EVOLUTION AND DIAGNOSTIC POTENTIAL OF MAJOR SURFACE PROTEASES FROM *TRYPANOSOMA VIVAX, T. BRUCEI BRUCEI AND T.CONGOLENSE*

By

Eunice M. Machuka (BSc.)

Reg. No. I56/13068/2005

A thesis submitted in partial fulfilment of the requirements for the Degree of Master of Science in Biotechnology of Kenyatta University

April 2009

DECLARATION

This thesis is my original work and has not been presented for a degree in any other University or any other award.

Eunice M. Machuka (BSc. Biochemistry)

Signed: _____ Date: _____

We confirm that the work reported in this thesis was carried out by the candidate under our supervision.

Dr. Daniel K. Masiga
Department of Biochemistry and Biotechnology
Kenyatta University

Signed: _____ Date: _____

Dr. Charity W. Gichuki
Department of Biochemistry and Biotechnology
Kenyatta University

Signed: _____ Date: _____

Dr. Johnson O. Ouma
Trypanosomiasis Research Centre
Kenya Agricultural Research Institute

Signed: _____ Date: _____

DEDICATION

To my family and friends for their support throughout my studies.

AKNOWLEDGEMENT

I give all praise and glory to the Almighty God for sustaining me and providing me with wisdom and strength throughout my life and my studies. I wish to convey many special thanks to my supervisors Drs Dan Masiga, Charity Gichuki and Johnson Ouma for their support during the study and critical analysis and discussions leading to this thesis write-up. My sincere gratitude goes to colleagues at the Molecular Biology and Biotechnology Department at the International Centre of Insect Physiology (*icipe*) and the Biochemistry Division at Kenya Agricultural Research Institute -Trypanosomiasis Research Centre for their technical assistance. I acknowledge the Biosciences eastern and Central Africa Network (BecANet) Secretariat for the BecA Graduate fellowship and World Federation of Scientists through a fellowship at the *icipe* for the financial and research support. I wish to thank *icipe* for the Dissertation Research Internship (DRIP) fellowship award. I would like to acknowledge my parents and friends who have supported me throughout my studies financially, spiritually and emotionally and for believing in my abilities.

ABSTRACT

Tsetse-transmitted trypanosomiasis is a disease unique to Africa, affecting both humans and animals. This disease occurs in about 10 million km^2 in 37 sub-Saharan countries corresponding approximately to one-third of Africa's total land area, and threatens an estimated 50 million people and 48 million cattle. The estimated annual losses in cattle production alone are in the range of 1.0-1.2 billion dollars. The diagnosis of trypanosomiasis is notoriously difficult as there are no specific clinical signs and the intermittent and usually low parasitaemia make detection of the difficult. The trypanosomes are infamous for their ability to evade the immune responses by periodically switching their major variant surface glycoprotein (VSG) in a phenomenon called antigenic variation and are likely to be only one of several mechanisms enabling these organisms to thrive in the face of the immune defences. The complexity of the trypanosome parasite's antigenic repertoire has made development of a vaccine based on the VSG coat unlikely hence focus is on identifying invariant surface trypanosome components such as the Major Surface Proteases (MSP) as potential targets for interrupting infection or infection-mediated disease. The MSP are a group Zinc metalloproteases belonging to the M8 metzincin family. MSP-like genes have been identified in the bloodstream-stage of African trypanosomes whose genome contains gene families encoding homologues of the MSP. It is a protein found on the cell surface that contributes to the ability of *Leishmania* to foil the mammalian immune system. However, very little is known about the MSP from *T. congolense, T. b. brucei* and *T. vivax*. The objective of this study was to clone, express and evaluate the evolution of the MSP variants with the aim of determining their possible use as diagnostic targets for African Animal Trypanosomiasis (AAT). We describe the identification and amplification of *T. congolense, T. b. brucei* and *T. vivax* MSP using an integrated approach of cloning, expression and phylogenetic analyses with the objective of evaluating and determining their possible use as diagnostic targets for AAT. Samples of *T. congolense, T. brucei* and *T. vivax* were obtained from diverse geographical locations throughout East Africa and sequenced for the MSP genes. The *MSP* genes of *T. congolense, T. brucei* and *T. vivax,* the causative agents of AAT, were successfully amplified, cloned and expressed in *E. coli* cells. The amplification of MSP by PCR yielded stable non-fragmenting products of approximately 1600bp on an agarose gel. The result of a 4-20% SDS–PAGE showed a bigger and broad band of the induced protein, suggesting full induction of the polypeptide. Separation of the proteins based on size on SDS-PAGE gel revealed induced products of approximately 45kDa in size from both the *T. congolense* and *T. b. brucei* samples. A protein signature search of the MSP sequences in this study showed characteristics of a typical metallopeptidase such as the peptidase domains and the Zinc finger domains, a catalytic Zinc atom and an active-site helix containing the two histidines of the Zinc proteinase sequence motif HEXXH. Evolutionary studies of these *MSP* variants showed regions of strong similarity amongst all the three species under study suggesting that they are conserved. Distance and maximum likelihood methods yielded phylogenetic trees of similar topology with indications that geographical placement has no effect on the phylogeny of the MSP. These conserved domains may form useful target sites for the development of a tool for detecting *T. congolense, T. brucei* and *T. vivax*.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACRONYMS AND ABBREVIATIONS

AAT	African Animal Trypanosomiasis
DTT	Dithiothreitol
ELISA	Enzyme-Linked Immunosorbent Assay
IPTG	Isopropyl β-D-1-thiogalactopyranoside
MAST	Motif Alignment and Search Tool
MEME	Multiple EM for Motif Elicitation
MSP	Major Surface Proteases
Ni-NTA	Nickel Nitro-acetic Acid
PBS	Phosphate buffered saline
PCR	Polymerase chain Reaction
SDS-PAGE	Sodium dodecyl sulphate Polyacrylamide gel electrophoresis
NEB	New England Biolabs
GO	Gene Ontology

CHAPTER ONE

1 INTRODUCTION

1.1 Background

African animal trypanosomiasis (AAT) is a disease complex caused by tsetse-fly-transmitted *Trypanosoma congolense* (Hoare, 1964), *T. vivax* (Chalmers, 1918), or *T. brucei brucei* (Luehe, 1906), or simultaneous multiple infections (Luckins, 1992). African animal trypanosomiasis is most important in cattle but can cause serious losses in pigs, camels, goats, and sheep (Luckins, 1992). Infection of cattle by one or more of the three African animal trypanosomes results in sub acute, acute, or chronic disease characterized by intermittent fever, anaemia, occasional diarrhoea, and rapid loss of condition and often terminates in death. In southern Africa, the disease is widely known as nagana, which is derived from a Zulu term meaning, "to be in low or depressed spirits".

1.1.1 Aetiology

AAT is caused by protozoa in the family *Trypanosomatidae* genus *Trypanosoma* (Gruby, 1843). *T. congolense* belongs to the subgenus *Nannomonas*, a group of small trypanosomes with medium-sized marginal kinetoplasts, no free flagella, and poorly developed undulating membranes. In East Africa, *T. congolense* is considered the single most important cause of AAT. This trypanosome is also a major cause of the disease in cattle in West Africa. Sheep, goats, horses, and pigs may also be seriously affected (Molyneux and Ashford, 1983). In domestic dogs, chronic infection often results in a carrier state.

T. vivax is a member of the subgenus *Duttonella*, a group of trypanosomes with large terminal kinetoplasts, distinct free flagella, and inconspicuous undulating membranes. *T. vivax* is a large (18-26 µm long) monomorphic organism that is very active in wet-mount blood smears. Cattle, sheep, and goats are primarily affected. Although *T. vivax* is considered less pathogenic for cattle than *T. congolense*, it is nevertheless the most important cause of AAT in West African cattle. This trypanosome readily persists in areas free of tsetse flies such as Central and South America and in the Caribbean), where it is transmitted mechanically by biting flies or contaminated needles, syringes, and surgical instruments (Mulligan, 1970).

T. brucei brucei belongs to the subgenus Trypanozoon. *T. b. brucei* is an extremely polymorphic trypanosome occurring as short, stumpy organisms without flagella, long slender organisms with distinct flagella, and intermediate forms that are usually flagellated. Horses, dogs, cats, camels and pigs are very susceptible to *T. b. brucei* infection. Infection of cattle, sheep, goats and sometimes pigs results in mild or chronic infection (Moulton and Sollod, 1976).

1.1.2 Host Range

Cattle, sheep, goats, pigs, horses, camels, dogs, cats, and monkeys are susceptible to AAT and may suffer syndromes ranging from subclinical mild or chronic infection to acute fatal disease. Rats, mice, guinea pigs, and rabbits are useful laboratory species. More than 30 species of wild animals can be infected with pathogenic trypanosomes, and many of these remain carriers of the organisms. Ruminants are widely known to be active reservoirs of the trypanosomes. Wild Equidae, lions, leopards, and wild pigs are susceptible and can serve as carriers of trypanosomes.

1.1.4 Geographic Distribution

The tsetse-fly-infested area of Africa extends from the southern edge of the Sahara desert (latitude 15° N.) to Angola, Zimbabwe, and Mozambique (latitude 20° S.). Of the three African animal trypanosomes, only *T. vivax* occurs in the Western Hemisphere in at least 10 countries in the Caribbean and South and Central America.

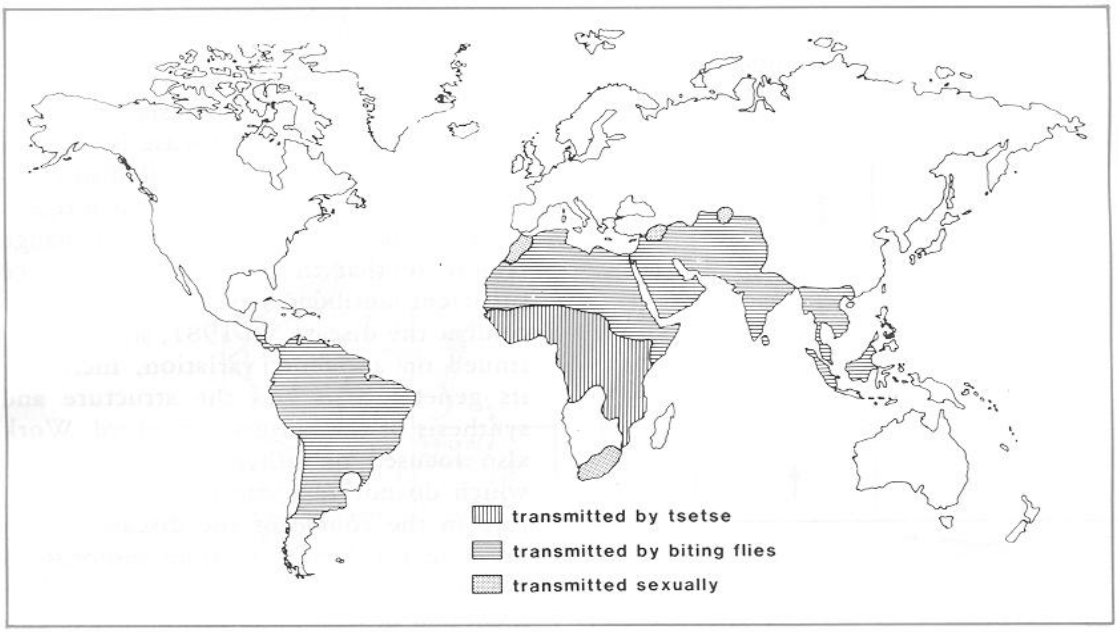

Figure 1: World distribution of animal Trypanosomiasis (Source: Annual Report of the International Laboratory for Research on Animal Diseases, ILRAD 1981).

1.1.5 Incubation Period

The incubation period for *T. congolense* varies from four (4) to twenty-four (24) days, *T. vivax*, from four (4) to forty (40) days and for *T. b. brucei*, from five (5) to ten (10) days.

1.1.6 Transmission

In Africa, the primary vector for *T. congolense*, *T. vivax*, and *T. b. brucei* is the tsetse fly. These trypanosomes replicate in the tsetse fly and are transmitted through tsetse fly saliva when the fly feeds on an animal. The three main groups of tsetse flies that transmit trypanosomes are *morsitans* group, which are found preferentially in the open savannah woodland; *palpalis* group prefers the shaded habitat immediately adjacent to rivers and lakes; and the *fusca* group favours the high, dense forest areas. Trypanosomiasis is also mechanically transmitted by tsetse and other biting flies through the transfer of blood from one animal to another. The most important mechanical vectors are flies of the genus Tabanus, but Haematopota, Liperosia, Stomoxys, and Chrysops flies have also been implicated. In Africa, both *T. vivax* and *T. b. brucei* have spread beyond the "tsetse fly belts" (Roder *et al.*, 1984), where transmission is principally by tabanid and hippoboscid flies.

T. vivax is transmitted mechanically, even in Africa, wherever sufficient biting flies occur, particularly (but not exclusively) tabanids and stable flies (*Stomoxys*). In the past *T. vivax* has also occurred on the island of Mauritius, in the absence of tsetse flies, but has apparently been eradicated there. *T. vivax* invaded the western hemisphere a long time ago, possibly with the introduction of West African cattle in the Caribbean region in the eighteenth and nineteenth centuries, and has since then established a firm foothold in the Americas. It has been found in the past as far south as Paraguay, and nowadays is known to occur at least from Bolivia in the south to El Salvador, in Central America, in the north, but the limits of its distribution are not well known. Cuba has also been reported to be infected (FAO, 1998).

The disease caused by the mechanically transmitted *T. vivax* does not appear to be different from that caused by the tsetse-transmitted parasite. Non-cyclical transmission is essentially mechanical transmission in which the trypanosomes are transferred from one mammalian host to another by the interrupted feeding of biting insects, notably tabanids and stomoxys. The trypanosomes in or on the contaminated proboscis do not multiply and die quickly so that cross contamination is only possible for a few hours (Urquhart *et al*., 1996). The importance of this mode of transmission is variable from place to place, depending on the numbers of hosts and biting insects present, and also on the species of trypanosome. This mode of transmission has proved to be sufficiently effective to maintain *Trypanosoma vivax* and *Trypanosoma evansi* in South and Central America, and the latter species in North Africa and Asia as well. No tsetse flies occur outside tropical Africa, apart from small tsetse pockets in the southwest of the Arabian Peninsula. The trypanosomes may also be transmitted through use of the same needle or surgical instrument on more than one animal, at sufficiently short intervals that the blood on the needle or instrument does not dry there are high chances of transmitting the parasites. It is not an uncommon occurrence when animals are vaccinated or treated by injection, or when blood is collected from several animals in a row, without changing or disinfecting needles or pins. It may also occur when several animals are subjected at short intervals to a surgical intervention such as dehorning and castration without properly disinfecting the instruments used (FAO, 1998). Finally apart from classical cyclical and non-cyclical transmission, dogs, cats and wild carnivores may become infected by eating fresh carcasses or organs of animals which have died of trypanosomiasis, the parasite penetrating oral abrasions (Urquhart *et al*., 1996).

1.1.7 Pathogenesis and Immunological Response

When the tsetse fly bites a host, metatrypamastigotes are transferred into the skin. A chancre develops at the site of the bite due to an acute inflammatory response by the host. The parasite then makes its way to the bloodstream via the lymph nodes that drain the site of the bite. The host begins to generate IgM and IgG against the variable surface glycoprotein (VSG), the primarily immunogenic protein covering the trypanosome surface. In most humans a specific form of High Density Lipoprotein (HDL), acts as an innate barrier to infection by one specific species of Trypanosome, *T. brucei brucei*. HDL can disrupt the membrane of this form of trypanosome resulting in lysis (Portela *et al*., 2000). Despite the largely successful immune response, a small number of the parasites, one out of every 10^5 to 10^6 parasites, will undergo a switch at the genetic level and begin to express a new VSG (Turner, 1982; Haag *et al*., 1998). These parasites will not be eliminated and will replicate unchecked until a new immune response is generated. This pattern is repeated again and again resulting in a fluctuating chronic parasitaemia as shown in Figure 2 below. This fluctuation is also seen symptomatically in the form of a relapsing and recurring fever.

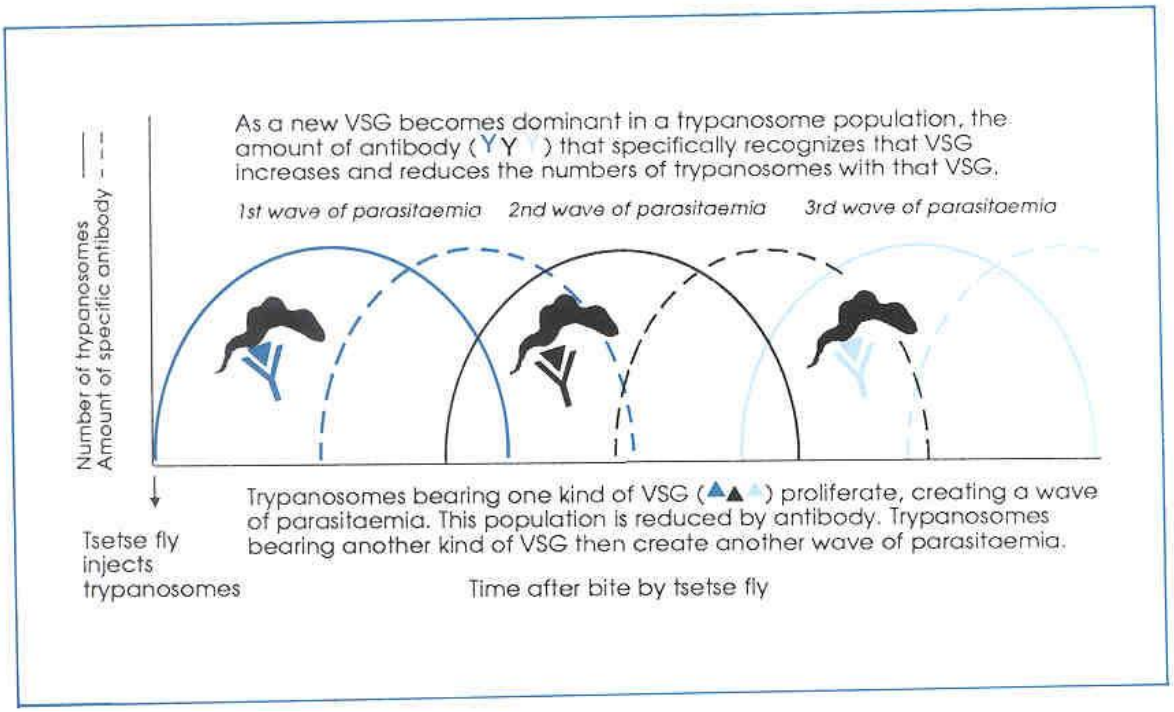

Figure 2: Diagram illustrating the process of antigenic variation, which enables trypanosomes to survive attack by the immune system of an animal host (ILRAD, 1990).

As the host continually tries to keep up with the parasite, the parasite eventually disseminates via the circulatory system into many areas of the body, including the brain. Along the way, it may become embedded in blood vessels causing damage that causes infiltration of cells and build-up of fluid in the vessel. Perivascular cuffing of the vessel occurs, restricting blood flow through that vessel. Once in the brain, it causes tissue damage that results in lethargy, confusion, and sleepiness. Eventually, the activities of parasites in the brain lead to coma and then death.

Initial replication of trypanosomes is at the site of inoculation in the skin resulting in a swelling. Trypanosomes then spread to the lymph nodes and blood and continue to replicate. *T. congolense* localizes in the endothelial cells of small blood vessels and capillaries. *T. brucei* and *T. vivax*

localize in tissue. Antibody developed to the glycoprotein coat of the trypanosome kills the trypanosome and results in the development of immune complexes. The antibody, however, does not clear the infection, for the trypanosome has genes that can code for many different surface-coat glycoproteins and change its surface glycoprotein to evade the antibody. Thus, there is a persistent infection that results in a continuing cycle of trypanosome replication, antibody production, immune complex development, and changing surface-coat. The most significant and complicating factor in the pathogenesis of trypanosomiasis is the profound immunosuppression that occurs following infection by these parasites, which lowers the host's resistance to other infections resulting in secondary disease, which greatly complicates both the clinical and pathological features of trypanosomiasis (Blood *et al*., 1989). Immunologic lesions are significant in trypanosomiasis, and it has been suggested that many of the lesions (e.g., anaemia and glomerulonephritis) in these diseases may be the result of the deposition of immune complexes that interfere with, or prevent, normal organ function (Blood *et al*., 1989).

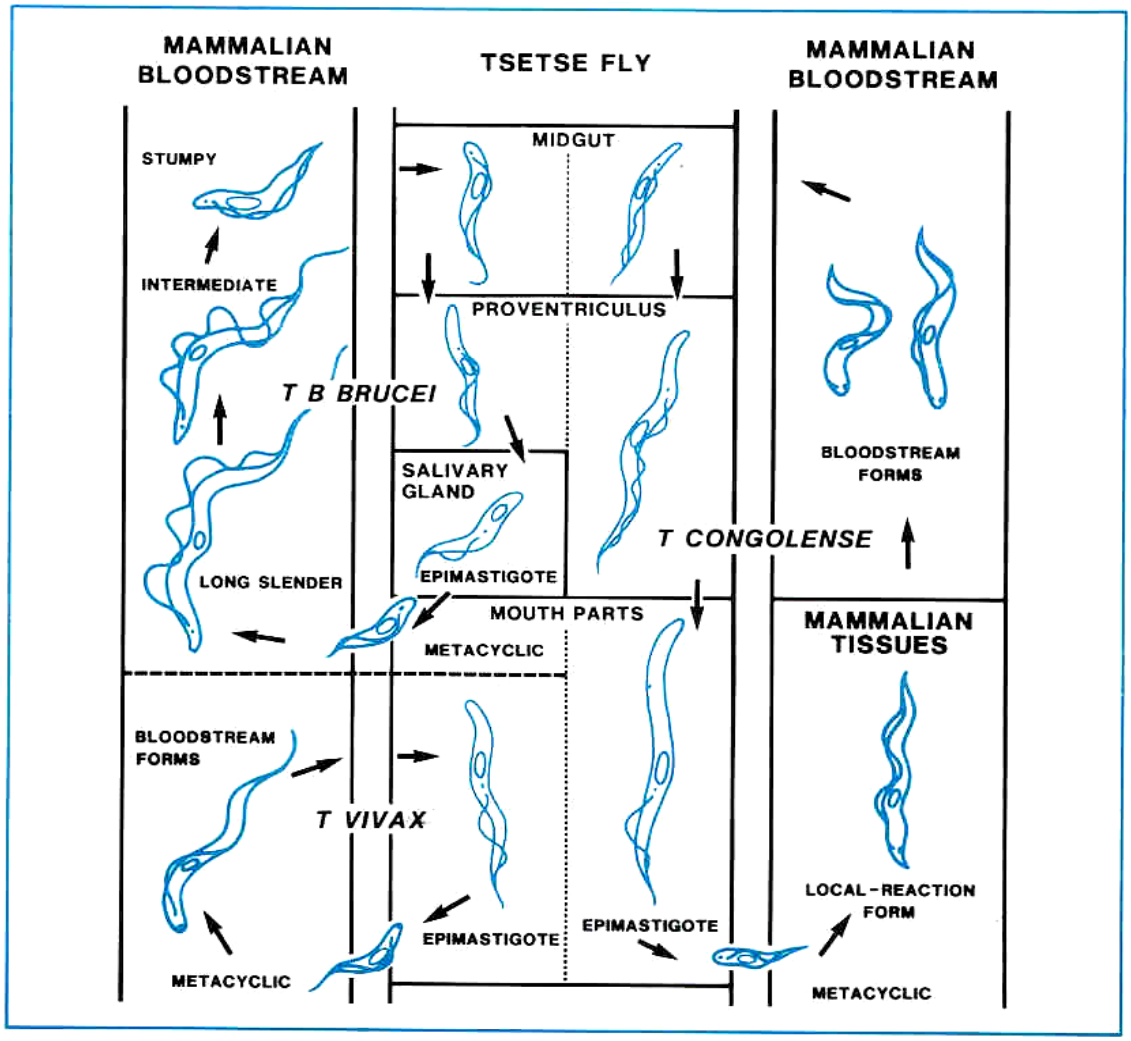

Figure 3: Lifecycle of *T. b. brucei*, *T. congolense* and *T. vivax* (Annual Report of the International Laboratory for Research on Animal Diseases ILRAD 1981).

1.1.8 Clinical Signs

Because simultaneous infections with more than one trypanosome species are very common (Nyeko *et al*, 1990), and simultaneous infection with trypanosomes and other hemoparasites (Babesia species, Theileria species, Anaplasma species, and Ehrlichia species) frequently occurs, it is difficult to conclude which clinical signs are attributable to a given parasite. Few adequately controlled studies have been made, and thus a "typical" clinical response to each trypanosome is difficult to reconstruct.

The cardinal clinical sign observed in AAT is anaemia. Within a week of infection with the haematic trypanosomes (*T. congolense* and *T. vivax*), there is usually a pronounced decrease in packed cell volume, haemoglobin, red blood cell, and white blood cell levels, and within two months these may drop to below 50 percent of their pre-infection values. Also invariably present are intermittent fever, oedema and loss of condition (Taylor & Authie, 2004). Abortion may be seen, and infertility of males and females may occur. The severity of the clinical response is dependent on the species and the breed of affected animal and the dose and virulence of the infecting trypanosome. Stress, such as poor nutrition or concurrent disease, plays a prominent role in the disease process, and under experimental conditions, where stress may be markedly reduced; it is difficult to elicit clinical disease.

T. congolense is a haematic trypanosome found only in the blood vessels of the animals it infects. It does not localize and multiply outside blood vessels. Infection with *T. congolense* may result in peracute, acute, or chronic disease in cattle, sheep, goats, horses, and camels. Pigs often develop a milder disease; chronic disease is common in dogs. The incubation period is followed by

intermittent febrile episodes, depression, lethargy, weakness, and loss of condition, anaemia, salivation, lacrimation, and nasal discharge. As the disease progresses, loss of condition and hair colour changes from black to metallic brown are seen. The back is often arched and the abdomen "tucked up." Accelerated pulse and jugular pulsation occur and breathing is difficult. Anaemia is a prominent sign. Early in the infection, the organisms are readily demonstrable in blood smears, but, as the disease progresses to its acute and chronic forms, organisms are most readily demonstrated in lymph node smears.

The period of infection by *T. vivax* is variable, depending on the host and the parasite isolate. In sheep and goats, the incubation period lasts from four to 12 days. In bovines, it ranges from nine to 14 days for virulent isolates, and from nine to 59 days in infections with less pathogenic isolates (Hoare, 1972). *T. vivax* parasitaemias exhibit irregular fluctuations, with some cases occurring at high levels in the morning and at lower levels in the afternoon on the same day. As a rule, however, an initial period of parasitaemia will last for several months before trypanosomes become undetectable in the blood. After that phase, periods of recurrence take place.

T. brucei brucei has a relatively short incubation period and causes severe to fatal infection in horses, camels, dogs, and cats. It usually causes mild, chronic, or sub clinical disease in cattle, sheep, goats, and pigs. A febrile response occurs in the horse 4-14 days after infection. This is followed by recurrent febrile reactions. The heartbeat and respiration may be accelerated, and loss of condition and weakness are seen, whereas the appetite remains good. Progressive anaemia and oedema of the ventral regions, especially the male genitalia, are characteristic. The organisms are not always easily perceived in blood smears and are best demonstrated in tissue smears or

sections, (e.g., lymph nodes). Infected animals die in a few weeks or several months, depending on the virulence of the strain of *T. b. brucei*.

1.1.9 Approaches towards Control and Eradication

Tsetse vector control methods relying on large scale bush clearing and aerial spraying methods are no longer used due to environmental concerns. Tsetse control currently relies on two bait systems: insecticide-treated traps and targets and insecticide treated livestock. Sterile Insect Technique (SIT) has also been used in efforts to eradicate tsetse flies in some areas. Because of the stability of tsetse populations and their low reproductive rate, little sustained mortality pressure (additional to natural mortality) needs to be exerted on a population to cause its extinction (Weidhaas and Haile, 1978).That makes them good candidates for traps and target control methods. Not to be forgotten though are the risk of reinvasion or immigration into an area already cleared of tsetse flies. The theory behind this control method is simple: the flies are visibly attracted to a trap or target; this attraction may be further helped by the use of olfactory attractants. When the tsetse lands on a trap or target they either receive a lethal dose of insecticide, or are caught in the trap and subsequently die (Leak, 1998).

The effectiveness of traps and targets depend on when the flies are active, how they move in their active state, whether they will move into the vicinity of a trap or target and finally, whether they are trapped or killed (Williams *et al*., 1992). Insecticide-treated livestock was developed as a method of tsetse control from the concept of baited traps and targets (Baylis and Stevenson, 1997). It is widely accepted by a majority of stockowners in Africa. The method has been used to control tsetse and trypanosomiasis with varying results in Kenya (Stevenson *et al*., 1991). Most

commonly used are the synthetic pyrethroids, and of these deltamethrin appears to be the most potent and it is also low in mammalian toxicity and has minimal environmental impact (Thompson *et al*., 1991). Extensive use of insecticides on cattle for tsetse control appears to have the potential to interfere with endemic stability/immunity of cattle to several tick borne diseases. Thus the long term use of these products may jeopardize control of tick borne diseases (FAO 2001).

1.1.10 Immunization

No vaccine is currently available for African animal trypanosomiasis.

1.1.11. Trypanotolerance

It has been recognized that certain breeds of African cattle are considerably more tolerant to African trypanosomiasis than others (Dwinger *et al.,* 1994). This is especially true of the West African short-horned cattle (Muturu, Baoule, Laguna, Samba, and Dahomey) and the N'Dama, which is also of West Africa. These cattle have existed in the region for over 5,000 years. Susceptibility studies have shown the N'Dama to be the most resistant breed followed by the smaller West African short-horned cattle, but the large Zebu is the most susceptible (Murray *et al*., 1979). The mechanisms of trypanotolerance have been extensively studied, and it is now well established that trypanotolerance has a genetic basis (Murray *et al*, 1984; Moulton and Sollod, 1976).

1.1.11 Evolution of Trypanosomes

With the advent of molecular techniques capable of elucidating evolutionary relationships from genes of extant species, the impasse in studies of kinetoplastid evolution, so long based on morphological and transmission characteristics (Baker, 1963; Hoare, 1972), appears to have been overcome. Since the first broad molecular study of eukaryote evolution which included only a single representative of the genus *Trypanosoma* (Sogin *et al*., 1986), phylogenetic analysis of kinetoplastid flagellates has become successively more focused. Initial studies concentrated on the origins of parasitism in the group (Lake *et al*., 1988; Fernandes *et al*., 1993) and latterly on detailed analyses of evolutionary relationships among *Trypanosoma* and *Leishmania* species (Maslov *et al*., 1996; Croan *et al*., 1997; Lukes *et al*., 1997; Haag *et al*., 1998; Stevens *et al*., 1998, 1999). As the level of focus has deepened, the number of species representing each genus has increased in successive studies and, significantly, there has been a progression of ideas concerning the evolutionary relationships between the species. In trypanosomes, where this process is particularly well marked the conclusions of initial studies, which included only limited numbers of species and which indicated the genus *Trypanosoma* to be paraphyletic, have been superseded by those of subsequent studies with increased numbers of taxa, such that the genus is now generally considered to be monophyletic (Stevens and Gibson, 1999).

1.2 Problem Statement and Justification

Tsetse-transmitted trypanosomiasis is a disease unique to Africa affecting both humans and animals (FAO, 2002). Livestock form an integral component of farming systems and thus contribute significantly to food and economic security in developing countries (Taylor and Mertens, 1999). Trypanosomiasis occurs in about 10 million km^2 in 37 sub-Saharan countries

corresponding approximately to one-third of Africa's total land area, and threatens an estimated 50 million people and 48 million cattle. Trypanosomiasis has a severe impact on African agriculture; estimated annual losses in cattle production alone are in the range of 1.0-1.2 billion dollars (FAO, 2002). In addition to this, the indirect negative effects engendered by trypanosomiasis on total crop production due to disease have an influence on where people decide to live, how they manage their livestock and the intensity of crop agriculture (FAO, 2002). The combined effects result in changes in land use and environment, affect human welfare and increase the vulnerability of agricultural activity. African animal trypanosomiasis (nagana) is a wasting and fatal disease in cattle (Aksoy, 2003). Definitive diagnosis is based on demonstration of parasites in blood or tissues via microscopy, a process that requires expertise and experience to differentiate the different species. This is further complicated by the fact that the clinical signs of trypanosomiasis are not pathognomonic and the standard techniques for the detection of trypanosomes are not sufficiently sensitive. Consequently, a high proportion of infections remain undetected. This situation is aggravated by the fact that the chronic and the more common form of the disease, is often aparasitaemic. Management of the different species requires different strategies and medication especially for the tissue invasive *T. brucei*. Therefore, specific and sensitive diagnosis methods are required if early and life-saving treatment for the disease is to be initiated. Gathering of accurate data regarding the epidemiology of trypanosomiasis is crucial for the development of good treatment and control strategies to protect livestock against trypanosomiasis (Geysen *et al*., 2003).

The cell surfaces of the parasites are covered in complex protein- or carbohydrate-rich coats that are required for parasite survival and infectivity in the insect vectors and mammalian hosts. Their

survival within these environments requires the regulated surface transport of highly abundant coat glycoproteins, glycolipids, plasma membrane transporters, surface enzymes, and receptors (McConville *et al*., 2002). The cell surface of the metacyclic *T. brucei* trypomastigote is covered with a dense glycocalyx composed of molecules of variant surface glycoproteins (VSG). The complexity of the parasite's antigenic repertoire has made development of a diagnostic tool based on the VSG coat unlikely. As a result, research is focused on identifying invariant trypanosome components as potential targets for detecting infection or infection-mediated disease (Taylor and Mertens, 1999). The glycans such as the Major Surface Proteases (MSP) have been proposed for studies aimed at detection of simultaneous infection of livestock with *T. vivax, T. b. brucei,* and *T. congolense,* where conventional methods manage to reveal only single infections, ultimately allowing cross infections to be diagnosed easily. Were the MSP proteins to be found on the surfaces of *T. congolense, T. vivax* and *T. brucei,* and a methodology to clone them *in vitro* was worked out, then they can be evaluated for their potential role as diagnostic targets and ultimately used for the development of a sensitive diagnostic method for *nagana* in livestock. Application of the diagnostic method would enable early diagnosis, effective treatment and would enhance livestock development in the rangelands currently hampered by the disease.

In the absence of a fossil record, the evolution of protozoa has until recently largely remained a matter for speculation. However, advances in molecular methods and phylogenetic analysis are now allowing interpretation of the "history written in the genes" (Stevens and Gibson, 1999). Determining the conservation pattern signatures as well as the phylogeography of the MSP across the *T. congolense, T. vivax* and *T. brucei* species shall be of great significance in assessing it’s viability as a potential diagnostic target. Due to the need for an easy universal and sensitive

diagnostic target that can be used to identify trypanosome infections in field samples coupled with species differentiation, characterising the MSP conserved domains across *T. congolense, T. vivax* and *T. brucei* species shall form the basis for identifying these potential target sites for diagnosis. Were these proteins to be found on the surfaces of *T. congolense, T. vivax* and *T. brucei*, and a methodology to clone them *in vitro* was worked out, then they could be used for the development of a sensitive diagnostic method for *nagana* in livestock. The aspect of trypanosome evolution can also be studied by building phylogenetic trees based on the major surface protein (MSP) gene sequences.

1.3 Research Questions

a) Can Major Surface Proteases (*MSPs*) be used to improve diagnostic sensitivity and specificity of *T. congolense, T. vivax* and *T. b. brucei* infections in domestic livestock?

b) How have the *MSPs* in *T. congolense, T. vivax* and *T. b. brucei* species from East Africa evolved?

1.4 Hypotheses

a) The MSPs on the trypanosomes do generate host immune response and can be used to develop an effective diagnostic assay for African Animal Trypanosomiasis.

b) MSPs from *T. congolense, T. vivax* and *T. b. brucei* species are conserved.

1.5 Objectives

1.5.1 General Objective

To determine the potential use of *MSPs* in diagnosis of *T. congolense, T. brucei brucei* and *T. vivax* infections in livestock.

1.5.2 Specific Objectives

a) To clone and express *MSP* genes from *T. congolense, T. brucei brucei* and *T. vivax* for diagnostic purposes,

b) To carry out a phylogenetic analysis of the trypanosome MSPs at both the DNA and protein levels.

1.6 Significance and Anticipated Outputs

- To build a phylogenetic tree based on MSP gene sequences for use in studying *T. congolense, T. brucei brucei* and *T. vivax* MSP variants' evolution.
- To express the MSP in bacterial cells and determine its potential use for trypanosome detection of AAT with the aim of ultimately developing a diagnostic method that is specific, sensitive, reproducible, simple, affordable and applicable in the field.

CHAPTER TWO

2 LITERATURE REVIEW

2.1 Glycans in Trypanosomes

The cell surfaces of the parasites are covered in complex protein- or carbohydrate-rich coats that are required for parasite survival and infectivity in the insect vectors and mammalian hosts. Their survival within these environments requires the regulated surface transport of highly abundant coat glycoproteins, glycolipids, plasma membrane transporters, surface enzymes, and receptors (McConville *et al*., 2002). Glycans can be found attached to proteins as in glycoproteins and proteoglycans and are referred to as either O-linked or N-linked glycans. The cell surface of the metacyclic *T. brucei* trypomastigote is covered with a dense glycocalyx composed of molecules of variant surface glycoproteins (VSG) while procyclic acidic repetitive proteins (PARPs) or procyclins are found on procyclic form of *T. brucei* and are less densely packed (Treumann *et al.*, 1997; Mehlert *et al.*, 1998).

Studies on a number of parasites indicate that immune responses to parasites in infected animals and humans are directed to glycan determinants within cell surface and secreted glycoconjugates, and that glycoconjugates are important in host-parasite interactions (Nyame *et al*., 2004). There is strong evidence that carbohydrate, rather than protein, antigens dominate the immune responses to many parasites. Complex carbohydrates are also important to interactions of protozoan parasites with their hosts. Furthermore, antigen processing cells (APCs), such as dendritic cells and macrophages, appear to specifically recognize many parasite-derived glycans, partly through Toll receptors and C-type lectins, which can lead to enhanced or attenuated responses to parasitic infection (Nyame *et al*, 2004).

2.1.2 Major Surface Proteases

The Major Surface Protease (MSP) of the protozoan parasite *Leishmania* is a highly abundant Zinc metalloprotease belonging to M8 metzincin family (Schlagenhauf *et al*., 1998). The characteristics of this class of proteins include a sequence motif HEXXHXXGXXH, and an N-terminal pro-peptide that renders the pro-enzyme inactive during translation, and is removed during maturation and activation (Yiallouros *et al*., 2002; Gong *et al*., 1998). MSP belongs to a peptidase family M8 (EC 3.4.24.36) according to the IUBMB Enzyme Nomenclature (http://www.iubmb.unibe.ch/). Due to its important roles in virulence and its high abundance, MSP has been the subject of several reviews (Bordier *et al*., 1987; Medina-Acosta *et al*., 1993; McMaster *et al*., 1994). The MSP is a protein found on the cell surface that contributes to the ability of *Leishmania* to foil the mammalian immune system. The major cell surface glycoprotein of *Leishmania* promastigotes (gp63) is a 63-kDa Zinc metalloprotease that is anchored to the cell surface via a myristic acid containing GPI anchor (Medina-Acosta *et al*., 1989). Each *Leishmania major* promastigote in stationary phase is estimated to have 500,000 copies of MSP, constituting about 1% of the organism's total protein (Yao *et al*., 2005). MSP homologues have also been detected in the monogenetic insect protozoa *Crithidia fasciculata* (Inverso *et al*., 1993) and *Herpetomonas samuelperssoai* (Elias *et al*., 2006), as well as the extracellular protozoan *Trypanosoma brucei* (LaCount *et al.*, 2003) and the digenetic protozoan *T. cruzi*. (Cuevas *et al*., 2003).

The genome of the African trypanosome *Trypanosoma brucei* (Tb) contains at least three gene families (TbMSP-A, -B, and -C) encoding homologues of the abundant major surface protease (MSP, previously called GP63), which is found in all *Leishmania* species (LaCount *et al.*, 2003). All three *TbMSP* families are expressed in bloodstream-stage trypanosomes, but *TbMSP-B* is expressed in the procyclic stage, whereas *TbMSP-A* and -*C* mRNAs are detected only in bloodstream organisms. RNA interference (RNAi)-mediated gene silencing has been employed to investigate the function of *TbMSP-B* protein. RNAi directed against *TbMSP-B* but not *TbMSP-A* ablated the steady state *TbMSP-B* mRNA levels in both procyclic and bloodstream cells but had no effect on the kinetics of cultured trypanosome growth in either stage. To determine whether *TbMSP-B* is responsible for this release, transgenic variant surface glycoprotein 117 (VSG117) has been expressed constitutively in *T. brucei* procyclic *TbMSP*-RNAi cell lines, and the amount of surface VSG117 was determined using a surface biotinylation assay. Ablation of *TbMSP-B* but not *TbMSP-A* mRNA resulted in a marked decrease in VSG release with a concomitant increase in steady state cell-associated VSG117, indicating that *TbMSP-B* mediates the surface protease activity of procyclic trypanosomes. These findings indicated that peptidomimetic collagenase inhibitors block the release of transgenic VSG from procyclic trypanosomes and is toxic for bloodstream but not procyclic organisms (LaCount *et al.*, 2003).

The MSP genes (*MSP*s) are located in tandem arranged multigene arrays in all *Leishmania* species studied (Victoir *et al*., 2005). Within the gene arrays, individual genes display different life cycle-specific expression patterns. In *Leishmania* sp., regions encoding the *MSP* catalytic

site show a conserved sequence, while regions encoding surface domains possibly involved in the host–parasite interaction are variable (Victoir *et al*., 2005). The corresponding MSP molecules in extracellular African trypanosomes do not contribute to parasite entry and survival in macrophages as they do for the intracellular *Leishmania*, so they might perform cellular functions for African trypanosomes. The procyclic organisms utilize the enzymatic activity of a cell-surface, Zinc-dependent metalloprotease to release the VSG molecules from the plasma membrane. GP63 has been shown to be proteolytically active against a number of substrates and thus may be involved in degradation of host macromolecules. It may also serve as a ligand for the macrophage receptor via complement components and protect the parasite against complement-mediated lysis (Alexander and Russell, 1992). Attempts to obtain mutants defective in GP63 by targeted gene deletion are made difficult by the fact that a multigene family encodes it. Targeted deletion of six out of seven GP63 genes does not affect growth of the parasite *in vitro* or prevent formation of disease in mice (Joshi *et al*., 2002). Hilley *et al.* 2000) generated a knockout of GPI8, the GPI: protein transaminidase that eliminated the expression of GP63 along with other GPI-anchored proteins. The knockout grew normally in culture, and its ability to infect macrophages *in vitro* was unaffected. The GPI8 mutant was able to establish infection in mice, suggesting that GP63 is not essential for growth or infectivity in mammals.

Due to its important roles in virulence and evidence of its high abundance in protozoan parasites elucidated above, the analysis of the role that these MSP proteins can play as potential diagnostic targets cannot be downplayed. The conserved sites in this protein shall form the basis upon which these studies can be achieved. The molecular cloning, expression and ultimately an

evolutionary analysis of MSP sequences at the protein level is the approach used in this study in order to ascertain their conservation across *T. brucei*, *T. vivax*, and *T. congolense* species.

2.2 Molecular Evolutionary Studies of Trypanosomes

In the absence of a fossil record, the evolution of protozoa has until recently largely remained a matter for speculation. However, advances in molecular methods and phylogenetic analysis are now allowing interpretation of the "history written in the genes" (Stevens and Gibson, 1999). Trypanosome evolution has been studied by building phylogenetic trees based on gene sequences. The absence of a fossil record and the limited number of distinctive morphological characters mean this is the only realistic approach. Trypanosomatid phylogenies have been constructed based on the genes for the small subunit ribosomal RNA [ssu rRNA] (Marché *et al*., 1995) and for glycosomal glyceraldehyde phosphate dehydrogenase [gGAPDH] (Briones *et al*., 1999; Wiemer *et al*., 1995). In the Trypanosomes initial studies, which included only limited numbers of species and which indicated the genus *Trypanosoma* to be paraphyletic, have been superseded by those of subsequent studies with increased numbers of taxa, such that the genus is now generally considered monophyletic (Stevens *et al.*, 1999).

Molecular evolutionary relationships within the protozoan order *Kinetoplastida* have been deduced from comparisons of the nuclear small and large subunit ribosomal RNA (rRNA) gene sequences (Fernandes *et al*., 1993). These studies show that relationships among the trypanosomatid protozoans differ from those previously proposed from studies of organismal characteristics or mitochondrial rRNAs. The genera *Leishmania, Endotrypanum*, *Leptomonas*,

and *Crithidia* form a closely related group, which shows progressively more distant relationships to *Phytomonas* and *Blastocrithidia*, *Trypanosoma cruzi*, and lastly *Trypanosoma brucei*. The rooting of the trypanosomatid tree was accomplished by using *Bodo caudatus* (family *Bodonidae*) as an out-group, a status confirmed by molecular comparisons with other eukaryotes. The nuclear rRNA tree agrees well with data obtained from comparisons of other nuclear genes ((Fernandes *et al*., 1993).

Analysis of nuclear small subunit ribosomal RNAs (SSU rRNAs) suggested a different evolutionary tree, in which *Trypanosoma brucei* and *Trypanosoma cruzi* diverge before the lineage leading to *Crithidia* and *Leishmania*, which are closely related. The implications of these various topologies to models concerning the age and origins of parasitism are profound. Early 18S ribosomal RNA gene studies were summarised by Maslov *et al*, (1995) in a phylogenetic tree, which included three trypanosome species, *T. brucei*, *T. cruzi* and a third species from a fish. In common with other early studies, this tree indicated the genus *Trypanosoma* to be paraphyletic. Subsequently, Maslov *et al*., 1996 increased the number of *Trypanosoma* species to seven; however, this still left *T. brucei* outside both the main trypanosome clade and the trypanosomatid clade containing *Leishmania* and *Crithidia*. The inclusion by Lukes *et al*. (1997) of four more trypanosome species demonstrated for the first time that the genus *Trypanosoma* might in fact be monophyletic and the addition of more outgroup taxa considerably strengthened this result. Subsequently, trees including 24 trypanosome species (Haag *et al*., 1998) and 47 trypanosome taxa (Stevens *et al*., 1999) have both supported monophyly of trypanosomes unequivocally and it seems unlikely that, at least for the 18S rRNA gene, addition of further taxa will alter this conclusion. The progressive definition of an “aquatic clade”, comprising

trypanosome species isolated from both marine and freshwater fish, amphibia and leeches, is also apparent in these trees.

While little information can be gleaned from the single isolate included in the study of Maslov & Simpson (1995), the study of Maslov *et al*. (1996), which includes seven trypanosome species, shows clearly the emergence of an aquatic clade. The *T. brucei* clade consists of the Salivarian tsetse-transmitted trypanosomes of African mammals; *T. evansi* and *T. equiperdum*, although non-tsetse transmitted and not restricted to Africa, also belong here by virtue of their close morphological and genetic similarity to *T. brucei* mitochondrial DNA (Borst *et al*., 1987) and isoenzymes (Gibson *et al*., 1983; Lun *et al*., 1992) points to *T. evansi* and *T. equiperdum* being comparatively recent mutants of *T. brucei*, which have been able to spread outside Africa because they no longer rely on tsetse transmission. Importantly, the *T. brucei* clade is also characterised by the phenomenon of antigenic variation (Haag *et al*., 1998) and these facts, taken together, suggest a distinct evolutionary history for the clade, initially confined to Africa. The *T. cruzi* clade, which includes *T. cruzi* and *T. rangeli*, contains a range of species originating from South American mammals and humans; interesting exceptions to this are three species of bat trypanosomes from Africa and Europe, and one as yet uncharacterized species of kangaroo trypanosome from Australia. Early trees, which showed trypanosomes to be paraphyletic (Maslov & Simpson, 1995; Maslov *et al*., 1996), suggested that parasitism and the digenetic lifecycle had arisen more than once in the trypanosome lineage. The unequivocal evidence of monophyly revealed by more recent trees clearly contradicts this, but still supports the idea that parasitism and digenetic lifecycles have evolved independently in several trypanosomatid lineages (see Figure 4). While the hypothesis of co-evolution of trypanosomatids and their vectors was not

supported by early trees (Maslov *et al*., 1996), later trees reveal obvious clade and vector associations; for example *T. brucei* clade taxa share transmission by tsetse flies. It is anticipated that analysis of additional trypanosome species from birds, reptiles and various mammals will begin to clarify the unresolved evolutionary relationships still evident in the lower half of the tree shown in Figure 4 below.

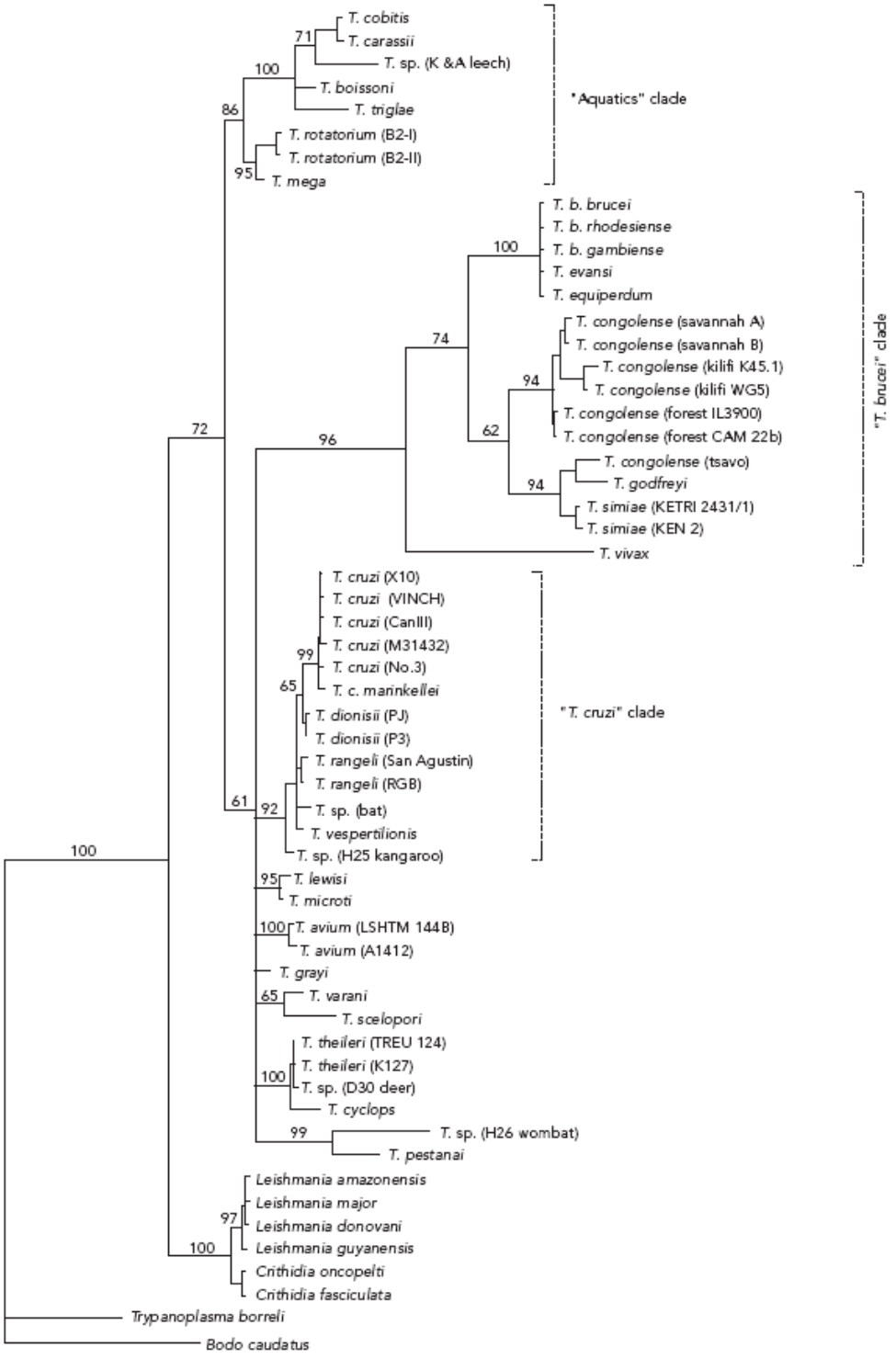

Figure 4: Phylogenetic tree based on bootstrapped maximum parsimony analysis of the 18S ribosomal gene (Stevens *et al*., 1999).

Certainly, the inclusion of additional taxa from bats and South American mammals in a study focusing on *T. rangeli* by Stevens *et al*. (1999) has allowed further clarification of the complex relationships of human infective trypanosomes within the *T. cruzi* clade. In particular, the study confirms unequivocally the close evolutionary relationship of *T. rangeli* with *T. cruzi* and a range of trypanosomes from South American mammals. In Figure 4 above, the ribosomal RNA data has not allowed the exact branching order of these groups to be determined and the tree shows an eight-way polytomy. The aquatic clade forms the first branch from the trypanosome lineage, providing evidence in support of host-parasite co-evolution, although the relatively low bootstrap value (61%) indicates that other hypotheses might also be considered. The polytomy and low bootstrap support suggest that the limit of the resolving power of the 18S ribosomal marker over this time scale may have been reached and other markers, e.g. glyceraldehyde phosphate dehydrogenase (GAPDH) and RNA polymerase, may be more informative (Stevens and Gibson, 1999). In addition, it appears that there may have been an explosive divergence of trypanosome species over a very short time period.

The conceptual and methodological assumptions underlying phylogenetic analysis have a significant bearing on the results of such analyses and hence the final evolutionary interpretations. These include alignment of sequences, the methods used in phylogenetic analysis, the out-groups and bootstrap support. Sequence alignment and the associated problem of identifying true homology between both variable sites and portions of sequences remains one of the most problematic areas in molecular phylogenetic analysis and the importance of sequence alignment on subsequent phylogenetic analyses is well recognised (Morrison & Ellis, 1997). Alignment can be performed by one or any combination of three main approaches: a) on the basis

of secondary structural and functional domains, e.g. secondary structure in ribosomal sequences (Neefs *et al*., 1990); b) using one of a range of specialist alignment programs with various weighting options and gap penalties, e.g. ClustalX (Thompson *et al*., 1999); c) by eye, often in relation to previously aligned sequences. Increasing the number of taxa may be accompanied by problems of hypervariability and saturation of nucleotide changes at some sites, resulting in a reduction of informative sites suitable for inclusion in phylogenetic analyses. Frequently, sites which are informative between closely related taxa may introduce 'noise' at higher phylogenetic levels as the frequency of non-evolutionary similarity (homoplasy) increases, resulting in a loss of definition and reduced bootstrap support. Such sites should be excluded from broad analyses, provided sufficient remain to be able to perform a meaningful analysis (Stevens amd Gibson, 1999).

The evolution of *Leishmania* gp63 has been studied (Medina-Acosta *et al*., 1993a). The mosaic evolution of *gp63* genes has been supported by the analysis of *gp63* evolutionary divergence and it has been found that different regions of the protein sequence exhibit different evolutionary histories (Medina-Acosta *et al.* 1993a). Studies have shown that relationships based on the N-terminal of the protein are consistent with the conventional taxonomical subdivisions. However, trees based on the C-terminal of the protein were very different. Mosaic genes are believed to be produced by intragenic recombination following mitotic crossover, which have been observed in genes encoding surface antigens in other pathogens (Deitsch *et al.,* 1997). The major surface protein of *Leishmania* promastigotes is evolutionarily conserved and is found in isolates of *L. donovani*, *L. major, L. tropica*, *L. mexicana*, and *L. braziliensis* (Etges *et al.,* 1986). In *Leishmania* species, regions encoding the *MSP* catalytic site show a conserved sequence, while

regions encoding surface domains possibly involved in the host–parasite interaction are variable (Victoir *et al*., 2005). These similarities and/or differences can be targeted for development of species-specific diagnostic approaches for African Trypanosomiasis. The diagnostic methods currently available for African Trypanosomiasis have got drawbacks that limit their use both in the field and in the laboratory. Most of these tests are unable to reveal single infections, hence cross infections cannot be diagnosed easily. The methods are also not sufficiently sensitive, specific and applicable in the field situation. Therefore, specific and sensitive diagnosis methods are required if early and life-saving treatment for the disease is to be initiated.

2.3 Current Diagnostic Methods for Trypanosomiasis

Examination of the blood by light microscopy is the most readily applied method for diagnosis of trypanosomiasis since it can be easily applied in the field (Luckins, 1992). Trypanosomes are seen either directly, moving between the blood cells, or indirectly, as they cause the blood cells to move. This method has advantages in that it is simple to use and inexpensive. However, microscopical diagnosis has a major disadvantage in that unless the animals are brought to the veterinary centre, or the blood (with an anticoagulant) can be taken quickly to the centre, a field microscope has to be taken to the herd, as the parasites lose their mobility after limited time duration. The technique has been modified to improve diagnostic sensitivity by concentrating the blood through centrifugation in a haematocrit tube: the haematocrit centrifuge technique or the dark ground buffy coat technique (Paris *et al*, 1982). Varying sensitivity of these diagnostic tests and the failure to detect trypanosomes if the number of parasites is too low, as is the case with chronic infections (Masake and Nantulya, 1990), illustrate the limitations of parasitological diagnosis and confirm the need for more reliable methods. The indirect fluorescent antibody test is

both specific and sensitive in detecting trypanosomal antibodies in infected cattle (Wilson, 1969; Luckins and Mehlitz, 1978) and camels (Luckins *et al*., 1979). The card agglutination test relies on the presence of a widely distributed variable surface antigen, is easy to carry out even in the field but requires experience in interpreting the results. The monoclonal antibody-based Enzyme-Linked Immunosorbent Assay (ELISA) for antigen detection has been proven to be unreliable (Eisler *et al*., 1998). ELISA antibody detection in trypanosomes (Nantulya and Lindqvist, 1989) is a great improvement regarding the sensitivity of the diagnosis; however it cannot differentiate between active and cured infections. Additionally, the test requires a well-equipped specialized laboratories as the test cannot be conducted in the field and the reagents are comparatively expensive (Masake and Nantulya, 1991; Luckins, 1997; Rebeski *et al*., 2000).

Molecular biology has provided tools for sensitive and specific diagnosis based on DNA sequence recognition and amplification. The polymerase chain reaction (PCR) permits identification of parasites at levels far below the detection limit of the commonly used parasitological techniques. PCR assays for trypanosome detection have been developed (Kukla *et al.,* 1987; Moser *et al.*; 1989 Desquesnes *et al*., 2001; Morlais *et al*., 1998) using species-specific DNA hybridisation probes (Majiwa and Webster, 1987; Kukla *et al*., 1987). This method requires either prior knowledge of the species to be found or the use of several probes for each sample to be tested. A 'pan-trypanosome' test based on the ITS 1 region of the ribosomal genes has been described by Desquesnes *et al.,* (2001) and Njiru *et al,* (2004) replacing the various PCRs with a single assay. However, the test is not applicable to the field situation as it lacks sensitivity, especially for *T. vivax*. The association of PCR with restriction fragment length polymorphism (RFLP) provides a sensitive and specific tool for use in the field that is sensitive, especially for *T. vivax*. Although

PCR and RFLP need a well-equipped laboratory, the collection of field samples on filter paper is very simple, and the sample transportation and conservation is easy hence it is well suited for large-scale surveys.

The popular PCR-based tool for isolate differentiation studies has been the Random Amplified Polymorphic DNA (RAPD) technique (Banuls *et al*., 1999). PCR allows the detection of a single specific sequence of DNA; consequently, a single parasite (0.1 pg of DNA), or even fewer can be detected when using satellite DNA, when the DNA is broken into pieces with specific DNAases. The enzymatic digestion of a DNA can lead to the detection of 0.01 trypanosome (Masiga *et al*., 1992) by hybridisation. The loop-mediated isothermal amplification (LAMP) reaction amplifies DNA with high specificity, efficiency, and rapidity under isothermal conditions using simple incubators. An added advantage of LAMP over PCR-based methods is that DNA amplification can be monitored spectrophotometrically and/or with the naked eye without the use of dyes (Kuboki *et al*., 2003).

Reliable DNA based methodologies to determine prevalence of trypanosome species in domestic livestock are rarely used to generate baseline data for control operations for these diseases in the field. Rather, such operations tend to rely on data, which can be generated using low technology methods such as direct observation of parasites by light microscopy. However, these methods require either prior knowledge of the species to be found or the use of several probes for each sample to be tested. These difficulties might be overcome by development of an appropriate method based on detection of MSPs, which are abundantly expressed on the surface of both the procyclic and the metacyclic forms the trypanosome.

CHAPTER THREE

MATERIALS AND METHODS

3.1 Test Samples

The samples for the three trypanosome species *Trypanosoma congolense, T. brucei brucei* and *T. vivax* to be studied were obtained from two sources namely: stabilates from the Trypanosome bank at Kenya Agricultural Research Institute - Trypanosomiasis Research Centre (KARI-TRC) in Muguga, Kenya (see Table1), blood samples collected from different locations and genomic DNA pellets stored at the KARI-TRC research laboratory. The parasites were identified by microscopic examination and had been maintained at the Trypanosome bank by inoculation into experimental mice and cryopreserved in liquid nitrogen.

Table 1: A list of the well-characterised stabilates obtained from the KETRI cryopreservation bank. Their origin, codes and isolation years are indicated.

Species	**Locality/Origin**	**Isolate Code**	**Year of Isolation**
T. congolense	Galana, Kenya	KETRI 2409	1978
		KETRI 2911	1983
	Central Nyanza, Kenya	KETRI 3343	1961
	South Nyanza, Kenya	KETRI 294	1962
	Maasai Mara, Kenya	KETRI 1598	1970
		KETRI 1755	1970
	Baringo, Kenya	KETRI 2765	1980
		KETRI 3288	1992
	Kiboko, Kenya	KETRI 2879	1983
	Coast, Kenya	KETRI 2784	1981
		KETRI 2800	1981
		KETRI 2831	1980
	Ngurumani, Kenya	KETRI 3280	1992
		KETRI 3284	1992
	Tanzania	KETRI 1117	1966
		KETRI 1163	1967
	Uganda	KETRI 209	1962
		KETRI 1345	1969
		KETRI 628	1963

		KETRI 644	1963
T. vivax	Ngurumani, Kenya	KETRI 3238	1992
		KETRI 3275	1992
	Baringo, Kenya	KETRI 3287	1992
	Samburu, Kenya	KETRI 3289	1998
	South Nyanza, Kenya	KETRI 3131	1988
		KETRI 3133	1988
	Coast, Kenya	KETRI 3214	1991
		KETRI 2369	1978
	Galana, Kenya	KETRI 2391	1978
		KETRI 2483	1980
	Kiboko, Kenya	KETRI 1481	1969
	Busia, Kenya	KETRI 2942	1987
		KETRI 2961	1987
	Meru, Kenya	KETRI 2009	1972
		KETRI 2066	1972
	Maasai Mara, Kenya	KETRI 1331	1969
		KETRI 1398	1969
	Germany	KETRI 2452	1979
	Uganda	KETRI 1441	1961
		KETRI 1288	1969
T. brucei	Central Nyanza, Kenya	KETRI 792	1964
		KETRI 794	1964
		KETRI 796	1964
		KETRI 807	1964
		KETRI 808	1964
		KETRI 809	1964
		KETRI 810	1964
		KETRI 862	1964
		KETRI 1051	1964
		KETRI 1067	1964
		KETRI 1068	1964
		KETRI 2831	-
		KETRI 2784	-
		KETRI 2773	-
		KETRI 2501	1964
		KETRI 3238	

3.2 Identification of MSP Homologues

The homologues for the Major Surface Proteases (MSP) were identified from among the sequences already deposited at the trypanosome genome database (available at The Sanger Institute website: http://www.genedb.org) and also using the previously determined *T. brucei*, *T. vivax* and *T. congolense* MSP DNA and protein sequences from GenBank (www.ncbi.nlm.nih.gov/blast/). The Basic Local Alignment Search Tool [BLAST] (Altschul *et al*., 1990) was used to find regions of local similarity between the sequences. Homologs in other species were also obtained and these were used for the DNA and protein sequence data analysis.

3.3 Design of PCR Primers

The primers for initial detection of MSP and for colony PCR were designed using PRIMER3 version 0.4.0 software (Steve *et al*., 2000) and in other instances, this was done manually and or in combination with the PRIMER3 software. In instances where the primers were designed manually, the melting temperatures (Tm), self compatibility, formation of hairpins were analysed using the OligoCalc online software (http://www.basic.northwestern.edu/biotools/oligocalc.html). Primers were obtained from MWG Biotech.

Table 2: List of primers designed for use in detection and cloning of MSP for the three species of trypanosomes under study. Restriction endonucleases recognition sites are underlined. The restriction enzymes mainly used were *BamHI*, *XhoI* and *NotI*.

Primer Name	**Primer Sequences (5' to 3')**	**Number of bases**	**Species**
TvivMSP3c Fwd	GCCCTAGGCTACTGGCGATAGAAA ACCACGTATAAACG	35	*T. vivax*

TvivMSP3c Rev	GCTCGAGATTCACATTTGACGAGGG TTCACGGCCG	35	
MSP2For	GTACAGCGGCCGCATGACGCCTATGCGGAG TTC	33	*T. congolense*
MSP2Rev	GCATGCGGCCGCCTAGCTACTCGACAC CGCCGCAAC	36	
pRSETATcMSP1For	GCGGATCCATGATACTCATCTCCCCAAATG	30	
pRSETATcMSP1/2Rev	GCCTCGAGCTACTAGCTACTCGACACCGCC GC	32	
pRSETATcMSP2For	GCGGATCCATGACGCCTATGCGGAGTTC	28	
pET28TbMSP For	CTGCGGCCGCATGAC(AG)GC)TGATTATGTT CC(CT)GC	40	*T. brucei*
pET28TbMSP Rev	CTGCGGCCGCCTAT(CT)ATAATAGGG(AG) AAA(AG)CAAGAAA	45	

3.4 Preparation of Samples

3.4.1 DNA Extraction

The samples for DNA extraction were collected from the Trypanosome Cryobank at the Trypanosomiasis Research Centre in Muguga, Kenya. The samples were contained in capillary tubes stored in liquid Nitrogen, and DNA extracted from the blood samples using the DNeasy Blood and Tissue Kit (Qiagen, Germany) as per the manufacture's instructions. The purified genomic DNA samples were stored at -20ºC till needed for use.

3.4.2 Sub-grouping of the Samples using the Internal Transcribed Spacer Primers

The samples obtained from the Trypanosome bank had been identified and grouped into the respective species based on morphology. However, there was need to confirm these identifications using the more accurate molecular tools such as the Internal Transcribed primers [ITS] (Njiru *et al*., 2005) This was carried out using the ITS-based primers ITS1 CF: 5'

CCGGAAGTTCACCGATATTG 3' and ITS1 BR: 5' TTGCTGCGTTCTTCAACGAA 3' (See Table 2) were available at the laboratory and were used for the purpose of detecting the pathogenic trypanosomes.

Table 3: A table of the ITS 1 primer sequences and their specific amplification size products for different trypanosome species.

Primer Name	**Primer Sequence (5'to 3')**	**Specificity**	**Amplification Size Product**	**Reference**
ITS 1 CF	CCGGAAGTTCACCGATATTG	*T. congolense,* savannah	700bp	Njiru *et al*., 2005
ITS 1 BR	TTGCTGCGTTCTTCAACGAA	*T. congolense,* kilifi	620bp	
		T. congolense, forest	710bp	
		T. vivax	250bp	
		T. b. brucei	480bp	

The PCR cycles were: initial denaturation step at 94˚C for 5 minutes, followed by 35 cycles of 94˚C for 40 seconds, 58˚C for 40 seconds, 72˚C for 90 seconds, and a final extension at 72˚C for 5 minutes. The amplification products were resolved in a 1.5% molecular grade agarose stained with Ethidium bromide (Njiru *et al*., 2005).

3.4.3 Sub-grouping of the *T. congolense* Samples using Specific Sub-species Primers

Another set of primers specific for *T. congolense* was used to identify the specific subgroups in which the *T. congolense* samples belonged. The primers used are listed in Table 4.

Table 4: List of *T. congolense* specifics primer sequences and their specific amplification size products (Ferreira *et al.,* 2008).

Primer Name	**Primer Sequence (5'to 3')**	**Specificity**	**Amplification Size Product**	**Reference**
TCS1 TCS2	CGAGAACGGGCACTTTGCGA GGACAAACAAATCCCGCACA	*T. congolense* savannah type	316bp	Majiwa *et al*., 1993
TCK1 TCK2	GTGCCCAAATTTGAAGTGAT ACTCAAAATCGTGCACCTCG	*T. congolense* kilifi type	294bp	Masiga *et al*., 1992
TCF1 TCF2	GGACACGCCAGAAGGTACTT GTTCTCGCACCAAATCCAAC	*T. congolense* forest type	350bp	

The PCR cycles were: initial denaturation step at 94˚C for 5 minutes, followed by 30 cycles of 94˚C for 60 seconds, 55˚C for 120 seconds, 72˚C for 120 seconds, and a final extension at 72˚C for 5 minutes (Picozzi *et al*., 2002). The amplification products were resolved in a 1.5% molecular grade agarose stained with Ethidium bromide.

3.4.1 DNA Amplification

The coding sequences of *MSP* were amplified via PCR from *Trypanosoma congolense*, *T. b. brucei* and *T. vivax* genomic DNA with the primers listed in Table 2. Before attaining the optimal PCR conditions to work with, the following considerations were made: the primer concentration, the amount of DNA polymerase enzyme, the $MgCl_2$ and dNTP concentrations, template DNA volumes to use and other additives such as Dimethylsulfoxide (DMSO). Optimising for the appropriate annealing temperatures was also carried out. Only a few of the samples were used during the optimisation period. A PCR master mix was prepared by assembling the following

components per reaction: 4.1µl Nuclease-free water, 0.4µl of 10Mm dNTP mix, 1µl upstream *MSP* primer, 1µl downstream *MSP* primer, 2µl of 10X Thermophilic DNA polymerase buffer without $MgCl_2$ (Promega, USA), 1.2µl of 25mM $MgCl_2$ (Promega, USA), 0.3µl Phusion High Fidelity DNA polymerase (Finnzymes Oy, Finland) alternatively *Taq* DNA polymerase (New England Biolabs, UK) was used. 5µl of the template DNA was added to 15µl master mix and mixed gently. Processing in the thermal cycler was done as follows: 3 minutes at 99° C initial denaturation followed by 35 cycles of 99ºC for 1 minute, 60ºC for 45 seconds, 72° C for 1 minute and 20 seconds with a final extension at 72° C for 8 minutes. 5µl of 6X loading dye was added to the PCR product and 5-10µl was loaded per lane on a 1% agarose gel containing 0.5µg/ml ethidium bromide. At least one lane of 1 kb DNA marker was included (New England Biolabs, UK).

3.4.2 Purification of PCR Products

The PCR products were purified using the QuickClean® PCR Purification Kit (Genscript, USA) following the manufacturer's instructions. The purified products were stored at -20ºC until needed for use. An aliquot of the PCR product was analysed on an agarose gel before use in the subsequent steps. In other instances, the bands of the appropriate size were cut out of the agarose gel and put in 1.5ml eppendorf tubes. The DNA was then purified from the gel using the QuickClean® Gel Purification Kit (Genscript, USA) following the manufacturer's instructions. An aliquot of the eluted sample was analysed on an agarose gel before use in the subsequent steps. The purified samples were stored at -20ºC until needed for use.

3.5 MSP GENE CLONING

3.5.1 Ligation

Cloning was undertaken using pGEMT-Easy vector (Promega, USA) following the manufacture's instructions. The MSP samples amplified with Phusion DNA Polymerase yield blunt-ended products hence have poly A tails have to be added to aid cloning into pGEMT-Easy vector. This was done by using the blunt-ended products as templates for a second round of PCR using a *Taq* DNA polymerase that gives products with poly A overhangs. Alternatively, the blunt-ended PCR products were incubated at with dATP final concentration of 0.2mM and *Taq* DNA polymerase 70° C for minutes 30minutes. The cloning was carried out using 20µl reaction volumes containing 5µl of 2X Rapid ligation buffer, 50ng of pGEMT Easy vector (Promega, USA), 3µl of each of the MSP PCR products, 3weiss units per µl of T4 DNA Ligase. The negative control consisted of distilled sterilized water in place of the PCR products. The mixture was vortexed once and kept at +4°C overnight to obtain maximum number of transformants.

3.5.2 Transformation of *Escherichia coli*

The 50µl of competent *E. coli* cells strain, DH5α used for transformation. The sets of ligation mixtures were added to the 1.5ml eppendorf tubes containing the competent cells. The tubes were left in ice for 20 minutes. The mixture was heat shocked at 42°C for 1 minute and put back on ice for 2 minutes. 950µl of SOC medium was added under sterile conditions. The tubes were incubated at 37°C for 2 hours. The reaction mixtures were centrifuged at 6000 rpm for 5 minutes. Luria bertani (LB) agar medium plates were made containing 80µg/ml of X-*Gal*/IPTG and 10µg/ml of ampicillin as the selectable marker. The excess SOC was poured out and approximately 100µl was left in the tube for use in resuspending the pelleted cells followed by

application onto plates, spreading using a sterilized glass spreader and left 30 minutes to dry. The plates were incubated at 37°C overnight.

3.5.3 Screening of Transformed Bacterial Colonies by PCR

The positive colonies for MSP were screened by PCR using the designed MSP-specific primers (Table 2). Standard PCR amplifications were carried out as described earlier but the DNA templates were the colonies. The PCR products were mixed with 1X loading dye and resolved on a 1% agarose gels stained with of 5µl ethidium bromide. A 5µl 100bp plus DNA ladder (MBI Fermentas, Lithuania) was included in every gel. The samples were electrophoresed for 1 hour at 80V and illumination by UV to check for the positive clones amplified in the range of 1500-2000bp. Colonies positive by PCR were picked from the plates using a sterile pipette tip, put in the falcon tubes containing Ampicillin, and incubated at 37°C overnight while shaking (Environ-Shaker 3597-1, Labline Instrument Inc., USA).

3.5.5 Plasmid Purification

Harvesting of cells for plasmid purification was carried out by centrifugation of the overnight cultures at 6000 rpm for 15 minutes at 4°C. Plasmid purification was carried out using QIAprep®Miniprep kit (QIAGEN, Germany) using the microcentrifuge method as per the manufacturer's instructions.

3.6 Sequencing

Aliquots of 10µl each of the samples were sent out for sequencing (BecANet sequencing platform at ILRI, Kenya and Macrogen Incorporation, Korea) using a procedure known as primer walking (Reeve *et al*., 2002).

3.7 Sub-Cloning of MSP into an Expression Vector

3.7.1 Insert and Vector Preparation

The test samples were MSP-pGEMT plasmids and *Trypanosoma congolense, T. brucei* and *T. vivax* MSP DNA pellet that had been amplified and stored at -20°C. The pET28a (Novagen, USA) and pRSET-A (Invitrogen, USA) vectors were prepared to receive the MSP fragments by restriction enzyme digestion. The MSP inserts were also obtained through restriction digestion using the same enzyme used to linearise the plasmids. The following components were assembled in a microcentrifuge tube: 10µl plasmid DNA, 5µl restriction enzyme buffer, 0.2µl acetylated Bovine Serum A (BSA), 0.4µl restriction enzyme, 34.4µl Nuclease-free water to volume. These were incubated at 37°C for 2–4 hours. A 3µl sample together with DNA Markers were run a on an Agarose gel to evaluate the extent of digestion. When digestion was complete, calf intestinal alkaline phosphatase was added directly to the remainder of the vector digestion followed by incubation at 37°C for 30 min. The gel sample buffer was added to the reaction and the entire sample loaded into a large well (0.5–1.0 cm wide) on a 1% Agarose gel containing 0.5 µg/ml ethidium bromide. Uncut vector DNA was electrophoresed in an adjacent lane to help distinguish undigested from linearised plasmid DNA. The DNA band was visualized with a long wave Ultra-Violet light source and the band excised from the gel using a clean razor blade. The DNA was

recovered from the gel slice using the Spin Prep™ Gel DNA kit (QIAGEN, Germany) as per manufacturer's instructions.

3.7.2 Ligation

The insert was ligated into the prepared expression vectors (pET28a/pRSET-A). The following components were assembled in a microcentrifuge tube per reaction: 1µl 10X Ligase Buffer, 1.5µl prepared pET28a/pRSET-A vector, 4.5µl prepared target MSP insert, 2.8µl Nuclease-free water to volume, 0.2µl T4 DNA Ligase. These were incubated at 16°C for 2 h to overnight. A control reaction was also set up in which the insert was omitted to check for non-recombinant background.

3.7.3 Transformation

The entire ligation reaction was added directly to 50 µl of competent *E. coli* strain, DN5α cells and stirred gently to mix. The tubes were incubated on ice for 5 minutes, and then heated for exactly 30 seconds in a 42°C water bath without shaking. The tubes were then placed on ice for 2 minutes. 950µl room temperature SOC medium (Appendix 6.1.3 7) was added to each tube. The tubes were kept on ice until all have received SOC. Selection for transformants was accomplished by plating on LB medium containing Kanamycin/Ampicillin. 5–50µl of each transformation was spread on LB agar plates containing Kanamycin/Ampicillin. The plates were set on the bench for 30 minutes to allow excess liquid to be absorbed, and then inverted and incubated overnight at 37°C.

3.7.4 Screening of Recombinants by PCR

The colony PCR (Shim and Kim, 2008) was achieved using vector-specific primers (Table 2). A colony was picked from an agar plate using a 200µl pipette tip. Colonies that were at least 1 mm in diameter was chosen. A copy of the colony desired was picked by touching the pipette tip to a plate before transferring to a 0.5-ml tube containing 50 µl of sterile water. The cells were dispersed by vortexing and centrifuged at 12,000 × g for 1 minute to remove cell debris.10 µl of the supernatant was transferred to a fresh 0.5 ml tube for PCR. A master mix was prepared for colony PCR by assembling the following components: Per reaction: 31.75 µl Nuclease-free water, 1 µl dNTP mix, 1 µl upstream primer, 1 µl downstream primer, 5 µl 10X Taq Buffer with $MgCl_2$, 0.25 µl *Taq* DNA Polymerase. 40 µl of the master mix was added to each 10µl sample and mixed gently. Processing in the thermal cycler for 35 cycles was as follows: 2 minutes at 94°C, 45 seconds at 55°C annealing temperature, 2 minutes at 72°C and 8 minutes final extension at 72°C. 5 µl 10X loading dye was added to the PCR product and 10–25 µl was loaded per lane on a 1% Agarose gel containing 0.5 µg/ml ethidium bromide. At least one lane of DNA Marker was included.

3.7.5 Plasmid DNA Preparation

Plasmid DNA was obtained using the QIAprep® Miniprep kit (QIAGEN, Germany) as per manufacture's instructions. To elute DNA, 50µl of deionised distilled water was added to the centre of each QIAprep Spin column and let to stand for 1 minute, followed by centrifugation for 1 minute. The purified plasmid DNA was stored at –20°C until required for use.

3.8 EXPRESSING THE MSP GENE

3.8.1 Expression Host Transformation

The expression hosts used were BL21 strain of *E. coli.* Transformation was done as follows: The entire ligation reaction was added directly to 50 µl of competent *E. coli* strain, BL21 (DE3) pLysS cells (Invitrogen, USA) and stirred gently to mix. The tubes were incubated on ice for 5 minutes, and then heated for exactly 30 seconds in a 42°C water bath without shaking. The tubes were then placed on ice for 2 minutes. 950µl room temperature SOC medium was added to each tube. The tubes were kept on ice until all had received SOC. Selection for transformants was accomplished by plating on LB medium containing Kanamycin/Ampicillin. 50µl of each transformation were spread on LB agar plates containing Kanamycin/Ampicillin. The plates were set on the bench for several minutes to allow excess liquid to be absorbed, and then inverted and incubated overnight at 37°C.

3.8.2 Induction of Protein Expression

A single colony was picked from a freshly streaked plate and inoculated to 50 ml LB containing Kanamycin/Ampicillin respectively in a 250 ml Erlenmeyer flask. A starter culture was prepared of the pET28a/pSETA recombinant in BL21DE3 (pLysS) cells [Genotype: F– *ompT hsdSB* (*rB*–*mB*–) *gal dcm* (DE3) pLysS] (Invitrogen, USA) by inoculating 3 ml of LB media (containing Kanamycin/Ampicillin) in a culture tube with a single colony from a plate or sterile loop of cells from a glycerol stock. This was followed by incubation at 37°C with shaking (Environ-Shaker 3597-1, Labline Instrument Inc., USA) at 250 rpm to an OD_{600} of approximately 0.5. The entire 3 ml culture was added to 100 ml medium containing Kanamycin/Ampicillin. The culture was shaken (Environ-Shaker 3597-1, Labline Instrument Inc., USA) at 37°C until the OD_{600} was

approximately 0.4–0.6. The OD_{600} was monitored during growth by removing aliquots aseptically. Just prior to induction, 100 ml culture was split into 2 × 50 ml cultures. IPTG was added to one of the 50 ml cultures and the other culture was used as an un-induced control. Both cultures were incubated at 37°C with shaking (Environ-Shaker 3597-1, Labline Instrument Inc., USA) for 6 hours. Aliquots of the induced culture were collected after every 2 hours.

3.8.3 Target Protein Verification

3.8.3.1 Determining Target Protein Solubility

10 ml LB medium containing 100µg/ml ampicillin or 25µg/ml Kanamycin was inoculated in a 50 ml flask. The cultures were grown overnight at 37°C with shaking (Environ-Shaker 3597-1, Labline Instrument Inc., USA). 50 ml of pre-warmed media (with antibiotics) was inoculated with 2.5 ml of the overnight cultures and grown at 37°C, with vigorous shaking [~200 rpm] (Environ-Shaker 3597-1, Labline Instrument Inc., USA), until the OD_{600} was 0.5–0.7. A 1 ml sample was taken immediately before induction (non-induced control). The cells were pelleted and resuspended in 50µl 1X SDS-PAGE sample buffer. IPTG to a final concentration of 1mM was added to induce expression. The cultures were grown for an additional 4–6 hours. A second 1 ml sample (induced control) was collected. The cells were pelleted and resuspended in 100µl 1X SDS-PAGE sample buffer. The cells were harvested by centrifugation at 5000 x *g* for 20 minutes.

3.9 Protein Extraction

The cell pellet from Section 3.8 was resuspended in 5 ml of lysis buffer (Appendix 6.1.6.1) for extraction under denaturing conditions for 1 hour. The lysate was centrifuged at 10,000 x *g* at 4°C for 20–30 minutes. The supernatant constituted of the crude extract A (soluble protein) was

decanted and saved on ice. The pellet was resuspended in 5 ml lysis buffer. This was a suspension of the insoluble matter (crude extract B, insoluble protein).

3.10 SDS-PAGE Analysis

5µl of 2 X SDS-PAGE sample buffers was added to 10µl of crude extracts A & B. These samples were heated, along with the frozen non-induced and induced cell samples at 95°C for 5 minutes after which they were centrifuged at 15,000 rpm for 1 minute. 15µl of the non-induced and induced cell samples and all of the extract samples were loaded on a 4-20% SDS-PAGE gel and run according to standard procedures (Sambrook *et al*., 1989). The gel was then stained using the Silver stain method to visualise the separated bands of protein.

3.11 Protein Purification

3.11.1 Ni-NTA Agarose Column Purification

Proteins were purified using the Qiagen Ni-NTA Agarose protein purification kit (QIAGEN, Germany) according to the manufacturer's instructions. The Ni-NTA spin column was equilibrated with 600µl buffer B followed by centrifugation for 2 minutes at 2000 rpm (approximately 700 x *g*).

The cleared lysate supernatant containing the 6XHis-tagged protein was loaded onto an equilibrated Ni-NTA spin column followed by centrifugation of the Ni-NTA spin column for 2 min at 2000 rpm, and the flow-through collected. The flow-through was saved for SDS-PAGE analysis. The Ni-NTA spin column was washed twice with 600µl buffer C (Appendix 6.1.6.2) and centrifuged for 2 min at 2000 rpm (approx. 700 x *g*). The flow-through (wash fraction) was

saved for analysis by SDS-PAGE to check the stringency of the wash conditions. The Ni-NTA spin column was washed twice with 600µl buffer D (Appendix 6.1.6.3) and centrifuged for 2 min at 2000 rpm. The flow-through was saved for analysis by SDS-PAGE. The protein was eluted with 2 x 200µl buffer E (Appendix 6.1.6.3) and centrifuge for 2 min at 2000 rpm, and the eluates collected. 2.5µl of 5 X SDS-PAGE sample buffer was added to 10µl aliquots of all samples, including the unbound fractions, and boiled for 5 min at 95°C followed by analysis of the samples by SDS-PAGE and staining using the Silver stain method.

3.11.2 Alternative Crude Method

3.11.2.1 Staining Side Strips of the Gel

After gel electrophoresis, a clean scalpel was used to cut off a strip on the right or left of the gel (to include the molecular weight marker lane and the first lane of protein sample). The strip was placed in a tray for staining; the rest of the gel was placed on a glass plate wrapped in plastic to prevent it from drying while staining the strip. The cut strip of gel was stained using the Silver Staining protocol. This strip will function as the "reference" gel strip. The stained strip of gel was aligned with the unstained gel portion and the band of gel that aligns with the stained protein of interest in the reference strip was cut out. Bands of gel just above and below the region presumed to contain the protein of interest were also excised and processed. The entire remaining gel was stained after excision of bands to determine the accuracy of excision.

3.11.2.2 Electro-elution of Proteins from Polyacrylamide Gel Pieces

The protein-containing gel pieces were placed in an electro elution chamber, where the proteins were eluted from the gel matrix into the SDS-PAGE tank buffer solution using an electrical field and captured against a dialysis membrane with a molecular weight cut off of 6000-8000 Daltons.

3.11.3 Target Protein Concentration

3.11.3.1 Dialysis

A dialysis membrane with a 6-8 kDa molecular mass cut off was used. The membrane was prepared by boiling in 100mM EDTA buffer for 10 minutes and rinsed thoroughly with distilled water. One end was sealed with clips or a string and the protein solution poured into the tube. The other end was also sealed. Polyethylene glycol 20,000 (PEG 20 kDa) was poured in a clean Tupperware box. Concentration was completed in 30 minutes. The time however depended on the starting volume of protein solution and the amount of PEG 20 kDa that was poured around the membrane. Once the concentration was finished, the outside of the tubing was gently washed with distilled water. One end of the tubing cut open and the concentrated solution of protein removed by aspiration with a micropipette.

3.11.3.2 Trichloroacetic acid (TCA) Precipitation

An equal volume 20% TCA to sample was added and mixed thoroughly by vortexing followed by incubation for 1 h at 4°C. The sample was centrifuged at 4°C for 30 minutes at 13,000 × g. The supernatant was decanted, taking care not to disturb pellet. 0.5ml of ice-cold acetone was added and mixed thoroughly by vortexing followed by incubation at –20°C for at least 30 minutes to overnight. Centrifugation was done at 4°C for 30 minutes at 13,000 × g and the

supernatant decanted, taking care not to disturb pellet. The pellet was let to dry thoroughly (about 1 hr) under a hood. The pellet was resuspended in 20 µl of 0.1 M NaOH, or deionised water.

3.11.3.3 Precipitation by Non-ionic Polymers

This form of precipitation involved the addition of a non-ionic polymer to the protein solution. This works because the addition of the polymer reduces the amount of water available to interact with the protein. Polyethylene glycol 20,000 (PEG 20 kDa) was poured in a clean Tupperware box containing a membrane with the proteins sealed inside and concentration was completed in about 30 minutes. The time, however, depended on the starting volume of the protein solution and the amount of PEG 20 kDa that was poured around the membrane. Due to its size, PEG 20 kDa will not diffuse into the protein solution.

3.12 Bioinformatics

The sequence data was analysed using the Desktop cDNA Annotation System (dCAS) version 1.3 (http://exon.niaid.nih.gov) that consists of several modules. The primer and vector sequences were removed from raw sequences, compared against the GenBank non-redundant (NR) protein database using the standalone BlastN program (Altschul *et al*., 1990) found in the executable package at ftp://ftp.ncbi.nlm.nih.gov/blast/executables/ and searched against the Conserved Domains Database (Marchler-Bauer *et al*., 2002) (found at ftp://ftp.ncbi.nlm.nih.gov/pub/mmdb/cdd/), which includes all Pfam (Bateman *et al*., 2000) and SMART (Schultz *et al*., 2000) protein domains. The predicted translated proteins were searched for a secretory signal through the SignalP server (Nielsen *et al*., 1997) and the data presented in the form of a table in a Microsoft Excel form. The STRIPPER program took as input a bunch of

the DNA sequence files with extension * .seq in one directory, or one FASTA formatted file (*.fsa), and trimmed high N content on beginning and end of files, found and removed primer sequences. It also Blasted (searched) the data against a vector database, and removed contaminating sequences and further Blasted the database against itself to indicate the number of new sequences to a growing database. The MSP DNA sequences were translated to protein sequences and these were submitted to InterPro (http://www.ebi.ac.uk/interpro/), an integrated protein database, in order to search for conserved domains and motifs within the peptide sequence. Other databases such as SMART (Schultz *et al*., 2000) were also used for these searches in the CAS software as described previously.

3.13 Evolutionary Analysis

3.13.1 Species Sampling

Representative samples of the species *Trypanosoma vivax, T. congolense* and *T. b. brucei* were included in this study. Samples were included from different geographically distant species for each species to have as representative a sample as possible. The species sampling included 3 species for all MSP genes studied. Specimens were stabilates available at the trypanosomes bank that had been collected from livestock in different geographical locations around East Africa. A representative sequence of *Leishmania braziliensis* gp63 gene sequences was included as an out-group for the evolutionary analyses.

3.13.2 Sequence Alignment and Phylogenetic Tree Drawing

Sequence fragments obtained as chromatograms were edited and assembled into contiguous alignments using the BioEdit software (Hall, 1999). Multiple sequence alignments (Corpet, 1998)

of both the DNA and protein sequences were performed using ClustalX software (Thompson *et al*., 1999) implemented in the BioEdit software package (Hall, 1999) using the default settings and then manually edited. In other instances, MEGA version 4 (Tamura *et al*., 2007) software was also used for the same purpose. In this program, a multiple protein sequence alignment was constructed by the Needleman and Wunsch algorithm, which maximizes similarities. Starting with the pair of species with the highest similarity score, the program inserted neutral elements to fill gaps, preserving the gap structures during the progressive alignment of additional species. From the resulting percentage difference matrix, the branching order and branch lengths were calculated for tree construction. These data were compared with sequences already deposited in the different databases available.

Phylogenetic and molecular evolutionary analyses were conducted using Phylip (Felsenstein, 1993) and MEGA (Tamura *et al*., 2007). Each gene sequence was analyzed separately before a combined analysis was performed in order to explore topological differences and possible sources of conflict between the signals of individual genes. Bootstrap analysis with 1000 replications was performed in the software ClustalX (Thompson *et al*., 1999), Phylip (Felsenstein, 1993) and in PhyML (Guindon and Gascuel, 2003). A phylogenetic tree was inferred using the models recommended by the PROTTEST software (Abascal *et al*., 2005).

CHAPTER FOUR

RESULTS

4.1 Identification of Genes Encoding Homologues of *Trypanosoma* MSPs

A total of five hundred and six (506) [as at November 20, 2009] MSP homologs were retrieved from the GeneDB trypanosome genome database [hosted at the Wellcome Trust Sanger Institute (Pathogen Sequencing Unit)] (http://www.genedb.org/genedb/) and used as databases for running a local BLAST exercise. Three fully annotated MSP (gp63) each from *Trypanosoma vivax, T. congolense* and *T. b. brucei* found great use during MSP primer design.

4.2 Trypanosome Samples Identification

4.3.1 Sub-grouping of the Samples using ITS 1 Primers

The amplification products resolved in a 1.5% molecular grade agarose stained with Ethidium bromide are shown in Figure 5.

The gel photo in Figure 5A illustrates the representative *T. congolense* samples that were amplified using the ITS primers. There was variation in the products obtained giving an indication that they belonged to different subspecies. The PCR products of approximately 700bp and 620bp were obtained. The *T. vivax* samples were also PCR amplified using the ITS 1 primers yielding products of approximately 250bp (See Figure 5B).

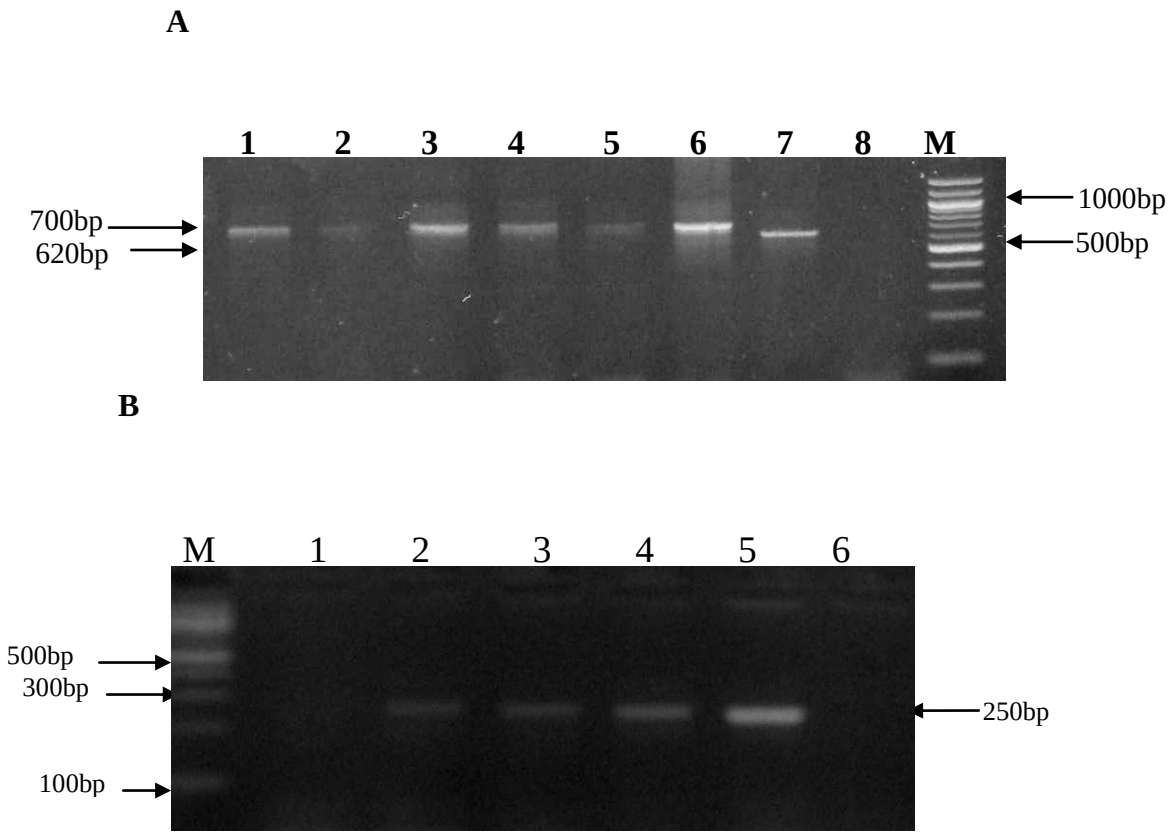

Figure 5: **A)**: A 1.5% agarose gel illustrating representative samples of *T. congolense* samples 1-7 with a 100bp marker (New England Biolabs, UK) on the first lane after PCR with the ITS 1 primers. Lane 8 contained the negative control. Samples 1-6 gave PCR products of 720bp long suggesting that they belong to the *T .congolense*, savannah subtype. Sample 7 gave a product of approximately 620bp characteristic of the *T. congolense*, kilifi subtype. **B)** A 1.5% agarose gel photo illustrating representative *T. vivax* samples 1-5 with a 100bp marker (NEB, UK) on the first lane after PCR with the ITS 1 primers. Products of approximately 250bp were obtained an indication that they are indeed *T. vivax* samples. Lane 6 contained the negative control.

4.3.3 Sub-grouping of the *T. congolense* Samples using Sub-species Specific Primers

Amplification of the *T. congolense* samples with the TCS1/2 primers gave PCR products of approximately 320bp as shown in Figure 6A. When the *T. congolense* samples were amplified with TCK primers products of approximately 294bp were obtained for three of the samples namely: KETRI 1117 and KETRI 2784 indicating that they belong to the subgroup. Amplification with TCF1/2 primers yielded no products (Figure 6c).

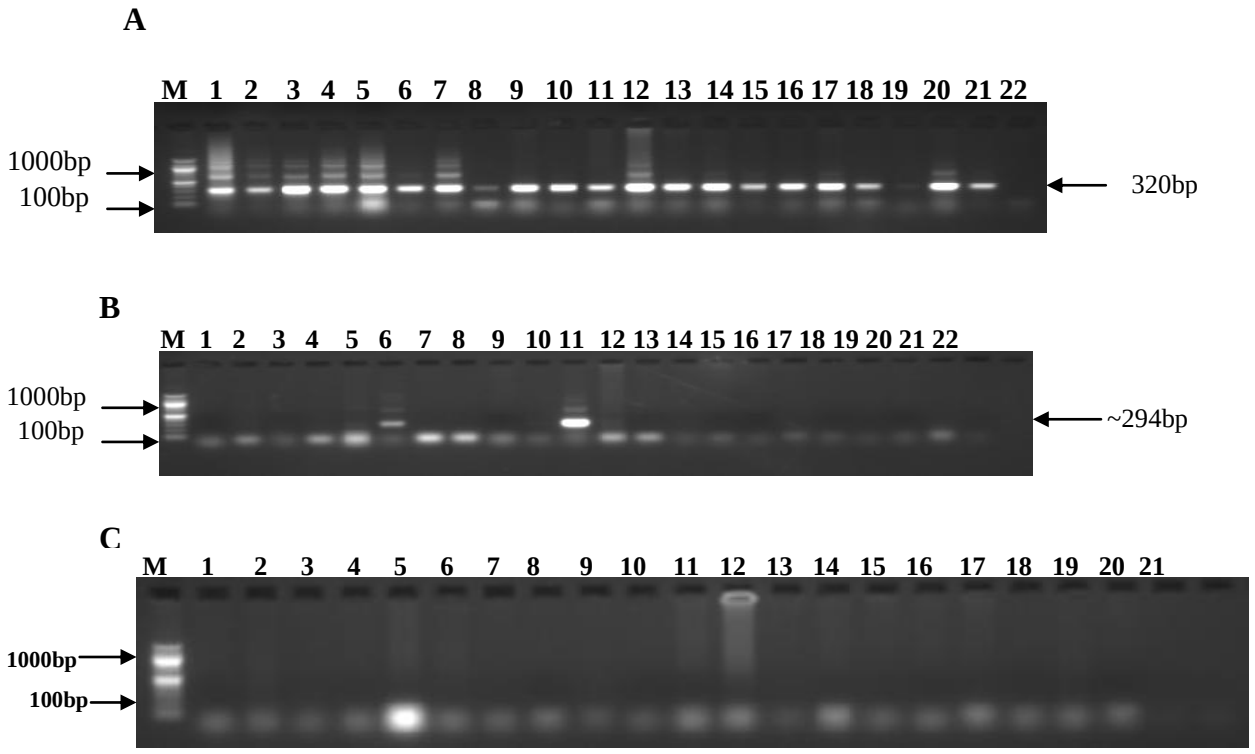

Figure 6: A) A 1.5% agarose gel of the PCR products of *T. congolense* samples, 1-20 using TCS1/2 primers, with a 100bp marker (NEB, UK) on the first lane and a positive and negative control on lanes 21 and 22 respectively. **B)** A 1.5% agarose gel of the PCR of products of *T. congolense* samples, 1-20 using TCK primers, with a 100bp marker (NEB, UK) on the first lane and a negative control on lane 22. Products of approximately 294bp were obtained confirming that two samples belong to the *T. congolense,* kilifi sub-group. **C)** A 1.5% agarose gel photo of the PCR products of *T. congolense* samples, 1-20 using TCF primers, with a 100bp marker (NEB, UK) and a negative control on the lane 21. There were no samples belonging to the subgroup *T. congolense*, forest as there were no observed products on the gel photo.

4.3.4 Amplification of MSP by PCR

The forward and reverse primers for the full length MSP gene synthesized (Table 1) were used to amplify MSP via PCR. The next step involved optimising the Magnesium Chloride ($MgCl_{2)}$ concentrations by titrating different concentrations of $MgCl_2$ while all the other PCR conditions remain standard. Resolution on a 1% agarose gel revealed the results shown below. The optimum concentration of $MgCl_2$ was at 1.5mM. The PCR products ranging in size between 1200-2000bp were obtained (See Figure 7).

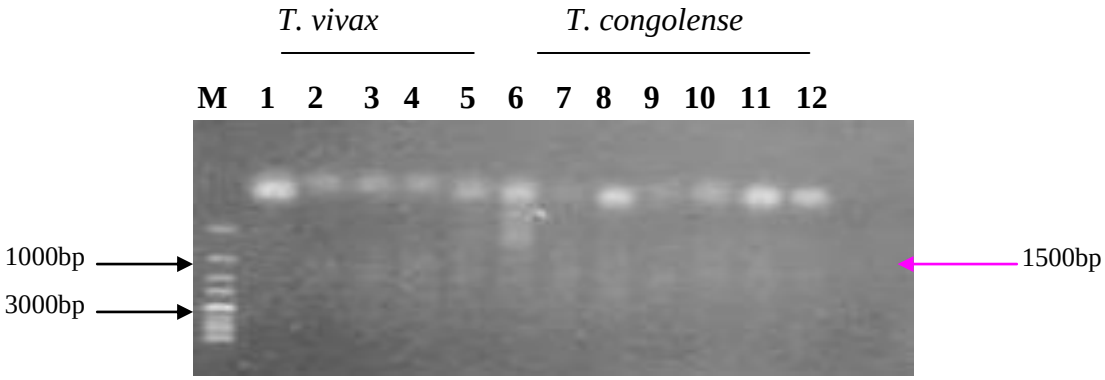

Figure 7: A 1% Agarose gel showing the *T. vivax* samples 1-5, a positive control on lane 6 and *T. congolense* samples 7-11 and a positive control on lane 12. The MSP2 detection primers were used. The products gave very faint bands of sizes between 1000-2000bp for the two species indicating the presence of MSP in the DNA samples used.

A nested PCR, followed by ethanol precipitation of the products and eventual application on a 1% agarose gel indicate that the amplification of MSP by PCR yielded stable non-fragmenting products of approximately 1600bp (See Figure 9) and 1500bp (See Figure 10). This was the correct size range of PCR products expected.

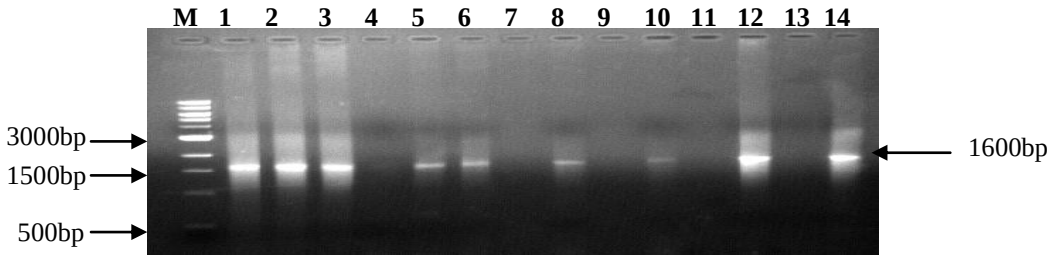

Figure 8: A 1% agarose gel photo of PCR products for *T. congolense* samples 1-14 after a nested PCR. Products of approximately 1500bp were obtained as expected.

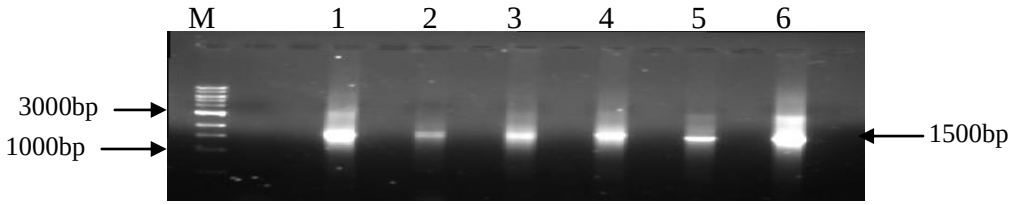

Figure 9: A 1% agarose gel photo of PCR products for *T. brucei* samples 1-14 obtained after a nested PCR in order to obtain sufficient concentrations. Products of approximately 1500bp were obtained.

The DNA was purified from the gel and an aliquot run on an agarose gel to confirm elution of the DNA samples from the purification column (See Figure 11).

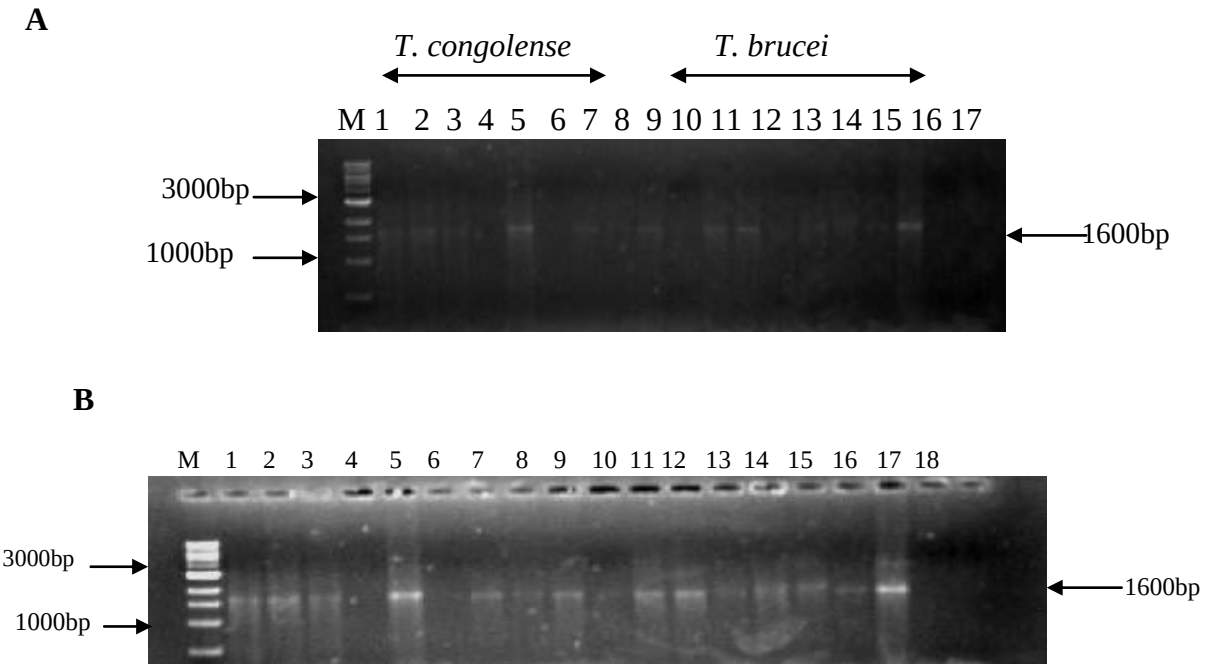

Figure 10: **A)** A 1% agarose gel photo of PCR products for *T. congolense* samples 1-9 and *T. brucei* samples 10-16 obtained after gel purification of the nested PCR products. A negative control was included in lane 17. Products of approximately 1600bp were obtained for the *T. congolense* samples and 1500bp for the *T. brucei* samples. **B)** A 1% agarose gel photo after a gel purification process of the PCR products for *T. congolense* and *T. brucei* samples 1-17 obtained after a nested PCR described above. A negative control was included in the last lane. Products of approximately 1600bp were obtained for the *T. congolense* samples and 1500bp for the *T. brucei* samples.

4.4 Cloning and Expression of the *Trypanosoma* MSP Gene

The cloning of MSP in pGEMT-Easy vector was successful followed by sub-cloning to expression vectors (pRSET-A and or pET28a).

4.4.1 Plasmid Purification and Screening

Recombinant plasmids were purified and screening using the restriction endonuclease digestion with the aim of removing the insert from its parent vector(s). The positive plasmids showed at least two distinct bands (Figure 13) of different sizes: one representing the pGEMT-Easy vector (3018bp) and the other the MSP insert (about 1600bp).

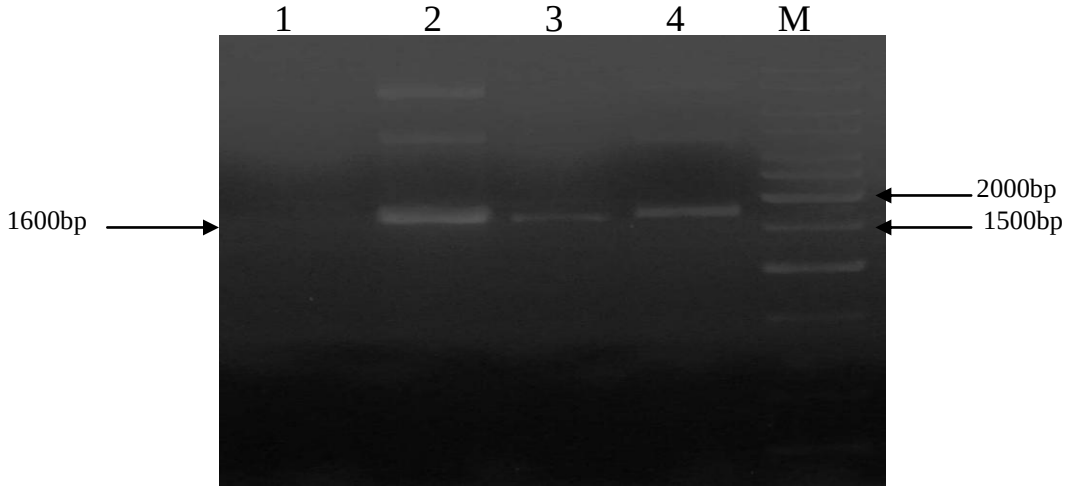

Figure 11: A 1% agarose gel for *MSP* samples 2-4 from *T. congolense* after restriction enzyme digestion with *BamHI* & *XhoI* restriction enzymes in order to release *MSP* from the parent vector, pGEMT-Easy (Promega, USA). A negative control is in the first lane and a 1kb ladder (Fermentas Life Sciences) in the last lane. As expected, a product of approximately 1600bp representing the MSP from *T. congolense* was released while the other product of approximately 3018bp represents the pGEMT vector.

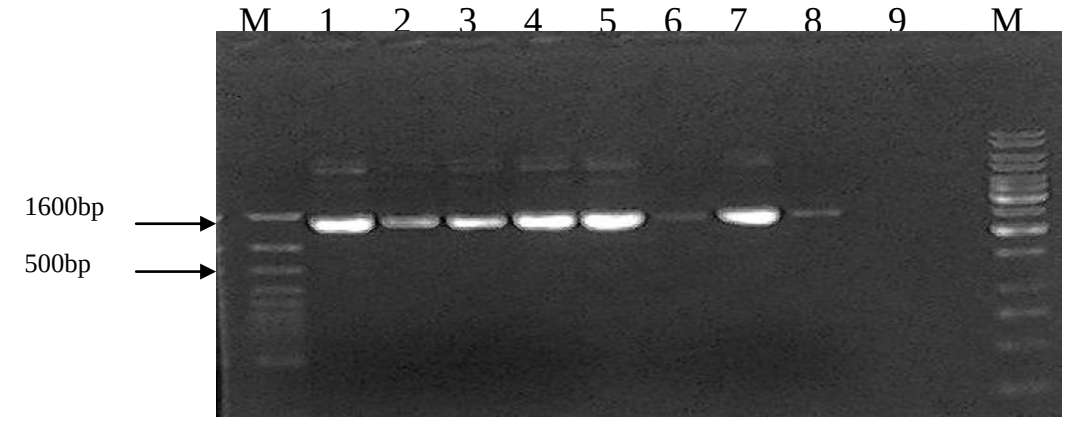

Figure 12: A 1% agarose gel to check pGEMT-*MSP* plasmids 1-8 after restriction digestion. A negative control (blank) was included in lane 9, a 100bp marker (Fermentas Life Sciences) in the first lane and a 1 kb ladder (Fermentas Life Sciences) on the last lane. As expected, a product of approximately 1600bp representing the MSP was obtained while the other product of approximately 3018bp represents the pGEMT vector.

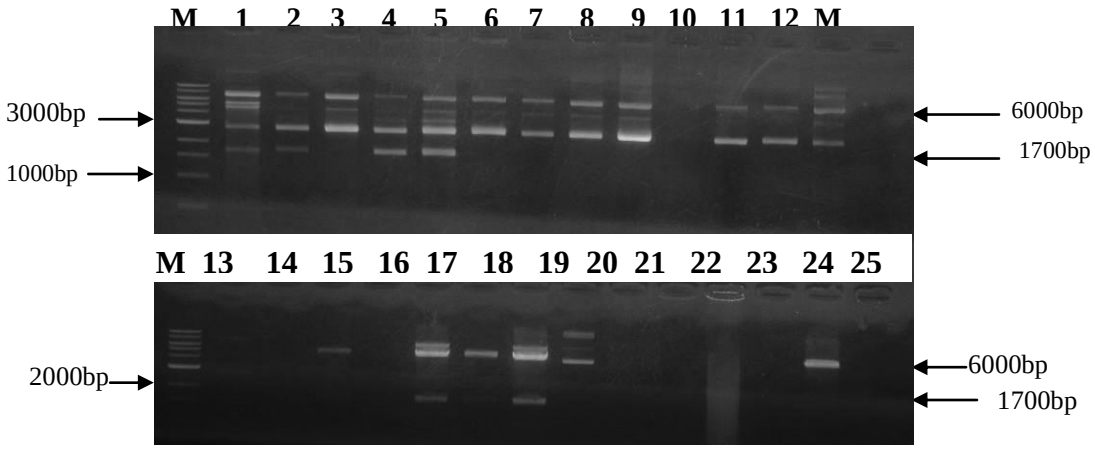

Figure 13: A 1% agarose gel for *MSP* samples 1-12 from *T. congolense* after restriction enzyme digestion with *BamHI* & *XhoI* to release *MSP* from the parent vector, pRSET-A (Invitrogen, USA). A 1kb ladder (NEB, UK) was included in the first lane. MSP products of approximately 1700bp were digested out as expected though not all plasmids seemed to have the insert. The other product of approximately 3000bp represents the pRSET-A vector. In the lower lane, Restriction digestion of the *T. brucei MSP*-pET28a plasmid samples 13-25 with restriction enzyme *NotI* yielded two products as expected: a 1700bp MSP product and a 6000bp pET28a vector product. A 1 kb Marker (NEB, UK) is in the first lane.

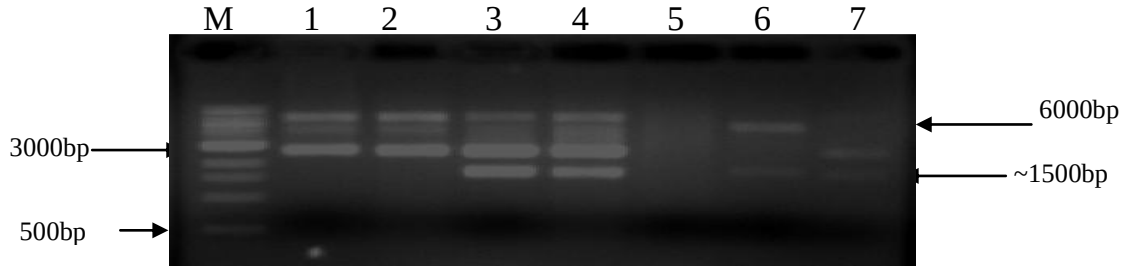

Figure 14: A 1% agarose gel for *MSP* samples 1-4 from *T. congolense* after a double restriction enzyme digestion reaction with *BamHI* & *XhoI* to release *MSP* from the parent vector, pRSET A (Invitrogen, USA) other *MSP* samples 6-7 from *T. brucei* after a restriction enzyme digestion with *NotI* to release *MSP* from the parent vector, pET28a (Novagen, USA). A negative control is in lane 5 and a 1kb ladder on the lane 1 (NEB, UK).

The other method used to screen the recombinant plasmids was through PCR. The vector-specific primers were used amplify the region of interest therefore indicating that the ligation process was successful as shown in the Figure 17 and 18 below. The insert obtained from the amplification process was approximately 1600bp long.

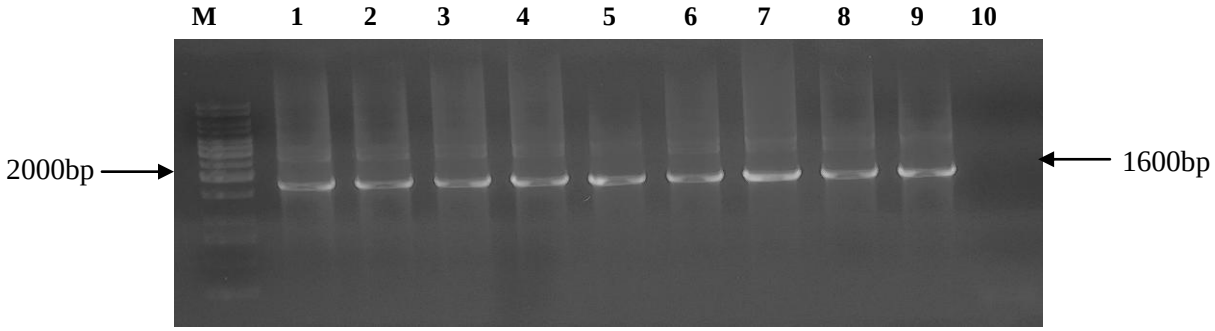

Figure 15: A 1% agarose gel of PCR products obtained after screening *T. brucei MSP*-pGEMT colonies labelled 1-9. A 1kb DNA ladder (Fermentas Inc., USA) was included in the first lane while a negative control was in the last lane. Products of approximately 1600bp were obtained as expected indicating that the MSP was present in the colonies screened.

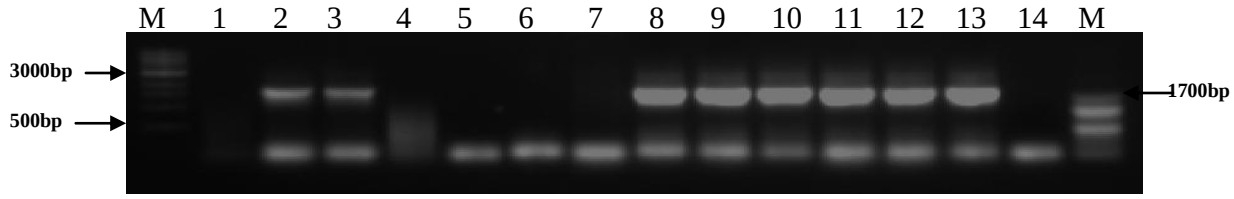

Figure 16: A 1% agarose gel to check for the presence of the insert in pRSET-A-*MSP* and pET28a-MSP plasmids after transformation of competent *E. coli* cells via PCR using insert-specific primers. Lanes 1-4 have the pET28a-MSP samples and 6-13 are pRSET-A-*MSP* samples. A negative control was included in lane 14. Also included on the gel is a 1 kb marker (NEB, UK) in the first lane and a 100bp ladder (NEB, UK) on the last lane. As expected, a product of approximately 1700bp representing the MSP was obtained.

4.5 Protein Analysis

4.5.1 SDS-PAGE Analysis

The result of a 12% SDS–PAGE showed a bigger and broad smear of the induced protein, which suggested that the polypeptide was induced. Separation of the proteins on SDS-PAGE gel revealed induced products of approximately 45kDa in size from both the *T. congolense* and *T. b. brucei* samples. The full-length polypeptides were predicted to be in the range of 60-70 kDa when fully expressed. However, this was 20kDa less the expected product size and indication that peptides of the MSP were expressed and not the full-length protein.

The samples were collected after 4 and 6 hours after induction with IPTG (Figure 17). The samples collected after six hours had a thicker band compared to the ones induce up to 4 hours an indication that induction was optimum after 6 hour. Induced MSP protein of approximately 45kDa was obtained from the *T. congolense* and *T. brucei* samples under study.

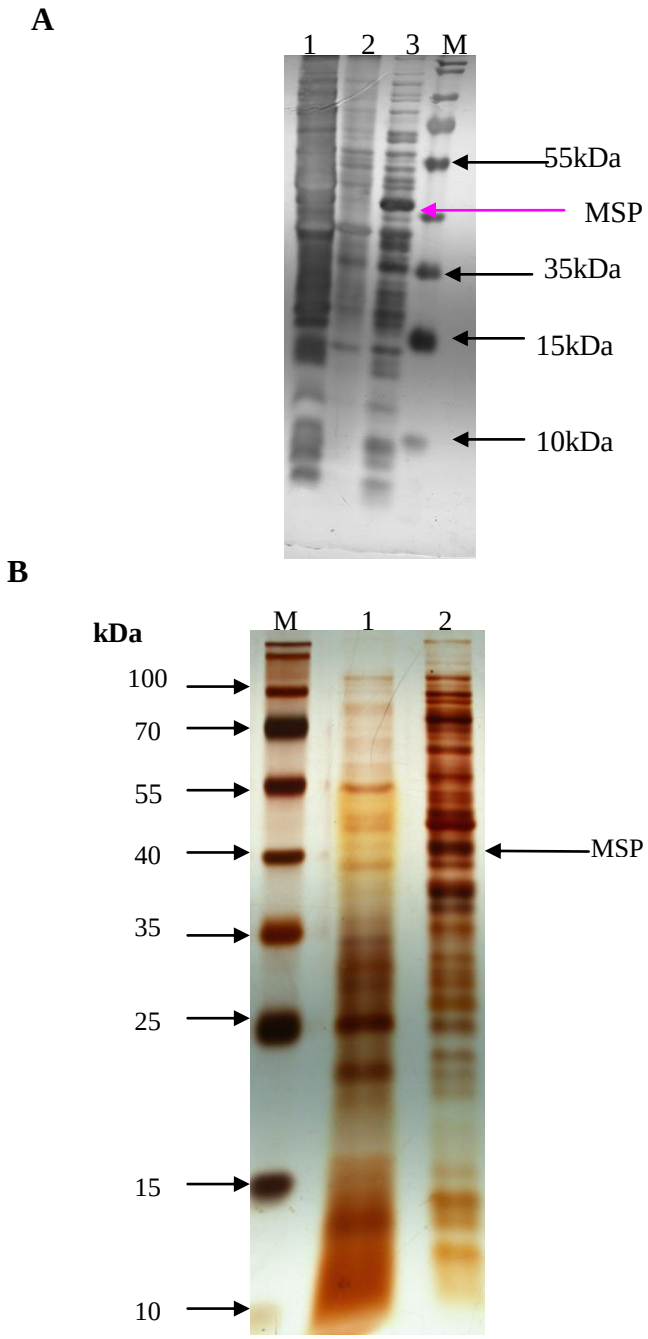

Figure 17: A): A 12% SDS-PAGE gel photo showing expression and induction of MSP 1 from *T. congolense* in BL21 DE3 *E. coli* cells after silver staining. On the first and second lanes are un-induced MSP sample, on lane 3 has the crude protein sample collected after 6hrs of induction with IPTG. On the last lane, (marked M) is the pre-stained Protein Marker (Fermentas AB, Lithuania). A broad band of approximately 45kDa marked MSP was obtained. **B)** A 12% SDS-PAGE gel photo showing expression and induction of MSP from *T. brucei* using the BL21 DE3 *E. coli* cells after silver staining. On the first lane (marked M) is the pre-stained Protein Marker (Fermentas AB, Lithuania), on the second lane is the un-induced crude protein while lane 3 has the crude protein sample collected after 4hrs of induction with IPTG. A broad band of approximately 45kDa marked MSP was obtained.

4.5.2 Protein Purification

The MSP proteins were purified on Ni-NTA Spin Columns which bind the 6X His-tag present in our fusion protein. The eluates were collected separately and concentrated though dialysis followed by precipitation in polyethylene glycol. A pure product of approximately 40kDa was obtained for both the *T. congolense* and *T. brucei* samples when analysed on a 4-20% SDS-PAGE gel (Figure 18).

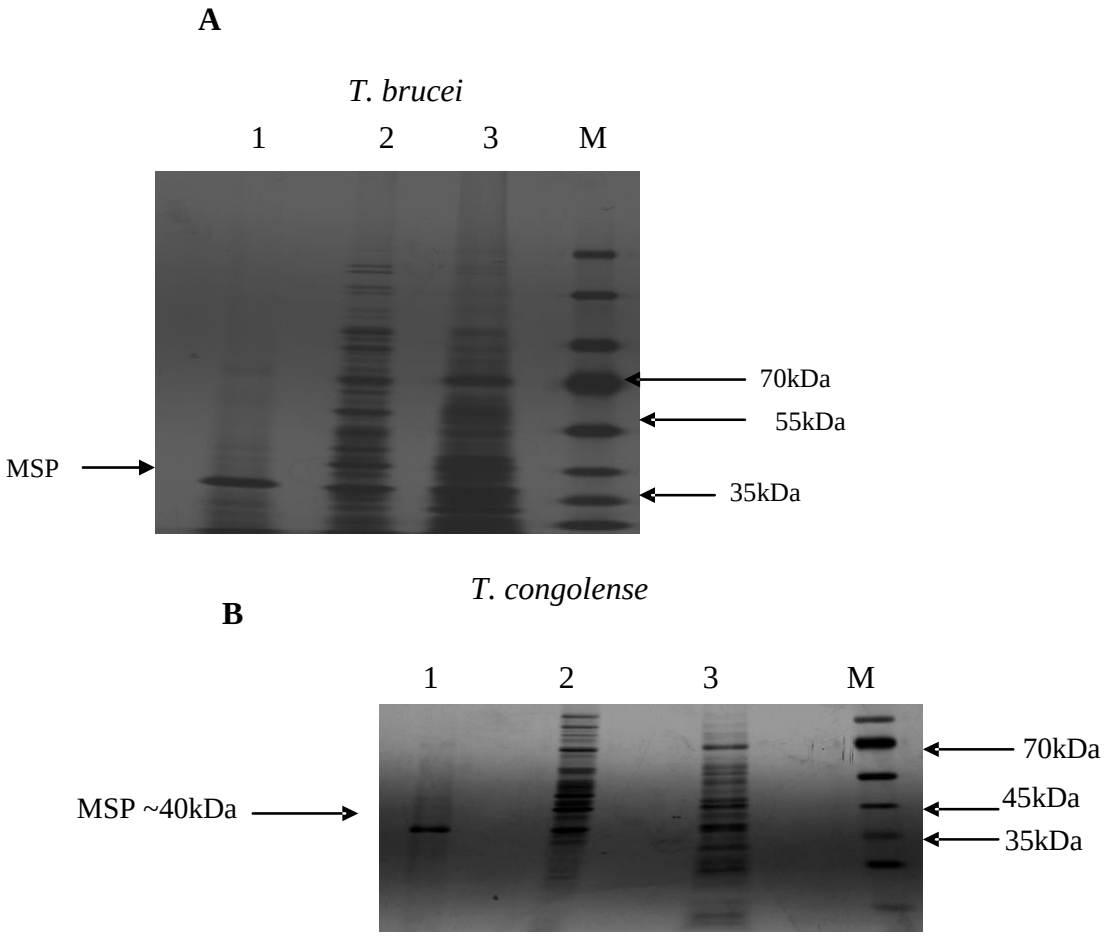

Figure 18: **A)** A 4-20% SDS-PAGE gel illustrating the product of Ni-NTA purification of *T. brucei* MSP protein. A pure product was shown on lane one, lanes 2 and 3 contain crude extracts of the protein prior to purification. A marker was included in the last lane. **B)** A 4-20% SDS-PAGE gel illustrating the product of Ni-NTA purification of *T. congolense* MSP protein. A pure product of approximately 40kDa was obtained (shown on lane one), lanes 2 and 3 contain crude extracts of the protein. A marker was included in the last lane (marked M).

4.6 Bioinformatics Analyses

4.6.1 Sequence Analysis

The DNA sequence fragments were obtained as chromatograms were viewed, edited and assembled into contiguous alignments on the BioEdit software (Hall, 1999). The DNA sequences were translated into the six possible reading frames and each sequence was inspected for long reading frames with an initial Methionine residue followed by at least 50 amino acid residues; these were submitted to the SignalP server for verification of a secretory signal peptide. The CAS (cDNA Annotation System) software was used to analyse the sequence data. The sequences were fed into the software, which consists of several modules that give an output with so much data about the sequences as shown in Tables 5 below.

The longest open reading frames were obtained with an initial Methionine residue followed by at least 50 amino acid residues; these were submitted to the SignalP server for verification of a secretory signal peptide. Of all the sequences analysed, the sequences analysed had indications of being related to membrane-anchored or cytoplasmic proteins, while the others gave no conclusive indication of a leader signal peptide, probably due to diminished sequence quality at the 5'- end. Notably, Fifty-nine (59) out of the sixty-one (61) sequences had Peptidase_M8 domains, characteristic of most metalloproteases such as the Leishmanolysin proteins. These sequences had significant matches to *T. congolense* and *T. brucei gambiense* and *Leishmania major* MSP (gp63) DNA sequences deposited at the GeneDB trypanosome database. When comparing all the MSP sequences known (retrieved as at October 20, 2008) with the complete MSP DNA sequences generated in the laboratory, using the BlastN flavour of Blast, 12 sequences had hits (matches) with a confidence value of 1E-18 or better, indicating they corresponded to the same or very

closely related DNA sequences. Eighteen sequences had significant matches to published Peptidase_M8 (PF01457 domain found in the Pfam protein database. Another dominant domain observed among most of the sequences was the ZnF_C2HC and ZnF_A20 which have a possible role in DNA binding (Gene ontology: 0003677), Zinc ion binding (Gene ontology: 0008270). The domain within the MSP amino acid sequence started at approximately position 140 and ended at position 163. Transmembrane and Peptidoglycan recognition protein (PGRP) domain domains were also obtained. The MSP sequencing output had more than 50% identity wit the *T. congolense* and *T. brucei* sequences that had been deposited at GenBank and GeneDB (www.genedb.org) that is hosted by the Sanger Institute (Pathogen Sequencing Unit). The final MSP nucleotide sequence comprised of fragments of lengths ranging between 900bp and 1400bp, and the predicted open-reading frames code for a protein of approximately 400 amino acids each with a predicted molecular mass of approximately 48kDa.

Table 5: A table of the CAS output containing the sequence data and the blast hits to several databases. The Expect (E) values are also shown alongside each database searched. SignalP software was also used to search for the signal peptides.

Number	Name of the Longest Sequence	Average Result	Best Match to GP63	E Value	Best Match to local DB	E Value	Best Match to Pfam	E Value
1	MSP10congo_Coast	IND					DUF1431, Protein of unknown function (DUF1431)	0.043
2	MSP11congo_Coast	CYT					Peptidase_M8, Leishmanolysin	4E-15
3	MSP4congo_Uganda	IND					Peptidase_M8, Leishmanolysin	1E-45
4	MSP13congo_Baringo	IND					Peptidase_M8, Leishmanolysin	3E-14
5	MSP53congo_2784	IND					Peptidase_M8, Leishmanolysin	2E-11
6	MSP55vivax_2501	IND					Peptidase_M8, Leishmanolysin	2E-33
7	MSP50brucei_CNyanza	SIG					Peptidase_M8, Leishmanolysin	3E-35
8	MSP46brucei_CNyanza	IND					Peptidase_M8, Leishmanolysin	4E-36
9	MSP42brucei_CNyanza	IND					Peptidase_M8, Leishmanolysin	3E-31
10	MSP33vivax_Coast	IND	tviv899g05.p1k_	2E-10	tviv899g05.p1k_8	2E-10	Peptidase_M8, Leishmanolysin	6E-36
11	MSP27vivax_Busia	IND					Peptidase_M8, Leishmanolysin	6E-34
12	MSP16congo_MMara	IND					Peptidase_M8, Leishmanolysin	2E-08
13	msp59congo_Uganda	IND					Peptidase_M8, Leishmanolysin	6E-19
14	MSP2congo_Uganda	IND	LbrM10_V2.165	5E-11	LbrM10_V2.1650	5E-11	Peptidase_M8, Leishmanolysin	4E-32
15	MSP14congo_Baringo	CYT					Peptidase_M8, Leishmanolysin	2E-15
16	MSP7congo_Ngurumani	IND					Peptidase_M8, Leishmanolysin	2E-36
17	msp57congo_Galana	CYT					Peptidase_M8, Leishmanolysin	1E-18
18	MSP51brucei_CNyanza	IND					Peptidase_M8, Leishmanolysin	9E-32
19	MSP29vivax_Busia	IND					Peptidase_M8, Leishmanolysin	2E-32
20	MSP8congo_Ngurumani	IND					Peptidase_M8, Leishmanolysin	4E-29
21	MSP52brucei_2831	IND					Peptidase_M8, Leishmanolysin	4E-32
22	MSP49brucei_CNyanza	IND					Peptidase_M8, Leishmanolysin	3E-28
23	MSP39vivax_Ngurumani	CYT					Peptidase_M8, Leishmanolysin	5E-31
24	msp58congo_CNyanza	IND					Peptidase_M8, Leishmanolysin	4E-23
25	MSP38vivax_Baringo	IND					Peptidase_M8, Leishmanolysin	1E-27
26	msp54congo_2773	IND					Peptidase_M8, Leishmanolysin	2E-23
27	MSP28vivax_Busia	IND	tviv899g05.p1k_	3E-12	tviv899g05.p1k_8	3E-12	Peptidase_M8, Leishmanolysin	1E-20
28	MSP37vivax_Samburu	IND	tviv899g05.p1k_	1E-14	tviv899g05.p1k_8	1E-14	Peptidase_M8, Leishmanolysin	2E-33
29	MSP32vivax_Galana	CYT	tviv899g05.p1k_	5E-11	tviv899g05.p1k_8	5E-11	Peptidase_M8, Leishmanolysin	1E-27
30	MSP26vivax_Meru	CYT					Peptidase_M8, Leishmanolysin	6E-30
31	MSP12congo_Kiboko	IND					Peptidase_M8, Leishmanolysin	3E-16
32	MSP1congo_Uganda	CYT					Peptidase_M8, Leishmanolysin	1E-44
33	MSP18congo_SNyanza	IND					Peptidase_M8, Leishmanolysin	4E-12
34	MSP9congo_Coast	IND					Peptidase_M8, Leishmanolysin	2E-11
35	MSP25vivax_MMara	CYT					Peptidase_M8, Leishmanolysin	1E-30
36	MSP3congo_Uganda	IND					Peptidase_M8, Leishmanolysin	1E-32
37	MSP17congo_SNyanza	IND					Peptidase_M8, Leishmanolysin	0.000004
38	msp19congo_Galana	CYT					MARCKS, MARCKS family	0.004
39	MSP5congo_Tanzania	IND					Peptidase_M8, Leishmanolysin	2E-31

***Continuation* of table 5**: The CAS output containing the sequence data and the blast hits to several databases such as GP63, Local and Pfam database. The Expect (E) values are also shown alongside each database searched. Signal peptides were also searched using the SignalP software.

Number	Name of the Longest Sequence	SignalP Average Result	Best Match to GP63	E Value	Best Match to local DB	E Value	Best Match to Pfam	E Value
40	MSP6congo_Tanzania	CYT	LbrM10_V2.157	1E-11			Peptidase_M8, Leishmanolysin	2E-33
41	MSP20congo_Galana	IND					Peptidase_M8, Leishmanolysin	4E-12
42	MSP34vivax_Coast	IND	tviv899g05.p1k_	2E-10			Peptidase_M8, Leishmanolysin	2E-26
43	MSP56brucei_3238	IND					Peptidase_M8, Leishmanolysin	3E-28
44	MSP36vivax_SNyanza	IND	tviv899g05.p1k_	2E-13			Peptidase_M8, Leishmanolysin	6E-29
45	MSP35vivax_SNyanza	CYT					Peptidase_M8, Leishmanolysin	5E-28
46	MSP21vivax_Uganda	IND					Peptidase_M8, Leishmanolysin	1E-20
47	MSP22vivax_Uganda	IND					Peptidase_M8, Leishmanolysin	1E-25
48	MSP31vivax_Galana	IND	tviv899g05.p1k_	5E-14			Peptidase_M8, Leishmanolysin	1E-22
49	MSP23vivax_Germany	IND					Peptidase_M8, Leishmanolysin	2E-13
50	MSP47brucei_CNyanza	CYT					Peptidase_M8, Leishmanolysin	4E-26
51	MSP44brucei_CNyanza	CYT					Peptidase_M8, Leishmanolysin	6E-10
52	MSP30vivax_Kiboko	IND					Peptidase_M8, Leishmanolysin	1E-20
53	MSP43brucei_CNyanza	CYT					Peptidase_M8, Leishmanolysin	0.006
54	MSP45brucei_CNyanza	IND			brucei gambiense|Tbgamb	0	Peptidase_M8, Leishmanolysin	1E-24
55	msp60congo_Baringo	IND					Peptidase_M8, Leishmanolysin	0.00006
56	MSP15congo_MMara	IND					Peptidase_M8, Leishmanolysin	1E-09
57	MSP24vivax_MMara	IND					TolA, TolA protein	0.0004
58	MSP40vivax_Ngurumani	CYT					TolA, TolA protein	0.0007
59	MSP41brucei_CNyanza	IND			brucei|Tb11	1E-122	Peptidase_M8, Leishmanolysin	0.0000003
60	MSP48brucei_CNyanza	CYT	brucei gambiens	0	brucei|Tb11	2E-69	Peptidase_M8, Leishmanolysin	0.00005
61	MSP79cogolense	SIG	congo843b03.q1	0	brucei|Tb11	6E-17	Peptidase_M8, Leishmanolysin	1E-132

Cont., **Table 5:** CAS output containing the sequence data and the blast hits to several databases. The Expect (E) values are also shown alongside each database searched.

Number	Name of the Longest Sequence	Blast Output for KOG	E Value	Best Match to SMART	E Value
1	MSP10congo_Coast	FGF receptor activating protein 1	0.064	DM6, Cysteine-rich domain currently specific to Drosophila	0.017
2	MSP11congo_Coast	Helicase of the DEAD superfamily	0.13	RasGAP, GTPase-activator protein for Ras-like GTPases; All alpha	0.004
3	MSP4congo_Uganda	FERM domain protein EHM2	0.074	THN, Thaumatin family; The thaumatin family gathers proteins rel	0.074
4	MSP13congo_Baringo	Thyroid hormone receptor-associated protei	0.1	PI3Ka, Phosphoinositide 3-kinase family, accessory domain (PIK d	0.005
5	MSP53congo_2784	Ca2+/H+ antiporter VCX1 and related prot	0.065	TSP1, Thrombospondin type 1 repeats; Type 1 repeats in thrombos	0.037
6	MSP55vivax_2501	Leishmanolysin-like peptidase (Peptidase M	0.0000004	ZnMc, Zinc-dependent metalloprotease; Neutral zinc metallopeptic	0.0004
7	MSP50brucei_CNyanza	Leishmanolysin-like peptidase (Peptidase M	0.000001	DM6, Cysteine-rich domain currently specific to Drosophila	0.009
8	MSP46brucei_CNyanza	Leishmanolysin-like peptidase (Peptidase M	2E-08	ZnMc, Zinc-dependent metalloprotease; Neutral zinc metallopeptic	0.014
9	MSP42brucei_CNyanza	Leishmanolysin-like peptidase (Peptidase M	2E-08	ZnMc, Zinc-dependent metalloprotease; Neutral zinc metallopeptic	0.014
10	MSP33vivax_Coast	Leishmanolysin-like peptidase (Peptidase M	2E-08	ZnMc, Zinc-dependent metalloprotease; Neutral zinc metallopeptic	0.0009
11	MSP27vivax_Busia	Leishmanolysin-like peptidase (Peptidase M	0.0000004	ZnMc, Zinc-dependent metalloprotease; Neutral zinc metallopeptic	0.0002
12	MSP16congo_MMara	Nuclear receptors of the nerve growth facto	0.006	MeTrc, Methyltransferase, chemotaxis proteins ; Methylates methy	0.24
13	msp59congo_Uganda	SWAP mRNA splicing regulator	0.001	FU, Furin-like repeats;	0.018
14	MSP2congo_Uganda	Activating transcription factor 4	0.47	Glyco_10, Glycosyl hydrolase family 10;	0.65
15	MSP14congo_Baringo	Aminopeptidases of the M20 family	0.19	RhoGAP, GTPase-activator protein for Rho-like GTPases; GTPase	0.097
16	MSP7congo_Ngurumani	Nidogen and related basement membrane p	0.022	RhoGAP, GTPase-activator protein for Rho-like GTPases; GTPase	0.053
17	msp57congo_Galana	Beta-dystrobrevin	0.007	BRLZ, basic region leucin zipper;	0.091
18	MSP51brucei_CNyanza	Thyroid hormone receptor-associated protei	0.0000001	PSN, Presenilin, signal peptide peptidase, family; Presenilin 1 and	0.00004
19	MSP29vivax_Busia	Thyroid hormone receptor-associated protei	5E-08	TLC, TRAM, LAG1 and CLN8 homology domains	7E-09
20	MSP8congo_Ngurumani	HMG-box transcription factor	0.056	DM6, Cysteine-rich domain currently specific to Drosophila	0.039
21	MSP52brucei_2831	Leishmanolysin-like peptidase (Peptidase M	0.00005	DM6, Cysteine-rich domain currently specific to Drosophila	0.002
22	MSP49brucei_CNyanza	Leishmanolysin-like peptidase (Peptidase M	0.00004	ZnMc, Zinc-dependent metalloprotease; Neutral zinc metallopeptic	0.008
23	MSP39vivax_Ngurumani	Leishmanolysin-like peptidase (Peptidase M	0.00008	ZnMc, Zinc-dependent metalloprotease; Neutral zinc metallopeptic	0.0002
24	msp58congo_CNyanza	Uncharacterized conserved protein	0.011	CLa, CLUSTERIN alpha chain;	0.051
25	MSP38vivax_Baringo	Leishmanolysin-like peptidase (Peptidase M	0.00007	ZnMc, Zinc-dependent metalloprotease; Neutral zinc metallopeptic	0.0006
26	msp54congo_2773	Uncharacterized conserved protein	0.002	DM8, Repeats found in several Drosophila proteins	0.029
27	MSP28vivax_Busia	Histone H1	1E-09	TLC, TRAM, LAG1 and CLN8 homology domains	6E-08
28	MSP37vivax_Samburu	Leishmanolysin-like peptidase (Peptidase M	0.001	ZnMc, Zinc-dependent metalloprotease; Neutral zinc metallopeptic	0.0001
29	MSP32vivax_Galana	Leishmanolysin-like peptidase (Peptidase M	0.00003	ZnMc, Zinc-dependent metalloprotease; Neutral zinc metallopeptic	0.0007
30	MSP26vivax_Meru	Leishmanolysin-like peptidase (Peptidase M	0.002	ZnMc, Zinc-dependent metalloprotease; Neutral zinc metallopeptic	0.0007
31	MSP12congo_Kiboko	p21-activated serine/threonine protein kinas	0.04	RHO, Rho (Ras homology) subfamily of Ras-like small GTPases; I	0.044
32	MSP1congo_Uganda	RNA-binding protein LARK, contains RRM	0.16	WSC, present in yeast cell wall integrity and stress response compo	0.008
33	MSP18congo_SNyanza	Chromatin remodeling complex SWI/SNF,	0.049	DM, Doublesex DNA-binding motif;	0.089
34	MSP9congo_Coast	Pyridoxamine-phosphate oxidase	0.009	AMOP, Adhesion-associated domain present in MUC4 and other p	0.024
35	MSP25vivax_MMara	Chromatin assembly factor-I	0.0004	ZnMc, Zinc-dependent metalloprotease; Neutral zinc metallopeptic	0.0002
36	MSP3congo_Uganda	Predicted membrane protein	0.081	TSPN, Thrombospondin N-terminal -like domains	0.088
37	MSP17congo_SNyanza	Cell surface glycoprotein STIM, contains S.	0.2	BAR, BAR domain	0.28
38	msp19congo_Galana	Splicing factor U2AF, large subunit (RRM	0.003	BRLZ, basic region leucin zipper;	0.044
39	MSP5congo_Tanzania	Predicted DNA-binding protein, contains S.	0.091	HTH_MARR, helix_turn_helix multiple antibiotic resistance prote	0.077

Cont., **Table 5:** of the CAS output containing the sequence data and the blast hits to several databases. The Expect (E) values are also shown alongside each database searched.

Number	Name of the Longest Sequence	Blast Output for KOG	E Value	Best Match to SMART	E Value
40	MSP6congo_Tanzania	Rhomboid family proteins	0.23	WSC, present in yeast cell wall integrity and stress response compc	0.076
41	MSP20congo_Galana	Transcription initiation factor TFIID, subun	0.088	DM6, Cysteine-rich domain currently specific to Drosophila	0.034
42	MSP34vivax_Coast	Ultrahigh sulfur keratin-associated protein	0.0003	ZnMc, Zinc-dependent metalloprotease; Neutral zinc metallopeptic	0.0005
43	MSP56brucei_3238	Leishmanolysin-like peptidase (Peptidase M	0.0002	CBF, CCAAT-Binding transcription Factor;	0.006
44	MSP36vivax_SNyanza	Fibrillins and related proteins containing Ca	0.0002	ZnMc, Zinc-dependent metalloprotease; Neutral zinc metallopeptic	0.0001
45	MSP35vivax_SNyanza	Ferric reductase-like proteins	0.00002	ZnMc, Zinc-dependent metalloprotease; Neutral zinc metallopeptic	0.0004
46	MSP21vivax_Uganda	Transcription factor Abd-B, contains HOX	0.00007	ZnMc, Zinc-dependent metalloprotease; Neutral zinc metallopeptic	0.0003
47	MSP22vivax_Uganda	HMG-box transcription factor	0.014	ZnMc, Zinc-dependent metalloprotease; Neutral zinc metallopeptic	0.0006
48	MSP31vivax_Galana	Thyroid hormone receptor-associated protei	0.014	ZnMc, Zinc-dependent metalloprotease; Neutral zinc metallopeptic	0.0001
49	MSP23vivax_Germany	Nidogen and related basement membrane pı	0.034	ZnMc, Zinc-dependent metalloprotease; Neutral zinc metallopeptic	0.0004
50	MSP47brucei_CNyanza	Nucleolar protein-like/EBNA1-binding prot	0.001	ZnMc, Zinc-dependent metalloprotease; Neutral zinc metallopeptic	0.01
51	MSP44brucei_CNyanza	RNA polymerase II, large subunit	0.001	ZnMc, Zinc-dependent metalloprotease; Neutral zinc metallopeptic	0.0004
52	MSP30vivax_Kiboko	Ultrahigh sulfur keratin-associated protein	0.0001	ZnMc, Zinc-dependent metalloprotease; Neutral zinc metallopeptic	0.001
53	MSP43brucei_CNyanza	Myosin class I heavy chain	0.059	GAL4, GAL4-like Zn(II)2Cys6 (or C6 zinc) binuclear cluster DNA	0.15
54	MSP45brucei_CNyanza	Ultrahigh sulfur keratin-associated protein	0.0003	DM6, Cysteine-rich domain currently specific to Drosophila	0.09
55	msp60congo_Baringo	Predicted alkaloid synthase/Surface mucin l	0.36	FU, Furin-like repeats;	0.29
56	MSP15congo_MMara	Thyroid hormone receptor-associated protei	0.008	PTPc_motif, Protein tyrosine phosphatase, catalytic domain motif;	0.036
57	MSP24vivax_MMara	Transcriptional corepressor Atrophin-1/DRl	0.016	ZnMc, Zinc-dependent metalloprotease; Neutral zinc metallopeptic	0.0006
58	MSP40vivax_Ngurumani	Kinesin-like protein	0.001	ZnMc, Zinc-dependent metalloprotease; Neutral zinc metallopeptic	0.0003
59	MSP41brucei_CNyanza	17 beta-hydroxysteroid dehydrogenase type	0.64	ZnMc, Zinc-dependent metalloprotease; Neutral zinc metallopeptic	0.002
60	MSP48brucei_CNyanza	Proteasome formation inhibitor PI31	0.092	ZnMc, Zinc-dependent metalloprotease; Neutral zinc metallopeptic	0.041
61	MSP79cogolense	Leishmanolysin-like peptidase (Peptidase M	7E-41	DM6, Cysteine-rich domain currently specific to Drosophila	0.022

4.6.1.1 Domains and Motifs Search

Peptidase_M8 domains were observed among most of the sequences studied and these were characteristic of the metalloproteinase family of proteins. Another dominant domain observed among most of the sequences was the Zn_Protease. Transmembrane regions were also found.

The graphical outputs obtained from the Integrated Protein (InterPro) database while searching for the domains in MSP are shown in Figure 20. The domains obtained are indicated by a coloured bar and an identity number depending on the database it was retrieved from and the name of the domain. For example in Figure 20 below, some of the domains obtained for the illustrated examples are: a Peptidase_M8 domain (InterPro ID: IPR001577), Zn_Protease domain (ID PS00142, metalloproteinases (‘zincins’) catalytic domains (ID: SSF55486) Transmembrane_regions were also obtained. The Interpro database contains several databases such as Protein family (Pfam),

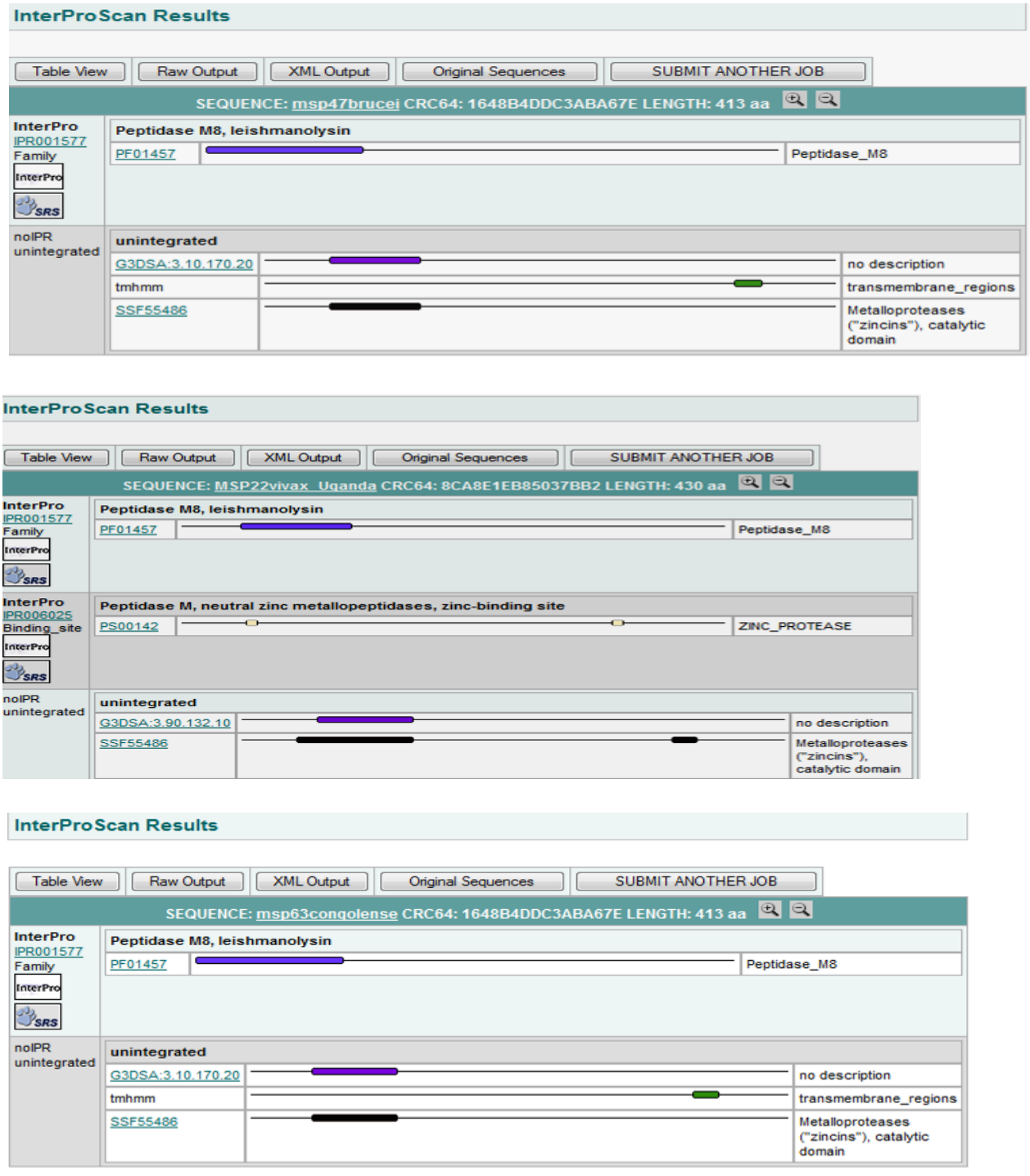

Figure 19: An example of the InterProScan output for the MSP47, *T. brucei*, MSP 22 and *T. vivax* MSP63 *T. congolense* protein sequences. The expected domain signatures were obtained using this method for all the MSP protein sequences in this study.

The protein sequences were subjected to an InterproScan database search in order to identify the domains and motifs present on these sequences. The different protein patterns, motifs and domains obtained for each of the MSP sequences analysed are tabulated below with the respective database with which it was compared; identity numbers and the Expect (E) values (see Table 6). In order to understand the approach and abbreviations used in Table 6, a detailed explanation of each database and function is elucidated. In InterProScan, the FPrintScan database scans against the fingerprints in the PRINTS database. These fingerprints are groups of motifs that together are more potent than single motifs by making use of the biological context inherent in a multiple motif method. HMMPfam scans the hidden markov models (HMMs) that are present in the PFAM (Protein families) database. HMMSmart scans the HMMs that are present in the SMART domain/domain families' database. ScanRegExp scans against the regular expressions in the PROSITE protein families and domains database while SuperFamily is a library of profile hidden Markov models that represent all proteins of known structure.

Table 6: A list of the different domains observed in the MSP protein sequences on InterProScan and their respective Expect (E) values, Identity numbers and descriptions.

	Query ID	Query Length	Database Name	ID	Description	E-value
1	MSP41brucei_CNyanza	425	HMMPfam	PF01457	Peptidase_M8	4.00E-10
		425	superfamily	SSF55486	Metalloproteases ("zincins"), cata	1.40E-08
		425	ScanRegExp	PS00142	ZINC_PROTEASE	NA
2	MSP42brucei_CNyanza	443	HMMPfam	PF01457	Peptidase_M8	6.90E-42
			FPrintScan	PR00782	LSHMANOLYSIN	1.20E-23
			FPrintScan	PR00782	LSHMANOLYSIN	1.20E-23
			FPrintScan	PR00782	LSHMANOLYSIN	1.20E-23
			superfamily	SSF55486	Metalloproteases ("zincins"), cata	1.50E-27
			ScanRegExp	PS00142	ZINC_PROTEASE	NA
			HMMPanther	PTHR10942	PROTEASE FAMILY M8 LEIS	5.50E-11
3	MSP43brucei_CNyanza	438	TMHMM	tmhmm	transmembrane_regions	NA
			ScanRegExp	PS00142	ZINC_PROTEASE	NA
4	MSP44brucei_CNyanza	422	HMMPfam	PF01457	Peptidase_M8	3.50E-08
			HMMPfam	PF01457	Peptidase_M8	4.90E-15
			FPrintScan	PR00782	LSHMANOLYSIN	1.20E-06
			FPrintScan	PR00782	LSHMANOLYSIN	1.20E-06
			superfamily	SSF55486	Metalloproteases ("zincins"), cata	1.80E-14
			ScanRegExp	PS00142	ZINC_PROTEASE	NA
5	MSP45brucei_CNyanza	450	superfamily	SSF55486	Metalloproteases ("zincins"), cata	5.20E-13
			superfamily	SSF55486	Metalloproteases ("zincins"), cata	0.0023
			Gene3D	G3DSA:3.90.132.1	no description	0.00046
			HMMPfam	PF01457	Peptidase_M8	2.70E-23
			SignalPHMM	SignalP	signal-peptide	NA
			ScanRegExp	PS00142	ZINC_PROTEASE	NA
6	MSP46brucei_Cnyanza	439	superfamily	SSF55486	Metalloproteases ("zincins"), cata	5.10E-35
			ScanRegExp	PS00142	ZINC_PROTEASE	NA
			HMMPfam	PF01457	Peptidase_M8	1.20E-47
			HMMPanther	PTHR10942	PROTEASE FAMILY M8 LEIS	6.60E-14
			TMHMM	tmhmm	transmembrane_regions	NA
7	MSP47brucei_CNyanza	447	superfamily	SSF55486	Metalloproteases ("zincins"), cata	2.50E-24
			ScanRegExp	PS00142	ZINC_PROTEASE	NA
			TMHMM	tmhmm	transmembrane_regions	NA
			HMMPfam	PF01457	Peptidase_M8	3.80E-08
8	MSP48brucei_CNyanza	401	HMMPfam	PF01457	Peptidase_M8	3.30E-06
9	MSP49brucei_CNyanza	426	HMMPanther	PTHR10942	PROTEASE FAMILY M8 LEIS	4.70E-07
			TMHMM	tmhmm	transmembrane_regions	NA
			FPrintScan	PR00782	LSHMANOLYSIN	9.40E-18
			Gene3D	G3DSA:3.90.132.1	no description	1.10E-15
			ScanRegExp	PS00142	ZINC_PROTEASE	NA
			FPrintScan	PR00782	LSHMANOLYSIN	9.40E-18
			SignalPHMM	SignalP	signal-peptide	NA
			HMMPfam	PF01457	Peptidase_M8	2.70E-05
			superfamily	SSF55486	Metalloproteases ("zincins"), cata	6.50E-28
10	MSP50brucei_CNyanza	443	HMMPfam	PF01457	Peptidase_M8	6.30E-45
			TMHMM	tmhmm	transmembrane_regions	NA
			HMMPfam	PF01457	Peptidase_M8	6.30E-45
			ScanRegExp	PS00142	ZINC_PROTEASE	NA
			HMMPanther	PTHR10942	PROTEASE FAMILY M8 LEIS	2.50E-11
			superfamily	SSF55486	Metalloproteases ("zincins"), cata	2.60E-27
			SignalPHMM	SignalP	signal-peptide	NA
			FPrintScan	PR00782	LSHMANOLYSIN	5.70E-24

Table 6 continued next page.

A continuation of Table 6

11	MSP51brucei_CNyanza	497	FPrintScan	PR00782	LSHMANOLYSIN	7.50E-17
			ScanRegExp	PS00142	ZINC_PROTEASE	NA
			HMMPanther	PTHR10942	PROTEASE FAMILY M8 LEIS	2.10E-07
			superfamily	SSF55486	Metalloproteases ("zincins"), cata	5.80E-26
			TMHMM	tmhmm	transmembrane_regions	NA
			HMMPfam	PF01457	Peptidase_M8	7.10E-34
12	MSP52brucei_2831	452	HMMPfam	PF01457	Peptidase_M8	1.90E-40
			HMMPfam	PF01457	Peptidase_M8	1.90E-40
			superfamily	SSF55486	Metalloproteases ("zincins"), cata	2.10E-26
			HMMPanther	PTHR10942	PROTEASE FAMILY M8 LEIS	1.20E-08
			PatternScan	PS00094	C5_MTASE_1	NA
			ScanRegExp	PS00094	C5_MTASE_1	NA
			ScanRegExp	PS00142	ZINC_PROTEASE	NA
13	MSP56brucei_3238	426	HMMPfam	PF01457	Peptidase_M8	3.60E-05
			ScanRegExp	PS00142	ZINC_PROTEASE	NA
			ScanRegExp	PS00142	ZINC_PROTEASE	NA
			FPrintScan	PR00782	LSHMANOLYSIN	6.40E-14
			HMMPanther	PTHR10942	PROTEASE FAMILY M8 LEIS	0.00016
			superfamily	SSF55486	Metalloproteases ("zincins"), cata	7.10E-28
14	MSP1congo_Uganda	440	superfamily	SSF55486	Metalloproteases ("zincins"), cata	1.30E-20
			SignalPHMM	SignalP	signal-peptide	NA
			superfamily	SSF55486	Metalloproteases ("zincins"), cata	1.30E-20
			ScanRegExp	PS00142	ZINC_PROTEASE	NA
			HMMPfam	PF01457	Peptidase_M8	1.70E-45
15	MSP2congo_Uganda	426	HMMPfam	PF01457	Peptidase_M8	3.40E-07
			superfamily	SSF55486	Metalloproteases ("zincins"), cata	3.50E-17
			ScanRegExp	PS00142	ZINC_PROTEASE	NA
			FPrintScan	PR00782	LSHMANOLYSIN	6.80E-12
16	MSP3congo_Uganda	434	superfamily	SSF55486	Metalloproteases ("zincins"), cata	2.00E-16
			ScanRegExp	PS00142	ZINC_PROTEASE	NA
			TMHMM	tmhmm	transmembrane_regions	NA
			HMMPfam	PF01457	Peptidase_M8	5.70E-05
17	MSP4congo_Uganda	445	HMMPfam	PF01457	Peptidase_M8	9.10E-28
			TMHMM	tmhmm	transmembrane_regions	NA
			superfamily	SSF55486	Metalloproteases ("zincins"), cata	2.60E-24
			SignalPHMM	SignalP	signal-peptide	NA
			ScanRegExp	PS00142	ZINC_PROTEASE	NA
18	MSP5congo_Tanzania	439	HMMPfam	PF01457	Peptidase_M8	4.00E-06
19	MSP5congo_Tanzania	439	HMMPfam	PF01457	Peptidase_M8	4.00E-06
			HMMPfam	PF01457	Peptidase_M8	2.80E-16
			FPrintScan	PR00782	LSHMANOLYSIN	1.80E-08
			FPrintScan	PR00782	LSHMANOLYSIN	1.80E-08
			superfamily	SSF55486	Metalloproteases ("zincins"), cata	4.10E-10
			SignalPHMM	SignalP	signal-peptide	NA
			ScanRegExp	PS00142	ZINC_PROTEASE	NA
20	MSP21vivax_Uganda	428	superfamily	SSF55486	Metalloproteases ("zincins"), cata	2.20E-18
			HMMPfam	PF01457	Peptidase_M8	NA
			ScanRegExp	PS00142	ZINC_PROTEASE	8.60E-19
21	MSP22vivax_Uganda	430	superfamily	SSF55486	Metalloproteases ("zincins"), cata	2.00E-18
			HMMPfam	PF01457	Peptidase_M8	1.40E-28
			Gene3D	G3DSA:3.90.132.1	no description	1.80E-06
			ScanRegExp	PS00142	ZINC_PROTEASE	NA
22	MSP23vivax_Germany	451	ScanRegExp	PS00142	ZINC_PROTEASE	NA
			HMMPfam	PF01457	Peptidase_M8	7.80E-20
			FPrintScan	PR00782	LSHMANOLYSIN	5.10E-12
			superfamily	SSF55486	Metalloproteases ("zincins"), cata	1.70E-18
23	MSP24vivax_MMara	413	superfamily	SSF55486	Metalloproteases ("zincins"), cata	0.00016
			HMMPfam	PF01457	Peptidase_M8	1.70E-06
			ScanRegExp	PS00142	ZINC_PROTEASE	NA
24	MSP25vivax_MMara	434	superfamily	SSF55486	Metalloproteases ("zincins"), cata	9.90E-24
25	MSP25vivax_MMara	434	Gene3D	G3DSA:3.90.132.1	no description	5.20E-12
			ScanRegExp	PS00142	ZINC_PROTEASE	NA
			HMMPfam	PF01457	Peptidase_M8	2.70E-07

4.6.1.2 Multiple Sequence Alignment

The MSP DNA sequences were aligned using the TCOFFEE software under the default parameters. The output is illustrated in Figure 20 below, whereby the conserved regions being the highlighted in red colour. The less conserved sequences are highlighted with an orange and green colour while the poorly aligned sequences are shaded blue.

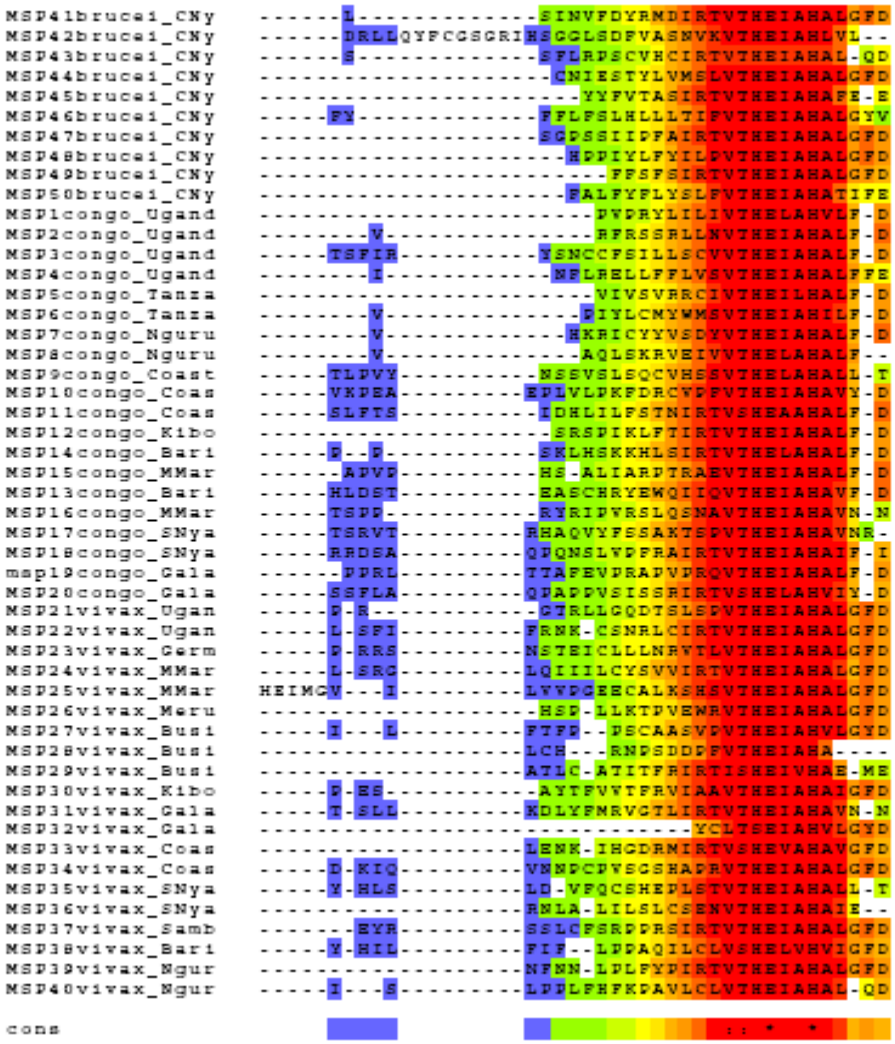

Figure 20: An illustration of the multiple sequence alignment of representative protein sequences from *T. brucei*, *T. vivax* and *T. congolense* MSP. The TCOFFEE software was used to produce this alignment. The conserved regions are highlighted in red colour; the less conserved sequences are highlighted with an orange and green colour while the poorly aligned sequences are shaded blue.

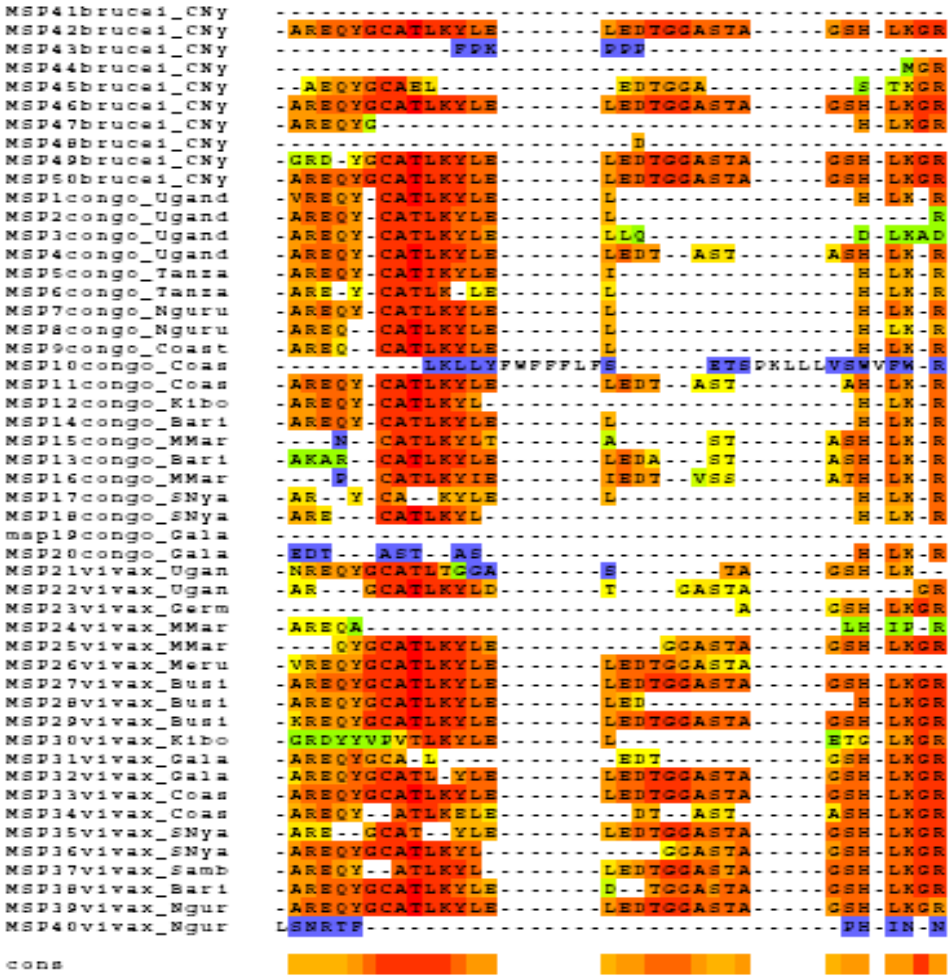

Figure 21: An illustration of the multiple sequence alignment of representative protein sequences from *T. brucei*, *T. vivax* and *T. congolense* MSP. The TCOFFEE software was used to produce this alignment. The conserved regions are highlighted in red colour; the less conserved sequences are highlighted with an orange and green colour while the poorly aligned sequences are shaded blue.

4.6.1.3 Sequence Variability

A comparison of the individual sequences amongst themselves reveals that there is greater similarity between the *T. brucei*, *T. vivax* and *T. congolense* MSP in certain regions as shown in Figure 22. However besides these regions the rest of the alignment is very diverse amongst all the

three species studied. The alignment between *T. vivax* and *T. congolense* MSP sequences show great differences between the two species.

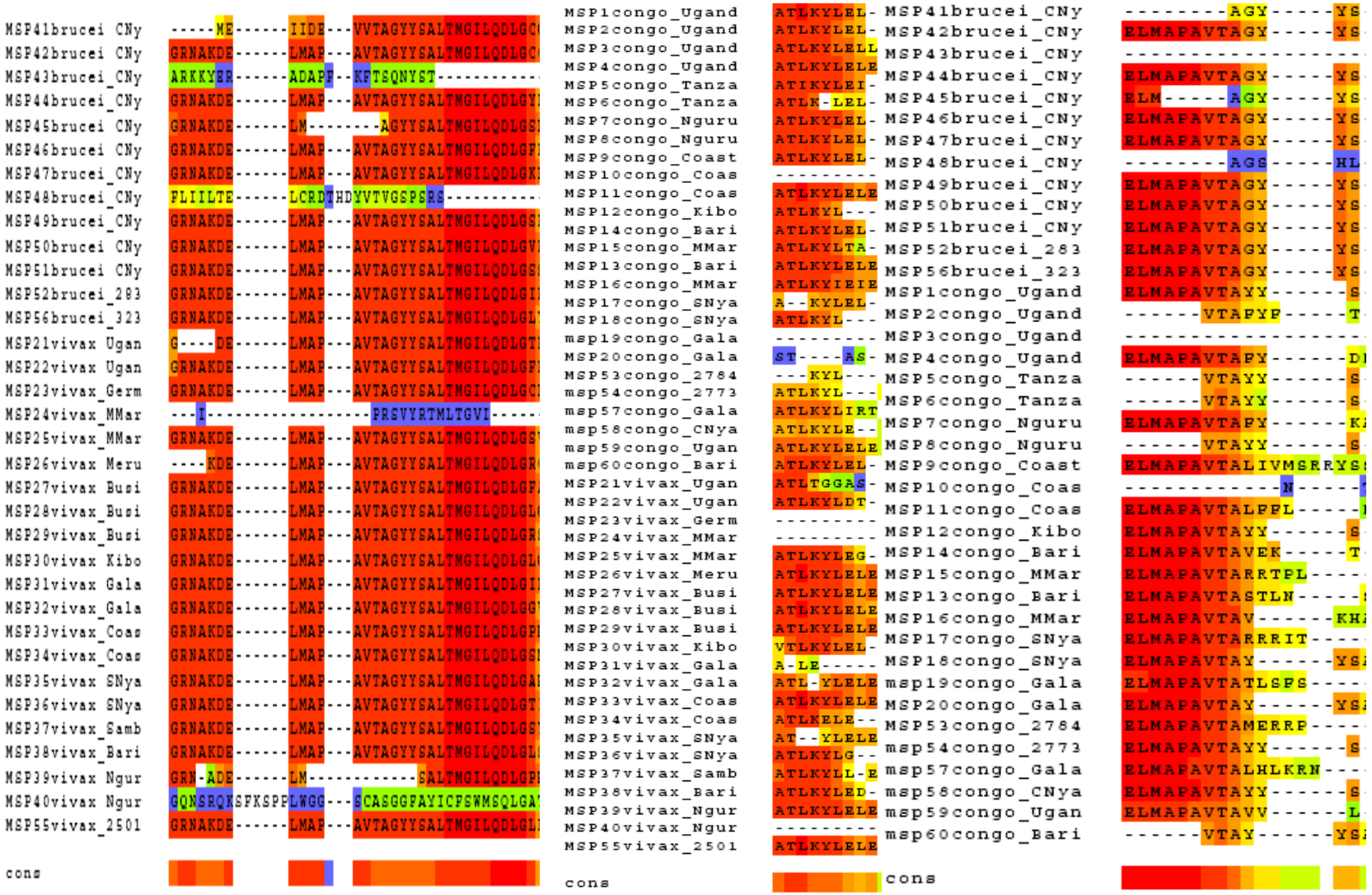

Figure 22: An illustration of the different alignments obtained when the trypanosome species under study are compared (aligned) two at a time instead of all three at once. The result of this is that *T. brucei* samples show good alignments (coloured red) with either *T. congolense* or *T. vivax* than *T. vivax* and *T. congolense* together do.

4.6.2 MSP Evolutionary Analyses

MSP sequences were aligned and phylogenetic trees reconstructed using the neighbour-joining (NJ) method 9 (Saitou and Nei 1987). The tree obtained from this alignment is supported by strong bootstrap values. The tree shows that the MSP sequences clade together according to their species regardless of their origin. The *T. brucei* MSP sequences clade together as well as those from *T. congolense* and *T. vivax*. There are a few stray MSP sequences from *T. brucei* and *T. vivax* that clustered with the other species as illustrated in the phylogenetic tree in Figure 24, though they are supported by very weak bootstrap values. The MSP sequence from the well studied *Leishmania major* (Systematic Name: LmjF10.0470) was used as an out-group and also to root the gene tree (Figure 24). The MSPs tend to clustered based mainly on the species types. Noticeably, the *T. brucei* MSP samples from Central Nyanza in Kenya had a propensity to cluster together in one clade. The geographical predisposition of these trypanosomes does not seem to have an impact on the clustering hence extrapolate that these proteins are conserved all through and retain their function regardless of their geographical location.

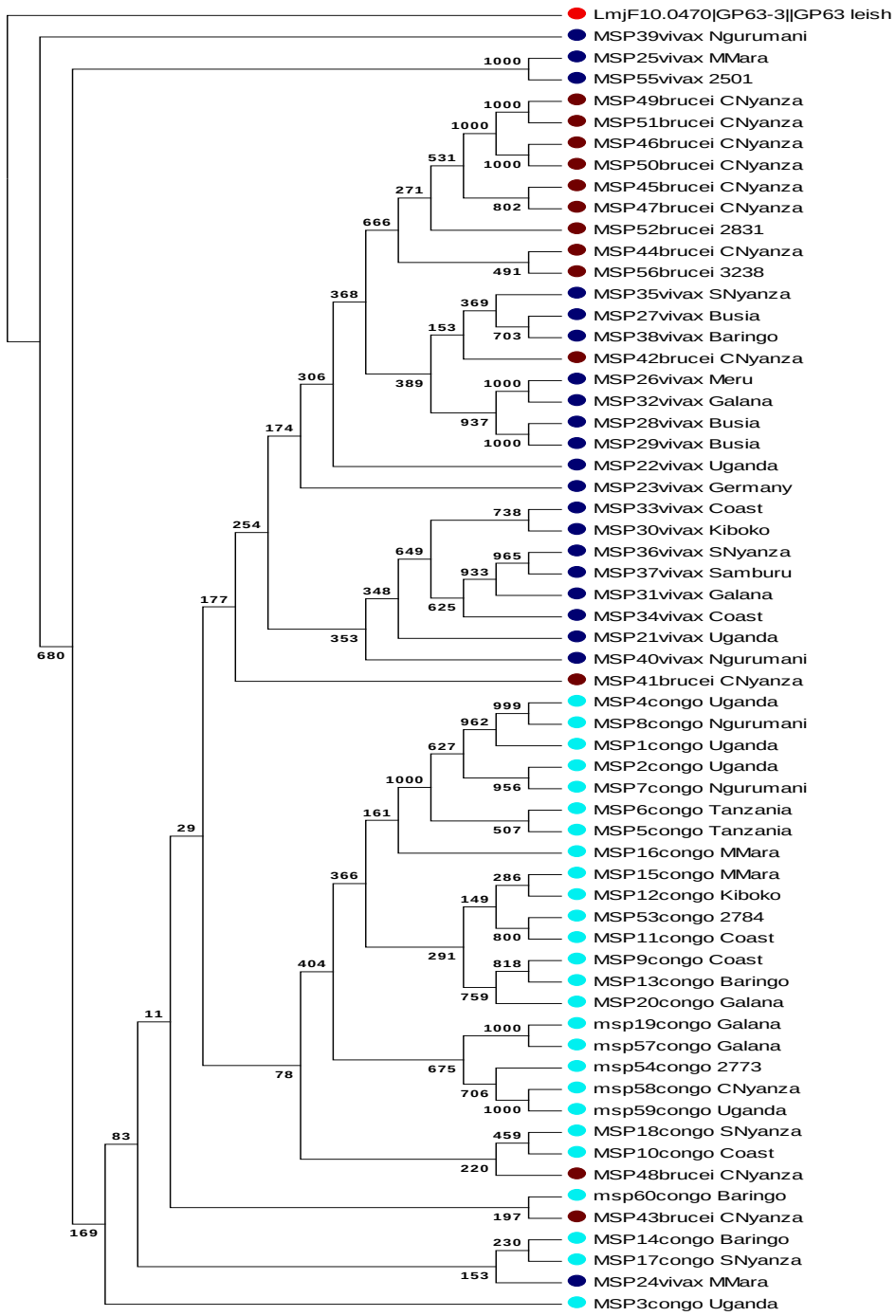

Figure 23: A Phylogenetic tree based on bootstrapped N-J distance method of the *T. congolense*, *T. b. brucei* and *T. vivax* protein sequences. The sequences are named with an indication of their geographical origin. The *Leishmania major* MSP was used to root the tree. The MSP sequences cluster according to species regardless of their geographical origin an indication that they are conserved. There were a few stray sequences that clustered with the other sequences.

CHAPTER FIVE

DISCUSSION

5.1 Identification of the *T. congolense* and *T. vivax* DNA Samples

This study was undertaken with the objective of unearthing the potential of MSP (leishmanolysin; gp63) as an alternative diagnostic target for AAT since the available methods are costly, non-specific and insensitive in cases of low parasitaemia. Blood samples from livestock collected from diverse locations within East Africa were used to obtain total genomic DNA that was used to amplify the gene of interest, the MSP. ITS-1 primers (Njiru *et al*., 2005) were used for the identification of all the samples used in the study. The DNA samples were grouped based on the variability of the ITS-1 region as described by Njiru *et al*., 2005. PCR products of approximately 700bp and 620bp for *T. congolense* were obtained while *T. vivax* samples gave products of approximately 480bp using the ITS 1 forward and reverse primers. Sub-typing primers and/or probes for *T. congolense* have been described (Majiwa *et al*., 1993; Masiga *et al*., 1992) and these were used to group the *T. congolense* samples in this study. The *T. congolense* samples studied showed indications of belonging to the subgroup *T. congolense*, savannah yielding products of approximately 320bp with the exception of two samples, KETRI 1117 and KETRI 2784, which gave products of approximately 294bp indicating that they belonged to the subgroup *T. congolense*, kilifi. None of the samples studied belonged to the subgroup *T. congolense*, forest. Once all the samples were correctly identified, grouped and sub-grouped, the gene of interest, the MSP, was amplified by way of PCR.

5.2 MSP Identification and Amplification

Previous studies have shown that promastigotes of all *Leishmania* species express an abundant surface glycoprotein, leishmanolysin (Etges, 1992), that is thought to be a ligand involved in the interaction of the parasite with the defensive systems of the host, including components of the complement system and the macrophage surface (Russell and Wilhelm,1986; Russell, 1987; Puentes *et al*., 1989). Leishmanolysin has therefore been thought of as an attractive vaccine candidate (Connell *et al*., 1993) and in this study, a potential diagnostic tool. The gene of interest in this study, the MSP (leishmanolysin) from *Trypanosoma vivax*, *T. b. brucei* and *T. congolense*, was amplified by way of PCR yielding stable products of approximately 1500-1700bp in size on a 1% agarose gel. The PCR products obtained illustrate that in deed the MSPs are found in the Kinetoplastid, *Trypanosoma* in abundance as they were amplified without much difficulty. The parasitic protozoa *Leishmania* species are evolutionarily related to *T. brucei* and occur as an extracellular form in their sand fly vector and an intracellular form in their mammalian hosts' macrophages. Studies on the possible role of MSP in *Leishmania* have shown that it provides resistance to complement-mediated lysis before *Leishmania* entry into the macrophage, participates in attachment and entry into the macrophage and it also supports survival in the macrophage after entry (Brittingham *et al*., 1995; Yao *et al*., 2003). Studies have shown that the *T. brucei* genome contains multiple genes encoding homologs of *Leishmania* MSP at different life stages but more predominantly the bloodstream stage, which have about 33% overall identity to one another and to *Leishmania* MSP (LaCount *et al*., 2003).

The MSP PCR products obtained from three different trypanosome species in this study, all of which infect livestock predominantly, indicate that they are a possible target for the development

of a diagnostic test for African Animal Trypanosomiasis. It was possible in this study to analyse the MSP from trypanosomes with reference to the well characterised structure of the MSP in *Leishmania* [gp63] (Schlagenhauf *et al*., 1998). Leishmanolysin molecules on the surface of promastigotes form a significant part of the interface between the invading parasite cell and the mammalian host at the time of infection, a necessary step of the *Leishmania* life cycle, and because comparatively few parasite cells are involved at this critical stage, there is likely to be strong selective pressure on leishmanolysin to maximize the probability of infection. Leishmanolysin has been thought to act before infection to ensure maximum numbers of parasites are produced and released from the insect or just after infection by facilitating binding and entry into target cells. The existence of a leishmanolysin homolog in *Crithidia fasciculata* (Elias *et al*., 2006), an insect parasite with no mammalian host, suggests it plays a primary role in the insect. Leishmanolysin antibodies have been shown to occur in infected individuals and trials of vaccines containing leishmanolysin have been carried out, some with promising results (Xu and Leiw, 1994; Rafati *et al*., 2001; Olobo *et al*., 1995). Reducing infection by stimulating immune responses to promastigotes has been complicated by the fact that promastigotes target macrophages, cells that phagocytose microorganisms and process antigens in the development of the immune response. The structure may be used to design peptide vaccines (Warburg and Schlein, 1986) that comprise just the appropriate antigenic parts of leishmanolysin, producing a specific protective response. Due to the great potential of the MSP (Leishmanolysin) as a diagnostic agent and vaccine target, the amplification of MSP was followed by the cloning and expression in bacterial cells.

5.3 MSP Protein Expression and Analysis

The cloning of MSP in pGEMT-Easy vector was followed by sub-cloning into the expression vectors (pRSET-A and or pET28a). The recombinant plasmids obtained were stable and screening via PCR and/or restriction endonuclease digestion revealed that the MSP was well ligated and in the correct orientation for expression. The expressed proteins were extracted under denaturing conditions. The release of MSP into the extracellular medium has been documented in several *Leishmania* species including both newly isolated clinical strains and laboratory-adapted strains (Yao *et al*., 2002; McGwire *et al*., 2002; Jaffe and Dwyer, 2003). Previous studies have shown a decrease in MSP release and an increase in MSP $T_{1/2}$ (protein stability), rather than changes in the rate of MSP translation, result in increased MSP expression during promastigote growth from logarithmic to stationary phase (Yao *et al*., 2005). Several factors affect the abundance of cell-associated proteins, including mRNA abundance, translation rate and protein stability. Previous studies showed that a difference in MSP stability, i.e. $T_{1/2}$, accounted for five-fold of the 14-fold increase in MSP abundance during promastigote growth from logarithmic to stationary phase, despite total *MSP* mRNA levels remaining stable (Wilson *et al*., 1993; Yao *et al*., 2002).

The result of a 12% SDS–PAGE showed a bigger and broad smear of the induced protein, which suggested that the polypeptide was induced. Separation of the proteins based on size on an SDS-PAGE gel revealed induced products of approximately 40kDa in size from both the *T. congolense* and *T. b. brucei* samples. The full-length polypeptides were predicted to be in the range of 60-70 kDa when fully expressed. However, this was 20kDa less the expected product size an indication that only short peptides of the MSP were expressed and not the full-length protein. Studies have

shown that expression of the *MSP* gene classes is differentially regulated after transcription during the promastigote growth from logarithmic to stationary phase (Roberts *et al*., 1993; McCoy *et al*., 1998). Great difficulties were experienced while expressing the MSPs especially for *T. vivax* samples which were eventually dropped out due to time constraints. The loss of MSP from the cell has been shown to occur primarily by the regulated release of MSP protein from the cell surface. Attenuated leishmania cell lines have been shown to lose their ability to regulate MSP protein abundance during cell growth (Brittingham *et al*., 2001). It is possible that this loss occurs because of changes in the ability of the parasite to remodel its plasma membrane, possibly membrane lipids, during growth (Yao *et al* 2005). Expression of fragments of MSP was possible because the N-terminal Methionine was intact and the presence of a mid-sequence stop codon meant that short fragments (peptides) than expected were expressed. These fragments are capable of undergoing less folding and therefore have their epitopes exposed favouring antigen-antibody reactions.

5.4 MSP Sequence Analyses

Amino acid sequences of surface antigens from organisms that infect mammals often show increased variation, reflecting changes occurring to overcome developing host immunity (Schlagenhauf *et al*., 1998). Well-characterized examples of antigenic variation in human pathogens include the evolution of viral coat protein sequences associated with influenza epidemics and the extreme variability of VSG sequences expressed during *T. brucei* infection. Although there is no evidence for similar antigenic variation in *Leishmania*, leishmanolysin sequences derived from different species do differ, and developmental changes in the expression of different genes in *Leishmania chagasi* has been reported (Streit, 1996). Despite the fact that

restriction fragment length polymorphism (RFLP) analysis of natural isolates indicates extensive polymorphism, amino acid sequence variability within the coding region of the mature protein is limited. *Leishmania* sequences are at least 60% identical and the mature proteins share the active site Zinc ligand and methionine-turn residues with the less conserved leishmanolysin homologs from *Crithidia* and *T. brucei* (Schlagenhauf *et al*., 1998). Surface residues have been shown to be less conserved than interior residues (Victoir *et al*., 2005), and that a surface representation of variability suggests sequence variability is correlated with structural flexibility (Schlagenhauf *et al*., 1998).

The MSP DNA sequences obtained as chromatograms after sequencing were viewed edited and assembled on BioEdit (Hall, 1999). The resulting sequences for each species were translated into the six possible open reading frames. The longest open reading frames were obtained starting from an initial Methionine residue followed by at least 50 amino acid residues; these were submitted to the SignalP server for verification of a secretory signal peptide. The sequences analysed had indications of being related to membrane-anchored or cytoplasmic proteins, while others gave no conclusive indication of a leader signal peptide, probably due to diminished sequence quality at the 5'- end. Notably, fifty-nine (59) out of the sixty-one (61) sequences had the Peptidase_M8 domain (ID: PF01457), characteristic of many metalloproteases such as the Leishmanolysin proteins. These sequences had significant matches to published *T. congolense* and *T. brucei gambiense* and *Leishmania major* MSP DNA sequences. When comparing all the major surface proteases sequences known (as at October 20, 2008) with the complete MSP DNA sequences generated in the laboratory, using the BlastN flavour of Blast, 12 sequences had hits

with a confidence value of 1E-18 or better, indicating they corresponded to the same or very closely related DNA sequences.

Fifty-nine (59) sequences had significant matches to the published Peptidase_M8 domain found in the Pfam protein database (ID: PF01457). This domain has a predicted function in proteolysis (Gene Ontology ID: 0006508), cell adhesion (Gene Ontology ID: 0007155), is a membrane component (Gene Ontology ID: 0016020), it also has a metalloendopeptidase activity (Gene Ontology ID: 0004222) and is Zinc ion binding (Gene Ontology ID: 0008270). Metalloproteases are the most diverse of the four main types of proteases, with more than 50 families identified to date. In these enzymes, a divalent cation, usually Zinc, activates the water molecule. The metal ion is held in place by amino acid ligands, usually three in number. The known metal ligands are Histidine, Glutamate, Asparagine or Lysine and at least one other residue is required for catalysis, which may play an electrophilic role. Of the known metalloproteases, about half contain an HEXXH motif, which has been shown in crystallographic studies to form part of the metal-binding site (Rawlings and Barret, 1995). The MSP sequences in this study showed characteristics of a typical metallopeptidase such as the Peptidase_M domains and the Zinc finger domains that are the Zinc binding domains. Transmembrane domains obtained on these sequences also indicate that the protein is indeed a membrane protein. A catalytic Zinc atom and an active-site helix containing the two histidines of the Zinc proteinase sequence motif HEXXH the characteristics of the Leishmanolysin N-terminal domain were observed among the sequences studied.

Another dominant domain observed among most of the sequences was the ZnF_C2HC and ZnF_A20 which have a possible role in DNA binding (GO: 0003677) and zinc ion binding (GO: 0008270). The domain within the MSP amino acid sequence started at approximately position 140 and ended at position 163. Zinc-binding motifs are stable structures, and they rarely undergo conformational changes upon binding their target. Transmembrane domains were also obtained bearing in mind that MSP is a membrane protein with regions spanning the cytoplasmic, membrane and surfaces regions. The other domain was the peptidoglycan recognition protein (PGRP) domain whose animal peptidoglycan recognition proteins are homologous to Bacteriophage T3 lysozyme (Kang *et al*., 1998). Transmembrane regions were also obtained bearing in mind that MSP is a membrane protein with regions spanning the plasma membrane as well as surfaces regions.

The MSP sequencing output had more than 50% identity with the *T. congolense* and *T. brucei* sequences that had been deposited in the GenBank and GeneDB (www.genedb.org) that is hosted by the Wellcome Trust Sanger Institute's Pathogen Genome Sequencing Unit. The final MSP nucleotide sequence comprised of fragments of lengths ranging between 1100bp and 1400bp, and the predicted open-reading frames for these sequences code for a protein of approximately 400 amino acids each and a predicted molecular mass of approximately 48kDa. As hypothesised that the trypanosome MSPs are conserved within the *T. vivax, T. congolense* and *T. b. brucei* trypanosome species we confirm that in deed there are conserved regions obtained in all the sequences analysed, especially the active-site helix containing the two histidines of the Zinc

proteinase sequence motif HEXXH (Figure 24) characteristic of the N-terminal domain of leishmanolysin.

Figure 24: An illustration of a conserved block obtained after a multiple sequence alignment of representative MSP protein sequences from *T. brucei*, *T. vivax* and *T. congolense*. TCOFFEE (Poirot *et al*., 2003) was used to produce this alignment. The regions highlighted in red indicate the highly conserved regions obtained from the MSP sequences analysed. This specific portion shows a HEXXH motif that is characteristic of the N-terminal domain of leishmanolysin (gp63).

5.5 MSP Evolutionary Analyses

The evolution of Leishmanial gp63 gene loci has been studied (Medina-Acosta *et al*., 1993a). The mosaic evolution of *gp63* genes has also been supported by the analysis of *gp63* evolutionary divergence and it has been found that different regions of the protein sequence exhibit different

evolutionary histories (Medina-Acosta *et al.* 1993a). Studies have shown that relationships based on the N-terminal of the protein are consistent with the conventional taxonomical subdivisions. Conversely, trees based on the C-terminal of the protein were very different. Mosaic genes are believed to be produced by intragenic recombination following mitotic crossover, which have been observed in genes encoding surface antigens in other pathogens (Deitsch *et al.,* 1997). The tandem organization of *gp63* genes and their localization in rearrangement-prone chromosomal extremities obviously represent an ideal combination for the production of mosaics. Levels of *gp63* sequence dissimilarity that could reach up to 45% among different isogenes of Leishmania have been reported (Steinkraus *et al.* 1993). Such heterogeneity was generated by synonymous and non-synonymous substitution (Alvarez-Valin *et al.* 2000).

Phylogenetic analyses of MSP were possible due to their highly conserved three-dimensional structure. To date, no work has been done on the phylogeny of the trypanosome MSPs hence most of the data available here is first hand and shall only be made in reference to the well studied gp63 in the *Leishmania* species (Medina-Acosta *et al*., 1993; Medina-Acosta *et al*., 1993a). Our goal in this study was not to redesign the evolution scheme of the trypanosomes and the matalloprotease family of proteins, but rather to track the evolution of selected trypanosome MSPs and their biogeographical distribution in order to determine how well conserved they are so as to recommend them as possible targets for AAT diagnostics. A multiple alignment on ClustalX software (Thompson *et al*., 1999) at the protein level showed blocks of conservation among the different MSP sequences under study with perfect matches in an alignment or substitutions of amino acids within the strong groups and the weak groups. In line with the hypothesis that MSPs from *T. congolense, T. vivax* and *T. b. brucei* species are conserved, phylogenetic trees obtained

in this study using the neighbour-joining method indicate that the MSPs are in deed conserved and that they diverged after speciation of as shown by the clustering (based on species type). The MSP sequences from the different geographical locations among the three species studied were clustering together regardless of their origin. The geographical predisposition of these trypanosomes therefore, does not seem to have an impact on the clustering hence it is possible to extrapolate that these proteins are conserved all through and retain their function regardless of their geographical location. The sequences clustered based mainly on the species types. Noticeably, the *T. brucei* MSP samples from Central Nyanza in Kenya tended to cluster together.

This study has shown that the MSP is a multi-domain protein. It is well known that proteins are composed of domains, recurrent protein fragments with distinct structure, function and/or evolutionary history. Protein domains may occur alone, as single-domain proteins, but many are found in combination with other domains in larger polypeptide chains. These multi-domain architectures are more frequent in eukaryotes than prokaryotes (Apic *et al*., 2000; Ekman *et al*., 2005; Gerstein and Levitt, 1998; Liu and Rost, 2004). During evolution, proteins with new functions or specificities have been invented through domain fusion and recombination as well as differentiation of existing domains. Domain fusion is a mechanism that allows the limited number of functional modules to be reused instead of being reinvented. The occurrences of domain families as well as the number of partner families follow a power-law distribution with a few very abundant and/or versatile domains (Apic *et al*., 2000; Ekman and Bjorklund, 2005; Gerstein and Levitt, 1998; Liu and Rost, 2004; Qian *et al*., 2001). However, the evolution of domain combinations is not purely stochastic, but depends upon selection of certain functions (Vogel *et al*., 2005). Often two or three domains in tandem have been reused in combination with other

domains. These supra-domains may have been selected because the function is dependent on the interface between them or because they are both necessary for proper function (Liu *et al* 2005).

It has also been shown in previous studies that some exon-bordering domains have unexpectedly many combination partners in animals. The addition of a domain to a protein is likely to alter its function, for example, it has been estimated that single-domain proteins from the same domain family have a 67% chance of having similar functions, whereas the corresponding number for two-domain proteins with just one of the domains in common is 35% (Hegyi and Gerstein, 2001). It has also been proposed that ancient enzymes with broad substrate specificities have evolved into more specific enzymes through gene duplication (Jensen, 1976), but enzymes often retain their biochemical function while gaining new substrate specificities or regulation mechanisms by the addition of a domain. As a matter of fact, enzymatic function is conserved down to 30% sequence identity for most single-domain enzymes and addition of a second domain rarely affects function (Todd *et al*., 2001). The evolution of multi-domain proteins can be studied by the use of phylogenetic trees built from sequence alignments. However, multi-domain proteins may cause problems when creating multiple alignments as these sequences may align poorly for distantly related proteins even if they share the same domain architecture (Åsa *et al*., 2005). Understanding the underlying mechanisms of protein evolution through domain rearrangements and sequence differentiation is crucial for understanding the development of new functionalities.

Since the MSP is a multi-domain protein, these domains are believed to be evolving at different rates and different times depending on the condition of the trypanosome. This observation is a major contributor to the great differences observed among the different MSP sequences in this

study. A comparison of the individual MSP sequences from same *Trypanosoma* species or from two different *Trypanosoma* species at a time revealed that there is greater similarity between the *T. brucei* and *T. congolense* MSP than with *T. vivax* and *T. congolense*.

CONCLUSIONS AND RECOMMENDATIONS

The MSP from *T. b. brucei* and *T. congolense* was successfully cloned and expressed in *E. coli* cells. A peptide of approximately 40kDa was expressed in *E. coli* BL21 cells. The MSP protein sequences in this study showed characteristics of typical metallopeptidase such as the Peptidase_M domains and the Zinc finger domains that are the Zinc binding domains. Transmembrane domains obtained on some these sequences also indicate that the protein could be a membrane protein. The characteristics of the Leishmanolysin N-terminal domain such as a catalytic Zinc atom and an active-site helix containing the two Histidines of the Zinc proteinase sequence motif HEXXH were observed among the sequences studied. The MSP protein shows conservation of certain regions among the species studied hence these can be exploited in the development of a sustainable diagnostic tool for AAT. The other regions on the MSP protein sequences show great variability upon multiple sequence alignment and can be used for analysing the differences between the three *Trypanosoma* species studied as a way of developing a diagnostic tool that is species-specific. This was the first study of its kind that attempted to compare the evolution of the MSP in relation to the geographical distribution of these proteins among *T. b. brucei, T. vivax* and *T. congolense* species from different geographical locations. The results indicate that the MSPs from the different locations actually do have conserved domains that define the main functions of this protein that can be targeted for AAT diagnosis and that evolutionary analyses also confirm that geographical predisposition has no effect on the phylogeny of the MSPs.

Propositions for further work on MSP shall include the sequencing of the full-length MSP DNA and critical sequence analysis for conserved domains and or motifs that can be exploited for

diagnostic purposes and even better, as potential vaccine targets. The differences in the MSP amino acid composition amongst the different trypanosome species can be exploited in the invention of a diagnostic method that is highly discriminative and specific thus make diagnosis, treatment and management of AAT much easier. The expressed pure MSP protein can also be used to develop an antigen-based ELISA protocol to detect AAT. The antibodies produced by the nagana-infected livestock would make it possible to conduct an antigen-ELISA for diagnosis of infections and also be used to illustrate the sensitivity and specificity of the MSP protein and ultimately confirm its probable use in diagnostics.

REFERENCES

Abascal, F., Zardoya, R. and Posada, D., 2005. ProtTest: Selection of best-fit models of protein evolution. *Bioinformatics*, **21**: 2104-2105.

Aksoy, S. 2003. Control of Tsetse flies and trypanosomes using molecular genetics. *Veterinary Parasitology*, **115**: 125-145.

Alexander, J. and Russell, D. G., 1992. The interaction of *Leishmania* species with macrophages. *Advances in Parasitology*, **31**: 175–254.

Altschul, S. F., Gish, W., Miller, W., Myers, E.W. and Lipman, D. J., 1990. Basic local alignment search tool. *Journal of Molecular Biology*, **215**: 403-410.

Alvarez-Valin, F. H., Tort, J. F. and Bernardi, G., 2000. Nonrandom spatial distribution of synonymous substitutions in the *GP63* gene from *Leishmania*. *Genetics* **155**: 1683–1692.

Apic, G., Gough, J. and Teichmann, S. A., 2001. Domain combinations in archaeal, eubacterial and eukaryotic proteomes. *Journal of Molecular Biology*, **310**: 311–325.

Åsa K. Björklund, Diana Ekman, Sara Light, Johannes Frey-Skött and Arne Elofsson J., 2005. Domain Rearrangements in Protein Evolution. *Molecular Biology*, **353**: 911–923.

Baker, J. R., 1963. Speculations on the evolution of the family Trypanosomatidae. *Experimental Parasitology*, **13**: 219-233.

Banuls, A. L., Hide M., and Tibayrenc M., 1999. Molecular epidemiology and evolutionary genetics of *Leishmania* parasites. *International Journal of Parasitol*ogy, **29**: 1137-1147.

Bateman, A., Birney, E., Durbin, R., Eddy, S. R., Howe, K. L., and Sonnhammer, E. L., 2000. The Pfam Protein Families Database. *Nucleic Acids Research*, **28**: 263-266.

Baylis, M. and Stevenson, P., 1997. Trypanosomiasis and tsetse control: Fact and fiction? *24th meeting of the International Scientific Council for Trypanosomiasis Research and Control, Maputo, Mozambique.*

Blood, D. C., Radostits, O. M. and Henderson, J. A., 1989. Diseases caused by protozoa. A textbook of the diseases of cattle, pigs, goats and horses. ELBS edition (7^{th} ed.), Oxford. Pages 1012-1015.

Bordier C., 1987. The promastigote surface protease of *Leishmania*. *Parasitology Today*, **3**: 151–153.

Borst, P., Fase-Fowler, F. and Gibson, W. C., 1987. Kinetoplast DNA of *Trypanosoma evansi*. *Molecular and Biochemical Parasitology,* **23**: 31-38.

Briones, M. R., Nelson, K.; Beverley, S. M., Affonso, H. T., Camargo, E. P. and Floeterwinter, L. M., 1992. *Leishmania tarentolae* taxonomic relatedness inferred from phylogenetic analysis of the small subunit ribosomal RNA gene. *Molecular and Biochemical Parasitology*, **53**: 121-127.

Brittingham, A., Morrison C. J., McMaster, W. R., McGwire, B. S., Chang, K. P. and Mosser, D. M., 1995. Role of the Leishmania surface protease gp63 in complement fixation, cell adhesion, and resistance to complement-mediated lysis. *Journal of Immunology*, **155**: 3102–3111.

Brittingham, A., Miller, M. A., Donelson, J. E. and Wilson, M. E., 2001. Regulation of GP63 mRNA stability in promastigotes of virulent and attenuated *Leishmania chagasi*. *Molecular and Biochemical Parasitology*, **112**: 51–59.

Connell, N. D., Medina-Acosta, E., McMaster, W. R., Bloom, B. R. and Russell, D. G., 1993. Effective immunization against cutaneous leishmaniasis with recombinant bacille Calmette-Guerin expressing the *Leishmania* surface proteinase gp63. *Proceedings of the National Academy of Sciences of the USA*, **90**: 11473-11477.

Corpet F., 1988. Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Research*, **16:** 10881-10890.

Croan, D. G., Morrison, D. A. and Ellis, J. T., 1997. Evolution of the genus *Leishmania* by comparison of DNA and RNA polymerase gene sequences. *Molecular and Biochemical Parasitology*, **89**: 149-159.

Cuevas, I. C., Cazzulo, J. J. and Sanchez, D. O., 2003. Gp63 Homologues in *Trypanosoma cruzi*: Surface Antigens with Metalloprotease Activity and a Possible Role in Host Cell Infection. *Infection and Immunity,* **71**: 5739–5749.

Deitsch, K. W, Moxon, E. R. and Wellems, T. E., 1997. Shared themes of antigenic variation and virulence in bacterial, protozoal and fungal infections. *Microbiology and Molecular Biology Reviews*, **61**: 281–293

Desquesnes, M. and Tresse, L., 1996. Evaluation of sensitivity of PCR for detecting DNA of *Trypanosoma vivax* with several methods of blood sample preparations. *Revue d'élevage et de médecine vétérinaire des pays tropicaux.* **49**: 322–327.

Desquesnes, M., McLaughlin, G., Zoungrana, A. and Davila, A. M., 2001. Detection and identification of *Trypanosoma* of African livestock through a single PCR based on internal transcribed spacer 1 of rDNA. *International Journal of Parasitology*, **31**: 610–614.

Dwinger, R. H., Agyemang, K., Snow, W. F., Rawlings, P., Leperre, P., Bah, M. L., 1994. Productivity of trypanotolerant cattle kept under traditional management conditions in the Gambia. *The veterinary Quarterly*, **16**, 81– 86.

Eisler, M. C., Lessard, P., Masake, R. A., Moloo, S. K. and Peregrine, A. S., 1998. Sensitivity and specificity of antigen-capture ELISAs for diagnosis of *Trypanosoma congolense* and *Trypanosoma vivax* infections in cattle. *Veterinary Parasitol*ogy, **79**: 187–201.

Ekman, D., Björklund, A. K., Frey-Skött, J. and Elofsson, A., 2005. Multi-domain proteins in the three kingdoms of life-orphan domains and other unassigned regions. *Journal of Molecular Biology*, **348**: 231–243.

Elias, C. G. R., Pereira, F. M., Silva, B. A., Alviano, C. S., Soares, R. M. A. and Santos, A. L. S., 2006. Leishmanolysin (gp63 metallopeptidase)-like activity extracellularly released by *Herpetomonas samuelpessoai*, *Parasitology*, **132**: 37-47.

Etges, R. 1992. Identification of a surface metalloproteinase on 13 species of *Leishmania* isolated from humans, *Crithidia fasciculata*, and *Herpetomonas samuelpessoai*, *Acta Tropica*, **50**: 205-217.

Etges, R., Bouvier, J. and Bordier, C. 1986. The major surface protein of *Leishmania* promastigotes is a protease. *Journal of Biology and Chemistry*, **261**: 9098-9101.

FAO, 1998. A field guide for the diagnosis, treatment and prevention of African Animal Trypanosomiasis .Uilenberg, G. FAO, Rome. Retrieved online 19th January 2008 from: http://www.fao.org/DOCREP/006/X0413E/X0413E00.HTM.

FAO, 2001. Integrating the sterile insect technique as a component of area-wide tsetse and trypanosomosis intervention. PAAT Technical and Scientific series by Feldman,U. and Hendricks,J. FAO, Rome. Retrieved online 19th January 2008 from: http://www.fao.org/docrep/004/y2022e/y2022e01a.htm.

FAO, 2002. Programme against African Animal Trypanosomiasis (PAAT). Twenty-second Regional Conference for Africa held in Cairo Egypt, 4-8 February, 2002. Retrieved on August 12, 2008 from: http://www.fao.org/DOCREP/MEETING/004/Y6057E.HTM

Felsenstein, J. 1993. PHYLIP (phylogeny inference package). *Cladistics*, **5**: 164-166.

Fernandes, A. P., Nelson, K. and Beverley, S. M., 1993. Evolution of nuclear ribosomal RNAs in Kinetoplastid protozoa: Perspectives on the age and origins of parasitism. *Proceedings of the National Academy of Sciences, USA*, **90**: 11608-11612.

Gerstein, M. and Levitt, M., 1998. Comprehensive assessment of automatic structural alignment against a manual standard, the scope classification of proteins. *Protein Science*, **7**: 445–456.

Gibson, W. C., Wilson, A. J. and Moloo, S. K., 1983. Characterisation of *Trypanosoma* (*Trypanozoon*) *evansi* from camels in Kenya using isoenzyme electrophoresis. *Research in Veterinary Science*, **34**: 114-118.

Gong W, Zhu X, Liu S, Teng M, Niu L., 1998. Crystal structures of acutolysin A, a three-disulfide hemorrhagic zinc metalloproteinase from the snake venom of *Agkistrodon acutus*. *Journal of Molecular Biology*, **283**: 657–668.

Geysen, D., Delespaux, V., Geerts, S., 2003. PCR-RFLP using Ssu-rDNA amplification as an easy method for species-specific diagnosis of *Trypanosoma* species in cattle. *Veterinary Parasitology*, **110**: 171–180.

Gong W, Zhu X, Liu S, Teng M, Niu L., 1998. Crystal structures of acutolysin A, a three-disulfide hemorrhagic zinc metalloproteinase from the snake venom of *Agkistrodon acutus*. *Journal of Molecular Biology*, **283**: 657–668.

Guindon S, and Gascuel, O, 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology*, **52**: 696–704

Grandgenett, P. M., Otsu, K., Wilson, H. R., Wilson, M. E. and Donelson, J. E., 2007. A function for a specific zinc metalloprotease of African trypanosomes. *PLoS Pathogens*, **3**: 150.

Haag, J., O'Huigin, C. and Overath, P., 1998. The molecular phylogeny of trypanosomes: Evidence for an early divergence of the Salivaria. *Molecular and Biochemical Parasitology*, **91**: 37-49.

Hall, T.A., 1999. BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids*, **41**: 95-98.

Hegyi, H. and Gerstein, M., 2001. Annotation transfer for genomics: measuring functional divergence in multi-domain proteins. *Genome Research*, **11**: 1632–1640.

Hilley, J. D., Zawadzski, J. L., McConville, M. J., Coombs, G. H., and Mottram, J. C., 2000. *Leishmania mexicana* mutants lacking glycosylphosphatidylinositol (GPI): Protein transaminidase provide insights into the biosynthesis and functions of GPI-anchored proteins. *Molecular Biology of the Cell*, **11**: 1183–1195.

Hoare C. A., 1972. The Trypanosomes of Mammals. *Oxford, Blackwell Scientific Publications*.

ILRAD, 1981. Annual report of the International Laboratory for Research on Animal Diseases. Retrieved on October 15, 2008 from http://www.ilri.org/InfoServ/Webpub/Fulldocs/Ilrad81/Trypanosomiasis.htm#TopOfPage

ILRAD, 1990. Annual report of the International Laboratory for Research on Animal Diseases. Retrieved on October 17, 2008 from: http://www.ilri.org/InfoServ/Webpub/Fulldocs/Ilrad90/Trypano.htm#P77_15231

Inverso, J. A., Medina-Acosta, E., O'Connor, J., Russell, D. G. and Cross, G. A., 1993. *Crithidia fasciculata* contains a transcribed leishmanial surface proteinase (*gp63*) gene homologue. *Molecular and Biochemical Parasitology*, **57**: 47–54.

Jaffe C. L. and Dwyer, D. M., 2003. Extracellular release of the surface metalloprotease, gp63, from *Leishmania* and insect trypanosomatids. *Parasitology Research*, **91**: 229–237.

Jensen, R. A., 1976. Enzyme recruitment in evolution of new function. *Annual Review of Microbiology*, **30**, 409–425

Joshi, P. B., Kelly, B. L., Kamhawi, S., Sacks, D. L., and McMaster, W. R., 2002. Targeted gene deletion in Leishmania major identifies leishmanolysin (gp63) as a virulence factor. *Molecular and Biochemical Parasitology*, **120**: 33-40.

Kang, D., **Liu, G.**, **Lundström, A.**, **Gelius, E.** and **Steiner, H., 1998.** A peptidoglycan recognition protein in innate immunity conserved from insects to humans. *Proceedings of the National Academy of Sciences of the United States of America*, **95**: 10078-10082.

Kuboki, N., Inoue, N., Sakurai, T., Di Cello, F., Dennis J. Grab, D.J., Suzuki, H., Sugimoto, C., and Igarashi, I., 2003. Loop-Mediated Isothermal Amplification for Detection of African Trypanosomes *Journal of Clinical Microbiology*, **41**: 5517-5524.

Kukla, B. A., Majiwa, P. A, Young, J. R., Moloo, S. K. and ole-MoiYoi, O. K., 1987. Use of species-specific DNA probes for detection and identification of trypanosome infection in tsetse flies. *Parasitology*, **95**: 1–16.

LaCount, D. J., Gruszynski, A. E., Grandgenett, P. M., Bangs, J. D. and Donelson, J. E., 2003. Expression and Function of the *Trypanosoma brucei* Major Surface Protease (GP63) Genes. *Journal of Biology and Chemistry*, **278**: 24658-24664.

Lake, J. A., De La Cruz, V. F., Ferreira, P. C. G., Morel, C. and Simpson, L., 1988. Evolution of parasitism: Kinetoplastid protozoan history reconstructed from mitochondrial rRNA gene sequences. *Proceedings of the National Academy of Sciences of the United States of America*, **85**: 4779-4783.

Leak, S.G.A., 1998. Tsetse Biology and Ecology. Their role in the Epidemiology and Control of Trypanosomiasis. CABI Publishing. Oxford and New York, page 568.

Liu, J. and Rost, B., 2004. Chop proteins into structural domain-like fragments. *Proteins: Structure, Function and Bioinformatics*, **55**: 678–688.

Liu, M., Walch, H., Wu, S. and Grigoriev, A., 2005. Significant expansion of exon-bordering protein domains during animal proteome evolution. *Nucleic Acids Research*, **33**: 95–105.

Luckins, A. G. 1992. Methods for diagnosis of trypanosomiasis in livestock *Senior Research Fellow, Centre for Tropical Veterinary Medicine, East Bush, UK*.

Luckins, A. G., Boid, R., Rae, P., Mahmoud, M. M., El Malik, K. H. and Gray, A. R., 1979. Serodiagnosis of infection with *Trypanosoma evansi* in camels in the Sudan. *Tropical Animal Health Production*, **11**: 1-12.

Luckins, A.G., 1997. Detection of antibodies in trypanosome-infected cattle by means of a microplate enzyme-linked immunosorbent assay. *Tropical Animal Health Production*, **9**: 53–62.

Luckins, A.G. and Mehlitz, D., 1978. Evaluation of an indirect fluorescent antibody test, enzyme-linked immunosorbent assay and quantification of immunoglobulins in the diagnosis of bovine trypanosomiasis. *Tropical Animal Health Production*, **10**: 149-151.

Lukes, J., Jirku, M., Dolezel, D., Kral'ova, I., Hollar, L. and Maslov, D. A., 1997. Analysis of ribosomal RNA genes suggests that trypanosomes are monophyletic. *Journal of Molecular Evolution*, **44**: 521-527.

Lun, Z. R., Allingham, R., Brun, R. and Lanham, S., 1992. The isoenzyme characteristics of *Trypanosoma evansi* and *Trypanosoma equiperdum* isolated from domestic stocks in China. *Annals of Tropical Medicine and Parasitology*, **86**: 333- 340.

Marchler-Bauer, A., Panchenko, A. R., Shoemaker, B. A., Thiessen, P. A., Geer, L. Y. and Bryant, S. H., 2002. CDD: a database of conserved domain alignments with links to domain three-dimensional structure. *Nucleic Acids Research*, **30**: 281-283.

Majiwa, P. A. and Webster, P., 1987. A repetitive deoxyribonucleic acid sequence distinguishes *Trypanosoma simiae* from *T. congolense*. *Parasitology*, **95**: 543-558.

Majiwa, P. A. O., Maina, M., Waitumbi, J. N., Mihok, S., Zweygarth, E., 1993. *Trypanosoma (Nannomonas) congolense*: molecular characterisation of a new genotype from Tsavo, Kenya. *Parasitology*, **106**: 151–162.

Masake R. A. and Nantulya V. M., 1991. Sensitivity of an antigen detection enzyme immunoassay for diagnosis of *Trypanosoma congolense* infection in goats and cattle. *Journal of Parasitology*, **77**, 231–236.

Masiga, D. K., Smyth, A. J., Hayes, P., Bromidge, T. J. and Gibson, W. C., 1992. Sensitive detection of trypanosomes in tsetse flies by DNA amplification. *International Journal of Parasitology*, **22**: 909–918.

Marché, S., Roth, C., Philippe, H., Dollet, M. and Baltz, T., 1995. Characterization and detection of plant trypanosomatids by sequence analysis of the small subunit ribosomal RNA gene. *Molecular and Biochemical Parasitology*, **71**: 15-26.

Maslov, D. A., Lukes, J., Jirku, M. and Simpson, L., 1996. Phylogeny of trypanosomes as inferred from the small and large subunit rRNAs: Implications for the evolution of parasitism in the trypanosomatid protozoa. *Molecular and Biochemical Parasitology*, **75**: 197-205.

Maslov, D. A. and Simpson, L., 1995. Evolution of parasitism in kinetoplastid protozoa. *Parasitology Today*, **11**: 30-32.

Maslov, D. A., Avila, H. A., Lake, J. A. and Simpson, L., 1994. Evolution of RNA editing in kinetoplastid protozoa. *Nature*, **368**: 345-348.

McConville, M. J., Mullin, K. A., Ilgoutz, S. C., and Teasdale, R. D., 2002. Secretory Pathway of Trypanosomatid**.** *Parasites'Microbiology and Molecular Biology Reviews*, **66**: 122-154.

McCoy, J. J., Beetham, J. K., Ochs, D. E., Donelson, J. E. and Wilson, M. E., 1998. Regulatory sequences and a novel gene in the msp (GP63) gene cluster of *Leishmania chagasi*. *Molecular and Biochemical Parasitology*, **95**: 251–265.

McGwire, B. S., O'Connell, W. A., Chang, K.P., Engman, D. M., 2002. Extracellular release of the glycosylphosphatidylinositol (GPI)-linked *Leishmania* surface metalloprotease, Gp63, is independent of GPI phospholipolysis: implications for parasite virulence. *Journal of Biology and Chemistry*, **277**: 8802–8809.

McMaster, W. R., Morrison, C. J., MacDonald, M. H., Joshi, P. B., 1994. Mutational and functional analysis of the *Leishmania* surface metalloproteinase GP63: similarities to matrix metalloproteinases. *Parasitology*, **108**: 29–36.

Medina-Acosta, E., Karess R. E., Schwartz, H. and Russell D. G., 1989. The promastigote surface protease (gp63) of *Leishmania* is expressed but differentially processed and localized in the amastigote stage. *Molecular and Biochemical Parasitology*, **37**: 263-273.

Medina-Acosta, E., Karess R. E., and Russell D. G., 1993. Structurally distinct genes for the surface protease of *Leishmania mexicana* are developmentally regulated. *Molecular and Biochemical Parasitology*, **57**: 31-45.

Medina-Acosta E, Beverley S. M. and Russell D. G., 1993a. Evolution and expression of the *Leishmania* surface proteinase (gp63) gene locus. *Infectious Agents and Disease Journal*, **2**: 25–34.

Mehlert, A., Zitzmann, N., Richardson, J. M., Treumann, A. and Ferguson, M. A. J., 1998. The glycosylation of the variant surface glycoproteins and procyclic acidic repetitive proteins of *Trypanosoma brucei*. *Molecular and Biochemical Parasitology*, **91**: 145-152.

Molyneux, D.H. and Ashford, R.W., 1983. African Animal Trypanosomiasis. The Biology of *Trypanosoma* and *Leishmania,* Parasites of Man and Domestic Animals. Taylor and Francis Ltd., London, pages 129-151.

Morrison, D. A. and Ellis, J. T., 1997. Effects of nucleotide sequence alignment on phylogeny estimation: A case study of 18S rDNAs of Apicomplexa. *Molecular Biology and Evolution*, **14**: 428- 441.

Morlais, I., Grebaut, P., Bodo, J.M., Djoha, S., Cuny, G. and Herder, S., 1998. Detection and identification of trypanosomes by polymerase chain reaction in wild tsetse flies in Cameroon. *Acta Tropica*, **70**: 109–117.

Moser, D. R., Cook, G. A., Ochs, D. E., Bailey, C. P., McKane, M. R. and Donelson, J. E., 1989. Detection of *Trypanosoma congolense* and *Trypanosoma brucei* subspecies by DNA amplification using the polymerase chain reaction. *Parasitology*, **99**: 57–66.

Moulton, J. E. and Sollod, A. E., 1976. Clinical, serological and pathological changes in calves with experimentally induced *Trypanosoma brucei* infection. *American Journal of Veterinary Research*, **37**: 791.

Mulligan, H. W, 1970. *The African Trypanosomiases* Ministry of overseas and Development, Allen and Unwin, London. Page 950.

Murray, M., Barry, J. D., Morrison, W. I., Williams, R. O., Hirumi, H., and Rovis, L., 1979. A review of the prospects for vaccination in African trypanosomiasis. *World Animal Review*, **32**: 913.

Murray, M., Trail, J. C. M., Davis, C. E., and Black, S. J., 1984. Genetic resistance to African trypanosomiasis. *Journal of Infectious Disease*, **149**: 311-319.

Nantulya, V. M. and Lindqvist, K. J., 1989. Antigen-detection enzyme immunoassays for the diagnosis of *Trypanosoma vivax, T. congolense and T. brucei* infections in cattle. *Tropical Medicine and Parasitology*, **40**: 267-272.

Nantulya, V.M., 1990. Trypanosomiasis in domestic animals: The problems of diagnosis. *Revised Science Techniques*, **9**: 357-367.

Nielsen H., Engelbrecht, J., Brunak, S. and Heijne, G., 1997. A neural network method for identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *International Journal of Neural Systems*, **8**: 581-599.

Neefs, J. M., Van Der Peer, Y., Hendricks, L. and De Wachter, R., 1990. Compilation of small ribosomal subunit RNA sequences. *Nucleic Acids Research*, **18**: 2237-2243.

Nei, M., 1991. Relative efficiencies of different tree making methods for molecular data. In: *Phylogenetic Analysis of DNA Sequences* (M. M. Miyamoto and J. Cracraft, eds.), pp. 90-128. New York: Oxford University Press.

Njiru, Z. K., Constantine, C. C., Guya, S., Crowther, J., Kiragu, J. M., Thompson, R. C. and Davila, A.M., 2005. The use of ITS1 rDNA PCR in detecting pathogenic African trypanosomes. *Parasitology Research*, **95**: 186-192.

Novagen, pET System Manual, TB055 11th Edition 01/06

Nyame, A. K., Kawar, Z. S. and Cummings, R. D., 2004. Antigenic glycans in parasitic infections: Implications for vaccines and diagnostics. *Archives of Biochemistry and Biophysics*, **426**: 182-200.

Nyeko, J. H. P., Ole-Moiyoi, O. K., Majiwa, P. A. O., Otieno, L. H., and Ociba, P. M., 1990. Characterization of trypanosome isolates from cattle in Uganda using species-specific DNA probes reveals predominance of mixed infections. *Insect Science and its Application*, **11**: 271-280

Olobo, J. O., Anjili, C. O., Gicheru, M. M., Mbati, P. A., Kariuki, T. M. and Githure, J. I., 1995. Vaccination of vervet monkeys against cutaneous leishmaniasis using recombinant *Leishmania major* surface glycoprotein (gp63). *Veterinary Parasitol*ogy, **60**: 199-212.

Paris, J., Murray, M. and McOdimba, F., 1982. A comparative evaluation of the parasitological techniques currently available for the diagnosis of African trypanosomiasis in cattle. *Acta Tropica*, **39**: 307-316.

Picozzi, K., Tilley, A., Févre, E. M., Coleman, P. G., Magona, J. W., Odiit, M., Eisler, M. C. and Welburn, S. C., 2002. The diagnosis of trypanosome infections: applications of novel technology for reducing disease risk. *African Journal of Biotechnology*, **1**: 39–45.

Portela, M. P. M., Raper, J. and Tomlinson, S., 2000. An investigation into the mechanism of trypanosome lysis by human serum factors. *Molecular Biochemical Parasitology*, **110**: 273-282.

Puentes, S. M., Dwyer, D. M., Bates, P. A. and Joiner, K. A. 1989. Binding and release of C3 from *Leishmania donovani* promastigotes during incubation in normal human serum. *Journal of Immunology*, **143**: 3743-3749.

Qian, J., Luscombe, N. M. and Gerstein, M., 2001. Protein family and fold occurrence in genomes: power-law behaviour and evolutionary model. *Journal of Molecular Biology*, **313**: 673–681.

Rafati, S., Salmanian, A. H., Taheri, T., Vafa, M., Fasel, N., 2001. A protective cocktail vaccine against murine cutaneous leishmaniaisis with DNA encoding cysteine proteinases of *L. major*. *Vaccine*, **19**: 3369-3375.

Rawlings, N. D. and Barrett, A. J., 1995. Evolutionary families of metallopeptidases. *Methods in Enzymology*, **248**:183-228

Rebeski, D. E., Winger, E. M., Okoro, H., Kowalik, S., Burger, H. J., Walters, D. E., Robinson, M. M., Dwinger R. H. and Crowther, J. R., 2000. Detection of *Trypanosoma congolense* antibodies with indirect ELISAs using antigen-precoated micro titre plates. *Veterinary Parasitology*, **89**: 187–198.

Reeve, M. A., Robinson, P. S. and Ball, S., 2002. Primer walking cycle sequencing. http://www.freepatentsonline.com/6376183.html , Patent Number: 6376183. Retrieved on October 16, 2008

Roberts, S. C., K. G. Swihart, M. W. A., Ramamoorthy, R., Wilson, M. E. and Donelson, J. E., 1993. Sequence diversity and organization of the *msp* gene family encoding gp63 of *Leishmania chagasi*. *Molecular and Biochemical Parasitology*, **62**: 157–172.

Roder, P. L., Scott, J. M., and Pegram, R. G. 1984. Acute Trypanosoma vivax infection of Ethiopian cattle in the apparent absence of tsetse. *Tropical Animal Health and Production*, **16**: 141-147.

Russell, D. G. and Wilhelm, H. 1986. The involvement of the major surface glycoprotein (gp63) of *Leishmania* promastigotes in attachment to macrophages. *Journal of Immunology*, **136**: 2613-2621.

Russell, D.G. 1987. The macrophage-attachment glycoprotein gp63 is the predominant C3-acceptor site on *Leishmania Mexicana* promastigotes. *European Journal of Biochemistry*, **164**: 213-221.

Saitou, N. and Nei, M., 1987. The Neighbour-joining Method: A New Method for Reconstructing Phylogenetic Trees. *Molecular Biology and Evolution*, **4**: 406-425.

Sambrook, J., Fritsch, E. F. and Maniatis, T., 1989. Molecular cloning: A laboratory manual. Second Edition, Cold Spring Harbour Laboratory Press, Cold Spring Harbour, N.Y.

Schlagenhauf, E., Etges, R., and Metcalf, P., 1998. The crystal structure of the *Leishmania major* surface proteinase leishmanolysin (gp63). *Structure*, **6**: 1035–1046.

Schultz, J., Copley R. R., Doerks T., Ponting C. P., and Bork P., 2000. Smart: A Web-Based Tool for the Study of Genetically Mobile Domains. *Nucleic Acids Research*, **28**: 231-234.

Sogin, M. L., Elwood, H. J. and Gunderson, J. H., 1986. Evolutionary diversity of eukaryotic small-subunit rRNA genes. *Proceedings of the National Academy of Sciences of the United States of America*, **83**: 1383-1387.

Steinkraus, H. B., Greer, J.M., Stephenson, D. C. and Langer, P. J. 1993. Sequence heterogeneity and polymorphic gene arrangements of the *Leishmania guyanensis* gp63 genes. *Molecular and Biochemical Parasitology*, **62**: 173-186.

Steve, R. and Skaletsky, H. J., 2000. Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds) *Bioinformatics Methods and Protocols: Methods in Molecular Biology*. Humana Press, Totowa, NJ, pp 365-386

Stevens, J., Noyes, H. and Gibson, W., 1998. The evolution of Salivarian trypanosomes. *Memórias do Instituto Oswaldo Cruz*, **93**: 669-676.

Stevens, J. R., Noyes, H. A., Dover, G. A., Gibson, W. C., 1999. The ancient and divergent origins of the human pathogenic trypanosomes. *Trypanosoma brucei* and *T. cruzi*. *Parasitology*, **118**: 107-116.

Stevens, J. R., Gibson, W., 1999. The molecular evolution of trypanosomes. *Parasitology Today*, **15**: 432-437.

Stevenson, P., Munga, L., Baylis, M. and Alushula, L., 1991. The control of tsetse and trypanosomiasis by delthamethrine treatment of ranch animals in Kenya. *Proceedings of 21st ISCTRC Meeting at Yamoussoukro,Ivory Coast*.pp.201.OUA/STRC,Nairobi.

Streit, J. A., Donelson, J. E., Agey, M. W. and Wilson, M. E., 1996. Developmental changes in the expression of *Leishmania chagasi* gp63 and heat shock protein in a human macrophage cell line. *Infection and Immunity*, **64**: 1810-1818

Tamura, K., Dudley, J., Nei M. and Kumar, S., 2007. *MEGA4*: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0**.** *Molecular Biology and Evolution*, **24**: 1596-1599.

Taylor, K. and Authie, E. M. L., 2004. Pathogenesis of Animal Trypanosomiasis. In: Maudlin, I., Holmes, P. H. and Miles, M. A. (eds.). The Trypanosomiasis. CABI Publishing, Wallingford, Oxfordshire OX10 8DE, UK.

Taylor, K. A. and Mertens B., 1999. Immune response of cattle infected with African trypanosomes. *Memorios Instituto de Oswaldo Cruz*, **94**: 239-44.

Thompson, J. D., Higgins, D. G. and Gibson, T. J., 1994. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, **22**: 4673-4680.

Thompson, J. W., Mitchel, M., Rees, R. B., Shereni, W., Schoenfeld, A. H and Wilson, A., 1991. Studies on the efficacy of deltamethrine applied to cattle for the control of tsetse flies (*Glossina* spp) in Southern Africa. *Tropical Animal Health and Production*, **23**: 221-226.

Todd, A., Orengo, C. and Thornton, J., 2001. Evolution of function in protein superfamilies, from a structural perspective. *Journal of Molecular Biology*, **307**: 1113–1143.

Treumann, A., Zitzmann, N., Hulsmeier, A, Prescott, A. R., Almond A., Sheehan, J., Ferguson, M. A., 1997. Structural characterization of two forms of procyclic acidic repetitive protein expressed by procyclic forms of *Trypanosoma brucei*. *Journal of Molecular Biology*, **269**: 529–547.

Turner, M. J., 1982. Biochemistry of the variant surface glycoproteins of salivarian trypanosomes. *Advances in Parasitology*, **21**: 69-153.

Urquhart, G. M., Armour, J., Duncan, J. L., Dunn, A. M. and Jennings, F. W., 1996. *Veterinary Parasitology*, second edition, pages 212-217.

Victoir, K., Arevalo, J., De Doncker, S., Barker, D. C., Laurent, T., Godfroid, E., Bollen, A., LeRay, D. and Dujardin, J. C., 2005. Complexity of the major surface protease (msp) gene organization in Leishmania (Viannia) braziliensis: evolutionary and functional implications. *Parasitology*, **131**: 207-214.

Vogel, C., Teichmann, S. and Pereira-Leal, J., 2005. The relationship between domain duplication and recombination. *Journal of Molecular Biology*, **346**: 355–365.

Warburg, A. and Schlein, Y. 1986. The effect of post-bloodmeal nutrition of *Phlebotomus papatasi* on the transmission of *Leishmania major*. *American Journal of Tropical Medicine and Hygiene*, **35**: 926-930.

Weidhaas, D. E. and Haile, D. G., 1978. A theoretical model to determine the degree of trapping required for insect population control. *Bulletin of the Entomological Society of America*, **24**: 18-20.

Wiemer, E. A. C., Hannaert, V., Van Den Ijssel, P. R. L. A., Van Roy, J., Opperdoes, F. R. and Michels, P. A. M., 1995. Molecular analysis of glyceraldehyde-3-phosphate dehydrogenase in *Trypanoplasma borelli*: An evolutionary scenario of subcellular compartmentation in Kinetoplastida. *Journal of Molecular Evolution*, **40**: 443-454.

Williams, B. G., Dransfield, R. D. and Brightwell, R., 1992. The control of tsetse flies in relation to fly movement and trapping efficiency. *Journal of applied Ecology*, **29**: 163-179.

Wilson, M. E., Paetz, K. E., Ramamoorthy, R., Donelson, J. E., 1993. The effect of ongoing protein synthesis on the steady state levels of Gp63 RNAs in *Leishmania chagasi*. *Journal of Biological Chemistry*, **268**: 15731–6.

Xu D, Leiw FY., 1994. Genetic vaccination against leishmaniasis. *Vaccine*, **12**: 1534-6.

Yao, C., Leidal, K. G., Brittingham, A., Tarr, D. E., Donelson, J. E. and Wilson, M. E., 2002. Biosynthesis of the major surface protease GP63 of *Leishmania chagasi*. *Molecular and Biochemical Parasitology*, **121**: 119–128.

Yao C, Donelson J. E. and Wilson, M. E., 2003. The major surface protease (MSP or GP63) of Leishmania species: Biosynthesis, regulation of expression, and function. *Molecular and Biochemical Parasitology*, **132**: 1–16.

Yao, C., Luo, J., Hsiao, C., Donelson, J. E. and Wilson, M. E., 2005. Internal and surface subpopulations of the major surface protease (MSP) of *Leishmania chagasi*. *Molecular and Biochemical Parasitology*, **139**: 173-183

Yiallouros, I., Kappelhoff, R and Schilling, O., 2002. Activation mechanism of pro-astacin: role of the pro-peptide, tryptic and autoproteolytic cleavage and importance of precise amino-terminal processing. *Journal of Molecular Biology*, **324**: 237–46.

APPENDICES

6.1 PROTOCOLS

6.1.1 Buffers for Preparing Competent *E. coli*

TFB1: 100 mM RbCl, 50 mM $MnCl_2$, and 30 mM potassium Acetate, 10 mM $CaCl_2$, 15% glycerol, pH 5.8; sterile-filter

TFB2: 10 mM MOPS, 10 mM RbCl, 75 mM $CaCl_2$, 15% glycerol, adjust to pH 6.8 with KOH, sterile filter

6.1.2 Preparation of Competent Cells

1. Inoculate 2ml of LB with a single DH5α colony. Incubate culture overnight at 37° C while shaking at 250 rpm.
2. The following morning, inoculate 500ml of LB with 1ml of saturated overnight culture. Incubate culture at 37° C while shaking at 250 rpm until OD_{600} = 0.5 (3-5 hours).
3. Transfer culture to 2 pre-chilled sterile 250ml centrifuge tubes. Pellet bacteria cells with a 5000 RPM spin for 10 minutes at 4° C. Discard supernatant. Place pellets on ice.
4. Re-suspend cells in 10ml cold $CaCl_2$ solution. Pool cells together into one pre-chilled 50ml Oakridge tube.
5. Pellet cells with a 2500 rpm spin for 5 minutes at 4° C. Discard supernatant and re-suspend cells in 10ml cold $CaCl_2$ solution. Set on ice 30 minutes.
6. Pellet cells with a 2500 rpm spin for 5 minutes at 4° C. Discard supernatant and re-suspend cells in 2ml cold $CaCl_2$ solution. At this point you can leave cells on ice overnight at 4° C.
7. Dispense cells into 50ul aliquot in pre-chilled sterile polypropylene tubes. Store cells at -80°C.

6.1.2.1 Test for Competency

8. Remove competent DH5α cells from the –80° C and immediately place on ice. Once thawed, add >10ng of plasmid DNA to a 50ul aliquot of competent cells. Place cells/DNA on ice for 3 minutes.
9. Heat shock cells at 42° C for 3 minutes.
10. Place cells back on ice for 3 minutes.
11. Add 1ml LB to cells/DNA. Tape tube onto shaking incubator platform and incubate cells/DNA for 1 hour at 37° C while shaking at 250 rpm.
12. Pellet cells with a quick spin. Remove 800µl of supernatant. Resuspend cells in the remaining supernatant.
13. Plate 100ul and 200ul of transformation onto 2 LB+ Ampicillin plate. Place plates inverted at 37° C overnight.

6.1.3 Bacterial Media and Solutions

6.1.3.1 LB medium: 10 g/litre tryptone, 5 g/litre yeast extract, 10 g/litre NaCl

6.1.3.2 LB agar: LB medium containing 15 g/litre agar

6.1.3.3 Kanamycin stock solution: 25 mg/ml in H_2O, sterile filter, store in aliquots at –20°C

6.1.3.4 Ampicillin 50mg/mL stock: 2.5g ampicillin with 50mL18 M. dH_2O

Store in 4°C with aluminium foil around 50mL conical. (Light sensitive)

6.1 3.5 IPTG (1 M): 238 mg/ml in H_2O, sterile filter, store in aliquots at –20°C

6.1 3.6 SOB

1. For 1L, mix the following ingredients in a glass container with a stir bar until dissolved.

   20.0g Bacto-tryptone

5.0g Bacto-yeast Extract

0.5g NaCl

~950mL 18 M. dH_2O

2. Add 833µL 3M KCl.

3. Adjust pH to 7.0 with 5M NaOH or HCl.

4. Adjust volume to 1L with 18 M. dH_2O.

5. Aliquot 250mL into 2L beakers. Label with date and cover with foam stoppers, aluminium foil, & autoclave tape.

6. Autoclaved and let cool before inoculation.

6.1.3 7 SOC (100 ml)

99 ml SOB medium

1 ml 2M Glucose Stock (Sterile)

6.1.4 Agarose Gel Electrophoresis Solutions

50 X TAE Buffer

1. Dissolve the following in ~800mL18 M. dH_2O. This may need to heat a little to dissolve completely.

242.0 g Tris Base

57.1 ml glacial acetic acid

100.0 ml 0.5M EDTA (pH 8.0)

The volume is raised to 1L and stored. Dilute 50-fold to use as DNA gel electrophoresis buffer

6.1.5 SDS-PAGE Solutions

10x Electrophoresis Buffer

30 g/L Tris Base

144 g/L Glycine

10 g/L SDS (or 100mL/L 10% SDS or 50mL/L 20% SDS)

0.09 M Tris·Cl, pH 6.8; 20% glycerol; 2% SDS;

0.02% bromophenol blue; 0.1 M DTT

6.1.6 Buffers for Protein Purification under denaturing conditions

6.1.6.1 Lysis buffers

Buffer A (1 liter):

100 mM NaH_2PO4 13.8 g $NaH_2PO4 \cdot H_2O$ (MW 137.99 g/mol)

10 mM Tris·Cl 1.2 g Tris base (MW 121.1 g/mol)

6 M GuHCl 573 g guanidine hydrochloride

Adjust pH to 8.0 using NaOH.

Buffer B (1 liter):

100 mM NaH_2PO4 13.8 g $NaH_2PO4 \cdot H_2O$ (MW 137.99 g/mol)

10 mM Tris·Cl 1.2 g Tris base (MW 121.1 g/mol)

8 M urea 480.5 g (MW 60.06 g/mol)

Adjust pH to 8.0 using NaOH.

6.1.6.2 Wash buffer

Buffer C (1 liter):

100 mM NaH2PO4 13.8 g $NaH_2PO4 \cdot H_2O$ (MW 137.99 g/mol)

10 mM Tris·Cl 1.2 g Tris base (MW 121.1 g/mol)

8 M urea 480.5 g (MW 60.06 g/mol)

Adjust pH to 6.3 using HCl.

6.1.6.3 Elution buffers

Buffer D (1 liter):

100 mM NaH_2PO4 13.8 g $NaH_2PO4 \cdot H_2O$ (MW 137.99 g/mol)

10 mM Tris·Cl 1.2 g Tris base (MW 121.1 g/mol)

8 M urea 480.5 g (MW 60.06 g/mol)

Adjust pH to 5.9 using HCl.

Buffer E (1 liter):

100 mM NaH_2PO4 13.8 g $NaH^2PO4 \cdot H_2O$ (MW 137.99 g/mol)

10 mM Tris·Cl 1.2 g Tris base (MW 121.1 g/mol)

8 M urea 480.5 g (MW 60.06 g/mol)

Adjust pH to 4.5 using HCl.

Due to the dissociation of urea, the pH of Buffers B, C, D, and E were be adjusted immediately prior to use.