Krone, P. H., Pearson, D. S. & Heikkila, J. J., 1996. Evaluation of stress-inducible hsp90 gene expression as a potential molecular biomarker in Xenopus laevis. *Cell stress chaperones*, 1,pp.62–69.

Ali, M.M.U., Roe, S.M., Vaughan, C.K., Meyer, P., Panaretou, B., Piper, P.W., Prodromou, C. & Pearl, L.H., 2006. Crystal structure of an Hsp90-nucleotide-p23/Sba1 closed chaperone complex, *Nature* 440(7087),pp.1013–1017.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J., 1990. Basic local alignment search tool. *Journal of molecular biology*, 215(3),pp.403-10.

Anderson D. E., Hurley J. H., Nicholson H., Baase W.A. & Matthews B. W., 1993. Hydrophobic core repacking and aromatic-aromatic interaction in the thermostable mutant of T4 lysozyme Ser 117-->Phe. *Protein Science*, 2(8), pp1285–1290.

Arlander, S.J.H. et al., 2006. Chaperoning checkpoint kinase 1 (Chk1), an Hsp90 client, with purified chaperones. *The Journal of biological chemistry*, 281(5), pp.2989–98.

Bailey, T.L. et al., 2009. MEME SUITE: tools for motif discovery and searching. *Nucleic acids research*, 37(Web Server issue), pp.W202–8.

Ballinger, C.A. et al., 1999. Identification of CHIP , a novel tetratricopeptide repeat-containing protein that interacts with heat shock proteins and negatively regulates chaperone functions. *Molecular and cellular biology*, 19(6), pp.4535-4545.

Banerjee, A.K., Manasa, B.P. & Murty, U.S., 2010. Assessing the relationship among physicochemical properties of proteins with respect to hydrophobicity: a case study on AGC kinase superfamily. *Indian journal of biochemistry & biophysics*, 47(6), pp.370–7.

Banumathy, G. et al., 2003. Heat shock protein 90 function is essential for Plasmodium falciparum growth in human erythrocytes. *The Journal of biological chemistry*, 278(20), pp.18336–45.

Becker, J. & Craig, E. a, 1994. Heat-shock proteins as molecular chaperones. *European journal of biochemistry / FEBS*, 219(1-2), pp.11–23.

Caplan, A.J., Mandal, A.K. & Theodoraki, M. a, 2007. Molecular chaperones and protein kinase quality control. *Trends in cell biology*, 17(2), pp.87–92.

Cavalier-Smith, T., 2004. Only six kingdoms of life. *Proceedings. Biological sciences / The Royal Society*, 271(1545), pp.1251–62.

Chen, B. et al., 2005. The HSP90 family of genes in the human genome: insights into their divergence and evolution. *Genomics*, 86(6), pp.627–37.

Chen, B., Zhong, D. & Monteiro, A., 2006. Comparative genomics and evolution of the HSP90 family of genes across all kingdoms of organisms. *BMC genomics*, 7, p.156.

Cohen-Saidon, C. et al., 2006. Antiapoptotic function of Bcl-2 in mast cells is dependent on its association with heat shock protein 90beta. *Blood*, 107(4), pp.1413–20.

Csermely, P., Schnaider, T., So˜ ti, C., Proha´szka, Z. & Nardai, G., 1998. The 90-kDa molecular chaperone family: Structure, function, and clinical applications. A comprehensive review, *Pharmacology and therapeutics*. 79(5), pp.129– 168.

Dao-Phan, H.P., Formstecher, P. & Lefebvre, P., 1997. Disruption of the glucocorticoid receptor assembly with heat shock protein 90 by a peptidic antiglucocorticoid. *Molecular endocrinology (Baltimore, Md.)*, 11(7), pp.962–72.

Date, A. a, Joshi, M.D. & Patravale, V.B., 2007a. Parasitic diseases: Liposomes and polymeric nanoparticles versus lipid nanoparticles. *Advanced drug delivery reviews*, 59(6), pp.505–21.

DeLano, W.L. The PyMOL Molecular Graphics System., 2002. DeLano Scientific, San Carlos, CA, USA. http://www.pymol.org

Dill, K.A., 1990. Perspectives in Biochemistry. , 29(31).

Dimmic, M.W. et al., 2002. rtREV : An Amino Acid Substitution Matrix for Inference of Retrovirus and Reverse Transcriptase Phylogeny.*Journal of molecular evolution*, 55, pp.65–73.

Dobson, C.M. & Karplus, M., 1999. The fundamentals of protein folding: bringing together theory and experiment. *Current opinion in structural biology*, 9(1), pp.92–101.

Dollins, D.E. et al., 2007. Structures of GRP94-nucleotide complexes reveal mechanistic differences between the hsp90 chaperones. *Molecular cell*, 28(1), pp.41–56.

Dutta, R. & Inouye, M., 2000. GHKL, an emergent ATPase/kinase superfamily. *Trends biochemical sci*ence, 25, pp.24–28

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research*, 32(5), pp.1792–7.

Edgar, R.C. & Batzoglou, S., 2006. Multiple sequence alignment. *Current opinion in structural biology*, 16(3), pp.368–73.

Fast, N.M. et al., 2002. Re-examining alveolate evolution using multiple protein molecular phylogenies. *The Journal of eukaryotic microbiology*, 49(1), pp.30–7.

Felts, S.J., 2000. The hsp90-related Protein TRAP1 Is a Mitochondrial Protein with Distinct Functional Properties. *Journal of Biological Chemistry*, 275(5), pp.3305–3312.

Grenert, J.P., 1999. The Importance of ATP Binding and Hydrolysis by Hsp90 in Formation and Function of Protein Heterocomplexes. *Journal of Biological Chemistry*, 274(25), pp.17525–17533.

Gupta, R. S., 1995. Phylogenetic analysis of the 90 kD heat shock family of protein sequences and an examination of the relationship among animals, plants, and fungi species. *Molecular biololgy and evolution*, 12,pp.1063–1073

Guruprasad, K., Reddy, B. V & Pandit, M.W., 1990. Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. *Protein engineering*, 4(2), pp.155–61.

Hagn, F. et al., 2011. Structural analysis of the interaction between Hsp90 and the tumor suppressor protein p53. *Nature structural & molecular biology*, 18(10), pp.1086–93.

Hainzl, O. et al., 2009. The charged linker region is an important regulator of Hsp90 function. *The Journal of biological chemistry*, 284(34), pp.22559–67.

Hartl, F.-U.,1996.Molecular chaperones in cellular protein folding. *Nature* 381, pp.571–580.

Hatherley, R., Blatch, G.L. & Bishop, O.T., 2013. Plasmodium falciparum Hsp70-x: a heat shock protein at the host-parasite interface. *Journal of biomolecular structure & dynamics*, (December), pp.37–41.

Hoare, C.A. & Wallace F., 1966. Developmental stages of trypanosomatid flagellates: a new terminology. *Nature* 212 (5068),pp.1385–6.

Honigberg, B. M.; Balamuth, E. C., Bovee, J. O., Corliss, M., Gojdics, R. P., Hall, R. R., Kudo, N. D., Levine, A. R., Lobblich, J. & Weiser W.,(1964). A revised classification of the phylum protozoa. *Journal of Eukaryotic Microbiology*11 (1): 7–20

Hubbard, J., Erlichman, C. & Toft, D.O., 2012. NIH Public Access. , 29(3), pp.473–480.

Ikai, a, 1980. Thermostability and aliphatic index of globular proteins. *Journal of biochemistry*, 88(6), pp.1895–8.

Jhaveri, K. et al., 2012. Advances in the clinical development of heat shock protein 90 (Hsp90) inhibitors in cancers. *Biochimica et biophysica acta*, 1823(3), pp.742–55.

Johnson, J.L. & Brown, C., 2009. Plasticity of the Hsp90 chaperone machine in divergent eukaryotic organisms. *Cell stress & chaperones*, 14(1), pp.83–94.

Johnson, K.S. et al., 1989. The 86-kilodalton antigen from Schistosoma mansoni is a heat-shock protein homologous to yeast HSP-90. *Molecular and biochemical parasitology*, 36(1), pp.19–28.

Kamikawa, R. et al., 2011. Split introns in the genome of Giardia intestinalis are excised by spliceosome-mediated trans-splicing. *Current biology : CB*, 21(4), pp.311–5.

Katoh, K. et al., 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic acids research*, 33(2), pp.511–8.

Keskin, O. et al., 2008. Principles of Protein − Protein Interactions : What are the Preferred Ways For Proteins To Interact ?. *Chemical reviews*, 108(301), pp.1225-1244.

Kimura, Y., Yahara, I. & Lindquist, S., 1995. Role of the protein chaperone YDJ1 in establishing Hsp90-mediated signal transduction pathways. *Science (New York, N.Y.)*, 268(5215), pp.1362–5.

Kosano, H., 1998. The assembly of progesterone receptor-hsp90 complexes using purified proteins. *Journal of Biological Chemistry*, 273(49), pp.32973–32979.

Krone, P. H. & Sass, J. B., 1994. Hsp90α and Hsp90β genes are present in the zebra fish and are differentially regulated in developing embryos. *Biochemical and biophysical research commununications*, 204, pp.746–752

Krukenberg, K. a et al., 2011. Conformational dynamics of the molecular chaperone Hsp90. *Quarterly reviews of biophysics*, 44(2), pp.229–55.

Krukenberg, K. a et al., 2009. pH-dependent conformational changes in bacterial Hsp90 reveal a Grp94-like conformation at pH 6 that is highly active in suppression of citrate synthase aggregation. *Journal of molecular biology*, 390(2), pp.278–91.

Kumar, R., Pavithra, S.R. & Tatu, U., 2007. Three-dimensional structure of heat shock protein 90 from Plasmodium falciparum: molecular modelling approach to rational drug design against malaria. *Journal of biosciences*, 32(3), pp.531–6.

Kuo, C.C., Liang, C.M., Lai, C.Y. & Liang, S.M., 2007. Involvement of heat shock protein (Hsp) 90 beta but not Hsp90 alpha in anti-apoptotic effect of CpG-B oligodeoxynucleotide, Journal of immunology, 178(12), pp.6100–6108

Kyte, J. & Doolittle, R. F., 1982. A simple method for displaying the hydropathic character of a protein. *Journal of molecular biology*, 157, pp.105–132.

Laudanski, K. & Wyczechowska, D., 2006. The distinctive role of small heat shock proteins in oncogenesis. *Archivum immunologiae et therapiae experimentalis*, 54(2), pp.103–11.

Leskovar, A., Wegele, H., Werbeck, N.D., Buchner J. & Reinstein, J..2008. The ATPase cycle of the mitochondrial hsp90 analog Trap1. *Journal of biological chemistry*,. 283,pp.11677–11688.

Lobry, J. R. & Gautier, C., 1994. Hydrophobicity, expressivity and aromaticity are the major trends of amino-acid usage in 999 Escherichia coli chromosome-encoded genes.*Nucleic acids research*, 22(15), pp.3174-3180.

Louvion, J.F., Warth, R. & Picard, D., 1996. Two eukaryote-specific regions of Hsp82 are dispensable for its viability and signal transduction functions in yeast. *Proceedings of the National Academy of Sciences of the United States of America*, 93(24), pp.13937–42.

Lundkvist, G.B. & Bentivoglio, M., 2004. Why trypanosomes cause sleeping sickness.*Physiology*, 19(10), pp.198–206.

Mayer, M.P. & Bukau, B., 2005. Hsp70 chaperones: cellular functions and molecular mechanism. *Cellular and molecular life sciences : CMLS*, 62(6), pp.670–84.

Mendel, D.B. & Ortí, E., 1988. Isoform composition and stoichiometry of the approximately 90-kDa heat shock protein associated with glucocorticoid receptors. *The Journal of biological chemistry*, 263(14), pp.6695–702.

Meyer, P. et al., 2003. Structural and functional analysis of the middle segment of hsp90: implications for ATP hydrolysis and client protein and cochaperone interactions. *Molecular cell*, 11(3), pp.647–58.

Minami, Y., Kawasaki, H., Miyata, Y., Suzuki, K. & Yahara, I., 1991. Analysis of native forms and isoform compositions of the mouse 90-kDa heat shock protein, HSP90. *Jornal of biological chemistry*, 266,pp.10099–10103.

Munro, S. & Pelham, H. R. B.,1987. A C-terminal signal prevents secretion of luminal ER proteins. *Cell*, 48, pp.899–907.

Nagamune, K., Yamamoto, K. & Honda, T., 1997. Intramolecular Chaperone Activity of the Pro-region of Vibrio cholerae El Tor Cytolysin. *Journal of biological chemistry*, 272(2), pp.1338–1343.

Orij, R. et al., 2009. In vivo measurement of cytosolic and mitochondrial pH using a pH-sensitive GFP derivative in Saccharomyces cerevisiae reveals a relation between intracellular pH and growth.*Microbiology*, 155(10), pp.268–278.

Pallavi, R. et al., 2010. Heat shock protein 90 as a drug target against protozoan infections: biochemical characterization of HSP90 from Plasmodium falciparum and Trypanosoma

evansi and evaluation of its inhibitor as a candidate drug. *The Journal of biological chemistry*, 285(49), pp.37964–75.

Park, S.J., Kostic, M. & Dyson, H.J., 2011. Dynamic Interaction of Hsp90 with Its Client Protein p53. *Journal of molecular biology*, 411(1), pp.158–73.

Patron, N. J., & Waller, R. F. (2007). Transit peptide diversityand divergence: A global analysis of plastid targetingsignals. *BioEssays*, 29(10), pp.1048–1058.

Pearl, L.H. & Prodromou, C., 2001. Structure, function, and mechanism of the Hsp90 molecular chaperone. *Advances in protein chemistry*, 59, pp.157-186

Pei, J., Kim, B.-H. & Grishin, N. V, 2008. PROMALS3D: a tool for multiple protein sequence and structure alignments. *Nucleic acids research*, 36(7), pp.2295–300.

Petersen, T.N. & Nielsen, H., Nature Methods SignalP 4.0: discriminating signal peptides from transmembrane regions. , pp.0–13.

Péroval, M., Péry, P. & Labbé, M., 2006. The heat shock protein 90 ofEimeria tenellais essential for invasion of host cell and schizont growth, *International journal of parasitolology*, 36(1), pp.1205–1215

Prodromou, C. et al., 1997. Identification and structural characterization of the ATP/ADP-binding site in the Hsp90 molecular chaperone. *Cell*, 90(1), pp.65–75.

Prohászka, Z. et al., 1998. The 90-kDa Molecular Chaperone Family : Structure , Function , and Clinical Applications . A Comprehensive Review. *Pharmacology and therapeutics*, 79(2), pp.129–168.

Qu, D., Mazzarella, R. a & Green, M., 1994. Analysis of the structure and synthesis of GRP94, an abundant stress protein of the endoplasmic reticulum. *DNA and cell biology*, 13(2), pp.117–24..

Quevillon, E. et al., 2005. InterProScan: protein domains identifier. *Nucleic acids research*, 33(Web Server issue), pp.W116–20.

Richter, K. et al., 2006. Intrinsic inhibition of the Hsp90 ATPase activity. *The Journal of biological chemistry*, 281(16), pp.11301–11.

Roy, N. et al., 2012. Heat shock protein 90 from neglected protozoan parasites. *Biochimica et biophysica acta*, 1823(3), pp.707–11.

Saha, J., Gupta, K. & Gupta, B., 2013. In silico characterization and evolutionary analyses of CCAAT binding proteins in the lycophyte plant Selaginella moellendorffii genome: A growing comparative genomics resource. *Computational biology and chemistry*, 47C, pp.81–88.

Sarma, K. et al., 2012. A comparative proteomic approach to analyse structure, function and evolution of rice chitinases: a step towards increasing plant fungal resistance. *Journal of molecular modeling*, 18(11), pp.4761–80.

Shahinas, D. et al., 2010. A repurposing strategy identifies novel synergistic inhibitors of Plasmodium falciparum heat shock protein 90. *Journal of medicinal chemistry*, 53(9), pp.3552–7.

Shiau, A.K. et al., 2006. Structural Analysis of E. coli hsp90 reveals dramatic nucleotide-dependent conformational rearrangements. *Cell*, 127(2), pp.329–40.

Soldano, K.L. et al., 2003. Structure of the N-terminal domain of GRP94. Basis for ligand specificity and regulation. *The Journal of biological chemistry*, 278(48), pp.48330–8.

Sõti, C. et al., 2005. Heat shock proteins as emerging therapeutic targets. *British journal of pharmacology*, 146(6), pp.769–80.

Spurrier, J.D., 2003. On the null distribution of the Kruskal–Wallis statistic. *Journal of Nonparametric Statistics*, 15(6), pp.685–691.

Stechmann, A. & Cavalier-Smith, T., 2003. Phylogenetic analysis of eukaryotes using heat-shock protein Hsp90. *Journal of molecular evolution*, 57(4), pp.408–19.

Stepanova, L. et al., 1996. Mammalian p50Cdc37 is a protein kinase-targeting subunit of Hsp90 that binds and stabilizes Cdk4. *Genes & Development*, 10(12), pp.1491–1502.

Street, T.O., Lavery, L. a & Agard, D. a, 2011. Substrate binding drives large-scale conformational changes in the Hsp90 molecular chaperone. *Molecular cell*, 42(1), pp.96–105.

Streit, J. A., Donelson, J. E., Agey, M. W. & Wilson, M. E., 1996. Developmental changes in the expression of Leishmaniachagasi gp63 and heat shock protein in a human macrophage cell line. *Infection and immunity*. 64(7): pp.1810–1818.

Terasawa, K., Minami, M. & Minami, Y., 2005. Constantly updated knowledge of Hsp90. *Journal of biochemistry*, 137(4), pp.443–7.

Thompson, J.D., Higgins, D.G. & Gibson, T.J., 1994. CLUSTAL W : improving the sensitivity of progressive multiple sequence alignment through sequence weighting , position-specific gap penalties and weight matrix choice. *Nucleic acids research*, 22(22), pp.4673–4680.

Trepel, J., 2010. Targeting the dynamic HSP90 complex in cancer. *Nature reviews*, 10, pp.537-550.

Tsutsumi, S. et al., 2008. A small molecule cell-impermeant Hsp90 antagonist inhibits tumor cell motility and invasion. *Oncogene*, 27(17), pp.2478–87.

von Heijne, G. (1990). The signal peptide. *Journal ofMembrane Biology*, 115(11), pp.195–201.

Walter, S. & Buchner, J., 2002. Molecular Chaperones – Cellular Machines for Protein Folding.*Angewandte Chemie International Edition*, 41, pp.1098-1113.

Waterhouse, A.M. et al., 2009. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics (Oxford, England)*, 25(9), pp.1189–91.

Wegele, H. et al., 2006. Substrate transfer from the chaperone Hsp70 to Hsp90. *Journal of molecular biology*, 356(3), pp.802–11.

Welch, W. J. & Brown, C. R., 1996. Influence of molecular and chemical chaperones on protein folding. *Cell stress chaperones*, 1, pp.1109–115.

Wiesgigl, M. &Clos, J. 2001. Heat shock protein 90 homeostasis controls stage differentiation inLeishmania donovani, *Molecular biology cell*, 12(2).pp3307–3316.

WHO., 2008. World malaria report 2008. Geneva, Switzerland. http://www.who.int/malaria/publications/atoz/9789241563697/en/index.html

Wirth, D.F., Rogers, W.O., Barker, R., Jr, Dourado, H., Suesebang, L. & Albuquerque R., 1986. Leishmaniasis and malaria: new tools for epidemiologic analysis. *Science* 234, pp.975-979.

Yamada, S. et al., 2002. A hydrophobic segment within the C-terminal domain is essential for both client-binding and dimer formation of the HSP90-family molecular chaperone. *European Journal of Biochemistry*, 270(1), pp.146–154..

Yamamoto, M. et al., 1991. Characterization of the hydrophobic region of heat shock protein 90. *Journal of biochemistry*, 110(1), pp.141–5.

Young, J.C., Schneider, C. & Hartl, F.U., 1997. In vitro evidence that hsp90 contains two independent chaperone sites. *FEBS Letters*, 418(1-2), pp.139–143.

Appendices

Appendix 1 – Physicochemical scripts

```
# The script analyses protein properties using sequences from a given file


from Bio.SeqUtils.ProtParam import ProteinAnalysis

from Bio.SeqUtils import ProtParamData

from Bio import SeqIO


# Kyte & Doolittle index of hydrophobicity

kd = {'A': 1.8, 'R':-4.5, 'N':-3.5, 'D':-3.5, 'C': 2.5,

    'Q':-3.5, 'E':-3.5, 'G':-0.4, 'H':-3.2, 'I': 4.5,

    'L': 3.8, 'K':-3.9, 'M': 1.9, 'F': 2.8, 'P':-1.6,

    'S':-0.8, 'T':-0.7, 'W':-0.9, 'Y':-1.3, 'V': 4.2 }

data = raw_input("Enter the name of the sequence file: ")

f = open(data, "r")


mr = []

aromaticity = []

inst_index = []

pI = []

flexibility = []

gravy1 = []

al = []

h = []


for rec in SeqIO.parse(f, "fasta"):

        myprot = ProteinAnalysis(str(rec.seq))

        mr += [int(myprot.molecular_weight())]

        aromaticity += [round(myprot.aromaticity(), 5)]

        inst_index += [round(myprot.instability_index(), 2)]

        pI += [round(myprot.isoelectric_point(), 2)]

        gravy1 += [round(myprot.gravy(), 4)]

                #Aliphatic index calculation
```

```
		aa_count = myprot.count_amino_acids()
		aa_no = len(rec.seq)

		A = aa_count['A']
		I = aa_count['I']
		L = aa_count['L']
		V = aa_count['V']

		Al = 100*(1.0*A/aa_no)
		Is = 100*(1.0*I/aa_no)
		Le = 100*(1.0*L/aa_no)
		Va = 100*(1.0*V/aa_no)

		Aliphatic_index = round(Al+(2.9*Va)+3.9*(Is+Le),2)
		al += [Aliphatic_index]
		#hydrophobicity calculations
		total = 0
		for a in str(rec.seq):
			if a in kd.keys():
				total += int(kd[a])
		#Normalizing
		total = total*1.0/len(str(rec.seq))
		h +=[round(total,5)]
f.seek(0)
#writing information to a file
species_name = []
for line in f:
		if line[0] == ">":
			a = line.index('\n')
			species_name += [line[1:a]]
new = open("prop_"+data, "w")
line1 = "Name" + " \t Arom" + " \tII" + " \tHydro" + " \t pI" + "\tGRAVY" + "\t Alip_I\n"
new.write(line1)
count = 0
```

```
for name in species_name:
        line = name + "\t"  + str(aromaticity[count]) + "\t" + str(inst_index[count]) + "\t"+ str(h[count]) + "\t" + str(pI[count])
        + "\t" + str(gravy1[count]) + "\t " + str(al[count])+ "\n"
        count += 1
        new.write(line)
```

Script A1.1.Physicochemical properties calculator

```
#!/usr/bin/Rscript
#Author: Ngonidzashe Faya
#Date: 25 August 2013

#NB: When calculationg the correlation, the boxplot command should be masked
#When creating boxplots the panel.pearson command should be masked....

dataA <- read.table("prop_edit_groupA.txt", header=TRUE, sep="\t")
dataB <- read.table("prop_edit_groupB.txt", header=TRUE, sep="\t")
dataC <- read.table("prop_edit_groupC.txt", header=TRUE, sep="\t")

attach(dataA)
attach(dataB)
attach(dataC)

data1<-data[order(pI),]

head(dataA)

jpeg(file = "./Cor.jpg")

panel.pearson <- function(x,y,...) {horizontal<-(par("usr")[1]+par("usr")[2])/2; vertical<-(par("usr")[3]
+par("usr")[4])/2; text(horizontal,vertical, format(abs(cor(x,y)), digits=2))}

pairs(dataA,lower.panel=panel.pearson)
boxplot(propertyofA, proprtyofB, propertyofC, ylab=expression(bold("Property name")))

dev.off()
```

Script A1.2: Boxplots and Peason correlation coefficients calculator

```
#!/usr/bin/Rscript
#Author: Ngonidzashe Faya
#Date: 20 November 2013

dataA <- read.table("prop_edit_groupA.txt", header=TRUE, sep="\t")
dataB <- read.table("prop_edit_groupB.txt", header=TRUE, sep="\t")
dataC <- read.table("prop_edit_groupC.txt", header=TRUE, sep="\t")

attach(dataA)
attach(dataB)
attach(dataC)

x <- c(dataA$physicochemical property)
y <- c(dataB$physicochemical property)
z <- c(dataC$physicochemical property)

kruskal.test(list(x, y, z))
```

Script A1.3. Kruskal-Wallis statistical test script.

Appendix 2 – MSA for human Hsp90 isoforms

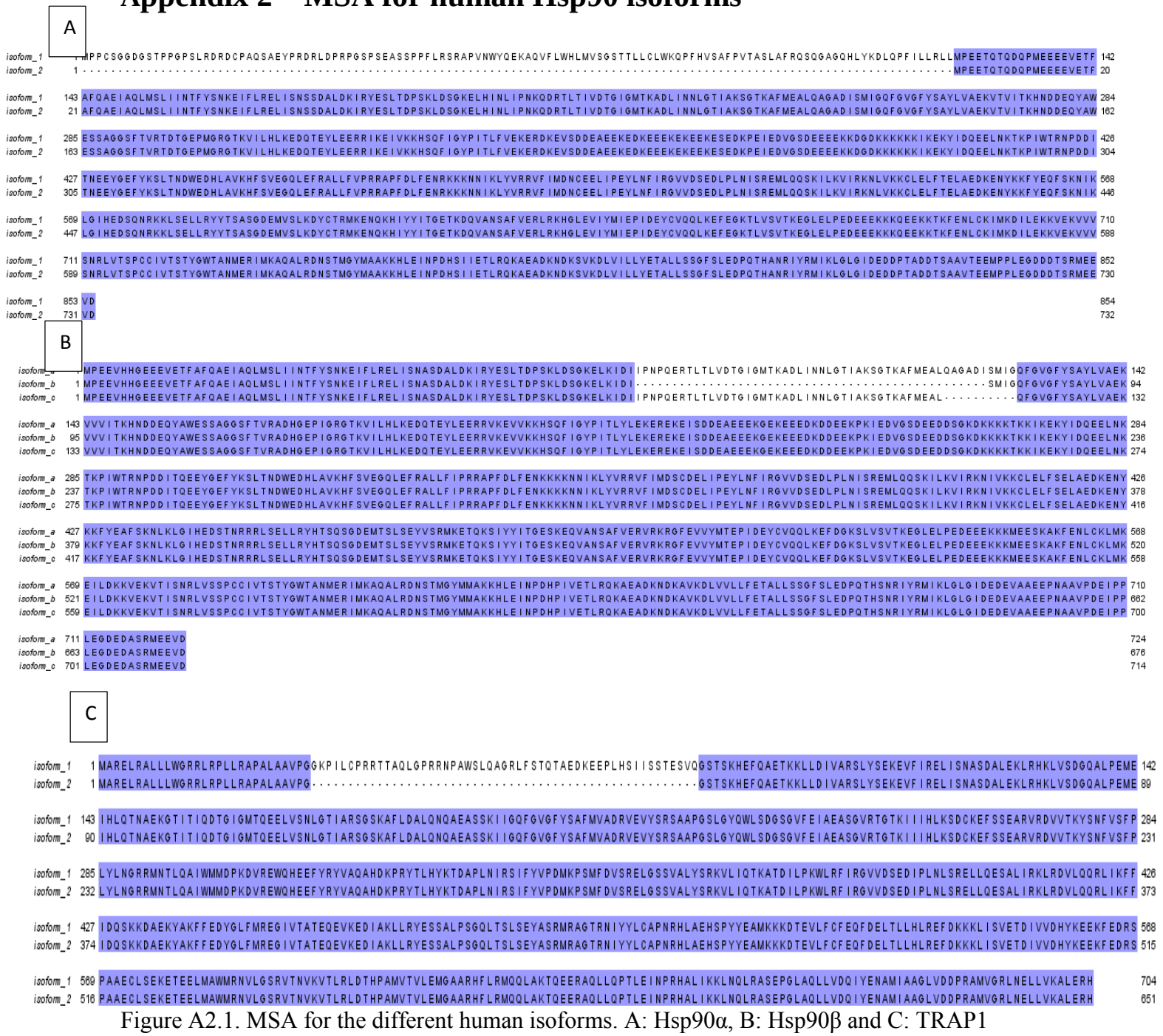


Figure A2.1. MSA for the different human isoforms. A: Hsp90α, B: Hsp90β and C: TRAP1

Appendix 3 – Signal peptides

Table A3.1 Length of group B signal peptides

Class	Name	Length (# of aa)
Host	HSA	21
Protozoa	BBO	27
	BEQ	28
	GIN	17
	LBR	24
	LIN	24
	LMA	24
	LME	24
	PFA	28
	PVI	22
	TGO	24
	TBR	30
	TCR	20
	ACA	23
	BMI	0
	BHO	0
	CHO	0
	CMU	21
	CPA	20
	EHI	16
	PBE	29
	PYY	29
Helminths	SJA	24
	SMA	20
	WBA	18
	BMA	20
	LLO	29
	TSP	19
Ectoparasite	PHC	23
Vectors	AAE	22
	AGA	22
	CQU	22
	FCA	21
	ISC	0

Appendix 4 – Sequence analysis scripts

```
#This script extracts domains from an alignment in fasta format
#Athor: Ngonidzashe Faya
#Date: 04 Nov 2013

print "\n*****Please ensure the alignment is in fasta format*****\n"

from Bio import SeqIO

filename = raw_input("Enter file name: ")
reg_no = int(raw_input("How many regions/domains do you want to extract? "))
region = raw_input("Enter the domain/region name: ")

counter = 0
while counter < reg_no:

    aln = open(filename, "r")
    domain = open(region+".txt", "w")

    a = int(raw_input("Enter start position of domain: "))
    b = 1 + int(raw_input("Enter end position of domain: "))

    for rec in SeqIO.parse(aln, "fasta"):
        domain.write(">"+str(rec.id) + "\n")
        domain.write(str(rec.seq[a:b])+"\n")

    if (reg_no-1) > counter:
        region = raw_input("Enter the domain/region name: ")
    counter += 1
```

Script A4.1. A photosnap of the script used in domain extraction. Bio indicates the Biopython module used to read alignments.

'''
Motif analysis script
Authors: Ngonidzashe Faya and Ozlem Tastan Bishop
Date: Date: 20/10/2013
'''

import operator as op
import matplotlib
import matplotlib.pyplot as plt
import numpy as np

```
from matplotlib import cm as CM

#Dataset information
dataset = raw_input("Enter the name of your dataset: ")
# Entering input files
meme_filename = raw_input("Enter meme.txt file name: ")
meme=open(meme_filename, "r")
t1=meme.readlines()

ask = "X"
while ask != "Y":
	ask = raw_input("Do you have MAST.txt output (Y/N)? ").upper()
	if ask == "Y":
		mast_filename = raw_input("Enter mast file name: ")
		mast=open(mast_filename, "r")
		t2=mast.readlines()
		mast_present = ask
	if ask == "N":
		mast_present = ask
		ask = "Y"


# Extracting seq_names and motif information from meme file
seq_names = []
count1 = 2
count2 = 2
count3 = 2
index1 = -3
index2 = -4
index3 = -3
motif = []
pos_data = []
reg_exp = []
motif_no = 1
M_no = []

for i, line in enumerate(t1, 1):
	# storing sequence names in a list
	if "Sequence name            Weight" in line:
		index1 = i

	if i == (index1 + count1):
		seq_names += [line[:14].strip(), line[40:56].strip()]
		count1 +=1


	if "COMMAND LINE SUMMARY" in line:
		del[seq_names[-8:]]
		index1 +=10000000000000
		if seq_names[-1] == "":
			del[seq_names[-1]]


	# storing motif width and number of sites
	if "MOTIF" in line and "E-value" in line:
		motif += [line[:8], int(line[18:21].strip()), int(line[31:35].strip())]
		M_no += [int(line[31:35].strip())]



	# storing the positions of the motifs in each sequence
	if "Sequence name            Start   P-value" in line:
		index2 = i

	if i == (index2 + count2) and (count2-2) < motif[-1]:
		pos_data += [(motif_no,line[:13].strip(), int(line[16:31].strip()), int(line[16:31].strip())+motif[-2])]
		count2 +=1

	if "Motif" in line and "block diagrams" in line:
```

```
                        index2 +=10000000000000000000000000000
                        count2 = 2
                        motif_no += 1

                #stroring the regular expression for each motif
                if "regular expression" in line:
                        index3 = i

                if i == (index3 + count3):
                        reg_exp += [line[:-2]]

                if "Time" in line:
                        index3 +=10000000000000

#calculating the number of motifs in each sequence
seq_names.sort()
motif_number = []
for name in seq_names:
        a = 0
        for data in pos_data:
                if name == data[1]:
                        a += 1
        motif_number += [a]
print motif_number
#Preparing data for writing
empty_pos_data = []
for x in range((len(motif)/3)): # divide by 3 because the list has 3 elements for each motif
        for name in seq_names:
                empty_pos_data += [(x+1, name, "0", "0")]

short_pos = []
for a in empty_pos_data:
        for b in pos_data:
                if a[0] == b[0] and a[1] == b[1]:

                        short_pos += [a]

for tup in short_pos:
        empty_pos_data.pop(empty_pos_data.index(tup))
pos_data += empty_pos_data



pos_data = sorted(pos_data, key=op.itemgetter(0, 1))
empty_pos_data = sorted(empty_pos_data, key=op.itemgetter(0, 1))

# Writing motif information to a file.
df = open("dataframe.csv", "w")


for name in seq_names:
        df.write(","+name)
df.write("\n")
mot_data = []
M = []
total = len(seq_names)
k1=0
k2=2
for i, mot in enumerate(motif, 1):
        if k1 < len(motif) and mot == motif[k1]:
                mot_data += [(int(mot[-3:].strip()), motif[k1+2]*1.0/len(seq_names))]
                k1 += 3
                M += ["M"+mot[-2:].strip()]
                df.write(mot)
                for pos in pos_data:
                        if mot[-2:].strip() == str(pos[0]):
                                df.write(","+str(pos[2])+'  '+str(pos[3]))
                df.write("\n")
df.write("\n")
```

```
#Writing the regular expressions to the file
x = 1
for exp in reg_exp:
	df.write("Motif "+str(x)+","+exp+"\n")
	x += 1

#Extracting information from a mast file if present
if mast_present == "Y":
	for line in t2:
		if "Removing motifs" in line:
			bad_motifs = line[17:-34]
			bad_motifs = bad_motifs.replace(',',").replace('and ',").split()
		if "No overly similar pairs" in line:
			bad_motifs = []


	#Writing the motifs that must be removed
	df.write("\n The following motifs should be removed basing on MAST output: \n")
	for z in bad_motifs:
		df.write(z+" ")



#Heat map
#preparing the dataset
d = []
for mot in mot_data:
	for data in pos_data:
		if mot[0] == data[0]:
			if int(data[2]) != 0:
				d += [mot[1]]
			else:
				d += [0]


d.append(1)
b = []
new_mot_data = []
for x in d:
	if len(b) != len(seq_names):
		b += [x]
	else:
		new_mot_data += [b]
		b = []
		b += [x]


#Removing bad motifs basing on MAST output
if mast_present == "Y":
	counter = 1
	for bad in bad_motifs:
		new_mot_data.pop(int(bad)-counter)
		M.pop(int(bad)-counter)
		counter += 1


#Heat map commands
column_labels = M
row_labels = seq_names
data = np.array(new_mot_data)
fig, ax = plt.subplots()
colors = [('white')] + [(CM.jet(i)) for i in xrange(40,250)]
new_map = matplotlib.colors.LinearSegmentedColormap.from_list('new_map', colors, N=300)
heatmap = ax.pcolor(data, cmap=new_map)
```

```
fig = plt.gcf()
fig.set_size_inches(8,11)

# turn off the frame
ax.set_frame_on(False)

# put the major ticks at the middle of each cell
ax.set_yticks(np.arange(data.shape[0])+0.5, minor=False)
ax.set_xticks(np.arange(data.shape[1])+0.5, minor=False)

# want a more natural, table-like display
ax.invert_yaxis()
ax.xaxis.tick_top()

ax.set_xticklabels(row_labels, minor=False)
ax.set_yticklabels(column_labels, minor=False)

# rotate the
plt.xticks(rotation=90)
ax.grid(False)

# Turn off all the ticks
ax = plt.gca()

for t in ax.xaxis.get_major_ticks():
    t.tick1On = False
    t.tick2On = False
for t in ax.yaxis.get_major_ticks():
    t.tick1On = False
    t.tick2On = False

cbar = plt.colorbar(heatmap)
cbar.ax.set_ylabel('# of motif sites/total # of sequences', rotation=270)
plt.savefig("heat_map.png")
plt.show()

import matplotlib.pyplot as plot
fig = plt.figure()
fig.suptitle(dataset+" motif information")
x = range(len(seq_names))
y = motif_number
labels = seq_names
plot.bar(x, y,width = 0.7, align="center")

# the x locations for the groups
plot.xticks(x, labels)
plot.xticks(rotation=90)
plot.axhline(min(motif_number), linewidth=3, color='r')
plot.axhline(max(motif_number), linewidth=3, color='r')
plot.xlabel("Sequence names")
plot.ylabel("Number of motifs")
fig.tight_layout()

plot.show()
```

Script A4.2. The script used for MEME analysis.

Appendix 5 – Multiple sequence alignments

Figure A5.1.:MSA result of group **A** Hsp90 using the PROMALS3D program.

Figure A5.2.Group B MSA obtained using the MAFFT program.

Figure A5.3.MSA for group C HSP90 obtained using the MAFFT program

Table A5.1: Domain positions in the MSA

Region	Group A	Group B	Group C
N-terminal Domain	1 – 231	1 – 391	1 – 505
Charged Linker Region	232 – 345	392 – 451	N/A
Middle Domain	346 – 681	452 – 846	506 – 890
C-terminal Domain	682 -	847 -	891 -

Appendix 6 – Motif analysis

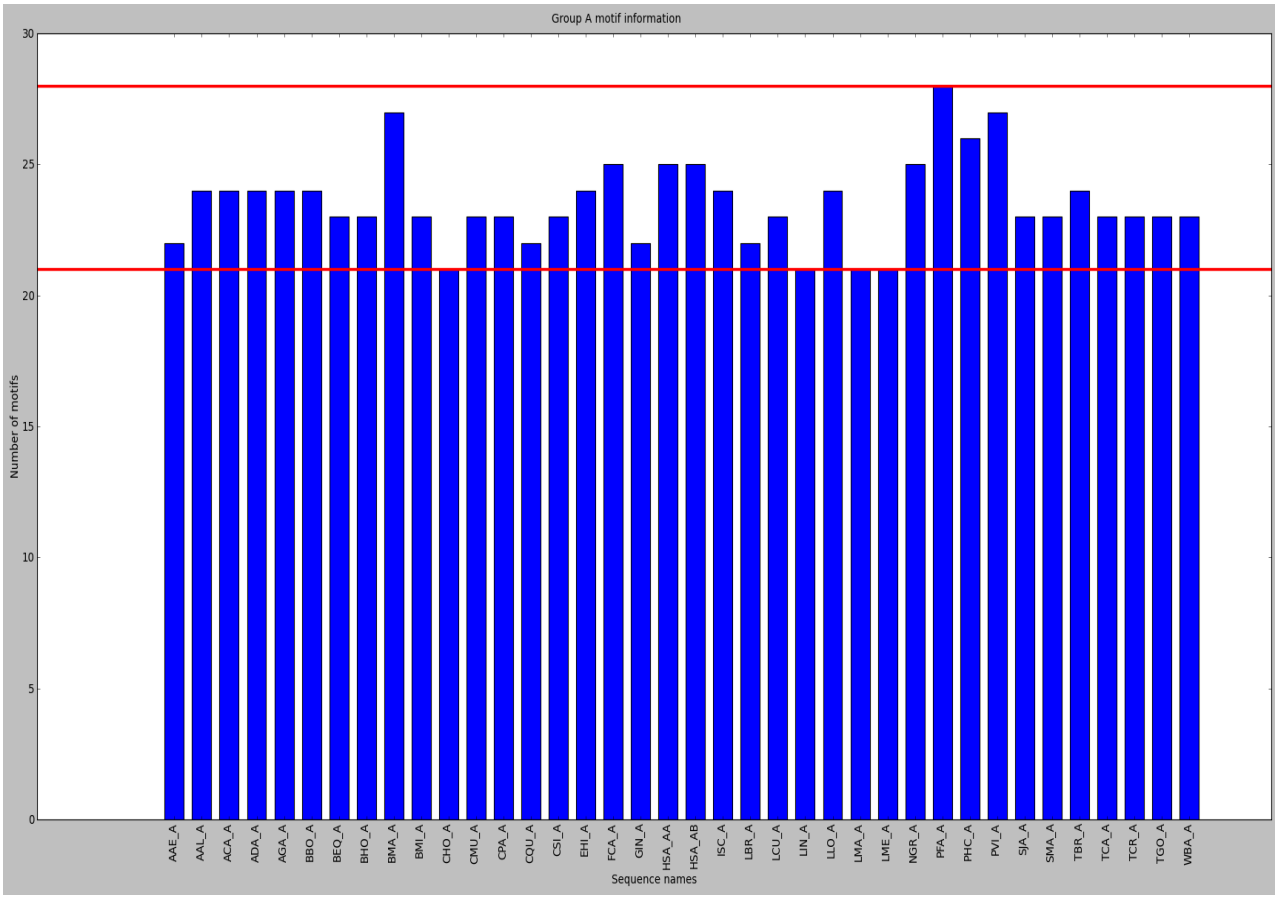


Figure A6.1. Bar graph showing the number of motifs per sequence in group A Hsp90. The red lines indicate the min and max values.

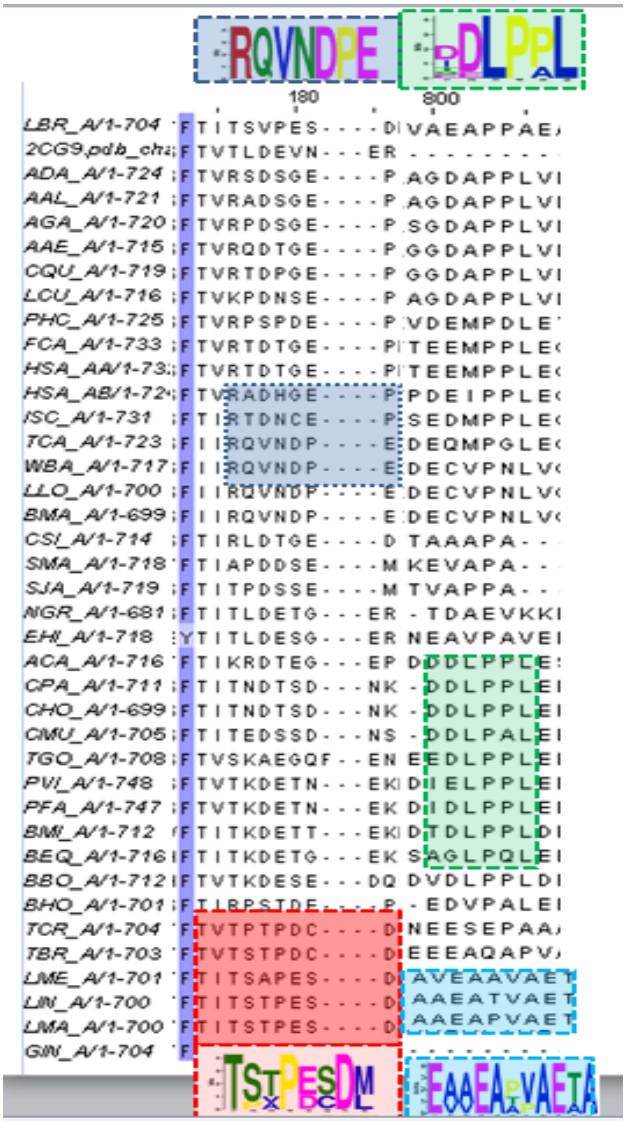


Figure A6.2 MSA showing the regions with unique motifs in group A Hsp90 sequences. In dark blue is motif 20, green is motif 26, red boundary is motif 25 and light blue: motif 28

Figure A6.3: The motifs unique to the group B *Leishmania* Hsp90.

Figure A6.4: Motifs unique to group B *Plasmodium* species

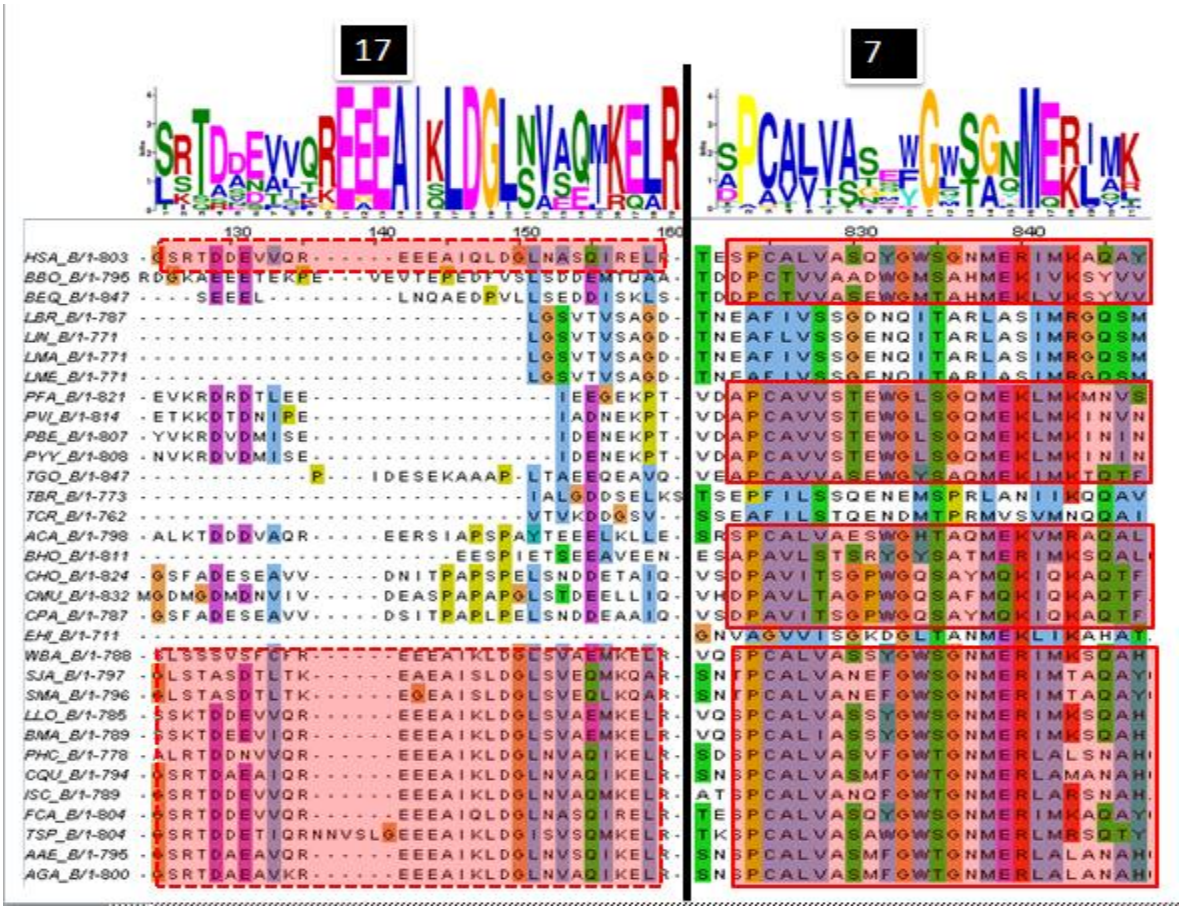


Figure A6.5. Sequence alignment showing the unique motifs 17 and 7 of group B Hsp90s.

Figure A6.6.Motifs unique to *Leishmania sp.*

Appendix 7– Phylogenetic tree analysis

Table A7.1: Top 5 models that were predicted at 95% partial deletion to produce a reliable tree in terms of BIC and AICc.

Model	Parameters	BIC	AICc	*lnL*	(+*I*)	(+*G*)	*f*(A)	*f*(R)
rtREV+G+I+F	218	88952.399	87005.239	-43283.769	0.05	1.29	0.055	0.049
rtREV+G+F	217	88996.625	87058.389	-43311.352	n/a	1.01	0.055	0.049
rtREV+G+I	199	89451.776	87674.187	-43637.385	0.05	1.34	0.065	0.045
WAG+G+I	199	89457.498	87679.909	-43640.246	0.05	1.60	0.087	0.044
rtREV+G	198	89501.498	87732.835	-43667.716	n/a	1.05	0.065	0.045

Table A7.2: Top 5 models that were predicted to construct a reliable tree at 100% deletion

Model	Parameters	BIC	AICc	*lnL*	(+*I*)	(+*G*)	*f*(A)	*f*(R)
rtREV+G+I+F	218	62643.555	60772.993	-30167.291	0.04	1.42	0.050	0.055
rtREV+G+F	217	62670.258	60808.266	-30185.938	n/a	1.08	0.050	0.055
rtREV+G+I	199	62929.314	61221.591	-30410.791	0.04	1.47	0.065	0.045
rtREV+G	198	62957.199	61258.048	-30430.029	n/a	1.14	0.065	0.045
WAG+G+I	199	63045.729	61338.006	-30468.998	0.05	1.78	0.087	0.044

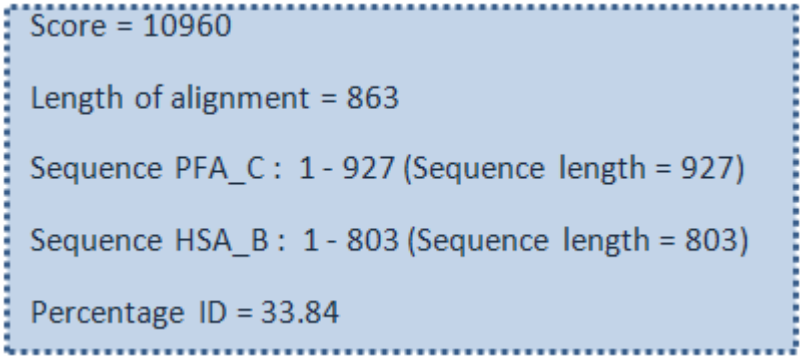


Figure A7.1: Sequence identity for human GRP94 and *P. falciparum* TRAP1.