Protein Discovery in African Trypanosomes: Studying Differential Protein Expression Throughout the Parasite life Cycle and Identification of Candidate Biomarkers for Diagnosing Trypanosome Infections

by

Brett Alexander Eyford
B.Sc., University of Victoria, 2007

A Dissertation Submitted in Partial Fulfillment
of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY

in the Department of Biochemistry & Microbiology

© Brett Alexander Eyford, 2013

University of Victoria

All rights reserved. This dissertation may not be reproduced in whole or in part, by photocopy or other means, without the permission of the author.

Supervisory Committee

Protein Discovery in African Trypanosomes: Studying Differential Protein Expression Throughout the Parasite life Cycle and Identification of Candidate Biomarkers for Diagnosing Trypanosome Infections

by

Brett Alexander Eyford
B.Sc., University of Victoria, 2007

Supervisory Committee

Dr. Terry W. Pearson (Department of Biochemistry & Microbiology)
Supervisor

Dr. Martin J. Boulanger (Department of Biochemistry & Microbiology)
Departmental Member

Dr. Caroline E. Cameron (Department of Biochemistry & Microbiology)
Departmental Member

Dr. Francis Y.M. Choy (Department of Biology)
Outside Member

Abstract

Supervisory Committee
Dr. Terry W. Pearson (Department of Biochemistry & Microbiology)
Supervisor

Dr. Martin J. Boulanger (Department of Biochemistry & Microbiology)
Departmental Member

Dr. Caroline E. Cameron (Department of Biochemistry & Microbiology)
Departmental Member

Dr. Francis Y. M. Choy (Department of Biology)
Outside Member

Research was undertaken to discover and study trypanosome proteins that may play important roles in host-parasite or vector-parasite interactions. The methods used mass spectrometry based proteomics ideally suited for analysis of low abundance molecules. First, isobaric tags were used to monitor changes in proteins expression throughout the life cycle of *Trypanosoma congolense*, an economically important livestock pathogen. This was the first large scale survey of protein expression in trypanosomes. In addition to generating protein expression data for approximately 2000 different parasite proteins, 6 previously undescribed *T. congolense* proteins were discovered. Several of the proteins with interesting expression trends were selected for molecular characterization and monoclonal antibody derivation.

Second, immunoenrichment and mass spectrometry were used to identify the cognate antigen recognized by a *T. congolense*-specific monoclonal antibody. The antigen, a flagellar calcium binding protein, was expressed as a recombinant protein and used to test its utility as a potential serodiagnostic antigen for diagnosis of *T. congolense* infections.

Third, a "deep-mining" protein discovery mass spectrometric method was used to identify trypanosome proteins present in the plasma of late-stage African sleeping sickness patients. A total of 254 trypanosome proteins were unequivocally identified by tandem mass spectrometry. These findings are unprecedented since never before have such a large number of pathogen proteins been discovered in human blood using a non-biased approach (i.e. without using a targeted assay). The proteins discovered provide insights into host-parasite interactions and are strong candidates as targets for new diagnostic assays.

Table of Contents

List of Tables

List of Figures

Abbreviations

AA,	amino acid
AAT,	animal African trypanosomiasis
ACN,	acetonitrile
AKB,	2-amino-3-ketobutyrate
AO,	alternative oxidase
AP,	alkaline phosphatase
ApoL1,	apolipoprotein L1
ATP,	adenosine triphosphate
BARP,	*brucei* alanine rich protein
BCA,	bicinchoninic acid
BLAST,	basic local alignment search tool
BSF,	bloodstream form
CAC,	citric acid cycle
CATT,	card agglutination test for trypanosomiasis
CCD,	charge coupled device
cDNA,	complementary DNA
CESP,	congolense epimastigote specific protein
CHAPS,	3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate
CID,	collision induced dissociation
CISSA,	congolense insect stage surface antigen
CoA,	coenzyme-A
CSF,	cerebrospinal fluid
Da,	Daltons
DAPI,	4', 6-diamidino-2-phenylindole
DEAE,	diethyl aminoethyl
D-MEM,	Dulbecco’s modified Eagles medium
DNR,	data not reliable
DPC,	delta-1-pyrroline-5-carboxylate
DTT,	dithiothreitol

EDTA,	ethylene diamine tetra acetic acid
EGTA,	ethylene glycol tetra acetic acid
ELISA,	enzyme linked immunosorbent assay
EMF,	epimastigote form
EST,	expressed sequence tags
FA,	formic acid
FBP,	fructose 1, 6-bisphosphate
FBS,	fetal bovine serum
FCA,	Freund's complete adjuvant
FCaBP,	flagellar calcium binding protein
FIA,	Freund's incomplete adjuvant
FP,	flagellar pocket
G3P,	glyceraldehyde-3-phosphate
G6P,	glucose-6-phosphate
GARP,	glutamic acid/alanine rich protein
Gly3P,	glycerol-3-phosphate
GPD,	glycerol-3-phosphate dehydrogenase
GPI,	glycosylphosphatidylinositol
HA,	haemagglutinin
HAT,	human African trypanosomiasis
HAT medium,	hypoxanthine, aminopterin, thymidine
HDL,	high density lipoprotein
HEPES,	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
Heptapep,	heptapeptide repeat protein
HPLC,	high performance liquid chromatography
HRP,	haptoglobin related protein
HRPO,	horseradish peroxidase
HSP,	heat shock protein
HT,	hypoxanthine, thymidine
IgG,	immunoglobulin G
IgM,	immunoglobulin M

IL-6,	interleukin 6
ILRAD,	International Laboratory for Research on Animal Diseases
ILRI,	International Livestock Research Institute
IP,	intraperitoneal
IPTG,	isopropyl β-D-1-thiogalactopyranoside
ISG,	invariant surface glycoprotein
iTRAQ,	isobaric tags for relative and absolute quantitation
IV,	intravenous
ka,	association rate
KD,	dissociation equilibrium constant
kd,	dissociation rate
kDa,	kiloDalton
kDNA,	kinetoplast DNA
LAMP,	loop-mediated isothermal amplification
LB,	Luria-Bertani
LC,	liquid chromatography
m/z,	mass to charge ratio
mAb,	monoclonal antibody
MALDI,	matrix assisted laser desorption ionization
MAP,	mitogen activated protein
MCF,	metacyclic form
MiSCREEN,	MALDI immunoscreen
MS,	mass spectrometry
MS/MS,	tandem mass spectrometry
MSP,	major surface metalloprotease
NADPH,	reduced nicotinamide adenine dinucleotide phosphate
NCBI,	national center for biotechnology information
NECT,	nifurtimox-eflornithine combination therapy
OD,	optical density
OPI,	oxaloacetate, pyruvate, insulin
ORF,	open reading frame

PAGE,	polyacrylamide gel electrophoresis
PBS,	phosphate buffered saline
PCF,	procyclic culture form
PCR,	polymerase chain reaction
PDC,	pyruvate dehydrogenase complex
PEG,	polyethylene glycol
PEM,	PIPES, EGTA, magnesium sulphate
PF,	procyclic form
PIPES,	1, 4-piperazinediethanesulfonic acid
PNPP,	para-nitro-phenylphosphate
PRS,	protease resistant surface molecule
PSG,	phosphate buffered saline glucose
PSSA,	procyclic stage surface antigen
PVDF,	polyvinylidene difluoride
RT,	room temperature
SCX,	strong cation exchange
SDS,	sodium dodecylsulphate
SPR,	surface plasmon resonance
TBS,	tris buffered saline
Tc3440,	TcIL3000.7.3440
TFA,	trifluoroacetic acid
TIM,	triose phosphate isomerise
TLF,	trypanosome lytic factor
TOF,	time of flight
VSG,	variant surface glycoprotein

Acknowledgments

I begin by expressing my gratitude to everyone from the Pearson Lab (Bianca Loveless, Morty Razavi, Matt Pope, Lee Haines, Tyler Brown, Martin Soste, Jessica Fudge, Rachelle Huot, and Richard Yip) for making my 5 years avoiding the real world an enjoyable yet productive success. Many thanks go to my numerous collaborators, both at home and abroad. Researchers do not work in a vacuum (at least not in our field) - your support has been instrumental. And finally, my most fervent gratitude goes to my lord and master, Terry Pearson. You have been, by far, the most influential player in my adult life.

Chapter 1. Introduction to Trypanosomes

1.1. Introduction to trypanosomes

Members of the genus *Trypanosoma* are protozoan parasites that can infect every known class of vertebrate [1-3]. The first trypanosomes were discovered in the blood of a trout by a Swiss physician and fisherman, Gabriel Valentin in 1841 [4, 5]. However, evidence of trypanosomiasis is much older [6]. Veterinary scrolls from ancient Egypt describe what appears to be trypanosomiasis of cattle (n'gana [7]) and in 1373, Sultan Mari Jata, Emperor of Mali, died after 2 years of being afflicted by a disease symptomatically similar to modern human African trypanosomiasis (HAT). Ancient trypanosomes have even been found preserved in amber in the Dominican Republic [8].

Trypanosomes are obligate parasites with life cycle stages in both a vertebrate host (mammals, fish, bird, reptiles) and a hematophagic vector (fleas, leeches, flies, and bats); although several exceptions to this is route of transmission do exist. Only two types of trypanosomes infect humans, African trypanosomes (discussed in section 1.3 and throughout this dissertation) and the American trypanosome, *Trypanosoma cruzi*, which is a distant relative of the African trypanosomes and is found throughout South and Central America. *T. cruzi* is a stercorarian parasite which is spread in the feces of the kissing bug (*Triatominae* sp), in contrast to the salivarian African trypanosomes that are primarily spread in the saliva of tsetse, the infamous insect vector. *T. cruzi* parasites cause Chagas' disease, a completely different malady [9], are not found in Africa and will not be discussed further in this dissertation.

1.2. Life cycle and biology of African trypanosomes

African trypanosomes that infect humans and other animals are spread by the bite of the infamous tsetse (*Glossina sp.*). Tsetse flies inhabit 8.7 million km^2 of sub-Saharan Africa known as the "tsetse belt". This area represents approximately a third of Africa, or to put it into perspective, an area greater than that of the entire Australian continent or of the United States of America. The best known African trypanosomes are *Trypanosoma brucei* [10] (the focus of Chapter 5), *T. congolense* [11] (the focus of Chapters 2-4), *T. simiae* [12], *T. suis* [3], *T. godfreyi* [11] *and T. vivax* [13]. However, *T. vivax* can be transmitted mechanically by biting insects so are not limited to the tsetse belt. Hominid primates, including humans, have broad innate resistance to African trypanosomes except for two subspecies of *T. brucei* (discussed further in section 1.3).

African trypanosomes are unicellular mono-flagellates of the class Kinetoplastida. This class is an early eukaryotic divergent and as such has some unusual characteristics. Each cell contains only a single large mitochondrion with its mitochondrial DNA organized into a large compact structure known as the kinetoplast (kDNA). The kDNA is comprised of series of concatenated maxi- and mini-circles. The maxi-circles encode all the mitochondrial genes while the mini-circles encode guide RNAs which are used to direct mRNA editing by insertion and removal of uracil residues [14].

The precise location of the kinetoplast changes depending on the parasite life cycle stage (and is used as the gold standard for life cycle stage determination) but is always found near the flagellar pocket (FP). The FP is an indentation in the plasma membrane near the posterior end of the cell from which the flagellum emerges (Figure 1). The basal body at the base of the flagellum and the kinetoplast are well known for their role in

orchestrating cell division [15]. The FP is also the only site of exo- and endocytosis. It is presumed that sequestering conserved receptor and transport proteins to this relatively inaccessible area is a defense strategy against the mammalian and insect immune systems.

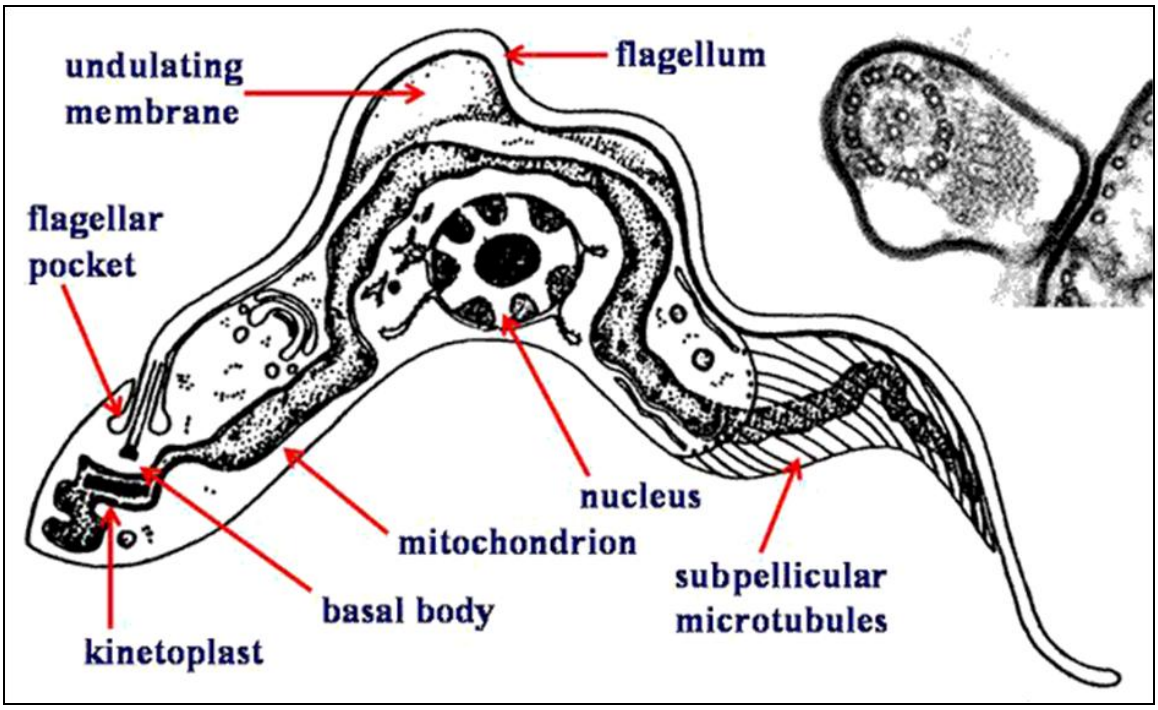


Figure 1. Cellular structure of an African Trypanosome.
Insert: electron micrograph showing a cross section of the membrane enclosed, 9+2 flagellum in close association with the cellular membrane. This figure is modified with permission from reference [16].

Trypanosome nuclear mRNA is transcribed polycistronically and maturation requires the trans-splicing of a conserved 35 base pair leader sequence [17]. Only one intron has been identified in a trypanosome genome (in the gene encoding a poly-A polymerase [18]). This genetic continuity has made genome analysis and cloning much easier than with other eukaryotic species. Gene expression is primarily controlled post-

transcriptionally through mRNA stability/degradation or by frequency of translation [19]. For this reason, studying mRNA abundance is of little use for assessing gene expression.

The African trypanosomes have four major life cycle stages. The procyclic form (PF), epimastigote form (EMF) and metacyclic form (MCF) all develop in tsetse while the bloodstream form (BSF) is found in the mammalian host (Figure 2). There are also minor life cycle stages such as the short stumpy BSF and mesocyclic forms [20, 21] which will not be discussed in detail in this dissertation.

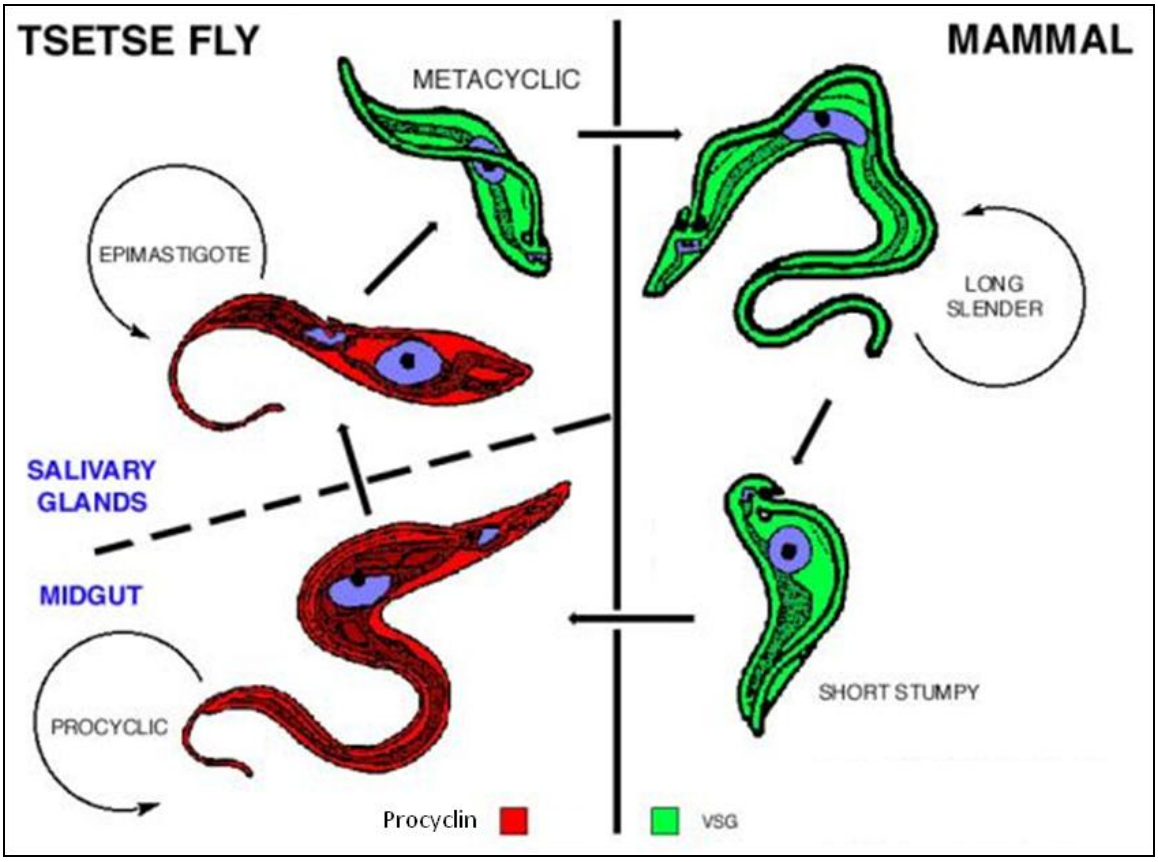


Figure 2. The *Trypanosoma brucei* life cycle.
Modified with permission from reference [22].

Infection of the mammalian host begins when MCF trypanosomes, in the saliva of an infected tsetse fly, are injected during a blood meal. The quiescent MCF trypanosomes

proliferate and differentiate into BSF at the site of infection. BSF rely entirely on glycolysis and substrate level phosphorylation to meet their energy requirements. In most organisms, glycolysis is carried out in the cytoplasm but trypanosomes compartmentalize the first seven steps of glycolysis into a membrane bound organelle called the glycosome [23]. BSF metabolism is incredibly wasteful. The end products of glycolysis (pyruvate and glycerol) are excreted and high energy electrons are disposed of through a mitochondrial alternative oxidase (AO) rather than the normal electron transport chain. The AO simply transfers high energy electrons to molecular oxygen to form water without generating a chemiosmotic gradient which can be coupled to ATP synthesis [24, 25]. Other than AO activity, the mitochondrion is almost entirely metabolically inactive during this life cycle stage [26].

The BSF parasites replicate by asexual binary fission in the interstitial fluids at the site of infection for 5-9 days before spreading to the circulatory system where they continue to replicate clonally, forming a peak of parasitemia (as high as 10^8 parasites per mL of blood). The BSF is covered with a dense surface coat of approximately 10^7 copies per cell of glycosyl-phosphatidylinositol (GPI) anchored variant surface glycoprotein (VSG). VSG molecules are famously involved in antigenic variation, allowing some members of the trypanosome population to avoid elimination by the host immune system [27]. As the parasite population increases clonally, a strong antibody response is generated against the predominant and highly immunogenic VSG. These antibodies facilitate parasite removal by agglutination and glomerular filtration, opsonisation and macrophage phagocytosis and most effectively, by complement mediated cell lysis. Although, trypanosomes delay complement lysis by rapid internalization and recycling of antibody bound VSG

molecules [28], eventually increased antibody concentration will overcome the rate of VSG internalization resulting in the lysis of the majority of the parasite population. However, trypanosomes have a large repertoire of VSG genes (~800) [29] and serve as the text book example of immune system evasion by antigenic variation. In a quorum sensing, yet poorly understood manner, a small subset of cells in the population will spontaneously express a different VSG, making them invisible to the antibody response directed against the VSG types in previous parasitemic waves. These survivors will then grow into a new parasitemic wave and continue the cycle (Figure 3).

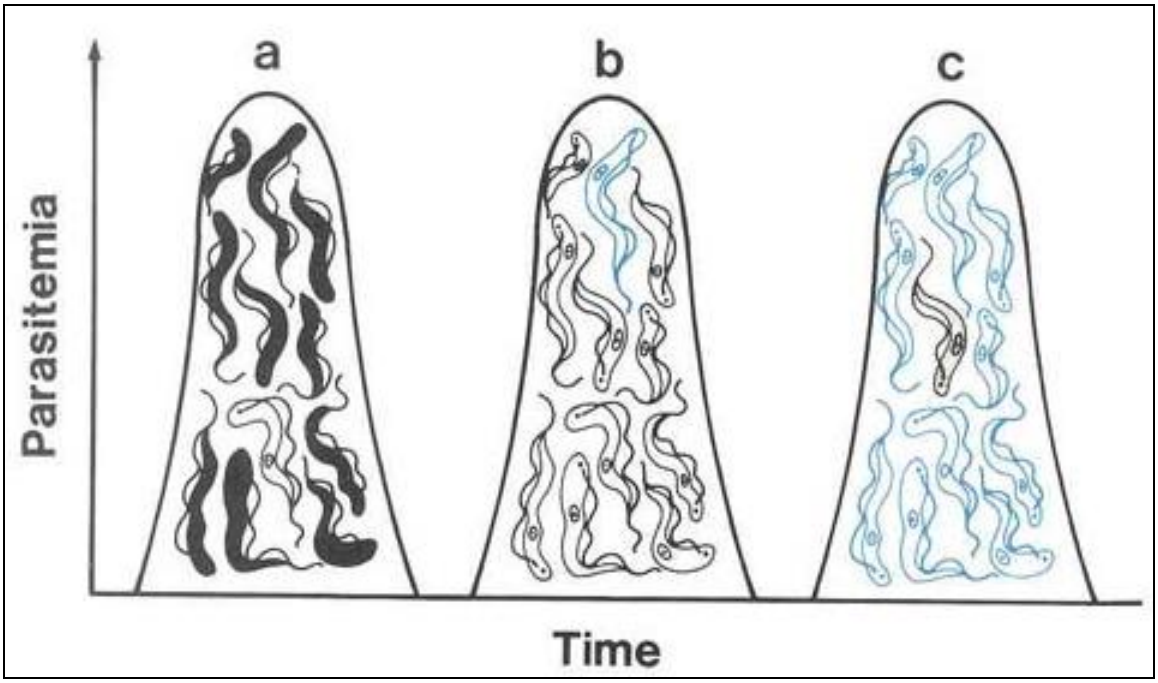


Figure 3. Highly simplified representation of antigenic variation and parasitemic waves observed during a trypanosome infection.
Figure used with permission from reference [30].

The trypanosome life cycle continues when BSF trypanosomes are consumed by a tsetse fly during a blood meal from an infected animal. The BSF trypanosomes enter the tsetse midgut where most will perish. However, a small proportion will differentiate into

the PF, a stage adapted to life in the midgut [31]. Differentiation from BSF to PF is accompanied by a switch to proline oxidation and oxidative phosphorylation for energy production. Procyclic trypanosomes can be easily cultured *in vitro* as procyclic culture forms (PCF) and are the most common source of this form of the parasite for research material [32]. PCF, like PF, use proline as their preferred energy source, although the typical media used to culture the parasites also contain glucose, which is absent or at very low and transient levels in the tsetse midgut. As a consequence, PCF still retain a basal level of glycolytic activity that is not present in their PF counterparts. Nevertheless, PCF are extremely similar to PF based on a variety of studies including comparison of their protein expression profiles [33]. Upon differentiation to procyclic forms, both *in vitro* and *in vivo,* the VSG coat is replaced by a set of invariant insect form specific surface molecules [34].

The most abundant surface proteins expressed by *T. brucei* are a set of (presumably) functionally redundant GPI-anchored proteins known as the procyclins [35]. Procyclins can be classified as either EP or GPEET forms, represented by the sequence of internal amino acid repeats. The relative proportion of EP and GPEET procyclins fluctuates temporally (reviewed in reference [34]). The function(s) of the procyclins remain a mystery but they are believe to protect underlying proteins from the fly's digestive enzymes and immune system [35]. Procyclins may also play a role in parasite tropism, migration and differentiation in the fly vector.

T. congolense PF express a different set of surface proteins. Based on structural and expression pattern similarities, the procyclins seem to be replaced by the heptapeptide repeat protein (Heptapep) [36]. Another abundant PF protein is the glutamic acid/alanine

rich protein (GARP, [37]). As with, what are presumed to be similar molecules expressed on the surface of *T. brucei* PF, the functions of these proteins are unknown.

After the PF have colonized the tsetse midgut, some parasites migrate to the salivary glands (*T. brucei*) or proboscis (*T. congolense*) where the parasites differentiate into adherent EMF [38]. The EMF of *T. brucei* and *T. congolense* are known to express on their surfaces, the *brucei* alanine rich protein (BARP; [39]) and congolense epimastigote specific protein (CESP; [40]) respectively. Both of these proteins are believed to be involved in the attachment of trypanosomes to the salivary gland epithelium (*T. brucei*) or chitinous labrum (*T. congolense*), although this has not been unequivocally demonstrated. The EMF population will grow for several days before asymmetric cell division begins, giving rise to non-dividing, motile, VSG coated MCF parasites which are ready to infect a new mammalian host.

1.3. Human African trypanosomiasis

HAT, also known as African sleeping sickness, is a disease caused by two subspecies of *T. brucei,* and is nearly always fatal if left untreated. Approximately 50 million people live in the tsetse belt of sub-Saharan Africa and are at risk of contracting this disease [7]. In 2009 the World Health Organization reported approximately 7,000 new cases of HAT but due to insufficient surveillance of rural populations, it is estimated that 30,000 people were infected [41]. This number is a significant improvement from the latest epidemic of HAT which peaked in the late 1990's with 25,000 new cases reported each year and estimates of infections of more than 300,000 [42].

Although this decline in HAT suggests the potential for eradicating, or at least controlling, the disease, it should be noted that in the 20th century Africa has experienced three major HAT epidemics. The first epidemic (1896-1906) is estimated to have killed between 300,000 [7] and 800,000 [43] people. This epidemic worried the European colonial powers and they invested heavily in combating the next epidemic which spanned the 1920's to 1940's. The second epidemic was eventually ended by screening and treatment of millions of people (Figure 4) by transmission control programs utilizing tsetse traps or insecticide sprays to kill tsetse and by culling of wild reservoir animals [7]. In the 1960's many African nations gained independence. The subsequent social upheaval, financial instability and apparent success of previous HAT suppression lead to many control programs being abandoned. This resulted in the re-emergence of HAT, peaking in the late 1990's (Figure 4).

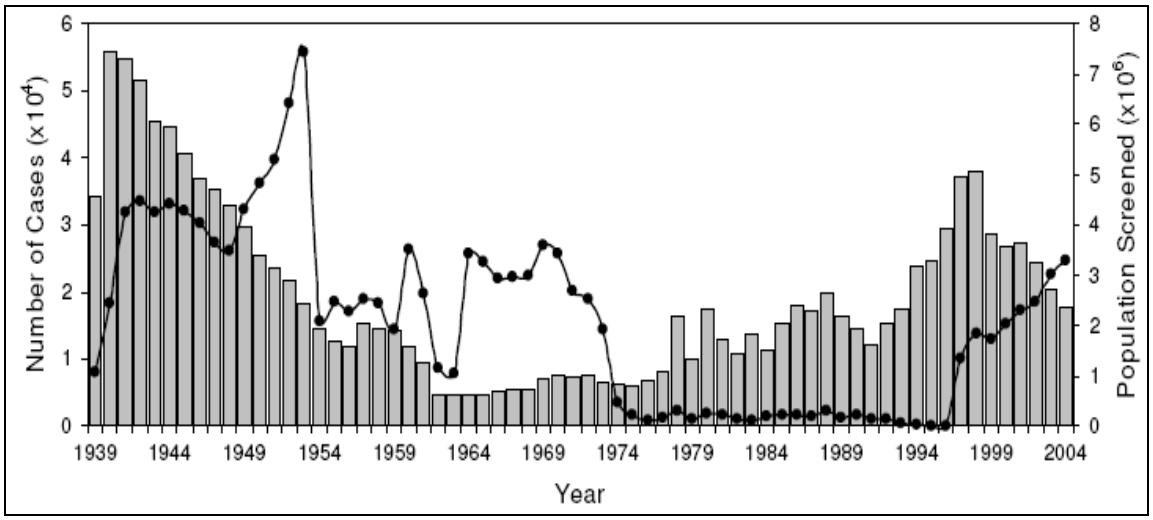


Figure 4. Reported cases of African sleeping sickness and numbers of people screened from 1939–2004.
Columns show the number of reported cases; Circles show the numbers of people screened. Figure taken with permission from reference [7].

As mentioned in section 1.1, humans are resistant to infection by most species of African trypanosomes. Non-infectious parasites are lysed rapidly by the "trypanolytic factor" (TLF), an innate component of human serum [44]. TLF activity has been ascribed to two proteins; apolipoprotein L1 (ApoL1) and haptoglobin-related protein (HRP), both of which are associated with high density lipoprotein. HRP is responsible for receptor binding and shuttling of ApoL1 into the parasite via haptoglobin-hemoglobin receptor mediated endocytosis [45, 46] where ApoL1 causes the disruption of the phago-lysosome membrane leading to cell lysis [47, 48]. The exact mechanism involved in lysosomal destruction remains under debate [49].

Only two subspecies of *T. brucei* are resistant to human TLF, *T. b. gambiense* and *T. b. rhodesiense*. The resistance of *T. b. rhodesiense* to TLF is due to the expression of the serum resistance associated (SRA) protein [50-52]. SRA appears to be a truncated VSG which co-localizes to endosomes along with TLF [53] and has been shown to bind strongly to ApoL1, presumably inhibiting lysosome disruption [48]. The ability of *T. b. gambiense* to resist TLF is less well understood although it appears to involve polymorphisms [54] or reduced expression [55] of the haptoglobin-hemoglobin receptor, both methods resulting in reduced uptake of TLF.

T. b. gambiense and *T. b. rhodesiense* cause symptomatically similar diseases which only differ in their severity and rate of progression. *T. b. rhodesiense* causes an acute disease and represents less than 10% of all HAT infections. This parasite is found in eastern and southern Africa and is fatal within weeks to months of infection [56] (Figure 5). Transmission of *T. b. rhodesiense* to humans by tsetse is typically via an animal reservoir with human-fly-human transmission less common.

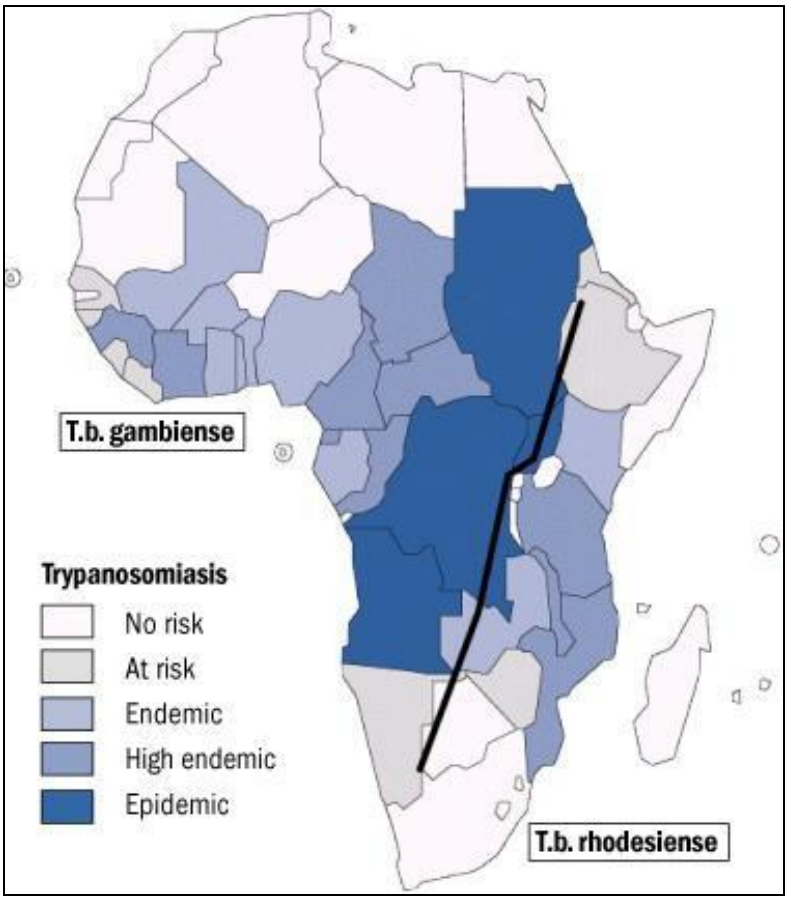


Figure 5. Distribution of human African trypanosomiasis, 1999.
The black line approximately divides the areas inhabited by *T. b. gambiense* and *T. b. rhodesiense.*
The figure was taken with permission from [57].

The chronic form of HAT is caused by *T. b. gambiense,* accounts for greater than 90% of all infections and is found throughout central and western Africa (Figure 5). The chronic form of HAT can be asymptomatic for several weeks to months after the fly bite. Death typically occurs after at least a year of infection [56]. Few animal species act as reservoirs for *T. b. gambiense* [58] but the longer duration of infection in humans means that human-fly-human is the usual cycle of transmission.

There is a recent report of patients from the Ivory Coast who were infected with *T. b. gambiense* yet refused drug treatment, thus the infections were allowed to proceed

naturally. Upon later testing it was found that many patients had spontaneously self-cured or controlled the infection and were parasite-free or asymptomatic, respectively [59]. This may indicate that HAT is not invariably fatal as previously believed. It is interesting to note that while the acute and chronic forms of HAT are well defined in native Africans, caucasians who become afflicted with HAT experience an acute form of the disease regardless of the infecting subspecies [60]. This would suggests that Africans have some level of innate resistance or tolerance to *T. b. gambiense* that is absent in non-Africans.

Both forms of the disease can progress through two phases. The first, early stage, is characterized by the presence of trypanosomes in the blood and lymph systems. The disease at this early stage manifests as a vague and general malaise and is often mistaken for malaria or influenza. The symptoms include headaches, fever, fatigue, joint pain, enlarged lymph nodes and occasionally, in the case of *T. b. rhodesiense*, a chancre around the tsetse bite site [61]. The second, late stage, begins when the parasites cross the blood-brain barrier and invade the central nervous system resulting in altered circadian rhythm, loss of motor control, psychosis, coma and ultimately death [7]. Once this stage has been reached the patient's condition declines rapidly and death soon ensues.

HAT affects the poorest people on the poorest continent. For this reason, pharmaceutical companies view it as an unprofitable disease and little effort has been put into developing new treatments. Nevertheless, there are five drugs currently being used to treat HAT. Unfortunately, all of the classic drugs for treatment of HAT must be administered by injection and most have serious side effects [62, 63].

Pentamidine is the drug used for treatment of early stage *T. b. gambiense* infections. Prolonged use can cause damage to the liver, kidneys and pancreas. However, for short term use the drug is well tolerated by the majority of patients and it was even used as a prophylactic agent in the 1950's and 1960's. Unfortunately, large-scale prophylaxis campaigns proved to be dangerous as well as financially and logistically impractical. Despite extensive use, pentamidine resistant parasites have not been observed [62].

Suramin (introduced in the 1920's) is the oldest anti-trypanosome drug still in use and it remains the treatment of choice for early stage of *T. b. rhodesiense* infections. However, suramin can cause nausea, vomiting, shock, allergic reactions and damage to the liver and kidneys [41, 62]. No significant resistance to suramin has been reported after 90 years of use.

Sometimes described as arsenic in antifreeze, melarsoprol is the only drug that can treat late stage disease caused by both *T. b. rhodesiense and gambiense*. Melarsoprol is arsenic based and practically insoluble in water, necessitating the use of propylene glycol as a solvent [62]. It has been observed that Melarsoprol/propylene glycol dissolves the medical tubing as it is injected intravenously (IV). Melarsoprol treatment can cause inflammatory encephalopathy which is lethal to 3-10% of patients [41]. For patients with the late stage of sleeping sickness the choices are either to face near certain death by infection or risk a potentially lethal treatment. To make matters worse, in the past decade there has been an increase in the number of melarsoprol resistant *T. b. gambiense* infections in Southern Sudan, Democratic Republic of Congo, Uganda and Angola [64].

Eflornithine is a failed cancer treatment drug which was later found to be trypanocidal. It is effective for treatment of the late stage disease but only against *T. b. gambiense.*

Eflornithine's application is problematic because it is costly to produce and requires 4 injections daily for 14 days [61, 62]. Despite its demanding administration schedule, eflornithine has had a major impact on patient outcome. It has been called the "resurrection drug" for its ability to revive comatose late stage patients who would not have otherwise survived.

Nifurtimox is a drug used to treat *T. cruzi* in the Americas and was also considered for use against African trypanosomes. On its own, nifurtimox is unable to cure HAT, but when applied as nifurtimox-eflornithine combination therapy (NECT) the frequency of injection can be reduced to twice daily for 7 days. An additional benefit of NECT is that the nifurtimox portion can be administered orally [63, 65-67].

There is no vaccine for HAT and until recently there were no vaccines against any eukaryotic parasite [68, 69]. Many researchers believe it is unlikely that a vaccine for African trypanosomiasis can be created. The large repertoire of VSG genes available for antigenic variation and the inaccessibility of underlying conserved membrane proteins makes a vaccine against African trypanosomiasis unlikely. However, several researchers are investigating the possibility of using low abundance conserved membrane proteins as vaccine targets [70, 71].

1.4. Animal African trypanosomiasis

Animal African trypanosomiasis (AAT) is a collection of symptomatically similar diseases caused by a number of different trypanosome species. The host range of the African trypanosomes differs from species to species of parasite. Some are limited to a single mammalian host while others can infect many animal species. For example, *T. suis* and *T. godfreyi* appear to be limited to suids (domesticated and wild pigs) while *T. simiae, T. congolense, T. b. brucei* and *T. vivax* can collectively infect nearly all wild and domestic African mammals (Table 1).

Table 1. Host range of pathogenic African trypanosomes and disease severity

Trypanosome	Human	Cattle	Goats	Pigs	Camels	Horses
T. b. brucei	R	+	+++	+	++	+++
T. b. gambiense	++	+	+++	+	++	+++
T. b. rhodesiense	+++	+	+++	+	++	+++
T. congolense	R	+++	++	+	++	++
T. simiae	R	R	+	+++	R	R
T. suis	R	R	R	+++	R	R
T. vivax	R	++	++	R	R	++

+++ Acute infection
++ Chronic infection
+ Mild or transient infection
R Resistant to infection
The table was generated using information from reference [72].

In cattle, AAT is caused by *T. vivax, T. b. brucei* and *T. congolense*. The most widespread and virulent of these species is *T. congolense* [72, 73]. In addition to cattle, *T. congolense* infects many other domesticated animals including sheep, pigs, goats, horses and camels causing a chronic wasting disease (cachexia); characterized by anemia, weight loss and immunosuppression. *T. congolense* is also the species which is most easily adapted to growth *in vitro* and recently has become the model organism of choice for studying the trypanosome life cycle.

Although Africa is fertile and has an abundance of natural resources, the human population of this continent is among the most impoverished in the world. The grasslands of Africa support thriving herds of large herbivores but raising domesticated livestock is hindered by zoonotic diseases. AAT, known locally as n'gana (from the Zulu word meaning "poorly"), has a particularly devastating impact on African agriculture. The Food and Agriculture Organization of the United Nations has acknowledged that AAT is a major impediment to the economic development of Africa [74, 75]. Every year approximately 40 million cattle are threatened and 3 million are killed by trypanosomiasis. The economic loss resulting directly from animal death is in the range of US$ 1.0 - 1.2 billion annually. When secondary losses such as reduced manure, milk and draft power and thus decreased crop yields are included, the total gross domestic product lost can be as much as $4.5 billion per annum [75, 76]. Due to the large impact of AAT on the livelihood of Africans, animal infective trypanosomes (especially *T. congolense*) are receiving increased attention with the focus on understanding the parasite life cycle in order to devise strategies for blocking transmission and controlling this economically important disease. Towards this goal, the proteins of *T. congolense* are the subject of the majority of this dissertation. The ability to culture, *in vitro*, all four major life cycle stages and the recent completion of the *T. congolense* IL3000 genome sequence opens the door for studying differential protein expression throughout the parasite life cycle. I have used a proteomics approach involving mass spectrometry for my research on defining the proteins expressed by *T. congolense*.

Chapter 2. Differential Protein Expression throughout the life cycle of *Trypanosoma congolense*

Work in this chapter was a collaborative effort involving three laboratories. *T. congolense* IL 3000 trypanosomes were grown in the laboratory of Dr. Noboru Inoue (Obihiro University, Hokkaido, Japan) and supplied to the Pearson lab at UVic as frozen material. iTRAQ mass spectrometry on *T. congolense* tryptic peptides was performed by Derek Smith at the UVic-Genome BC proteomics Centre. The anti-CESP monoclonal antibody used in some of my work was supplied by Bianca Loveless, formerly of the Pearson lab, now with Dr. Marty Boulanger's lab at UVic. Experimental planning was performed by Brett Eyford and Terry Pearson and iTRAQ data analysis and non-iTRAQ experimentation was performed by Brett Eyford.

The work presented in this chapter has been published as:

Eyford BA, Sakurai T, Smith D, Loveless B, Hertz-Fowler C, Donelson J, Inoue N, Pearson TW. 2011. *Differential protein expression throughout the life cycle of Trypanosoma congolense, a major parasite of cattle in Africa*. Molecular and Biochemical Parasitology, 177(2), 116-125.

2.1. Introduction

2.1.1. The life cycle of *Trypanosoma congolense*

The life cycle of *T. congolense* forces the parasite to interact with a variety of challenging environments in both the tsetse vector and mammalian host. These interactions influence parasite differentiation, survival, maturation and infectivity. The adaptations of the parasites to these stresses are critical for parasite transmission and perpetuation of disease. In the mammal, the parasites live at 37 °C in a neutral pH, nutrient rich environment where they must cope with the mammalian immune system and hydrodynamic forces encountered by life in the bloodstream. Several trypanosome molecules involved in the adaptation to life in the mammalian host have been well studied, including metabolic enzymes and some surface molecules, especially VSGs in BSF (reviewed in [77]).

While surviving in the mammalian blood stream, BSFs must also be prepared to rapidly differentiate into procyclic forms if taken up in a tsetse blood meal. In the bloodstream, *T. brucei* sspp., long-slender BSF differentiate into short stumpy BSF which appear to be pre-adapted to life in the fly. However, analogues to short stumpy BSF have not been identified in *T. congolense* [3]. In the tsetse midgut, PF experience an alkaline pH, ambient temperature, transient and unpredictable nutrient availability, insect immune factors and digestive enzymes. The aerobic metabolism of proline in PF has been well studied [78] and recently the surface molecules of PF that are involved in tsetse-trypanosome interactions have received more attention [79]. Heptapep, GARP and a glycolipid known as the protease resistant surface molecule (PRS) are all believed to protect the parasites from the harsh environment in the tsetse's digestive tract [80]. They

have also been hypothesized to play a role in parasite migration and tissue tropism [80]. Cyclical transmission places extreme constraints on trypanosomes in the tsetse vectors [38, 79] and 90% of flies that feed on a trypanosome infected animal will resist the establishment of a midgut PF infection [81, 82]. In addition to the stringent selection during parasite establishment in the tsetse midgut, a similar bottleneck occurs during differentiation to the EMF which colonize the tsetse's mouth parts [38]. At this stage, nutrients are limited to what the parasites can scavenge from a passing blood meal or obtain directly from the tsetse fly. EMF are slow growing forms and not free swimming, unlike parasites in the other life cycle stages. The parasites are still motile but adhere tightly to the fly's mouth parts through hemi-desmosome like structures [83]. The surface protein CESP is expressed only in EMF and is believed to be involved in adherence but this has not been unequivocally demonstrated [40]. To complete the *T. congolense* life cycle, the EMF divide asymmetrically to give rise to non-dividing, free swimming, VSG coated MCF that survive in the tsetse's saliva but must also be capable of infecting a mammal when the tsetse takes a blood meal.

2.1.2. Research Objectives and Experimental Design

T. congolense causes an economically and socially important disease. However, beyond some general information about enzymes involved in metabolism and the description of the few surface coat proteins, little is known about protein expression during the *T. congolense* life cycle, especially when compared to the heavily studied *T. brucei* sspp. Based on the widely disparate environments where the parasites live, it is

hypothesized that appropriate expression of essential, life cycle stage specific, proteins is important for parasite survival.

Transcriptome analysis has been performed on *T. congolense* and has provided some insight into the differential gene expression in the four major life cycle stages [84]. However, protein expression in trypanosomes is strongly regulated post-transcriptionally [19], thus, changes in mRNA abundance are not necessarily reflected at the protein level. To study protein expression directly, protein mass spectrometry was selected as the tool of choice. In addition, *T. congolense* IL3000 [85, 86], a fly transmissible strain, was chosen for this work because all four major life cycle stages of this parasite can be grown *in vitro.* This is not possible for any other species of trypanosome, thus *T. congolense* is the parasite of choice for this type of expression study. Fortunately, *T. congolense* IL3000 was also chosen as the type strain from the *Trypanosoma* subgenus nannomonas for genome sequencing which was being performed during my protein expression analysis and which was completed soon thereafter.

To analyze protein expression changes throughout the *T. congolense* life cycle, the method called isobaric tags for relative and absolute quantitation (iTRAQ) was used in conjunction with tandem mass spectrometry (MS/MS). This method can be used to identify and determine the relative amounts of proteins from different sources in a single experiment. The method is based on the covalent labelling of tryptic peptides, digested from protein samples of interest, with isobaric mass tags. As with standard MS/MS, the peptide sequence can be determined by analyzing the fragmentation pattern and can be assigned to a parent protein after database searching. To quantify the abundance of each peptide, and by extension, its parent protein, each unique protein sample is labeled with a

different iTRAQ reagent (tag). The different tags all have the same mass while intact but are isotopically labeled so that, when dissociated inside a mass spectrometer, the resulting fragments will have unique and distinguishable masses (Figure 6A). The relative amounts of each unique fragment can be used to determine the relative quantitation of the peptides and thus their parent proteins in each sample. A schematic overview of the method is shown in Figure 6B.

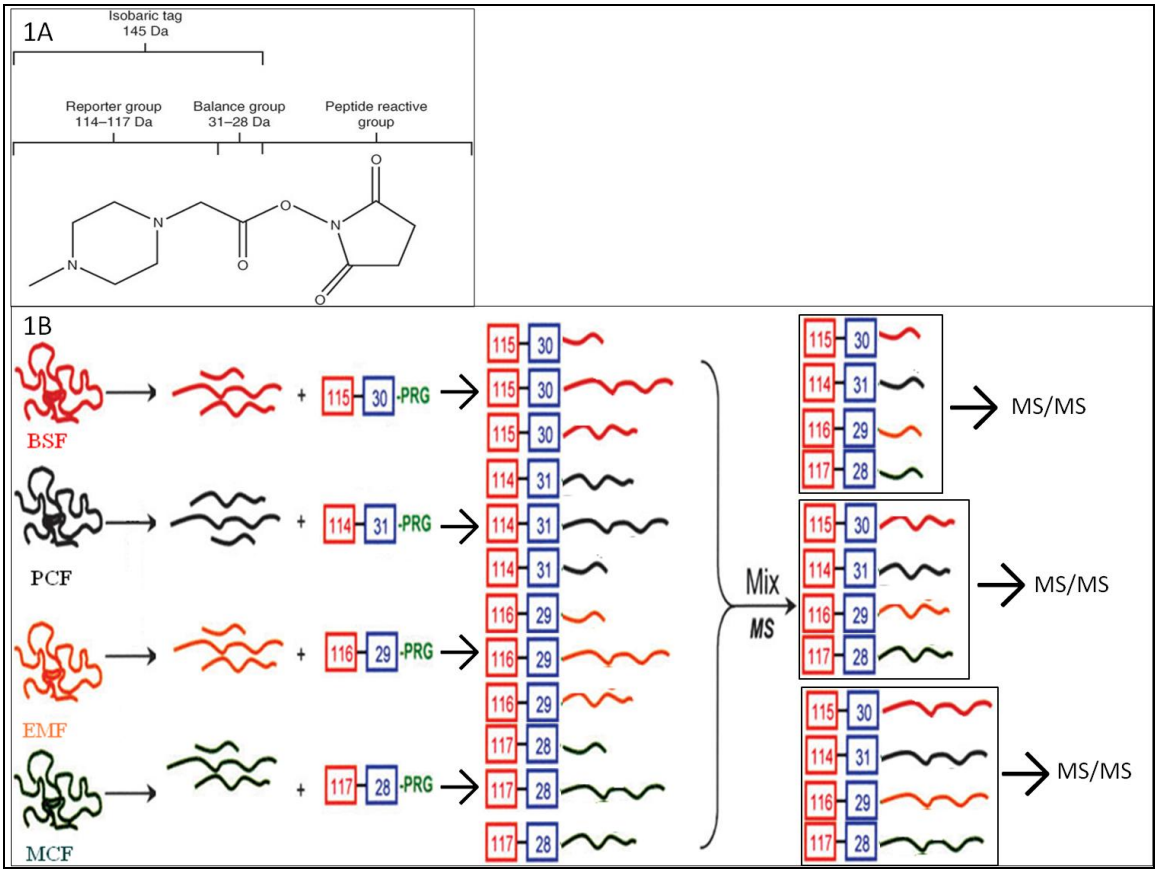


Figure 6. The iTRAQ mass tags and an experimental work flow.
A: The iTRAQ reagent is an isobaric mass tag consisting of a charged reporter group, a peptide-reactive group and a neutral balance portion. The reporter groups range in mass from 114 to 117 Da, whereas the balance groups range in mass from 28 to 31 Da, such that the combined mass remains constant (145 Da) for each of the four reagents. Figure 7A was taken with permission from reference [87]. B: A simplified schematic overview of the iTRAQ sample preparation and MS/MS analysis. Proteins from each of the four major life cycle stages of *T. congolense* were digested with trypsin and the peptides were labeled with different isobaric tags in parallel. Samples were mixed and analyzed by MS. Figure 7B was modified with permission from reference [88].

The mass spectrometric fragmentation of the attached iTRAQ tags generates low molecular mass reporter ions (from 114-117 Da) that can be used to determine the relative amounts of the peptides and thus the relative abundance of the proteins from which the peptides originated (Figure 7).

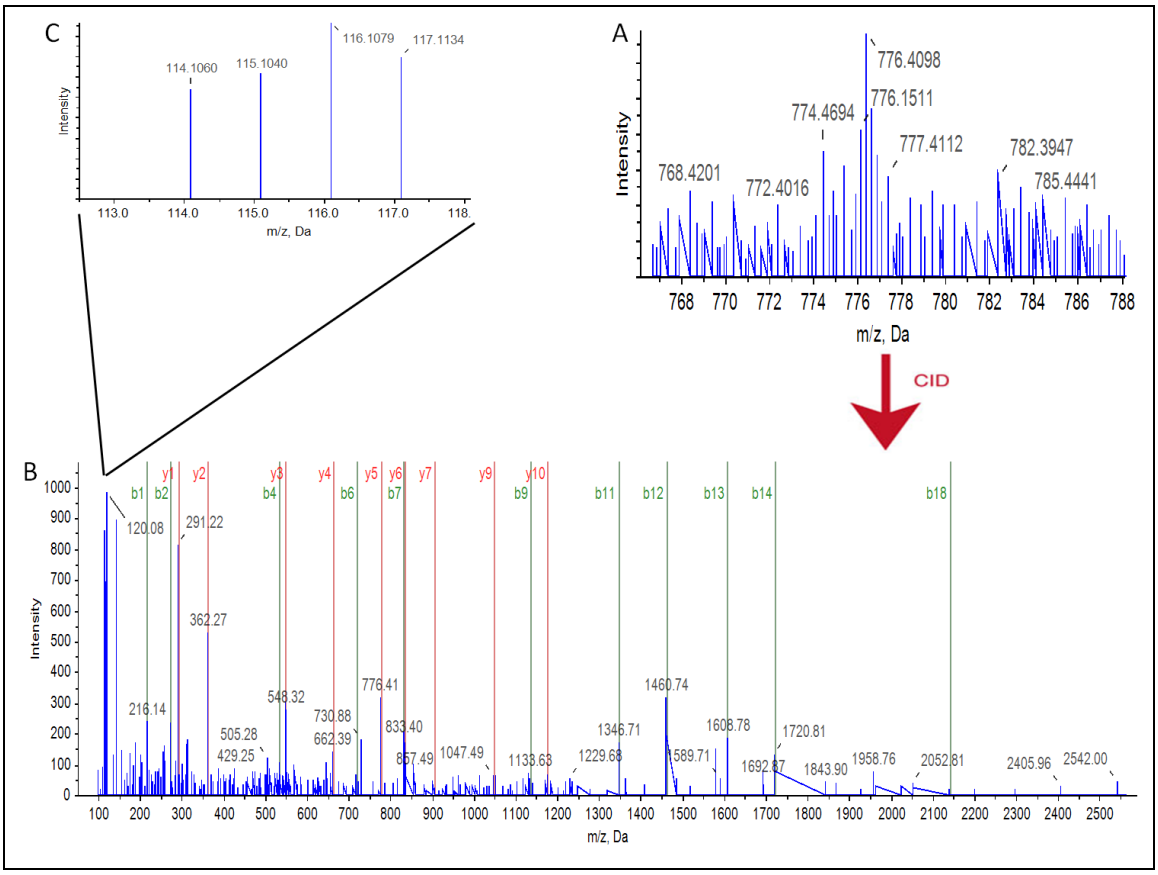


Figure 7. Example of a mass spectrum for a single iTRAQ labeled peptide
A: Precursor ion scan (MS) of a peptide labeled with the iTRAQ reagents. The differentially labeled forms are indistinguishable and contribute to the same peak. B: Product ion scan (MS/MS) of the fragmented iTRAQ-labeled peptide. The b and y ions are indicative of the peptide sequence. C: Enlarged region of the product ion spectrum, with the iTRAQ reporter ions indicating the relative levels of peptide in the four original samples. The figure is an example and was compiled from the spectra 16.1.1.3838.2 from iTRAQ replicate number one. The source protein is β-tubulin.

2.2. Methods

2.2.1. Trypanosomes and culture conditions

The four major life cycle stages of *T. congolense* clone IL3000 were prepared essentially as previously described [85] in the laboratory of Dr. Noboru Inoue, Obihiro University, Hokkaido, Japan. BSF of a cloned population of *T. congolense* IL3000 [85, 86] were obtained from the International Livestock Research Institute (ILRI, Nairobi, Kenya), formerly the International Laboratory for Research on Animal Diseases (ILRAD), and were stored in liquid nitrogen at the National Research Center for Protozoan Diseases, Obihiro University of Agriculture and Veterinary Medicine, Obihiro, Hokkaido, Japan. To grow BSF, frozen parasites were thawed and aliquots were inoculated intraperitoneally (IP) into five, 8 week old, female BALB/c mice. Infections were monitored microscopically using tail blood samples. At the first peak of parasitemia the mice were bled by cardiac puncture and trypanosomes were purified using diethyl aminoethyl (DEAE) cellulose anion exchange column chromatography [89]. All animal experiments were performed according to the standards for Care and Management of Experimental Animals at Obihiro University of Agriculture and Veterinary Medicine (No. 21-88).

Cultures of PCF, EMF and MCF trypanosomes were produced *in vitro* from the IL3000 BSF grown *in vivo* essentially following the methods of Hirumi and Hirumi [90]. BSF were adjusted to 3×10^6 cells/ml in Eagle's minimum essential medium containing 20% heat inactivated fetal bovine serum (FBS, Cat. No. SH30396.03, HyClone Laboratories Inc.) containing 2 mM L-glutamine and 10 mM L-proline. Ten mL of this suspension were incubated in T-25 type culture flasks at 27° C for approximately one week to allow

differentiation to PCF. These cultures were maintained for several weeks to allow complete differentiation of BSF to PCF. For production of EMF, cultures of PCF were allowed to become slightly acidic and overgrown and after 1 to 2 months, adherent clusters of parasites began to appear on the surfaces of the flasks. Once these adherent cells were confluent, they were washed to remove non-adherent parasites before harvesting the EMF. As judged by light microscopy and the position of the kinetoplast, these parasites were clearly EMF although a few contaminating PCF were observed. After a few weeks of sustained incubation of flasks with confluent parasites, non-adherent, VSG-expressing MCF began to appear in the supernatant.

For use in iTRAQ experiments, PCF were collected from suspension cultures of PCF growing in log-phase. EMF were collected from flasks containing confluent monolayers by first gently washing away loosely bound cells (PCF and emerging MCF) with 3 applications of 10 mL of phosphate-buffered saline containing 1% glucose (PSG), followed by gently scraping the adherent cells from the culture flasks and resuspending them in PSG. MCF parasites were purified from EMF culture supernatants by anion exchange chromatography as described above for BSF. Finally, the purified MCF parasites were used to inoculate BALB/c mice (IP, 10^5 parasites per animal). Approximately seven days after infection, the infected mouse blood was collected by cardiac puncture and the BSF purified as described above. These BSF were used to continue the life cycle through two more complete cycles in the same manner as described above. A total of 12 cell isolates (4 life cycle stages, 3 replicates of each) were obtained over a one year period. At harvest, parasites from all four life cycle stages were concentrated by centrifugation at 1,300 x g for 10 minutes at room temperature (RT) and

were washed three times with PSG to remove serum proteins present in the various media. The washed cells were examined microscopically for determination of cell counts, viability and cell morphology, prior to solubilization for iTRAQ processing (see section 2.2.2). PCF and EMF forms of *T. congolense* Il3000 were also grown in the Pearson lab at UVic for use in gel electrophoresis, immunoblotting, immunofluorescence and flow cytometry experiments.

2.2.2. Parasite collection and protein solubilization for iTRAQ analysis

Parasites in each washed cell pellet were suspended in PSG to 1,010 µL. Ten µL of the cell suspension were used for counting the parasite numbers using a Neubaur hemocytometer. The remaining parasite suspension was centrifuged (5,000 x g, 5 minutes at 4 °C) and the supernatant removed. The pellets were resuspended to a total volume of 1.0 mL in chilled lysis buffer (4.5 M urea, 0.2% sodium dodecylsulphate (SDS)) and sonicated on ice for 1 min. Lysates were frozen at -80 °C until all cell samples had been collected for iTRAQ-MS/MS analysis.

2.2.3. Protein quantitation, tryptic digestion and peptide labeling

The protein concentration of each solubilized trypanosome sample was measured using a bicinchoninic acid (BCA) kit (Cat. No. BCA1-1KT, Sigma-Aldrich) using bovine serum albumin as a protein standard. One hundred µg of protein were precipitated by adding 9 volumes of ice cold acetone and incubating overnight at 4° C. The precipitate was pelleted by centrifugation at 16,000 x g for 10 minutes and after removal of the acetone, the proteins were re-suspended in 30 µL of 0.5 M triethyl ammonium

bicarbonate containing 0.2 % SDS. The proteins were then reduced, alkylated and digested with sequencing grade modified porcine trypsin (Cat. No. V5111, Promega) as described in the iTRAQ Reagents Multiplex Kit (Cat. No. 4352135, AB SCIEX). The resulting tryptic peptides were labeled with isobaric tags according to the instructions of the manufacturer (BSF peptides were labeled with iTRAQ tag 115; PCF peptides with tag 114, EMF peptides with tag 116 and MCF peptides with tag 117) and the labeled peptides from each of the four life cycle stages were combined to form a single biological replicate. This was repeated for each set of the four life cycle stages to yield 3 peptide mixtures called replicates 1, 2 and 3, representing three complete biological replicates of the four major life cycle stages of *T. congolense.*

2.2.4. Strong cation exchange and high performance liquid chromatography

iTRAQ reagent labeled peptides were subjected to strong cation exchange (SCX) high performance liquid chromatography (HPLC) prior to MS/MS analysis. A Vision Workstation (AB SCIEX) was equipped with a polysulfoethyl A, 100 mm X 4.6 mm, 5 μm, 300 A SCX column (Poly LC). iTRAQ labeled peptide mixtures were suspended to a total volume of 2 mL in buffer A (10 mM KPO_4, 25% acetonitrile (ACN), pH 2.7) and injected onto the column. The column was allowed to equilibrate for 20 minutes in buffer A before a buffer gradient (0-35% buffer B; 10 mM KH_2PO_4, 25% ACN, 0.5 M KCl) was applied over 30 minutes at a flow rate of 0.5 mL/min. Fractions were collected at one minute intervals. The collected fractions were then reduced in volume to 125 μL in a Speed-Vac concentrator and transferred to autosampler vials (LC Packings).

2.2.5. Reverse phase liquid chromatography and tandem mass spectrometry (LC-MS/MS)

LC-MS/MS analysis was performed using an integrated Famos autosampler, Switchos II switching pump and UltiMate micro pump (LC Packings) system coupled to a Hybrid Quadrupole-time of flight (TOF) LC-MS/MS Mass Spectrometer (QStar Pulsar i; AB SCIEX) equipped with a nano-electrospray ionization source (Proxeon) and fitted with a 10 μm fused silica emitter tip (New Objective). Chromatographic separation of peptides was achieved on a 75 μm x 15 cm C18 PepMap Nano LC column (3 μm, 100 Å, LC Packings). A 300 μm x 5 mm C18 PepMap guard column (5 μm, 100 Å, LC Packings) was in place before switching in line with the analytical column and the MS. The mobile phase (solvent A) consisted of 2% ACN and 0.05% formic acid (FA) for sample injection and equilibration on the guard column at a flow rate of 100 μL/min. A linear gradient was created upon switching the trapping column inline by mixing with solvent B which consisted of 98% ACN and 0.05% FA and the flow rate was reduced to 200 nL/min for high resolution chromatography and introduction into the mass spectrometer. Twenty-seven μL of each Speed-Vac concentrated sample were injected in 95% solvent A and allowed to equilibrate on the trapping column for 10 minutes to wash away any contaminants. Upon switching inline with the MS, a linear gradient from 95% to 40% solvent A was developed for 40 minutes and in the following 5 minutes the composition of mobile phase was decreased to 20% A before increasing to 95% A for a 15 minute equilibration before the next sample injection. MS data were acquired automatically using Analyst QS 1.0 software Service Pack 8 (ABI MDS SCIEX). An information dependent acquisition method was used, consisting of a 1 second TOF-MS survey scan of mass range 400-1200 m/z and two 2.5 second product ion scans of mass range 100-1500

m/z. The two most intense peaks over 20 counts, with charge state 2-5 were selected for fragmentation and a 6 m/z window was used to prevent the peaks from the same isotopic cluster from being fragmented again. Once an ion was selected for MS/MS fragmentation it was put on an exclude list for 180 seconds. Nitrogen was used as the collision gas and the ionization tip was set to 2700 V. If the observed absorbance at 215 nm was greater than 0.1 for any fraction collected during the SCX, a 2.5 hour gradient (95-50% solvent A) was used to compensate for the higher peptide concentration in that fraction.

2.2.6. Data processing and analysis

Using ProteinPilot V2.0.1 (AB SCIEX), a semi-annotated *T. congolense* translated genome database of 13,485 ORFs [91] was trypsin digested *in silico* and the MS/MS fragmentation patterns were predicted for all resulting peptides. The tryptic peptides (and parent proteins) were identified by comparing the predicted MS/MS peptide fragmentation patterns with those observed during the iTRAQ experiments. Proteins were included in the results file if they had a total score representing ≥95% confidence of a correct identification. Further confidence about the accuracy of protein identification was gained when a protein was identified in more than one biological replicate. For more information about the ProteinPilot software see reference [92]. Technical variation among iTRAQ replicates has been well studied and is known to be minimal when properly performed [93]. The sequences for all trypanosome proteins described in this chapter can be found by their accession number at http://www.tritrypdb.org. The ProteinPilot software calculated the ratios for the iTRAQ reporter ion fragment masses for each peptide. The relative abundance of each reporter ion was used to calculate the

relative abundance of the peptide from each life cycle stage. The cumulative strength of the reporter ions from all the peptides assigned to a given protein represents the relative abundance of that protein in one life cycle stage when compared to another.

All expression changes shown in the tables and appended Excel files were normalized against the previous life cycle stage. For example, a value of 2 in a column labeled BSF→PCF means that the protein was 2 fold more abundant in PCF than BSF.

2.2.7. Gel electrophoresis and immunoblotting

Immunoblotting experiments were performed for relative quantitation of selected proteins in each life cycle stage of trypanosomes in order to support the iTRAQ data. Parasite proteins in lysates (grown as described above and representing a 4th biological replicate) were separated by SDS-polyacrylamide gel electrophoresis (PAGE) followed by transfer to polyvinylidene difluoride (PVDF) transfer membrane (Immobilon TM-P, Millipore) as previously described [94]. The primary antibodies used were: 1:2,000 dilution of ascites fluids containing murine mAbs: anti-trypanosome β-tubulin (TWP lab, unpublished) and anti-major lysosomal membrane protein p67 [95] (CLP007A, Cedarlane Laboratories, Burlington, ON) or 1:1 dilutions of hybridoma supernatants containing mouse mAbs: anti-trypanosome glycerol-3-phosphate dehydrogenase (GPD) [96], anti-GARP [94], anti-CESP [97] or anti-flagellar calcium binding protein (BE and TWP, unpublished; see Chapter 4), originally described as a *T. congolense* specific protein of unknown identity [98]. The secondary antibody used was a 1/20,000 dilution of horseradish peroxidase (HRPO) conjugated goat anti-mouse IgG/IgM (H+L) (Cat. No. 1858413; Pierce). The substrate used was SuperSignal West Dura (Cat. No. 34075,

Thermo Scientific). Kodak Biomax XAR film (Cat. No. 165 1454, Sigma-Aldrich) was used to detect chemiluminescence. After development of the autoluminograms, proteins were stained on the PVDF membrane with 0.2% nigrosin in phosphate buffered saline (PBS).

2.3. Results

2.3.1. Identification of proteins by iTRAQ

In the experiments reported in this dissertation, four iTRAQ tags were used, a different one for each life cycle stage of *T. congolense*. The labelled peptide mixtures from each of the four life cycle stages were then pooled, fractionated by SCX-HPLC and analyzed by MS/MS. Database searching was performed using the MS/MS fragmentation data to identify the corresponding source proteins. As a database, a semi-annotated proteome library, derived from the pre-publication *T. congolense* IL3000 genome, was used. The database was initially obtained through a research collaboration with Dr. Christiane Hertz-Fowler (Sanger Institute, Hinxton, UK) but since the genome sequence has been completed, it can now be accessed through reference [91]. As part of the iTRAQ differential protein expression analysis, a global list of expressed proteins was collected. Since few proteins from *T. congolense* had previously been identified or characterized, the larger list of proteins obtained by my iTRAQ experiments was used to aid the annotation of the *T. congolense* genome by providing direct protein evidence for the expression of particular open reading frames (ORF). In addition, differential protein expression data from all four *T. congolense* life cycle stages were obtained.

All four *T. congolense* major life cycle stages were produced on three separate occasions over a one year period, representing three complete, sequential life cycles (biological replicates). With each replicate, the parasite proteins were solubilized and quantitated. One hundred µg of protein from each life cycle stage were reduced, alkylated and digested with trypsin to produce peptides for iTRAQ labeling. After peptide separation by SCX and LC-MS/MS analysis, a total of 138,787 spectra were collected and 61,410 peptide sequences were determined. For each peptide sequence, a score was calculated that represents the confidence with which that peptide had been identified correctly. These peptides were then assigned to parent proteins by searching the *T. congolense* protein database and a confidence score for each protein was assigned based on the cumulative score/confidence of all the peptides assigned to it. Only those proteins with a cumulative peptide score representing ≥95% confidence were included in the iTRAQ analysis results. In this manner, 1,561, 1,400, and 1,249 proteins were identified in replicates one, two and three respectively. Of these, 831 proteins were identified in all 3 replicates, 460 were present in 2 of the 3 replicates and 797 were only observed in a single replicate. A schematic breakdown of the numbers of identified proteins is shown in Figure 8. A total of 2,088 unique protein sequences were identified, representing 23% of the ~9,000 proteins predicted for the *T. congolense* proteome.

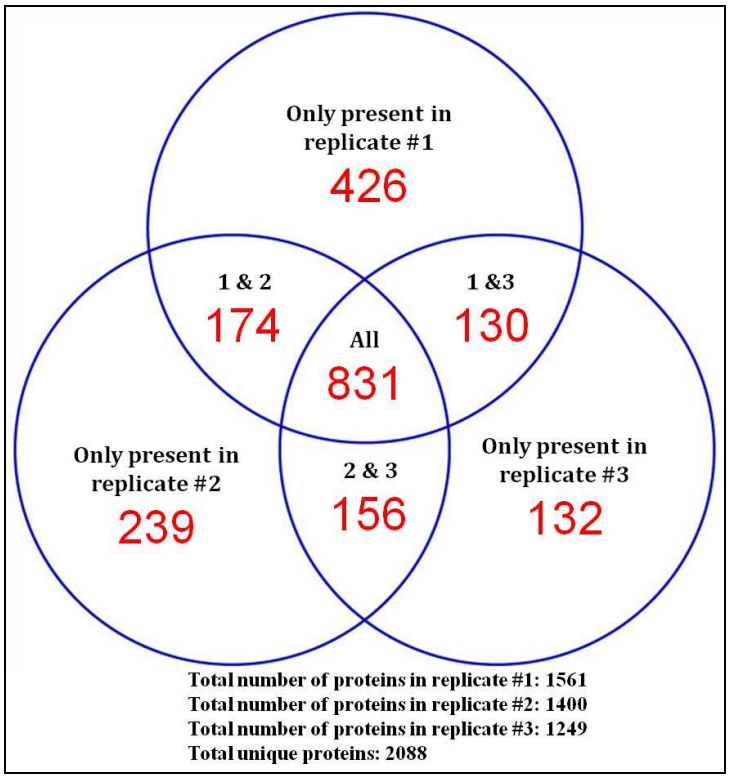


Figure 8. Distribution among the three biological replicates of iTRAQ identified *T. congolense* proteins.

For analysis of the MS/MS data by the ProteinPilot software, the DNA ORF sequences in the *T. congolense* genome database were translated into protein sequences. The genome sequence had undergone preliminary annotation thus some of the proteins identified by the iTRAQ experiments were already annotated. However, most ORFs were labelled as encoding "undefined products." To further the annotation of the *T. congolense* proteome and genome, the protein sequences identified by iTRAQ were queried by BLAST searching against the global non-redundant sequence database. Not surprisingly, several proteins drew strong hits with known, characterized proteins of kinetoplastids including *T. congolense*, *T. brucei*, *T. cruzi*, and *Leishmania*. However, many of these proteins drawing hits from other kinetoplastids, were only labeled as "hypothetical proteins," an annotation presumably resulting from genome sequencing

efforts. Peptides from such proteins were of course directly observed by my iTRAQ-MS/MS experiments meaning that the designation "hypothetical" no longer applies. The convention in the literature has been to identify a given protein in two separate iTRAQ experiments performed on biological replicates (each at ≥ 95% confidence) before protein identification can be considered significant. Therefore, to ensure that my results were reliable, only proteins that were observed in 2 or 3 replicates were considered for in depth analysis in the work reported in this dissertation. A total of 1,291 proteins, representing approximately 14% of the *T. congolense* proteome, met this criterion. These proteins are the focus of the sections described below.

Because of the large number of proteins identified and the expression data generated, only select categories of proteins will be discussed in this chapter. An Excel file with a full list of the proteins can be found in Appendix 1 (electronic form). Tab 1 of the file contains all 2,088 identified proteins, divided into lists from each of the three different biological replicates. Tab 2 is an amalgamated list of only those proteins that were identified in at least 2 biological replicates, ordered by accession number. Expression data in this tab have been averaged. Tab 3 contains the same list of proteins and data as in tab 2 but the proteins are organized by their putative functions.

2.3.2. Validation of iTRAQ expression data

Before the iTRAQ expression data were analyzed in depth, an independent method was used to determine whether the expression trends observed by iTRAQ could be corroborated. Immunoblotting was chosen to confirm some of the iTRAQ expression results because well characterized mAbs were available that specifically recognize

several of the proteins identified by the iTRAQ analysis. The immunoblot results are shown in Figure 9. The immunoblot results paralleled the iTRAQ expression data for all five proteins examined.

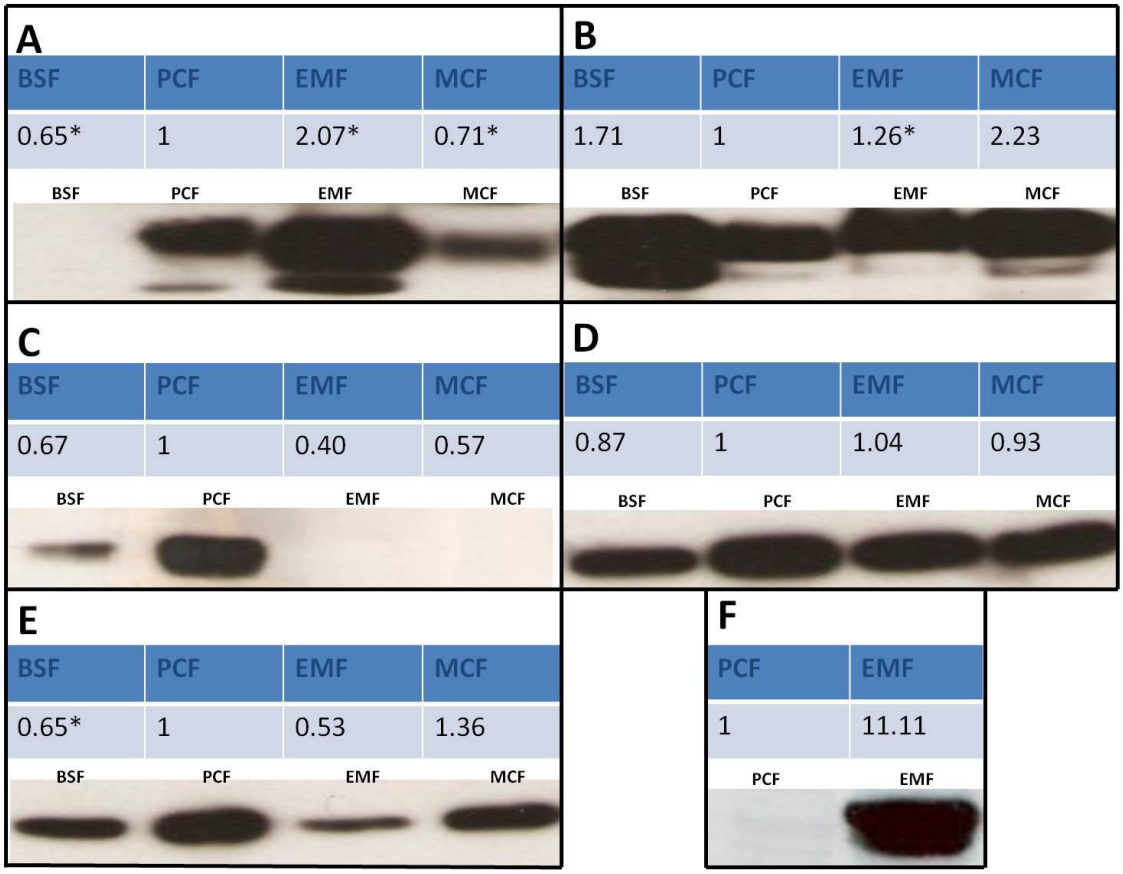


Figure 9. Immunoblot analysis of *T. congolense* cell lysates using protein specific anti-trypanosome mAbs.

The trypanosome lysates used for these blots were prepared from a fourth biological replicate of the *T. congolense* life cycle stages. Averaged iTRAQ data are included for comparison (in some cases data with a P-value > 0.05 were included; averages that have some uncertain data included are marked with an asterisk) and normalized so that the expression level in PCF always equals 1.

The mAbs used were specific for: A: glutamic acid/alanine rich protein (GARP). B: major lysosomal membrane protein, p67. C: glycerol-3-phosphate dehydrogenase (GPD). D: β tubulin. E: flagellar calcium binding protein. F: congolense epimastigote specific protein (CESP). The multiple or smeared bands observed in panels A, B and F are possibly due to glycosylation or other post-translational modifications of the proteins.

The major *T. congolense* surface molecule GARP [94, 99] was shown by iTRAQ analysis to be most strongly up regulated in EMF followed by PCF and this was confirmed by immunoblotting (Figure 9A). Transcriptome analysis [84] has shown that GARP exhibits high expression levels in EMF with lower expression in PCF and this was clearly observed here. It is interesting that the MCF continued to express low levels of GARP, perhaps indicating that although these parasites already expressed VSG they were not fully differentiated from EMF.

The major lysosomal membrane protein, p67 [95], was more highly expressed in the mammal-infective BSF and MCF parasites according to both iTRAQ and immunoblot methods (Figure 9B). The mAb used in this immunoblot recognizes a carbohydrate specific to p67. In *T. brucei* the core polypeptide is expressed in both BSF and PCF. However the antibody binds only the *T. brucei* BSF version apparently due to changes in glycosylation patterns upon differentiation to PCF. The fact that this antibody is able to bind to an antigen in *T. congolense* PCF suggests that this example of life cycle stage specific differential glycosylation is not evident in this species of trypanosome.

By iTRAQ analysis, GPD was expressed in all four life cycle stages but was expressed at a level 1.5 – 2.5 fold higher in PCF than in any other stage (Figure 9C). Only PCF and to a lesser extent, BSF, expressed significant amounts of GPD as detected by immunoblotting. Although the trends observed by iTRAQ and immunoblot are consistent with each other, these results are somewhat unexpected. GPD is an enzyme known to help maintain glycosomal redox balance during glycolysis in BSF. Although GPD has been described as having high activity in BSF, previous reports have also indicated that it is present in all life cycle stages at roughly equal levels [96].

Trypanosome β-tubulin was expressed at approximately equal levels in all four major life cycle stages as determined by both iTRAQ and immunoblotting (Figure 9D). Although the different trypanosome life cycle stages show different sizes and shapes, perhaps with different amounts of β tubulin per cell, this structural protein remains relatively constant when based on total protein as was calculated for the current iTRAQ analysis.

By both immunoblot and iTRAQ (Figure 9E), the *T. congolense* flagellar calcium binding protein was present in all 4 life cycle stages with highest levels in the PCF and MCF (see Chapter 4 for more detail about this protein).

The surface protein, CESP [40] was strikingly up regulated in EMF as determined by both iTRAQ (average of 11 fold higher in EMF compared to PCF) and by immunoblotting where it was not detected at all in the immediately preceding PCF stage (Figure 9F) nor in BSF and MCF stages (data not shown). CESP has also been measured using flow cytometry and immunofluorescence on all four life cycle stages and was found to be expressed only in EMF parasites [97].

Another, albeit indirect, means for validating the iTRAQ results can be achieved by comparison of the iTRAQ expression data with previously well established information about the VSGs which are considered a hallmark of MCF and BSF parasites that are infective for mammals. VSGs constitute the major surface coat on these life cycle stages. MCF parasites of both *T. congolense* [100] and *T. brucei* [101] express one of about 12 – 15 different VSGs (mVSGs), whereas BSF parasites can sequentially express hundreds of different VSGs, enabling them to evade the host immune response [102]. The sequence of the IL3000 VSG (for which this strain of *T. congolense* is named) and several other

mVSGs have been deduced previously from expressed sequence tag (EST) libraries of the four life cycle stages [84]. The iTRAQ data showed that eleven different VSGs were expressed in at least one of the three biological replicates used in these experiments (Table 1). Four were confidently determined to be expressed at higher levels in MCF than in BSF, two of which correspond to mVSGs 3 and 6 previously identified from ESTs in a MCF EST library [10]. Five more VSGs also showed increased expression in MCF but these data were obtained in only one replicate and may be less reliable. One of these VSGs, (TcIL3000.0.12850) is the same as the previously identified mVSG10 [84]. Two VSGs were confidently expressed at higher levels in BSF and importantly, one of which was the expected IL3000 VSG (Table 2). The other BSF VSG (TcIL3000.0.41540) has not been detected previously. The data for this VSG are believable since the VSG peptides were detected in all 3 independent biological replicates. In addition to the patterns described above, 5 VSGs showed moderately increased expression (1.45 to 1.76 fold) in EMF compared to PCF. This observation may indicate that the EMF populations used for our iTRAQ experiments were not homogenous or alternatively, that some VSG expression had started to occur in the EMF prior to the asymmetric cell division which gives rise to the MCF. Since trypanosome differentiation through the different life cycle stages occurs as a continuum, the latter is a distinct possibility.

Table 2. Expression of variant surface glycoproteins (VSGs) in *T. congolense*

***T. congolense* Accession No.**	**BLAST hit Accession No.**	**BLAST hit protein name**	**BSF → PCF**	**PCF → EMF**	**EMF → MCF**	**MCF → BSF**
TcIL3000.0.17670	emb | CAA55938.1	VSG[1] (mVSG6)	DNR[2]	DNR	8.09	0.19
TcIL3000.0.18160	gb | AAA16436.1	VSG	DNR	1.59	6.50	0.13
TcIL3000.0.24660	emb | CAA55938.1	VSG (mVSG3)	0.87	1.76	6.52	0.10
TcIL3000.0.25970	gb | AAA16962.1	VSG	0.12	DNR	8.61	DNR
TcIL3000.0.28420	gb | AAA16959.1	VSG (IL3000)	0.16	1.45	0.72	8.01
TcIL3000.0.41540	emb | CAA55938.1	VSG	0.10	1.75	1.66	4.61
TcIL3000.0.12850[3]	emb | CAA55938.1	VSG (mVSG10)	DNR	1.71	5.98	0.22
TcIL3000.0.40130	emb | CAJ16171.1	VSG	1.50	0.20	5.50	0.60
TcIL3000.0.58720	gb | AAA16961.1	VSG	0.60	DNR	6.00	DNR
TcIL3000.0.58930	emb | CAQ57406.1	VSG	0.80	0.50	2.50	0.85
TcIL3000.0.58960	emb | CAA55938.1	VSG	0.15	DNR	3.70	DNR

[1]VSG = variant surface glycoprotein. The name of the VSG is indicated in parentheses if it has been detected previously [84].
[2]DNR = data not reliable
[3]The VSGs in the shaded areas were confidently identified (detected in two or three biological replicates) but the differential expression data were reliable only for one of the replicates.

2.3.3. Discovery of new *T. congolense* proteins

Six different proteins identified by iTRAQ analysis showed no significant sequence similarity with proteins in the non-redundant database (Table 3). One of these, TcIL3000.0.22490 showed no hit at all. Unfortunately there were no reliable stage-expression data associated with this protein. A second novel protein (TcIL3000.0.28430), showed a 17.9 fold increase when MCF differentiated to BSF and then a 14-fold drop when the BSF transformed to PCF. This protein deserves examination as it is so highly expressed in BSF. Expression levels of the third protein (TcIL3000.0.38630) increased 8.5 fold during differentiation from EMF to MCF, then decreased 1.6 fold in BSF and 6.7 fold in PCF. Thus this protein appears to be mainly expressed in MCF with lowest expression in PCF. In contrast, a fourth protein (TcIL3000.0.51090) was most highly expressed in PCF. It dropped 5.3 fold during

differentiation from PCF to EMF, dropped 6 fold further in MCF and then rose 5.8 fold in BSF and 5.1 fold more in PCF. No stage-specific expression data were obtained for the fifth previously unidentified protein (TcIL3000.2.410). Finally, the sixth protein (TcIL3000.7.3440), was more highly expressed in EMF (5.44-fold up regulated from PCF) than in any other life cycle stage. All six of these molecules offer possibilities for study as none have been previously identified and several showed strong changes in expression in different life cycle stages.

Table 3. Newly identified *T. congolense* proteins

T. congolense **Accession No.**	**BLAST hit Accession No.**	**BSF → PCF**	**PCF → EMF**	**EMF → MCF**	**MCF → BSF**
TcIL3000.0.22490	No Hits[1]	DNR[2]	DNR	DNR	DNR
TcIL3000.0.28430	No Significant Hits	0.07	DNR	DNR	17.91
TcIL3000.0.38630	No Significant Hits	0.15	DNR	8.47	0.62
TcIL3000.0.51090	No Significant Hits	5.05	0.19	0.17	5.75
TcIL3000.2.410	No Significant Hits	DNR	DNR	DNR	DNR
TcIL3000.7.3440	No Significant Hits	DNR	5.44	0.18	DNR

[1]No sequence similarity with any protein sequence in the non-redundant database.
[2]DNR = data not reliable

2.3.4. Membrane proteins

It is not surprising that several of the *T. congolense* surface membrane proteins known to exhibit stage-specific expression based on immunological detection also showed stage-specific expression revealed by iTRAQ analysis. First, in MCF and BSF trypanosomes that are infective for mammals, a total of 11 different VSGs were detected (see section 2.3.2 and Table 2). In addition to VSG expression, four proteins with sequence similarity to invariant surface glycoprotein (ISG) 65 and ISG75 [103] of *T. b. brucei* were found (the two with reliable expression data are shown in Table 4). Analysis of the *T. b. brucei* molecules revealed that both amino acid sequences predict polypeptides with N-terminal signal sequences and large extracellular domains [103]. These glycoproteins have no sequence similarity to known proteins in non-trypanosomatids and after VSGs are the most abundant surface proteins in mammal infective forms. The *T. brucei* ISGs have been shown to be spread over the entire cell surface, although they are present at only 50,000 - 70,000 copies per cell, approximately 0.5% of the VSG copy number [103]. These glycoproteins have been proposed to be dispersed between VSG molecules on the trypanosome surface [104]. The genes encoding ISG75 in at least one strain of *T. b. brucei* exist as seven copies at two loci, suggesting that different isoforms may be important for the survival of the parasite [104]. By iTRAQ study, the *T. congolense* ISGs were revealed to be most up regulated in MCF. This finding is novel. *T. b. brucei* MCF have not been previously examined for ISG expression and neither had *T. congolense* MCF.

Table 4. iTRAQ expression data for select *T. congolense* cell surface proteins.

***T. congolense* Accession No.**	**BLAST hit Accession No.**	**BLAST hit protein name**	**BSF → PCF**	**PCF → EMF**	**EMF → MCF**	**MCF → BSF**
TcIL3000.0.06380	Tb927.8.1610	GP63[1]	DNR[2]	5.03	0.20	DNR
TcIL3000.0.15620	dbj|BAF91109.1	CESP[3]	DNR	11.11	0.17	DNR
TcIL3000.0.22060	Tb927.2.3300	ISG[4] 65	0.32	DNR	8.48	0.29
TcIL3000.0.45520	gb|ABB22183.1	ISG 75	DNR	DNR	4.73	0.64
TcIL3000.0.56730	Tb927.1.2880	pteridine transporter	DNR	10.51	0.18	DNR
TcIL3000.10.9440	gb|AAA30244.1	PSSA-2[5]	2.23	4.61	0.59	0.30
TcIL3000.10.2930	Tb927.6.440	Hp-Hb receptor[6]	DNR	11.46	0.17	0.86
TcIL3000.10.13970	Tb927.4.3990	amino acid transporter	DNR	17.43	0.05	DNR

[1]Major surface metalloprotease GP63
[2]DNR = data not reliable
[3]CESP = congolense epimastigote specific protein
[4]ISG = invariant surface glycoprotein
[5]PSSA-2 = procyclic stage surface antigen
[6]Hp-Hb = haptoglobin-hemoglobin

Of the previously known *T. congolense* surface molecules, the most strikingly up regulated (in EMF) was CESP. Supporting the original report that this is an epimastigote specific protein [40], the iTRAQ data shows that CESP was up regulated more than 11 fold after differentiation to EMF from PCF (Table 4). It has been hypothesized that CESP may be involved in adherence of EMF to the mouthparts of tsetse [40]. The high levels of CESP observed in EMF and its detection on the cell surface lend support to this idea and suggest that CESP structure-function relationships should be examined as a priority for understanding tsetse-trypanosome interactions.

The iTRAQ data indicated that another well known surface molecule of insect forms of *T. congolense*, GARP [94, 99] was up-regulated more than two fold in EMF although this was only seen in one replicate. Considering that by immunoblotting (Figure 9A) GARP was easily detectable in both PCF and EMF (and most highly expressed in EMF), the poor quality of the iTRAQ results imply that the peptides either were not released by

trypsin cleavage or did not ionize well enough for reliable mass spectrometric analysis. The pattern of GARP expression observed with both methods was consistent with data reported previously [37] showing that GARP was expressed early after differentiation to PCF from BSF after which it decreased on established PCF, then was expressed to even higher levels in EMF.

In *T. b. brucei*, Jackson et al. [105] used cDNA expression to identify a protein they called procyclic stage surface antigen 2 (PSSA-2), also annotated as "surface glycoprotein". This new and previously unidentified stage-specific surface antigen had the features of a typical transmembrane glycoprotein but with an unusual cytoplasmic extension composed of a proline-rich tandem repeat. My iTRAQ analysis revealed that a similar molecule was expressed in *T. congolense* (TcIL3000.10.9440) and that this molecule was up regulated in EMF (2.23 fold increase in PCF from BSF and 4.61 fold increase in EMF from PCF) (Table 4). The *T. congolense* version of this protein has 61% sequence identity and 72% similarity to the *T. b. brucei* molecule. This protein is very unusual. Throughout the sequence there are numerous instances of the same amino acid repeated two (VV, FF, CC, GG, SS, DD, RR, TT, EE, QQ, LL, AA) or three (VVV, RRR, SSS, GGG) times in a row. Similar to the *T. b. brucei* protein, the final fifth of this *T. congolense* protein is proline rich and has several, di-, tri- and tetra-proline repeats. This protein has been named congolense insect stage surface antigen (CISSA) and is discussed further in Chapter 3.

Another family of trypanosome surface proteins previously reported in both *T. congolense* [106] and *T. brucei* sspp. [107] are the major surface metalloproteases (MSP), also called GP63. MSP have been most extensively studied in *T. b. brucei*, which has a

minimum of three classes of MSP genes, MSP-A, -B and -C, based on sequence differences and differential expression during the *T. b. brucei* life cycle [108]. All three classes have been shown by northern blot analysis to be transcribed in BSF, whereas only MSP-B RNA was detected in PCF. MSP-B protein is absent in BSF, but occurs early in BSF→PCF differentiation and is known to function in removal of VSG during this transition [109]. The functions of MSP-A and C are unknown. The *T. congolense* genome has at least five MSP gene classes (MSP-A to –E), the first three of which correspond to the three known classes in *T. b. brucei* [106]. The sequence differences in the MSP classes of both organisms suggest that each class has distinct substrates and biological functions. The MSP identified by iTRAQ in *T. congolense* falls into the MSP-B group and was, surprisingly, most highly expressed in EMF where it was 5.0 fold higher than in PCF. This is the first demonstration of a MSP in the EMF of either trypanosome species and represents an interesting area for further investigation.

Two surface transport molecules showed a striking expression profile. A putative amino acid transporter (TcIL3000.10.13970) and a putative pteridine transporter (TcIL3000.0.56730) were both drastically up-regulated in EMF (17 and 11 fold respectively) followed by a sharp decrease upon transition to MCF (20 and 6 fold respectively). This increase in expression of these transporters in the EMF may reflect the parasite's response to the scarce supply of nutrients in the tsetse mouthparts.

Another previously identified surface molecule is the *T. congolense* procyclin analogue, heptapep, that has been reported to be expressed continuously on procyclic forms in the tsetse midgut [36]. At first it was surprising that this abundant protein was not detected by the iTRAQ experiments since it was clearly expressed on the surface of

both PCF and EMF as determined by flow cytometry and immunofluorescence microscopy using a mAb that recognized the heptapeptide repeat [97]. However, in *T. congolense* IL3000 this protein contains only four trypsin cleavage sites. Additionally, trypsin cleavage would only release two peptides with the appropriate masses for MS detection. Both peptides are quite hydrophobic and each has a glutamate that could cancel the charge from the arginine/lysine potentially inhibiting their ionization. The absence of this well known protein from my MS results highlights one of the shortcomings of this technique. If a protein cannot be efficient digested into peptides amenable to ionization and subsequent detection by MS, then it will not be identified.

The iTRAQ data showed a marked up regulation of a haptoglobin-hemoglobin receptor in EMF (11.5 fold increase from PCF – Table 3). This can perhaps be explained by the requirement for this form of the parasite to scavenge heme in the nutrient scarce environment of the tsetse mouthparts. In *T. brucei* sspp., this receptor normally binds the haptoglobin-hemoglobin complex at high affinity, allowing uptake of heme and supporting growth of the parasites. In humans infections, this receptor also binds HRP that is part of the HDL particles containing trypanolytic ApoL1 [45]. Thus this haptoglobin-hemoglobin receptor binds HDLs and contributes to killing of susceptible parasites. Its role in *T. congolense* has not been studied but presumably, if expressed in bloodstream forms as expected, it would explain in part why this cattle pathogen does not infect humans.

2.3.5. Metabolic enzymes

The iTRAQ experiments described here were successful in detecting many metabolic enzymes. A detailed discussion of trypanosome metabolism is beyond the scope of this dissertation but discernible trends in protein expression will be noted. Also, *T. congolense* metabolism is not well studied, thus for this reason the known protein expression trends discussed in this section are based primarily on studies from *T. brucei* sspp. [78, 110].

Most glycolytic enzymes were observed to be up regulated in BSF and MCF (Table 5) although several exceptions to this trend were found. The first example is hexokinase, which catalyzes the first, although non-committed step of glycolysis. This enzyme was at highest levels in PCF, an observation that might be explained by the presence of the pentose phosphate pathway which is also compartmentalized in the glycosome. It is possible that hexokinase in PCF parasites provides substrates for that pathway rather than glycolysis. Three other glycolytic proteins which are expressed at higher levels in PCF than BSF are triose phosphate isomerise (TIM), an isoform of glyceraldehyde 3-phosphate (G3P) dehydrogenase and phosphoglycerate kinase. The reason for this discontinuity with the rest of the glycolytic enzymes is unknown but it is important to remember that the names attached to these proteins come from sequence similarity by BLAST searching and the assigned functions are entirely putative.

Table 5. Putative glucose metabolism enzymes identified by iTRAQ

T. congolense Accession No.	BLAST hit Accession No.	BLAST hit protein name	BSF → PCF	PCF → EMF	EMF → MCF	MCF → BSF
TcIL3000.0.38930	Tb10.70.5820	hexokinase	1.77	0.42	1.76	DNR[1]
TcIL3000.0.38940	emb|CAC69958.1	hexokinase	1.66	0.53	1.74	DNR
TcIL3000.0.29570	Tb927.1.3830	G6P[2] isomerase, glycosomal	0.31	1.52	1.68	1.26
TcIL3000.0.42400	Tb927.3.3270	phosphofructokinase	DNR	0.42	1.89	1.19
TcIL3000.10.4760	Tb10.70.1370	FBP[3] aldolase glycosomal	0.54	DNR	1.32	1.37
TcIL3000.11.5810	Tb11.02.3210	triose phosphate isomerase	2.16	0.56	DNR	DNR
TcIL3000.6.3720	gb|AAX79281.1	G3P[4] dehydrogenase, glycosomal	1.48	0.56	DNR	1.27
TcIL3000.10.5910	Tb10.6k15.3850	G3P dehydrogenase, cytosolic	0.7	0.46	2.52	1.24
TcIL3000.1.220	gb|AAC37222.1	phosphoglycerate kinase	0.49	DNR	3.17	0.73
TcIL3000.1.230	gb|AAC37222.1	phosphoglycerate kinase	4.33	0.21	1.74	DNR
TcIL3000.11.2200	Tb11.22.0003	phosphoglycerate kinase	DNR	DNR	DNR	DNR
TcIL3000.10.6800	Tb10.6k15.2620	phosphoglycerate mutase	DNR	0.39	1.59	2.66
TcIL3000.10.6810	Tb10.6k15.2620	phosphoglycerate mutase	DNR	0.33	1.88	2.05
TcIL3000.10.2500	Tb10.70.4740	enolase	DNR	0.34	2.15	1.13
TcIL3000.10.12020	emb|CAA41018.1	pyruvate kinase	0.45	DNR	1.92	1.39
TcIL3000.8.3450	Tb927.8.3530	Gly3P[5] dehydrogenase, glycosomal	1.54	0.4	1.46	1.22
TcIL3000.0.55170	dbj|BAD06715.1	glycerol kinase	0.65	0.54	1.43	2.02

[1]DNR = data not reliable
[2]G6P = glucose-6-phosphate
[3]FBP = fructose-bisphosphate
[4]G3P = glyceraldehyde 3-phosphate
[5]Gly3P = glycerol-3-phosphate

All components of the pyruvate dehydrogenase complex (PDC) were identified (Table 6). It was expected that they would be up regulated in PCF and EMF and down regulated in MCF and BSF. This was hypothesized because pyruvate is an excreted end product of glycolysis for MCF and BSF. Only PCF and EMF convert pyruvate to acetyl-CoA for entry into the citric acid cycle (CAC) and oxidative phosphorylation. The expression data for the PDC were surprising. The results show exactly the opposite of what was supposed. For those proteins with reliable data, expression was lowest in PCF and rose through the rest of the life cycle to peak in the BSF. The reason for this is unknown but merits further study.

Three of the four enzymes involved in the oxidative, NADPH generating, phase of the pentose phosphate pathway were identified (Table 7) [110]. Little expression data were obtained for these enzymes. The only reliable information indicates that the first committed step of the pathway (glucose-6-phosphate to 6-phosphogluconolactone catalyzed by glucose-6-phosphate dehydrogenase) was down regulated when MCF differentiated to BSF. Only two enzymes from the non-oxidative phase (transketolase and transaldolase) were identified. They were both up regulated when BSFs differentiated to PCFs but were down regulated when PCFs differentiated to EMFs and then further down regulated when MCFs transformed to BSFs. The non-oxidative phase is involved in recycling 6 molecules of ribose-5-phosphate (the product of the oxidative phase) to 5 molecules of glucose-6-phosphate. Trypanosomes typically scavenge nucleotides from the environment so the up regulation of transketolase and transaldolase may represent a need for NADPH for reductive anabolism rather than using ribose-5-phosphate for *de novo* nucleotide synthesis.

Table 6. Putative components of the pyruvate dehydrogenase complex identified by iTRAQ

T. congolense Accession No.	BLAST hit Accession No.	BLAST hit protein name	BSF → PCF	PCF → EMF	EMF → MCF	MCF → BSF
TcIL3000.3.890	Tb927.3.1790	pyruvate dehydrogenase E1 beta	DNR[1]	DNR	DNR	DNR
TcIL3000.3.910	Tb927.3.1790	pyruvate dehydrogenase E1 beta	DNR	DNR	DNR	DNR
TcIL3000.10.10810	Tb10.389.0890	pyruvate dehydrogenase E1 alpha	0.49	1.3	1.53	1.25
TcIL3000.10.1950	Tb10.70.5380	dihydrolipoamide acetyltransferase	DNR	1.72	DNR	1.85
TcIL3000.10.6510	Tb10.6k15.3080	dihydrolipoamide acetyltransferase	0.51	1.45	DNR	1.35
TcIL3000.11.16620	Tb11.01.8470	dihydrolipoyl dehydrogenase	0.7	DNR	DNR	DNR
TcIL3000.8.7590	Tb927.4.5040	dihydrolipoamide dehydrogenase	DNR	DNR	DNR	DNR

[1]DNR = data not reliable

Table 7. Putative enzymes of the pentose phosphate pathway identified by iTRAQ

T. congolense Accession No.	BLAST hit Accession No.	BLAST hit protein name	BSF → PCF	PCF → EMF	EMF → MCF	MCF → BSF
TcIL3000.10.2120	Tb10.70.5200	G6P[1] dehydrogenase	DNR[1]	DNR	DNR	0.63
TcIL3000.9.4990	Tb09.211.3180	phosphogluconate dehydrogenase	DNR	DNR	DNR	DNR
TcIL3000.0.48580	Tb11.01.0700	ribose 5-phosphate isomerase	DNR	DNR	DNR	DNR
TcIL3000.8.6020	Tb927.8.6170	transketolase	5.14	0.52	0.54	DNR
TcIL3000.0.10000	Tb927.8.5600	transaldolase	5.59	0.29	1.6	0.43

[1]G6P = glucose-6-phosphate
[2]DNR = data not reliable

All enzymes of the CAC were observed by iTRAQ analysis (Table 8). This includes the 8 classical enzymes as well as an NADH-dependent fumarate reductase which catalyzes the conversion of fumarate to succinate, in the opposite direction of the normal flow of the cycle. All enzymes with an associated confidence expression change were consistently up regulated in PCF and EMF in agreement with previous literature. Some enzymes, such as an isocitrate dehydrogenase and succinate dehydrogenase were strongly up regulated (13.6 fold in EMF and 8.8 fold in PCF respectively). These enzymes are interesting by virtue of their strong expression profiles. The only exception to this was succinyl-CoA ligase which was 1.5 times lower in PCF than BSF. Enzymes required to channel proline and threonine into the CAC were also observed to be up regulated in the PCF and EMF (Table 8). All available expression data show that these enzymes for amino acid catabolism were up regulated in the PCF and EMF as expected for the life cycle stages that rely on proline as the primary energy source.

Seventeen proteins involved in aerobic respiration and the electron transport chain were identified by iTRAQ, including the alternative oxidase mentioned in section 1.2. As described in previous reports, the alternative oxidase showed peak expression in the BSF where it transfers high energy electrons, generated by glycolysis, directly to oxygen with no coupled energy capture or ATP synthesis [24, 25]. The remaining 16 proteins were all up regulated in the amino acid oxidizing PCF and EMF. Twelve proteins are putative members of the electron transport chain and 4 are components of the ATP synthase complex.

Table 8. *T. congolense* enzymes involved in the citric acid cycle and amino acid metabolism identified by iTRAQ

***T. congolense* Accession No.**	**BLAST hit Accession No.**	**BLAST hit protein name**	**BSF → PCF**	**PCF → EMF**	**EMF → MCF**	**MCF → BSF**
TcIL3000.10.11540	Tb10.05.0150	citrate synthase	DNR[1]	2.72	DNR	DNR
TcIL3000.10.11690	Tb10.61.2880	aconitase	2.4	2.46	DNR	0.25
TcIL3000.0.43620	Tb927.8.3690	isocitrate dehydrogenase	DNR	DNR	DNR	DNR
TcIL3000.0.50590	Tb11.03.0230	isocitrate dehydrogenase	DNR	13.61	0.49	0.19
TcIL3000.11.10480	Tb11.01.1740	2-oxoglutarate dehydrogenase E1	2.27	1.72	0.54	0.45
TcIL3000.11.12290	Tb11.01.3550	2-oxoglutarate dehydrogenase E2	2.23	2.73	0.48	0.47
TcIL3000.11.1340	Tb11.47.0004	2-oxoglutarate dehydrogenase subunit	2.61	2.72	0.36	0.52
TcIL3000.3.1350	Tb927.3.2230	succinyl-CoA synthetase alpha subunit	DNR	DNR	DNR	DNR
TcIL3000.10.6380	Tb10.6k15.3250	succinyl-CoA ligase beta-chain	0.64	1.22	DNR	1.35
TcIL3000.8.6450	TcCLB.511909.40	succinate dehydrogenase flavoprotein	8.76	1.99	0.44	0.22
TcIL3000.9.2010	Tb09.160.4380	succinate dehydrogenase	4.26	DNR	0.53	0.4
TcIL3000.11.5170	Tb11.02.2700	fumarate hydratase class I	DNR	DNR	DNR	DNR
TcIL3000.10.2190	TcCLB.506195.110	malate dehydrogenase	2.62	1.45	DNR	0.34
TcIL3000.0.22370	Tb927.5.940	NADH-dependent fumarate reductase	3.98	0.32	DNR	0.54
TcIL3000.0.07190	Tb927.7.210	proline oxidase	DNR	1.82	0.56	DNR
TcIL3000.10.2660	Tb10.70.4280	DPC[2] dehydrogenase	2.46	1.55	0.48	0.6
TcIL3000.9.1990	Tb09.160.4310	glutamate dehydrogenase	DNR	2.28	0.5	0.7
TcIL3000.6.2270	Tb927.6.2790	L-threonine 3-dehydrogenase	6.04	0.31	0.5	DNR
TcIL3000.8.5850	Tb927.8.6060	AKB coenzyme A ligase	DNR	DNR	DNR	DNR

[1]DNR = data not reliable
[2]DPC = delta-1-pyrroline-5-carboxylate
[3]AKB = 2-amino-3-ketobutyrate

Table 9. Proteins involved in mitochondrial electron transport and oxidative phosphorylation identified by iTRAQ

T. congolense Accession No.	BLAST hit Accession No.	BLAST hit protein name	BSF → PCF	PCF → EMF	EMF → MCF	MCF → BSF
TcIL3000.10.6090	dbj|BAE16575.1	alternative oxidase	0.13	1.77	2.42	2.85
TcIL3000.9.6080	Tb09.211.4700	reiske iron-sulfur protein	3.85	DNR[1]	DNR	0.45
TcIL3000.0.51330	Tb927.5.1470	NADH-cytochrome b5 reductase	DNR	DNR	DNR	DNR
TcIL3000.11.15610	Tb11.01.7190	NADH-cytochrome b5 reductase	DNR	1.95	DNR	DNR
TcIL3000.7.2050	Tb927.7.2700	NADH-cytochrome b5 reductase	2.06	DNR	0.56	DNR
TcIL3000.8.1840	Tb927.8.1890	cytochrome c1, mitochondrial	5.28	1.69	0.84	0.26
TcIL3000.8.4910	Tb927.8.5120	cytochrome c	5.86	1.57	DNR	0.22
TcIL3000.1.1680	Tb927.1.4100	cytochrome C oxidase subunit IV	2.52	1.49	0.76	0.41
TcIL3000.0.52640	Tb09.160.1820	cytochrome c oxidase subunit V	3.63	1.49	DNR	0.35
TcIL3000.10.140	Tb10.100.0160	cytochrome C oxidase subunit VI	DNR	DNR	DNR	DNR
TcIL3000.3.630	Tb927.3.1410	cytochrome c oxidase VII	3.44	1.56	DNR	0.28
TcIL3000.4.4020	Tb927.4.4620	cytochrome c oxidase VIII	1.92	DNR	DNR	0.65
TcIL3000.10.7180	Tb10.6k15.2180	cytochrome c oxidase subunit IX	3.05	DNR	DNR	0.28
TcIL3000.7.6050	Tb927.7.7420	ATP synthase alpha chain	2.59	1.49	DNR	0.32
TcIL3000.3.590	Tb927.3.1380	ATP synthase beta chain	2.66	1.48	0.65	0.36
TcIL3000.10.50	Tb10.100.0070	ATP synthase F1 subunit gamma	2.41	1.59	0.87	0.34
TcIL3000.6.4420	Tb927.6.4990	ATP synthase, epsilon chain	DNR	DNR	DNR	0.39

[1]**DNR = data not reliable**

2.4. Discussion

The work described in this chapter was aimed at determining protein expression through the life cycle of *T. congolense*. It is not feasible to isolate procyclic forms, epimastigote forms or metacyclic forms of trypanosomes of any species from infected tsetse in numbers sufficient for analysis of their proteins by quantitative mass spectrometry. Thus *T. congolense* IL3000 was used since all four life cycle stages can be grown *in vitro* in quantities sufficient for analysis by protein mass spectrometry. By using iTRAQ-MS/MS analysis it was possible to reliably identify a total of 1291 trypanosome proteins, representing ~14% of the predicted *T. congolense* proteome. The data revealed changes in the expression level of many conserved metabolic proteins, corroborating previous knowledge and revealing unexpected new trends. Much more interesting was the identification of several previously unknown proteins and the determination of drastic differential expression of several surface proteins and glycoproteins, both of known and unknown function. This study represents the first large scale sequence-driven proteomic analysis of all four major life cycle stages for any trypanosome species. A similar tagging and MS approach (stable isotope labeling by amino acids in cell culture, SILAC) is being used to study PCF and BSF of *T. b. brucei* but is yet unpublished (personal communication, K. Gunasekera, University of Bern, 2012). It will be interesting to find any similarities or differences between *T. congolense* and *T. b. brucei* protein expression patterns once this work is published.

Only a few of the proteins identified by my research have been discussed in this dissertation. Also observed were large groups of other metabolic proteins, ribosomal proteins, chaperones, proteases/peptidases, nucleic acid binding proteins,

kinases/signaling proteins, structural and motor proteins. Uncharacterized "hypothetical proteins" account for approximately half of all the proteins identified. As a starting point for future thought, Table 10 contains a list of all proteins (which have not already been included in other tables) that show at least a 10-fold change in expression between any two life cycle stages. Ultimately, any researcher wishing complete information should view the supplementary Excel sheet (Appendix 1). The proteins discussed in the above sections are those of most interest to our lab or that showed large expression changes. However, many proteins that exhibit much lower differential expression may be of interest and of some biological significance. The data will provide a valuable resource for exploring trypanosome differentiation and/or host-parasite interactions.

Table 10. *T. congolense* proteins showing at least a 10-fold expression change by iTRAQ.

***T. congolense* Accession No.**	**BLAST hit Accession No.**	**BLAST hit protein name**	**BSF → PCF**	**PCF → EMF**	**EMF → MCF**	**MCF → BSF**
TcIL3000.0.02370	Tb927.7.440	hypothetical protein	DNR[1]	13.11	0.07	DNR
TcIL3000.0.50300	Tcclb.511253.31	kinesin-like protein	DNR	18.53	0.15	DNR
TcIL3000.10.10170	Tb10.389.1810	kynurenine aminotransferase	10.81	0.16	DNR	DNR
TcIL3000.11.8970	Tb11.01.0110	hypothetical protein	DNR	12.62	0.14	DNR
TcIL3000.7.5580	Tb927.7.6920	hypothetical protein	DNR	12.58	0.11	DNR

[1]DNR = data not reliable

2.4.1. Future Work

The data presented in this chapter are intended to provide both identification and quantification data for *T. congolense* proteins which can be used as a resource for more detailed study of individual proteins and a better understanding of *T. congolense* biology. Several of the proteins mentioned in this chapter have been studied in more depth and are the topics of Chapter 3 and Chapter 4.

Chapter 3. Characterization of Selected *Trypanosoma congolense* Cell Surface Proteins

The protein purification and X-ray crystallography described in this chapter were performed by Sean Workman, Michelle Tonkin and Dr. Marty Boulanger from the University of Victoria. All other experiments and monoclonal antibody derivation were designed by Brett Eyford and performed with the assistance of Jessica Fudge, Richard Yip and Bianca Loveless.

The X-ray crystallography work has been published as:

Tonkin ML, Workman SD, Eyford BA, Loveless BC, Fudge JL, Pearson TW, Boulanger MJ. 2012. Purification, crystallization and X-ray diffraction analysis of Congolense Insect Stage Surface Antigen (TcCISSA) from *Trypanosoma congolense*. Acta Crystallographica F68 (12).

3.1. Introduction

In Chapter 2, I presented data that can be used by any trypanosome researcher both as supporting evidence for current research and as a starting point for new lines of study. For example, the Pearson lab has a particular interest in the surface membrane proteins of trypanosome insect stages (PCF and EMF) and continuing work in this domain is anticipated. Notable achievements from this lab include the discovery of the first *T. brucei* procyclin [111] and discovery of GARP, a major surface molecule of *T. congolense* PCF and EMF and originally described as the procyclin equivalent [94, 99]. In addition, epitope mapping of the GARP molecule and its crystal structure were determined in collaboration with Dr. Marty Boulanger's lab at UVic [112]. To carry on this general area of research into discovery and characterization of trypanosome molecules, I searched the iTRAQ data obtained in my dissertation research for novel or understudied predicted surface proteins that showed differential expression between life cycle stages of *T. congolense*. In this chapter I describe studies on two such proteins: one that I have named, congolense insect stage surface antigen (CISSA) and the novel protein, TcIL3000.7.3440 (abbreviated Tc3440), both of which were shown by iTRAQ analysis to be up regulated in insect stages of *T. congolense*.

CISSA (TcIL3000.10.9440, Table 4 in Chapter 2) drew my attention as a potential surface protein due to its relatively non-descriptive database annotation of "surface glycoprotein". CISSA is also interesting because, by iTRAQ, it reaches the highest levels of expression in PCF and EMF and was one of the few proteins that had reliable expression data for all four *T. congolense* major life cycle stage transitions. Table 11 is an excerpt of Table 4, showing the iTRAQ expression data for CISSA.

Table 11. iTRAQ expression data for CISSA

T. congolense Accession No.	BLAST Hit Accession No.	BLAST Hit Protein Name	BSF → PCF	PCF → EMF	EMF → MCF	MCF → BSF
TcIL3000.10.9440	gb|AAA30244.1	PSSA-2[1] *(T. brucei)*	2.23	4.61	0.59	0.30

[1]**PSSA-2 = procyclic stage surface antigen**

CISSA had no useful, conserved structural motifs as determined by InterProScan [113] (data not shown) so functional prediction through sequence homology analysis provided little insight. Based on its primary amino acid sequence, CISSA is predicted to be an integral membrane protein. The protein appears to have a signal peptide for the secretory pathway, an N-terminal extracellular domain, a single transmembrane helix and a C-terminal cytoplasmic domain (Figure 10). Closer inspection of the amino acid sequence of CISSA showed this to be an unusual protein. There are numerous di- and tri-amino acid repeats (DD, EE, GG, QQ, RR, SS, TT, GGG, RRR, SSS, and VVV) found throughout the N-terminal region. The C terminal domain is mostly comprised of proline rich repeats; four $GAAYGQ(P)_4$ and three $YGQ(P)_{2\text{-}4}$ (Figure 10).

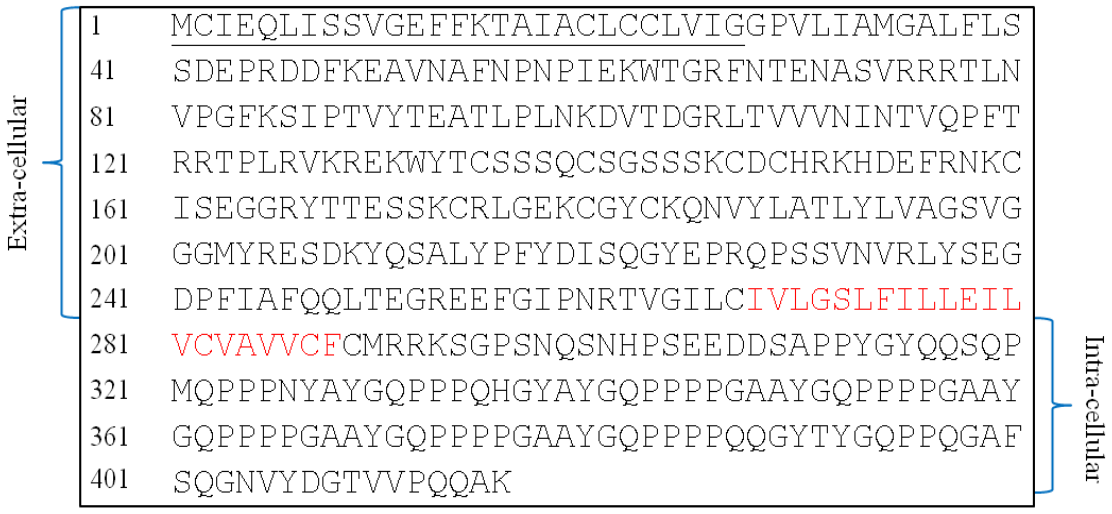


Figure 10. CISSA protein sequence
The predicted transmembrane helix is highlighted in red and the predicted signal peptide is underlined.

CISSA shares 61% identity and 72% similarity to the "surface glycoprotein" (also known as PSSA-2) from *T. brucei* (Figure 11). Only two research articles have been published regarding PSSA-2. The first report, from 1993, describes the discovery of PSSA-2 by screening an expressed *T. brucei* PCF cDNA library with antiserum made against *T. brucei* PCF plasma membrane extracts [105]. At that time, the authors were unable to purify or further characterize PSSA-2 but did find that mRNA for this gene was only detectable in PCF and not BSF, mirroring the iTRAQ results for CISSA as an insect stage protein.

Figure 11. Amino acid sequence alignment between *T. brucei* PSSA-2 and *T. congolense* CISSA.
Accession numbers for these proteins are: *T. brucei* PSSA-2 (AAA30244.1) [105] and *T. congolense* CISSA (TcIL3000.10.9440). Shared amino acids are shown in blue and similar amino acids are shown in grey.

In 2009, Fragoso et al., [114] studied PSSA-2 using immunoassays with genetically modified parasites. It was confirmed that PSSA-2 was expressed in PCF and EMF of

T. b. brucei but not in the mammal infective stages (BSF or MCF). The predicted topology of the protein (extra-cellular N terminal, intra-cellular C terminal) was also supported. It was shown that the cytoplasmic domain was essential for the transit of PSSA-2 from the endoplasmic reticulum to the cell surface. This localization was attributed to the MAP kinase-4 phosphorylation of a threonine-proline motif at residue 305-306. There was evidence to suggest that this phosphorylation directs a further, unidentified, post-translational modification and possibly multimerization. It should be noted that in CISSA, the threonine-proline phosphorylation site has been replaced with a pair of glutamic acids. Therefore, if CISSA localizes in a similar manner to PSSA-2, the glutamates may act as a functional mimic to phosphorylation.

Beyond the localization experiments, Fragoso et al. also studied the phenotype resulting from PSSA-2 knock out and C-terminal tail deletions. Both types of mutation did not apparently impact the growth, motility or morphology of these cells. Neither mutation had an effect on the parasite's ability to establish midgut infections although it was found that the mutation/knock out of PSSA-2 did have an impact on progression to the EMF, with mutants giving rise to fewer and weaker salivary gland infections. It is not known whether the mutations resulted in decreased migration from the midgut to the salivary gland, in decreased PCF to EMF differentiation or in decreased survival rates of established EMF. While the function of PSSA-2 is unknown, the fact that it is an integral membrane protein with both intra- and extracellular domains suggests that this protein may act as cytoskeletal scaffold protein or a receptor/sensor.

The antibodies used in the previous study [114] to study PSSA-2 were raised against the haemagglutinin tag (HA) added to the recombinant molecule. Therefore all

measurements of PSSA-2 required genetic modification in order to introduce the HA-PSSA-2 fusion protein. This method is time consuming and limited to laboratories with expertise in the genetic manipulation of trypanosomes. In addition, this work will require follow up validation experiments to show that the recombinant fusion protein is truly representative of the natural protein. Antibodies which recognize the native CISSA protein would be preferred tools for studying this protein in order to avoid having to genetically modify the trypanosomes.

The newly identified *T. congolense* specific protein Tc3440 was also discovered in the iTRAQ differential expression experiments outlined in Chapter 2. There is no previous literature describing this protein. As with CISSA, protein Tc3440 reached its highest expression in EMF (Table 12 is an excerpt from Table 3, Chapter 2) and has an N-terminal signal peptide directing it to the secretory pathway. It also has a predicted GPI anchor single sequence at the C-terminus (Figure 12). Tc3440 has no significant primary sequence similarity to any protein in the non-redundant BLAST database, including proteins from other trypanosomatids. The best BLAST hit was an undefined protein from *Ciona intestinalis* (sea squirt) with an E-value of 4.1. However when secondary structure prediction was included, the top result was a very weak hit with the surface molecule GARP from *T. congolense* (E-value = 18). The E-value indicates that this is not a reliable hit but it is interesting that a *T. congolense* GPI anchored surface protein, which is also known to be highly expressed in EMF, was the closest match to Tc3440.

Table 12. iTRAQ expression data for TcIL3000.7.3440

T. congolense Accession No.	BLAST hit Accession No.	BSF → PCF	PCF → EMF	EMF → MCF	MCF → BSF
TcIL3000.7.3440	No Significant Hits	DNR[1]	5.44	0.18	DNR

[1]**DNR = Data not reliable**

```
1    MAPRHMHTSLASLVPRLLLALMLFGGLAARSNACSLDLEY
41   GVLKKCTQVMCRVATQLQGLRNPCRNLSSNIEKNKGAVSY
81   YFGATARIYENATKASAAANASGTAELVEELAAAHARARA
121  ALVESRKALFNILHTSNYFHTKSKSLDESFGVAAGISKAQ
161  YTSAKGANCTAEENPLGKRGSREDCLEARIAALSCGDGSE
201  DTTTESLNQASLALNDYLLSQGAGATREQQALAALVGDAA
241  GIFTELHVVSDDADEARVNASIANSMAEEVRELVNEHVPV
281  QEDNGEGPQDEPTAGGSTVTNNMLLLPFLLFACIQGH
```

Figure 12. Amino acid sequence of the protein TcIL3000.7.3440
The predicted secretory signal peptide and GPI anchor sequence are shown underlined and in red respectively.

In this chapter I describe the cloning and recombinant expression of the extracellular domain of CISSA and the full-length protein Tc3440. Both recombinant molecules were used to immunize BALB/c mice for production of monoclonal antibodies (mAbs). Additionally, in collaboration with the Boulanger lab at the University of Victoria, the extracellular domain of CISSA was purified and crystallized for structural determination by X-ray crystallography.

3.2. Materials and Methods

3.2.1. Gene constructs and recombinant expression of *T. congolense* proteins

DNA constructs of both CISSA (AA 40-264) and Tc3440 (AA 35-279) were synthesized by GenScript (Piscataway, NJ) and cloned via NcoI-NotI into the pET-28a vector (Novagen; EMD Millipore). In the pET-28a vector the inserted genes are expressed with an N-terminal hexa-histidine tag and thrombin protease cleavage site. Coincidentally, the CISSA and Tc3440 constructs from the pET28a vector were nearly identical in mass (28.2 and 28.4 kDa respectively). CISSA was also cloned using NcoI-NotI into a modified pET-32a vector (Boulanger Lab). This construct contained an N-terminal thioredoxin and hexa-histidine tags separated from CISSA by a thrombin cleavage site. Both constructs were designed to align with the predicted start of the mature proteins (after signal peptide removal) and to terminate near the transmembrane helix (CISSA) or GPI anchor signal peptide (Tc3440).

Both proteins were expressed by introducing the pET-28a constructs into *E. coli* BL21*DE3. The bacteria were grown overnight at 37 °C with shaking in Luria-Bertani (LB) medium containing 40 μg/mL kanamycin. One hundred μL of the overnight culture were used to inoculate 5 mL of fresh LB medium and the bacteria were grown for 2 h. Protein expression was induced by adding 1 mM isopropyl β-D thiogalactopyranoside (IPTG) for 2 h. The bacterial cells were harvested by centrifugation and resuspended in one mL of water containing 1x protease inhibitor cocktail V (Cat. No. 539137, Calbiochem) and lysed by sonication on ice. Protein expression was assessed by observation in 1-D polyacrylamide gels of new protein bands of the correct mass in

lysates of *E. coli* induced with IPTG and that were not present in lysates of non-induced *E. coli*.

3.2.2. Monoclonal antibody derivation

Monoclonal antibodies against CISSA and Tc3440 were derived according to the Pearson lab standard protocol [97]. First, crude *E. coli* lysates containing the expressed proteins were used for immunizations. For each of the expressed proteins, two BALB/c mice, aged six to eight weeks, were immunized with *E. coli* lysate. The priming injection consisted of 100 μg of immunogen (total protein) in Freund's complete adjuvant (FCA) delivered intraperitoneally (IP). Four booster injections, each containing 50 μg of protein in Freund's incomplete adjuvant (FIA), were administered IP at 2 week intervals. Two weeks after the final boost, sera were collected and antibody titres were assessed by indirect ELISA using *E. coli* lysates as antigen (described in 3.2.3). The mice were given a final boost (10 μg in saline, IV) and sacrificed three days later by cervical dislocation. Cardiac blood was collected and spleens were removed. Antibodies in the serum prepared from the cardiac blood were also tested by ELISA (section 3.2.3).

To prepare for the cell fusion, the spleens were aseptically transferred to a Petri dish containing 10 mL serum-free Dulbecco's Modified Eagles Medium (D-MEM, Cat. No. 23700-057, Gibco Life Technologies) containing 25 mM HEPES, 44 mM $NaHCO_3$, 2.86 mM 2-mercaptoethanol, 2 mM L-glutamine, 50 units/ mL penicillin-streptomycin, and 2 mM sodium pyruvate, and teased apart to release splenocytes. The splenocyte suspension was passed through a 40 μm nylon cell strainer (Cat. No. 352340, Becton-Dickinson) into a 50 mL conical tube, topped up to 50 mL with serum-free medium and centrifuged at 200 × g for 6 min. The pellet was resuspended in 3 mL serum-free medium, slowly

layered onto 4 mL of RT Ficoll-Paque (Cat. No. 17-1440-02, Amersham Biosciences) and centrifuged at 200 × g for 6 min, then 12 min at 650 × g. Lymphocytes were collected from the buffy coat, transferred to a new tube and the suspension was topped up to 10 mL with serum-free medium. Purified lymphocytes were counted by using a hemocytometer. Cultured parental myeloma cells were also counted at this time. A ratio of one SP2/0 parental myeloma cell to five lymphocytes was used (ideally with 10^8 lymphocytes). The required volumes of myelomas and lymphocytes were added to separate tubes, topped up to 50 mL with serum-free medium and centrifuged for 6 minutes at 200 × g. Both pellets were resuspended in residual medium by flicking the tubes. The lymphocytes and parental myelomas were combined, topped up to 50 mL with serum-free medium, mixed gently by inversion and then centrifuged for 5 minutes at 200 × g. The supernatant was removed completely and the pellet was loosened by flicking the tube. The tube was then placed in a 37 °C water bath and the cell suspension was stirred continuously while 1 mL of polyethylene glycol 1500 (PEG 1500, Cat. No. 783 641, Roche Diagnostics) was added over 1 min, followed by addition of 1 mL, 3 mL, and 10 mL serum-free medium over 1 min, 3 min, and 2 min, respectively. The fused cell mixture was incubated for 5 minutes at 37 °C, then removed from the water bath, topped up to 50 mL with complete medium (serum-free medium containing 10% FBS). The fused cells were centrifuged for 6 minutes at 200 × g and the pellet was resuspended in 5 mL recovery medium composed of 60 % serum-free D-MEM, 20 % FBS, 20 % fresh SP2/O conditioned medium, 1x IL-6 (Cat. No. 200-06, Peprotech Inc. Rocky Hill, NJ), 0.1x OPI (Cat. No. O-5003, Sigma-Aldrich, St. Louis, MO) and 20 μg/ mL gentamicin. The fused cells were added drop-wise to a 25 cm^2 non-tissue culture treated flask

containing 45 mL recovery medium and incubated at 37 °C overnight. Single step selection and cloning of hybridomas was performed using the ClonaCell-HY system (StemCell Technologies Inc. Vancouver BC). The cell fusion mixture was centrifuged for 6 minutes at 1.67 × g. The supernatant was removed and the pellet was gently resuspended in residual medium and added to 100 mL of MethoCult semi-solid medium (Cat. No. M03134, StemCell Technologies Inc.) containing 20% FBS, 36% serum-free medium, 1x IL-6, 1x OPI, 1x GlutaMAX-I (Cat. No. 35050, Gibco Life Technologies), 1x HAT selection medium (Cat. No. H-0262, Sigma-Aldrich), and 10 μg/ mL gentamicin. Hybridomas were incubated in a 37 °C, 5 % CO_2 incubator for at least 1 hour prior to plating in the ClonaCell-HY semi-solid selective medium. For plating, 9 mL of the hybridoma-containing MethoCult medium (StemCell Technologies Inc.) were aliquoted into twelve 60 x 15 mm FalconR tissue culture Petri dishes (Cat. No. 353002, Becton Dickinson, Mississauga, ON). To retain humid conditions, six more 60 x 15 mm FalconR tissue culture Petri dishes, with their lids removed, were filled with dH_2O. Two MethoCult-filled and one dH_2O-filled Petri dishes were placed into each of six large (150 x 15) mm Petri dishes. These were then placed, lid on, into a 5 % CO_2, 37 °C incubator for growth and selection of hybridomas. Approximately ten days later, 200 colonies from the CISSA immunized mice and 400 colonies from the Tc3440 mice were picked and transferred to 96 well tissue culture microplates in 200 μL of clone picking medium (75% serum-free medium 20% FBS, 1x GlutaMAX-I, 1x IL-6, 1x HT (Cat. No. H0137, Sigma-Aldrich)) and 10 μg/ mL gentamicin.

3.2.3. Enzyme linked immunosorbent assay

Indirect ELISAs were performed as previously described [115]. Antigens used to coat the ELISA plate wells were *E. coli* lysates (unmodified DH5α *E. coli* or *E. coli* BL21*DE3 expressing recombinant CISSA or Tc3440), *T. congolense* IL3000 PCF lysates or purified recombinant CISSA (from section 3.2.5). Parasite lysates from the equivalent of 5 x 10^5 trypanosome cells/well, 1 μg/well purified recombinant CISSA, or 1 μg/well total protein from *E. coli* lysate were diluted in dH_2O and 100 μL dried onto polystyrene ELISA plates. Primary antibodies used were either diluted mouse antiserum or undiluted hybridoma culture supernatant. The secondary antibody was goat α-mouse IgG/M (H+L) – alkaline phosphatase (AP) at 1:2,000 dilution (Cat. No. 31328, Thermo Scientific, Rockford IL). Cleavage of the substrate (para-nitrophenylphosphate, PNPP) was monitored by absorbance at 405 nm.

3.2.4. Polyacrylamide gel electrophoresis and immunoblotting

SDS-PAGE and immunoblots were performed as described in section 2.2.7. Specifically, primary antibodies (hybridoma supernatant) were diluted 1:1 in blocking buffer and the secondary antibody was a 1:20,000 dilution of HRPO-conjugated goat anti-mouse IgG/M (Cat. No. 1858413, Pierce; Thermo Scientific).

3.2.5. Protein expression and purification for crystallography

(This protocol was obtained from Dr. Marty Boulanger and represents his lab's standard operating procedure.)

For X-ray crystallography studies, the CISSA extracellular domain construct in a modified pET-32a vector, was transformed into *E. coli* Rosetta-gami 2 (DE3) cells (Novagen) which were grown in ZYP-5052 auto-induction medium containing 40 μg/mL chloramphenicol and 60 μg/mL ampicillin. Following 4 h of growth at 37 °C and 64 h at 16 °C, the cells were harvested by centrifugation and the cell pellet was resuspended in 10 mM Tris pH 8.5, 1 M NaCl, 30 mM imidazole. Cells were lysed using a French press and insoluble material was removed by centrifugation. The soluble fraction was incubated with Ni-chelated Sepharose beads for 1 h at 4 °C. Bound CISSA protein was eluted with 10 mM Tris pH 8.5, 1 M NaCl, 250 mM imidazole. The thioredoxin and hexa-histidine tags were removed by thrombin cleavage overnight at 18 °C in Tris buffered saline (TBS; 10 mM Tris pH 7.5, 150 mM NaCl) with 2.25 mM $CaCl_2$. CISSA was purified by gel filtration chromatography on a Superdex 75 16/60 HiLoad column (GE Healthcare) in TBS. Fractions containing monomeric CISSA were pooled and further purified by cation exchange chromatography on a HiTRAP SP FF column (GE Healthcare) in 20 mM HEPES pH 6.8, 10 mM NaCl. CISSA was eluted with an increasing concentration of NaCl and pooled fractions were concentrated to 25 mg/mL in TBS using Amicon Ultra (Millipore) centrifugal filter devices. The purity of the CISSA construct was determined by SDS-PAGE at each stage and protein concentrations were analyzed by absorbance at 280 nm based on an extinction coefficient of 29380 M^{-1} cm^{-1}.

3.2.6. Protein crystallization and X-ray diffraction

Initial crystallization trials for the extracellular domain of CISSA were performed using commercial screens (Wizard I/II from Emerald Biosystems, and Crystal Screen I/II and PEG/Ion from Hampton Research) in 96-well Intelli-Plates (Hampton Research) using a Crystal Gryphon (Art Robbins Instruments, Sunnyvale, CA). Dr. Marty Boulanger's standard lab protocol was employed. Two ratios of protein: reservoir solution were explored in sitting drops consisting of 0.2 μL or 0.3 μL protein solution (25 mg/mL in TBS) and 0.2 μL reservoir solution and were equilibrated against 55 μL reservoir solution. A single crystal form of CISSA was observed in 20% PEG 3350/0.2 M ammonium sulfate after overnight incubation at 18 °C and the crystals developed equally well in both drop ratios.

A single CISSA crystal was looped, stepped into a final cryoprotectant of reservoir solution supplemented with 25% glycerol and flash-frozen in liquid nitrogen. Diffraction data were collected on beamline 7-1 at the Stanford Synchrotron Radiation Laboratory (SSRL) using an ADSC Quantum 315r CCD detector. Overall, 320 images were collected with 0.5° oscillation and selected data were processed to 2.7 Å. Diffraction data for CISSA were processed using iMOSFLM [116] and Aimless from the CCP4 suite of programs [117].

3.3. Results

3.3.1. Derivation and characterization of monoclonal antibodies specific for *T. congolense* recombinant proteins Tc3440 and CISSA

Initially, His-tagged constructs of Tc3440 and the N terminal domain of CISSA were expressed as recombinant proteins in *E. coli* from the vector pET-28a (Figure 13). *E. coli* lysate, containing these expressed proteins were used as immunogen for monoclonal antibody derivation. At a later date, CISSA was also expressed from pET-32a for purification and structural determination by X-ray crystallography in collaboration with the Boulanger lab. Purified protein from the extra-cellular domain of CISSA was also used for screening of hybridoma supernatants and as immunogens for mAb derivation.

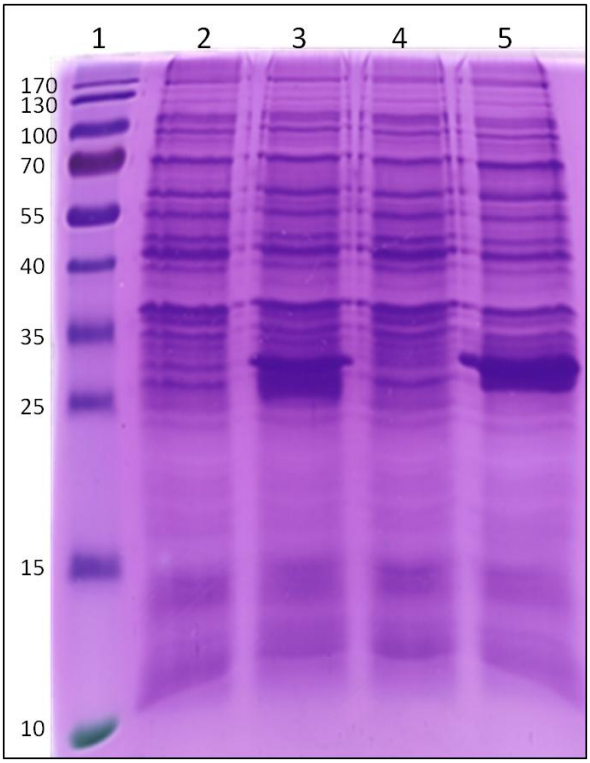


Figure 13. Coomassie brilliant blue stained polyacrylamide gel of *E. coli* expressing recombinant CISSA and Tc3440 constructs.
Lane 1: molecular weight ladder (kDa)
Lane 2: *E. coli* BL21*DE3 transformed with CISSA in pET-28a, not induced
Lane 3: *E. coli* BL21*DE3 transformed with CISSA in pET-28a, induced with IPTG
Lane 4: *E. coli* BL21*DE3 transformed with Tc3440 in pET-28a, not induced
Lane 5: *E. coli* BL21*DE3 transformed with Tc3440 in pET-28a, induced with IPTG

At the time of immunization, the Boulanger lab's HPLC required for purification of Tc3440 and CISSA expressed in *E. coli* lysates was in use. Therefore, mice were immunized with *E. coli* lysate from cells expressing the trypanosome proteins. The resulting hybridomas (~ 400 for Tc3440 and ~ 200 for CISSA) were screened by ELISA on the immunizing lysate. To eliminate hybridomas that secreted antibodies against normal *E. coli* proteins, the hybridoma supernatants were also screened against lysate from *E. coli* DH5α, a strain of *E. coli* which lacks the polymerase required to express proteins from the pET vectors. After this double screening, four hybridomas secreting mAbs specific for Tc3440 and one for CISSA were identified. These five antibodies were also positive on lysates of *T. congolense* PCF lysate by indirect ELISA (data not shown). These antibodies presumably recognized the native forms of Tc3440 or CISSA.

By the time these hybridomas had been derived, purified recombinant CISSA (section 3.2.5) was available and was used as antigen in immunoblots with the lone anti-CISSA mAb, 2B1. No reactivity was seen with the purified recombinant protein or with any molecule in the *T. congolense* PCF lysate. As can be seen in Figure 14 lane 3, mAb 2B1 bound to a low molecular weight smear in lysates of *E. coli* BL21*DE3 expressing CISSA. The observed lack of binding to *E. coli* DH5α (Figure 14, lane 2) explains how this mAb made it through the initial ELISA screening. The antigen recognized by mAb 2B1 was present in *E. coli* BL21*DE3 transformed with CISSA in a pET-28a vector but was absent from the DH5α strain.

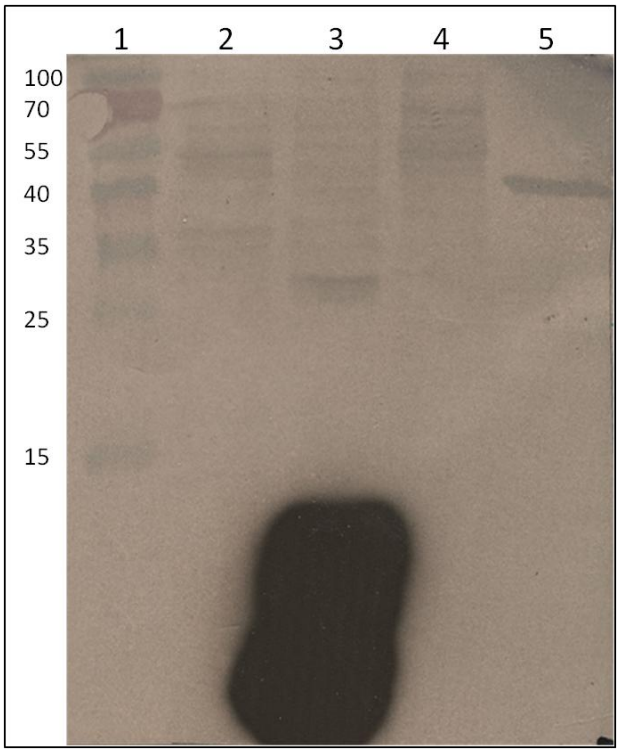


Figure 14. Immunoblot analysis of mAb 2B1against lysates of *E. coli and T. congolense* PCF and purified recombinant CISSA
The PVDF membrane was counterstained with nigrosin after development and the two images were overlaid in this figure.
Lane 1: protein standards (kDa)
Lane 2: *E. coli* DH5α, induced with IPTG
Lane 3: *E. coli* BL21*DE3 transformed with CISSA in pET-28a, induced with IPTG
Lane 4: *T. congolense* IL3000 PCF
Lane 5: purified recombinant CISSA
The CISSA in lane 5 was expressed from pET-32a and has both thioredoxin and His tags. This accounts for its higher apparent molecular weight compared to the His tagged version expressed from pET-28a in lane 3.

In a series of tangential experiments the source of the antigen recognized by mAb 2B1 was determined. I hypothesized that this antigen originated from either the plasmid vector carrying CISSA (possibilities include the lac repressor or kanamycin resistance conferring transacetylase) or is inherent to the BL21*DE3 strain of *E. coli*. Further immunoblots were performed using mAb 2B1 against lysate from IPTG induced, non-transformed *E. coli* BL21*DE3 or BL21*DE3 transformed with an empty (i.e. no transgene) pET-28a vector (data not shown). An immunoreactive smear was observed in

all BL21 cells indicating that the antigen is strain specific and not associated with plasmid transformation. The list of genetic differences between BL21 and DH5α strains is large so there is no clear candidate antigen. Further characterization of this mAb-antigen interaction, perhaps using immunopurification and mass spectrometry is beyond the scope of this dissertation but it is possible that a strain specific anti-*E. coli* mAb may be useful to other researchers.

The four anti-Tc3440 mAbs were also assessed by immunoblotting. However no pure Tc3440 protein was available and thus specific binding to Tc3440 binding could not be unequivocally confirmed. The first mAb (1A2) did not recognize any protein expressed by *E. coli* (including the recombinant Tc3440) but did bind to a ~ 40 kDa molecule in *T. congolense* PCF lysate (Figure 15). The unmodified Tc3440 polypeptide has a calculated mass of ~33 kDa but with the removal of the signal peptides and addition of a GPI anchor it is possible that the mature proteins could migrate at this higher (40 kDa) mass. Two mAbs (5E12 and 5H8) showed no reactivity with any *E. coli* or trypanosome derived protein by immunoblot (data not shown). The failure of these two mAbs to bind native or recombinant Tc3440 in immunoblot does not inherently mean that they are not specific for Tc3440, only that they these mAbs do not work in immunoblots. Further experiments with pure recombinant Tc3440 will be needed for clarity. The final mAb (3G3) generated against Tc3440 recognized a low molecular mass molecule that ran with the dye front of wells loaded with *E. coli* lysate but not *T. congolense* lysate (data not shown). It would appear that mAb 3G3 is specific for another bacterial molecule that made it through the selection process.

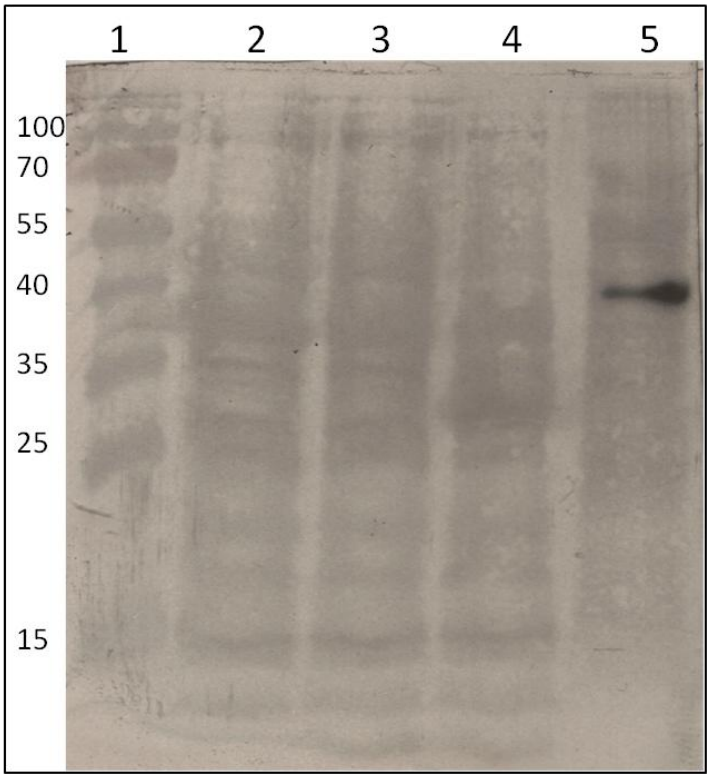


Figure 15. Immunoblot analysis of mAb 1A2 against cell lysates of *E. coli* and *T. congolense* The PVDF membrane was counterstained with nigrosin after development and the two images were overlaid in this figure.
Lane 1: protein standards (kDa)
Lane 2: *E. coli* BL21*DE3, induced with IPTG
Lane 3: *E. coli* BL21*DE3 transformed with Tc3440 in pET-28a, not induced
Lane 4: *E. coli* BL21*DE3 transformed with Tc3440 in pET-28a, induced with IPTG
Lane 5: *T. congolense* IL3000 PCF

3.3.2. Protein crystallization and X-ray diffraction

Data and figures presented in this section were generated by Michelle Tonkin, Sean Workman and Marty Boulanger. These data are presented in this dissertation since the expressed molecule was first identified by me and selected for further characterization. Structural analysis of selected trypanosome proteins is part of a collaborative effort between the Pearson and Boulanger labs.

In consultation with Dr. Marty Boulanger, I designed a gene fragment encoding the N-terminal domain of CISSA. The gene custom synthesized by Genscript (Piscataway, NJ)

and upon receipt, I cloned the gene into a modified pET-32a vector via NcoI/NotI restriction sites. In this vector CISSA was expressed with an N-terminal thioredoxin and histidine tag separated from the CISSA protein coding sequence by a thrombin cleavage site. This construct was made with the intent to use it for expression and purification of CISSA for protein crystallization. The thioredoxin tag was included to allow proper disulphide bond formation and protein folding. The CISSA-pET-32a vector construct was given to the Boulanger lab for protein expression, purification and X-ray crystallization studies. The CISSA N-terminal domain construct was expressed in *E. coli* Rosetta-gami 2 cells to utilize supplemental tRNAs that overcome typical *E. coli* codon bias and for their less reducing cytoplasmic environment. Following cleavage of CISSA from its fusion tag, gel filtration chromatography showed that it eluted as single monomeric peak in comparison to a set of globular standards (Figure 16a). However, residual thioredoxin co-eluted with CISSA (Figure 16a, inset) necessitating a cation exchange chromatography step to obtain sufficiently pure CISSA suitable for crystallization (Figure 16b).

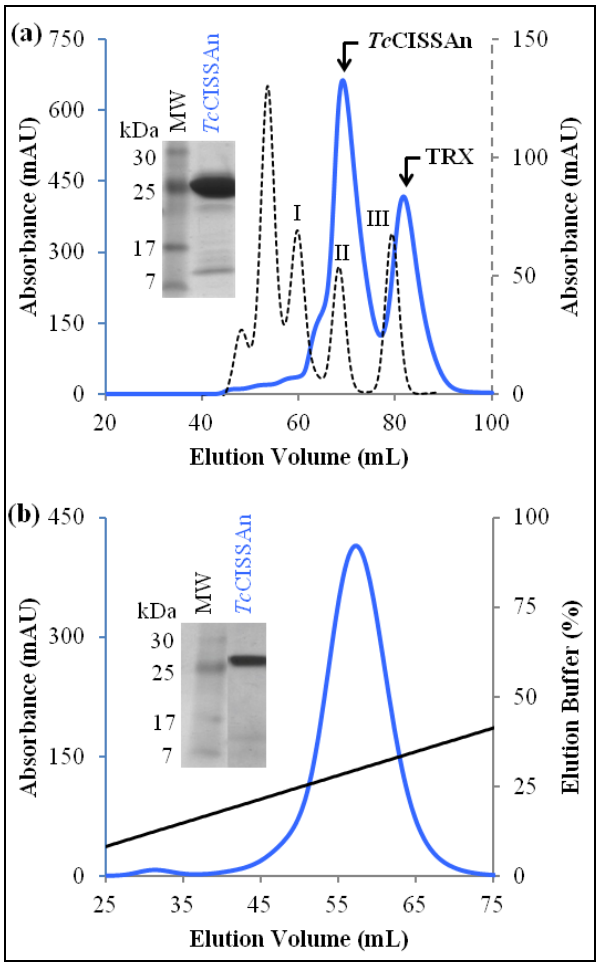


Figure 16. Purification of recombinant CISSA
(Figure created by M. Tonkin and S. Workman)
Panel A: Purification of CISSA by gel filtration chromatography. Standards for gel filtration chromatography (plotted against secondary axis, dotted curve): peak I, ovalbumin (43 kDa); peak II, carbonic anhydrase (29 kDa); peak III, ribonuclease A (13.7 kDa). CISSA (solid blue curve) eluted as a monomer.
Inset – polyacrylamide gel of elution fractions expected to contain CISSA.
Panel B: Further purification of CISSA by cation exchange chromatography,
Inset – polyacrylamide gel showing pure CISSA.

Crystallization trials of CISSA yielded fused clusters of thick crystal plates that grew to a maximum size of approximately 0.1 x 0.35 x 0.05 mm (Figure 17). Dissected single plates diffracted to a maximum of 2.7 Å resolution on beamline 7-1 at the SSRL, with an average mosaicity of 0.52° (Figure 18). Processing of these data revealed that CISSA crystallized in a primitive orthorhombic unit cell ($P2_12_12_1$), with the highest probability

Matthew's coefficient of 2.15 $Å^3$ Da^{-1} corresponding to four molecules in the asymmetric unit and a solvent content of 42.71% [118]. Scaling and merging of the data resulted in overall R_{merge} of 5.5% (40.0 % in the highest resolution shell).

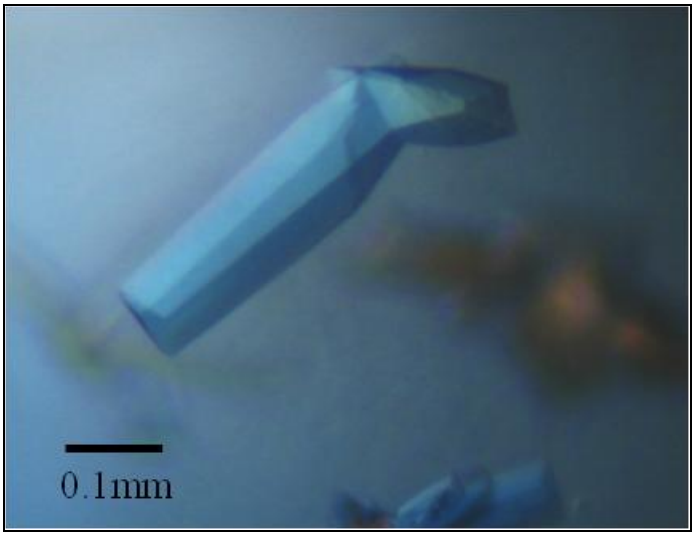


Figure 17. Photograph of a recombinant CISSA crystal
(Taken by M. Tonkin and S. Workman)
CISSA crystals were grown in a sitting drop with 20% PEG 3350 and 0.2 M ammonium sulfate as the reservoir solution.

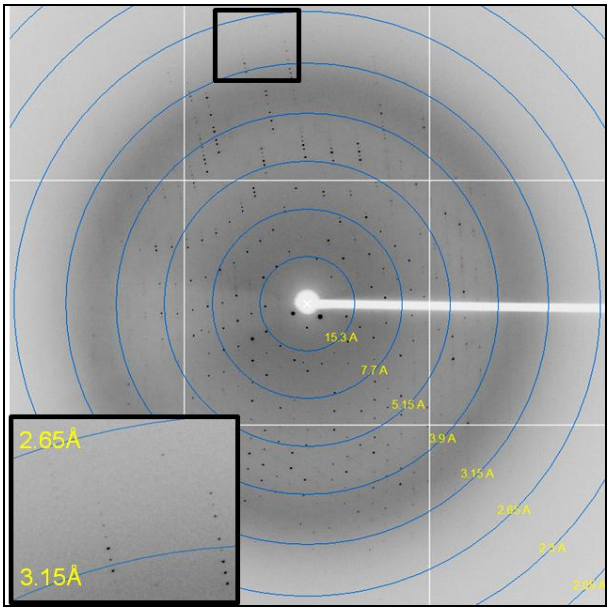


Figure 18. X-ray diffraction image from a CISSA crystal.
(Figure created by Dr. M. Boulanger)
Inset – high-resolution area; data were processed to 2.7 Å resolution.

3.4. Discussion

Two crude *E. coli* lysates, each containing an expressed, unique trypanosome protein were used as immunogens to generate mAbs. Since the entire lysates were used to immunize mice, *E. coli* proteins dominated the samples so it is no surprise that few of the derived hybridomas produced mAbs specific to the molecules of interest (none for CISSA and potentially three for Tc3440). With the production of purified recombinant CISSA by the Boulanger lab for their crystal structure studies, a defined immunogen is now available. Immunization of two BALB/c mice with pure CISSA began in August 2012 and in late October; the mice were given an IV boost 3 days prior to fusion. Unfortunately both mice both died during the 3 day period when they were not being observed, thus a new immunization program was initiated on November 2, 2012.

DNA encoding the Tc3440 protein has also been cloned into pET-32a for protein purification and crystallization. Unfortunately, Tc3440 expressed from pET-32a in *E. coli* resulted in a low yield of soluble protein (i.e. not enough for purification). In an attempt to improve the amount and quality of soluble Tc3440 recombinant protein, expression of Tc3440 from an insect cell system is currently underway. Recombinant soluble Tc3440 protein will be used as immunogen for mAb derivation and for crystallization.

Despite crystallization of the N-terminal domain of CISSA and collection of X-ray diffraction data, the protein structure has not been elucidated. This is because one key variable required for calculating the three-dimensional protein structure, the phase of the protein within the crystal lattice, is unknown. Phasing is usually solved in one of two ways, either by molecular replacement or anomalous dispersion. Molecular replacement

essentially uses information from a similar previously solved protein structure as a template for calculating the new protein structure. Since CISSA has no identifiable domain structure and only has sequence similarity to PSSA-2, which has not been crystallized, molecular replacement is not possible here. Anomalous dispersion uses higher atomic weight elements, incorporated into the protein structure to induce anomalous X-ray diffraction patterns. By comparing the data gathered from the native protein with the heavy atom labelled protein, the phases can be solved. To introduce heavy elements into CISSA, the protein will be expressed by a methionine auxotroph of *E. coli* in the presence of seleno-methionine. This ensures that all expressed proteins will be labelled with selenium. Unfortunately, the CISSA construct I designed (AA 40-264), and was used for all the work reported in this chapter, has a single methionine which limits the likelihood that a seleno-methionine approach will produce a strong enough anomalous signal for phase determination. To overcome this, a new CISSA construct (AA 27-264) has been designed and synthesized by the Boulanger lab. This new construct will include a second methionine which was excluded from the first construct. The new construct will be expressed with seleno-methionine, purified, crystallized and the phasing data provided by the selenium will be used to finally solve the CISSA protein structure.

This chapter outlined work performed to characterize two selected trypanosome proteins identified by the iTRAQ experiments described in of the previous chapter. These two were selected as they represent new identified trypanosome proteins that have a strong likelihood of being expressed on the surface membrane of T. *congolense* epimastigote forms.

3.4.1. Future Work

Work is underway towards deriving monoclonal antibodies against CISSA. These antibodies will be used to study the sub-cellular localization of the protein and to validate the iTRAQ expression data (both on cultured parasites and in the tsetse fly). Once Tc3440 has been expressed and purified from insect cells, it will also be the focus of crystallization studies and further monoclonal antibody derivation.

Several other proteins revealed by my iTRAQ results are also interesting and are being studied although results for these are not yet ready to be reported. Table 13 is a list of some of these proteins currently being cloned, expressed or characterized. A *T. brucei* protein (*brucei* alanine rich protein; Accession No. DQ246439 [39]) has also been selected for expression and characterization based on its similarity to CESP. Characterization of these molecules is beyond the scope of this dissertation.

Table 13. Additional iTRAQ identified proteins selected for expression and characterization

***T. congolense* Accession No.**	**BLAST hit Accession No.**	**BLAST hit protein name**	**BSF → PCF**	**PCF → EMF**	**EMF → MCF**	**MCF → BSF**
TcIL3000.0.06380	Tb927.8.1610	GP63[1]	DNR[2]	5.03	0.20	DNR
TcIL3000.0.15620	dbj|BAF91109.1	CESP[3]	DNR	11.11	0.17	DNR
TcIL3000.0.28430	No Significant Hits	No Significant Hits	0.07	DNR	DNR	17.91
TcIL3000.0.38630	No Significant Hits	No Significant Hits	0.15	DNR	8.47	0.62
TcIL3000.0.51090	No Significant Hits	No Significant Hits	5.05	0.19	0.17	5.75
TcIL3000.10.2930	Tb927.6.440	Hp-Hb receptor[4]	DNR	11.46	0.17	0.86

[1]Major surface metalloprotease GP63
[2]DNR = data not reliable
[3]CESP = congolense epimastigote specific protein
[4]Hp-Hb = haptoglobin-hemoglobin

Chapter 4. Molecular Characterization and Diagnostic Potential of the *T. congolense* Flagellar Calcium-Binding Protein, Calflagin

I was responsible for the overall experimental design of the work presented in this chapter and performed many of the experiments with technical assistance from Richard Yip and Rachelle Huot. Surface plasmon resonance experiments were performed by Matt Pope, the MiSCREEN mass spectrometry experiment was performed by Morteza Razavi, tandem mass spectrometry was performed by Darryl Hardie at the UVic-Genome BC Proteomics Centre and confocal immunofluorescence microscopy was performed with help from Dr. Robert Burke. Dr. Marty Boulanger and Laura Kaufman aided with the recombinant protein construct design and protein purification respectively. Dr. Jude Uzonna at the University of Manitoba provided sera from *T. congolense* infected mice.

The data reported in this chapter have been submitted for publication as:

Eyford BA, Kaufman L, Pope ME, Boulanger MJ, Burke RD, Uzonna JE, Pearson TW. Identification and characterization of calflagin, a flagellar calcium binding protein of trypanosomatids, as a potential antigen for use in serodiagnosis of *Trypanosoma congolense* infection. Journal of Parasitology (submitted Dec 2012).

4.1. Introduction

4.1.1. Diagnosis of animal African trypanosomiasis (AAT)

Diagnostic methods for AAT are less well developed than those for human disease despite the fact that AAT continues to be a major impediment to the agricultural and economic development of Africa. Improved diagnostic tests will aid in studying the epidemiology of the disease and in reducing the burden of AAT. Current diagnostic methods rely on direct microscopic observation of parasites, detection of anti-trypanosome antibodies [119, 120] or detection of trypanosome DNA by polymerase chain reaction (PCR) or loop-mediated isothermal amplification (LAMP) [121, 122]).

Due to the insensitivity of parasitological techniques and technical demands of current DNA based assays, serological detection of trypanosome-specific antibodies is the most easily applicable to wide-scale testing in the field [119, 120]. Such tests traditionally use molecularly undefined lysates of parasites as antigens and thus lack reproducibility and precision. Diagnostic tests employing defined antigens for detection of antibodies in infected animals show promise but have limitations. For example, a serological test for *T. evansi* infections of camels [123], based on detection of antibodies to a specific variant surface glycoprotein (VSG), is not applicable to animals infected with *T. congolense* and *T. vivax*, the most serious pathogens of livestock. A serodiagnostic test with wider applicability detects antibodies to a trypanosome-specific portion of *T. congolense* HSP70 [124], and has now been developed into an inhibition ELISA [125]. Surprisingly, *T. brucei* and *T. vivax* infections in cattle elicited only weak responses to the recombinant antigen used, possibly suggesting that a *T. congolense* specific epitope is involved in the detection system [125]. Thus, although defined antigens have been used in several

serological test formats, there is a need for more candidate molecules for development of a panel of serodiagnostic tests for African trypanosomes, for species-specific and pan-specific applications.

Of the trypanosomes that infect cattle (*T. b. brucei*, *T. vivax* and *T. congolense*), the latter has received much attention recently because it is the most easily studied *in vitro* and it is also considered to be the most economically important cattle pathogen. For this and for epidemiological reasons, it is important to develop diagnostic tests that can detect *T. congolense* infections. Although *T. congolense*-specific mAbs have been described and used to develop antigen-detection assays [126-128], these assays have not been successfully applied in the field and the relevant antigens have not been identified.

I sought an antigen that could be used to detect *T. congolense* infections of cattle. The known non-VSG surface molecules of *T. congolense* were possible candidates as they are relatively abundant. However, they are not known to be expressed in bloodstream forms that are present in infected animals and are only expressed in insect forms that inhabit the tsetse vector. Thus I decided to try to identify a *T. congolense* antigen that has long evaded description, a molecule that is recognized by the *T. congolense*-specific mAb (Tc6/42.6.4) [98]. This mAb showed very strong specific binding to *T. congolense* but was completely unreactive with *T. brucei*, *T. vivax* or *Leishmania braziliensis*. It was previously determined that the antigen recognized by mAb Tc6/42.6.4 was a 26-31 kDa protein which appeared by immunofluorescence to be distributed throughout the cytoplasm and along the plasma membrane, particularly the flagellar membrane. Many attempts over more than 25 years to identify the antigen recognized by mAb Tc6/42.6.4 were unsuccessful, thus precluding assessment of its serodiagnostic potential.

Interestingly, the antigen appears to show a degree of immunodominance since mAb Tc6/42.6.4 [98] and other mAbs that reacted with a protein band at approximately 26-31 kDa were produced in several experiments when mice were immunized with live *T. congolense* BSF (TWP unpublished observations). In this chapter I describe experiments used to identify the *T. congolense* antigen as a flagellar calcium-binding protein (FCaBP), also called calflagin.

Homologues of calflagin are found throughout the Trypanosomatidae family [129]. Most work on this group of proteins has been performed in *T. cruzi* and *T. b. brucei* where calflagin is thought to act a calcium sensor. Calflagin is di-acylated (palmitoylated and myristoylated) and is thus capable of close association with the plasma membrane, and has been observed to have the highest concentration along the flagellar membrane, exactly as originally reported by Parish et al [98], although this group did not know the identity of the antigen. Calcium binding (through EF hand domains) is required for membrane localization. When calcium chelators are added, calflagin becomes cytoplasmic [130]. This calcium dependent localization to the flagellar membrane could indicate that its signalling effect is directed at the flagellum, possibly involving the regulation of flagellar assembly or function.

4.2. Methods

4.2.1. Trypanosomes and cell culture

T. congolense IL3000 savannah strain [86], *T. congolense* K45/1, Kilifi strain [131], and *T. simiae* CP-11 [132] were originally obtained as BSF stabilates from the International Livestock Research Institute (ILRI; Nairobi, Kenya). *T. congolense* BSF variant antigenic type TC13 were passaged in CD-1 mice as previously described [133]. *T. b. brucei* YTaT 1.1 parasites [134] were obtained from Dr. Serap Aksoy (Yale University). PCF trypanosomes were originally produced from the BSF in the Pearson lab (as described in 2.2.1) and were cryopreserved until needed. Parasite lysates were prepared as described in section 2.2.1.

4.2.2. Monoclonal antibody from hybridoma Tc6/42.6.4

The hybridoma Tc6/42.6.4 that secretes a *T. congolense*-specific monoclonal antibody (also called Tc6/42.6.4) was derived from BALB/c mice that were immunized with intact, irradiated *T. congolense* BSF [98]. The mAb is an IgG_{2b}, kappa, and for the work reported here was used as diluted ascites fluid.

4.2.3. Immunoenrichment of the *T. congolense* antigen recognized by mAb Tc6/42.6.4

The molecule recognized by mAb Tc6/42.6.4 was enriched from *T. congolense* lysate by pull-down experiments and was identified by MS/MS analysis of the trypsin-digested, gel-separated protein bands. In brief, 50 µL of goat anti-mouse IgG Dynabeads (Cat. No. 110.33; Invitrogen) were rinsed with 50 µL sterile cold PBS and then resuspended in 500 µL PBS. Ten µL of Tc6/42.6.4 ascites fluid were added and the bead slurry was mixed

overnight at 4 °C. The beads were magnetically pelleted and washed 3 times with sterile PBS before incubation with trypanosome lysates prepared as follows: *T. congolense* IL3000 PCF (10^7 cells) were pelleted by centrifugation (5 minutes at 10,000 x g), washed once with sterile PBS, resuspended in 1 mL sterile PBS with 1x protease inhibitor cocktail V (Cat. No. 539137, Calbiochem) and lysed by sonication on ice. The trypanosome lysate was added to the Tc6/42.6.4 coated Dynabeads and mixed overnight at 4°C. The beads were washed 3 times with 1 mL PBS + 0.03% CHAPS. Bound molecules were eluted from the beads by heating to 60 °C in 50 µL of 2x Laemmli SDS sample buffer for 15 minutes. Eluted proteins were resolved by SDS-PAGE and stained with colloidal Coomassie Brilliant Blue R250. Proteins in parallel gel lanes were used in immunoblotting experiments to indicate which stained bands corresponded to the antigen recognized by mAb Tc6/42.6.4. The relevant protein gel band was cored for trypsin digestion and peptide analysis by mass spectrometry.

4.2.4. Gel electrophoresis and immunoblotting

Electrophoresis and immunoblotting were performed as described in Chapter 2, section 2.2.7. The primary antibody used was mAb Tc6/42.6.4 (ascites fluid) diluted 1:2,000 and the secondary antibody was either a 1:20,000 dilution of HRPO-conjugated goat anti-mouse (Cat. No 1858413; Pierce Chemical Co.) or a 1:2,000 dilution of HRPO-conjugated Clean Blot IP Detection Reagent (Cat. No. 21230, Thermo Scientific).

4.2.5. Triton X-114 extraction of trypanosome proteins

Extraction of aqueous and detergent-soluble proteins of 10^8 *T. congolense* IL3000 PCF was performed essentially according to the method described by Bordier [135]. Proteins in the aqueous and detergent phases were then subjected to chloroform/methanol extraction and the *T. congolense* antigen was detected by immunoblotting using mAb Tc6/42.6.4 as primary antibody.

4.2.6. In-gel trypsin digestion and peptide extraction

The protein band of interest was cored and sliced into small (~ 1 mm^3) pieces and transferred to an ethanol cleaned tube. Gel pieces were de-stained by alternating two wash solutions, 20% ACN/ 1M ammonium bicarbonate and 50% methanol/ 5% acetic acid, at 1 hour intervals until Coomassie brilliant blue stain was no longer visible. The de-stained gel pieces were completely dehydrated with ACN then sequentially reduced and alkylated for 30 minutes at RT with 10 mM dithiothreitol (DTT) and 100 mM iodoacetamide (both in 100 mM ammonium bicarbonate). Slices were dehydrated with ACN, rehydrated with 200 mM ammonium bicarbonate and dehydrated again. Slices were rehydrated on ice for 10 minutes with 100 µL of 20 ng/µL sequencing grade porcine trypsin (Cat. No. V5111, Promega) then incubated overnight at 37 °C (Note, for epitope mapping experiments 50 ng/ µL Glu-C, Cat. No. V1651, Promega, was used in place of trypsin). Tryptic peptides were extracted from the gel by adding 100 µL of extraction buffer (60% ACN/ 1% trifluoroacetic acid (TFA)) at 37 °C. The solution was transferred to a new tube and the process was repeated by adding another 100 µL of extraction buffer to the gel pieces. Peptide extraction was repeated until the gel slices were completely

dehydrated. All extraction samples were pooled and concentrated to ~50 μL using a Savant Speed-Vac vacuum concentrator (Thermo).

4.2.7. Mass spectrometry

Matrix assisted laser desorption ionization time of flight (MALDI-TOF) MS/MS and LC-MS/MS were used to analyse peptides from the tryptic digests of gel bands. An Applied Biosystems MDS Sciex TOF/TOF 4800 Mass Analyzer was used. The 25 most intense peaks in the mass range of 800-4000 m/z were selected for MS/MS fragmentation. For nano-HPLC-MS/MS analyses, an Applied Biosystems/MDS Sciex QSTAR Pulsar I Hybrid Quadrupole-TOF LC-MS/MS Mass Spectrometer was used. MS/MS spectra were acquired selecting the top 2 most intense eluting ions in the 400-1600 m/z range with a 2+ to 4+ charge state. MALDI-TOF MS/MS and LC-MS/MS data were searched against the *T. congolense* protein database (described in Chapter 2) resulting in two peptides tracing to 4 nearly identical *T. congolense* genes which when queried by BLAST were annotated as flagellar calcium binding proteins.

4.2.8. Gene cloning and recombinant protein expression

The calflagin gene was cloned and expressed as a recombinant protein in two different forms. First, PCR primers were designed to anneal within the termini of the protein coding region (overlapping the endogenous start and stop codons of TcIL3000.0.43820 or TcIL3000.05280) to obtain the full-length gene. The primers used were 5'-GGC TCA TAT GGG TTG CTC TGG ATC AA-3' and 5'-CGC GGA TCC CTA TCA GTG GTA GGG GTC T-3'. *T. congolense* IL3000 PCF were used as the source of the template

DNA. The PCR product (634 bp encoding 204 amino acids) was cloned into pET-24a via NdeI/BamHI. Calflagin was expressed as described in section 3.2.1. The expressed calflagin was detected by immunoblotting using mAb Tc6/42.6.4. This form of calflagin expressed from pET-24a represents the full length protein without any truncations or added affinity tags.

A second, slightly smaller form of calflagin was also created (AA 12–202). This was obtained by PCR using the primers 5'-CGT CAT ATG TCC AAG GGC TCT GCG TG-3' and 5'- GCT GGA TCC CTA GGG GTC TCC GAA CGC-3'. The PCR product was cloned into pET-28a by NdeI/BamHI where it was expressed (using the method described in section 3.2.5) fused to an N-terminal His tag. Recombinant, His-tagged calflagin was purified as previously described [136]. The His tag was not removed by thrombin cleavage because the thrombin reaction buffer contains calcium which is known to alter the folding state of calflagin thus exposing unstructured hydrophobic regions [129, 130]. In several experiments, I observed that the presence of calcium caused protein precipitation, thus 5 mM EDTA was added to all purification buffers to chelate trace amounts of calcium.

4.2.9. Monoclonal antibody derivation

Several new anti-calflagin mAbs were generated in BALB/c mice as described in Chapter 3, section 3.2.2 except purified recombinant calflagin was used as the immunogen.

4.2.10. Enzyme-linked immunosorbent assay (ELISA)

To detect anti-trypanosome antibodies, indirect ELISAs were performed on parasite lysates or on recombinant purified calflagin as described in section 3.2.3. Additionally, antigen-capture (sandwich) ELISAs were performed. For antigen-capture ELISAs, the microplate wells were first coated with 1:50 diluted primary antibody (hybridoma ascites fluid containing the mAbs described in section 4.2.9) overnight at 37 °C in a humidified incubator. Antigen was either 1 µg of purified recombinant calflagin or lysate from $5x10^5$ *T. congolense* PCF. The secondary antibody used was a 1:1000 dilution of AP-conjugated mAb Tc6/42.6.4.

4.2.11. Confocal immunofluorescence microscopy

One mL of a suspension of log phase *T. congolense* IL3000 PCF was centrifuged for 1 minute at 10,000 g and the pelleted parasites were washed once with ice-cold, sterile PBS. Cells were fixed for 10 minutes at RT by resuspension in 1 mL of either 4% para-formaldehyde in PBS or 1 part formalin to 9 parts PEM buffer (100 mM PIPES, 5 mM EGTA, 2 mM $MgSO_4$, pH 6.8). The fixed parasites were pelleted by centrifugation and blocked with 1 mL of Super Block Blocking Buffer (Cat. No. 37515, Thermo Scientific) containing 0.3% Triton X-100. Fifty µL of the fixed cells were added to poly-L-lysine coated microscope slides and allowed to adhere for 15 minutes at RT. Primary antibodies used were a 1:1000 dilution of mAb Tc6/42.6.4 ascites fluid and a 1:500 dilution of rat antiserum specific for trypanosome para-flagellar rod protein [137]. Both antibodies were diluted in blocking buffer and were incubated with parasites overnight at 4 °C. Slides were then washed three times by immersion for 5 minutes in PBS. The secondary

antibodies, goat α-mouse IgG (H+L)-AlexaFluor 488 (Cat. No. A11029, Invitrogen) and goat α-rat-AlexaFluor 568 (Cat. No. A11077, Invitrogen) were diluted 1:1000 in blocking buffer and 50 μL were added to the slides for 1hr at RT followed by washing as described above. Specimens were covered with Slow-Fade Gold containing the DNA-staining reagent 4', 6-diamidino-2-phenylindole (DAPI) (Cat. No. S36938, Invitrogen) and examined with a Zeiss LSM 700 confocal laser scanning microscope using a Zeiss 63 x oil-immersion objective lens. A series of optical sections of specimens were collected. Maximum intensity projections were prepared using Zen software (2009, 5.05.00; Carl Zeiss).

4.2.12. Surface plasmon resonance analysis

A Biacore 3000 surface plasmon resonance (SPR) instrument (GE Healthcare) was used to determine the kinetics of binding of mAb Tc6/42.6.4 to purified recombinant *T. congolense* calflagin. A high-resolution multi-concentration analysis was performed by first capturing mAb Tc6/42.6.4 using an affinity-purified sheep anti-mouse IgG antibody that was covalently coupled to the Biacore chip. The purified, His-tagged recombinant calflagin (section 4.2.8) was passed over the Biacore affinity chip at concentrations ranging from 62.5 nM to 500 nM. Data were double referenced and fit globally using a 1:1 Langmuir binding model. The general method has been described previously [138].

4.2.13. Epitope mapping

Two variations of the trypsin digestion protocols were used in an attempt to generate peptides for mapping the calflagin epitope recognized by mAb Tc6/42.6.4. The first method used was described by Kiselar and Downard, 1999 [139]. Briefly, six hundred μg of purified calflagin were precipitated overnight at 4 °C with 9 volumes of acetone. The acetone was removed and the pellet was air dried before resuspension in 60 μL of 25 mM ammonium bicarbonate/ 6M urea. Calflagin was reduced and alkylated by adding 60 μL of 10 mM DTT in 25 mM ammonium bicarbonate for 45 minutes followed by the addition of 60 μL of 40 mM iodoacetamide for 45 minutes, both at 37 °C. The reaction was quenched by adding an additional 60 μL of DTT. Calflagin was digested overnight at 37 °C by adding 350 μL of 20 ng/μL trypsin. The resulting peptide mixture was desalted using an Oasis column (Waters Inc.) and quantitated by measuring the OD at 280 nm. At this point, a fraction (~10 μL) of the calflagin tryptic peptide mixture was removed and analyzed by MALDI-MS to establish a standard profile of the peptides present. MAb Tc6/42.6.4 was added to the remaining peptide mixture at a 1:1 molar ratio and allowed to bind at 4 °C for 2 h. The antibody-peptide mixture was directly spotted onto a MALDI plate and analyzed by MALDI-TOF MS using a matrix with no added acid to prevent the dissociation of antibody-peptide complexes.

The second method used relies on antibody binding to shield the appropriate epitope from trypsin cleavage. Calflagin was allowed to bind to bead-immobilized mAb Tc6/42.6.4 as described in section 4.2.3. Rather than eluting the bound calflagin-mAb complexes, 100 μL of 20 ng/ μL trypsin or Glu-C (Cat. No. 90054, Thermo Scientific) were added and the beads were mixed at 37 °C. After 10 minutes (and at 15 minutes

intervals thereafter) 15 μL of the bead slurry were removed from the mixture (while the remainder continued to be digested). The beads were pelleted and washed 3x with PBS. Bound peptides were then eluted with 5% acetic acid and analyzed by MALDI-TOF MS. In this way, a total of 6 fractions were collected.

4.2.14. MALDI immunoscreening (MiSCREEN) analysis

Synthetic peptides suspected to contain the epitope recognized by mAb Tc6/42.6.4 (representing calflagin AA 12-21, 14-30, and 18-35) were custom synthesized by JPT Peptides (Berlin, Germany). The peptides were assessed for binding to mAb Tc6/42.6.4 in solution using the MiSCREEN method described by Razavi et al. [140].

4.2.15. Measurement of anti-calflagin antibodies in sera from trypanosome infected mice

Mice were infected with *T. congolense* BSF and at intervals after infection sera were prepared for measurement of anti-calflagin antibodies. These mouse experiments were performed in Dr. Jude Uzonna's lab at the University of Manitoba. Six to eight week old female C57BL/6 and CD-1 mice were purchased from the Central Animal Care Services breeding facility at the University of Manitoba. All mice were maintained in a specific pathogen free environment at the University of Manitoba and were used according to the guidelines of the Canadian Council for Animal Care.

Cryopreserved *T. congolense* BSF, variant antigenic type TC13 [133] were isolated from the blood of CD-1 mice, 3 days after passage, by DEAE-cellulose anion exchange chromatography [89]. For infection, C57Bl/6 and lymphotoxin deficient (knockout) mice

on C57BL/6 background [133] were injected IP with 10^3 *T. congolense* TC13 BSF in 0.1 ml TBS supplemented with 5 mM glucose and 10% FBS.

To monitor parasitemias, a drop of blood from the tail vein of each mouse was collected on a glass slide and the number of parasites in three or more fields was counted using a light microscope with a 40 x objective. At specified time points, blood was collected by cardiac puncture. Blood samples were kept at 4 ºC for 4 h**,** centrifuged at 600 g for 10 minutes and serum was collected and stored at -20 °C until used.

Anti-calflagin antibodies (both IgM and IgG) were measured in sera by indirect ELISA using purified recombinant calflagin as antigen (from section 4.2.8). Prebleed and test bleed sera were diluted 1:100 for use as a source of primary antibodies. For measurement of IgM and IgG levels, two different secondary antibodies were used: AP labeled goat α-mouse IgG (Cat. No. LM30108, Caltag Laboratories) and AP labeled goat α-mouse IgM (Cat. No. M31508, Caltag Laboratories), both at a 1:3000 dilution.

4.3. Results

4.3.1. Species and life cycle stage specificity of mAb Tc6/42.6.4

MAb Tc6/42.6.4 was shown previously to be specific for *T. congolense* (subgenus Nannomonas) as it showed no reactivity with the two other major trypanosome species that infect African livestock, *T. brucei* ssp. (subgenus Trypanozoon) or *T. vivax* (subgenus Duttonella) [98]. To extend the characterization of mAb Tc6/42.6.4, *T. simiae* CP11 (another member of the subgenus Nannomonas and the closest relative to *T. congolense*) was tested and found not to react with mAb Tc6/42.6.4 by ELISA (data not shown). The *T. congolense* species can be divided into 3 groups, the Kilifi, savannah and

forest groups. The original report of Tc6/42.6.4 only tested savannah strains of *T. congolense* [98]. To expand the original specificity studies, ELISAs were performed on parasite lysates and showed that mAb Tc6/42.6.4 also reacted with the PCF of a Kilifi clone of *T. congolense* (strain K45/1).

It was previously reported that the antigen recognized by mAb Tc6/42.6.4 was present in both BSF and PCF of *T. congolense* [98]. In Chapter 2, immunoblots were performed on all four life cycle stages of *T. congolense* using various mAbs, including Tc6/42.6.4. The results for Tc6/42.6.4 are replicated in Figure 19. An immunoreactive band at ~26 kDa was observed in all four life cycle stages, with both PCF and MCF showing the greatest intensity and EMF the lowest.

Figure 19. Immunoblot of the four major life cycle stages of *T. congolense* IL3000 using mAb Tc6/42.6.4.
This figure was modified from Figure 9.

4.3.2. Identification of the antigen recognized by mAb Tc6/42.6.4

The antigen recognized by mAb Tc6/42.6.4 appeared as a strong immunoreactive band of ~26 kDa when detected by immunoblotting on whole *T. congolense* PCF lysates (Figure 20A, lane 3). After enrichment by pull-down with mAb Tc6/42.6.4, the antigen band at ~26 kDa was easily detected (Figure 20A, lane 2). A parallel gel containing the enriched protein was stained with colloidal Coomassie Brilliant Blue R-250 (Figure 20B). To identify the protein antigen, the ~26 kDa band (highlighted in red, Figure 20B, lane 5)

was excised and subjected to in-gel trypsin digestion followed by peptide extraction and sequencing by MALDI-TOF-MS/MS and LC-MS/MS.

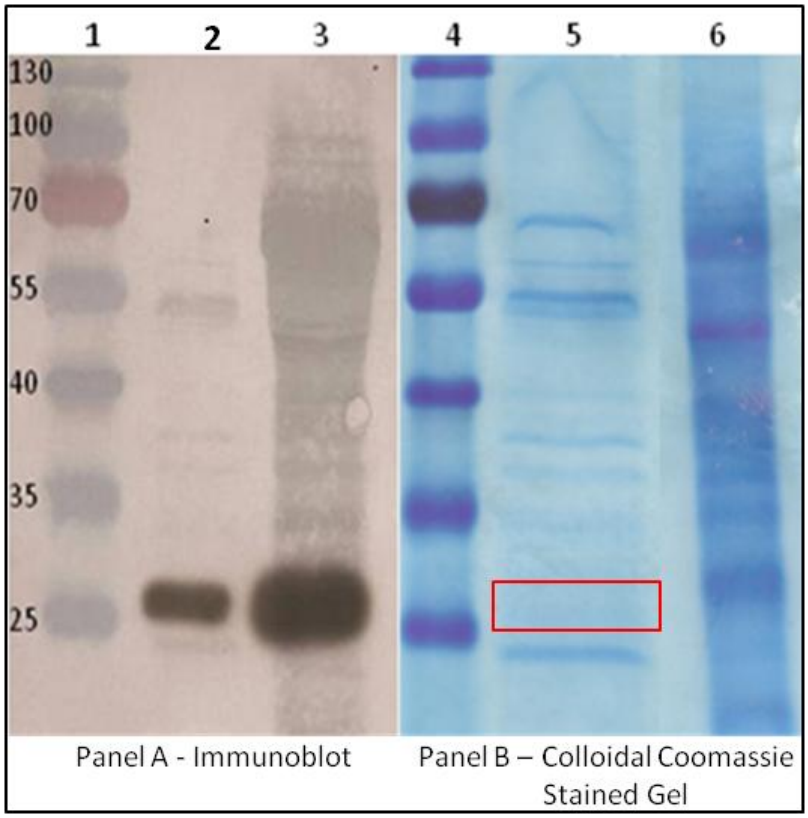


Figure 20. Stained polyacrylamide gel and immunoblot of the *T. congolense* antigen recognized by mAb Tc6/42.6.4.
The PVDF membrane was counterstained with nigrosin after development of the autoluminogram and the two images were overlaid in panel A.
Panel A. Immunoblot using mAb Tc6/42.6.4.
Lane 1: Molecular weight standards.
Lane 2: Protein immunoenriched from *T. congolense* PCF.
Lane 3: *T. congolense* PCF.
Panel B. Colloidal Coomassie Brilliant Blue staining of trypanosome proteins. The boxed band shows the ~26 kDa protein band that was cored for mass spectrometric analysis.
Lane 4: molecular weight standards.
Lane 5: Protein immunoenriched from *T. congolense* PCF.
Lane 6: *T. congolense* PCF.

The MALDI-MS/MS data (Figure 21) and LC-MS/MS data from the tryptic digest resulted in identification of two peptides of trypanosome origin when searched against the *T. congolense* proteome database [91]. These two peptides (m/z 1227.72 and 1457.51

from Figure 21) were determined to have sequences VLQMHELTTR and LSFNEVCSGCER. These two peptides were also identified by LC-MS/MS (data not shown).

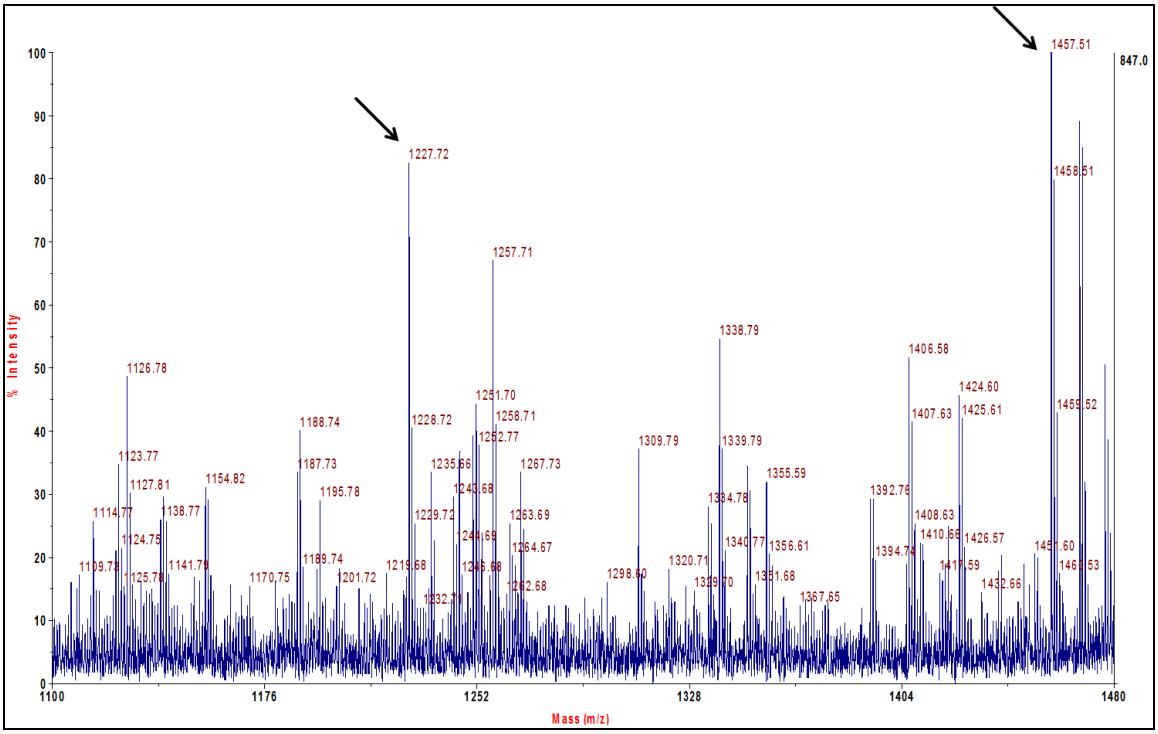


Figure 21. MALDI-TOF mass spectrum of the trypsin-digested ~26 kDa gel band recognized by mAb Tc6/42.6.4.
The peak at 1227.72 m/z corresponds to the peptide VLQMHELTTR and the peak at 1457.51 m/z corresponds to the peptide LSFNEVCSGCER (with two carbamidomethylations).

These two peptides were traced back to proteins encoded by four nearly identical *T. congolense* ORFs annotated as "undefined product" in the *T. congolense* database (Figure 22). The two MS/MS peptides used to identify these proteins are highlighted. The protein sequences were used to query the non-redundant database by BLAST for clues regarding their function. These proteins drew strong hits with the Tb-24 family of flagellar calcium binding proteins (FCaBP or calflagin) from *T. brucei*, *T. cruzi* and other trypanosomatids. The *T. congolense* proteins have 55% identity and 69% similarity to

the *T. brucei* calflagin and showed 59% identity and 75% similarity to *T. cruzi* calflagin (Figure 23).

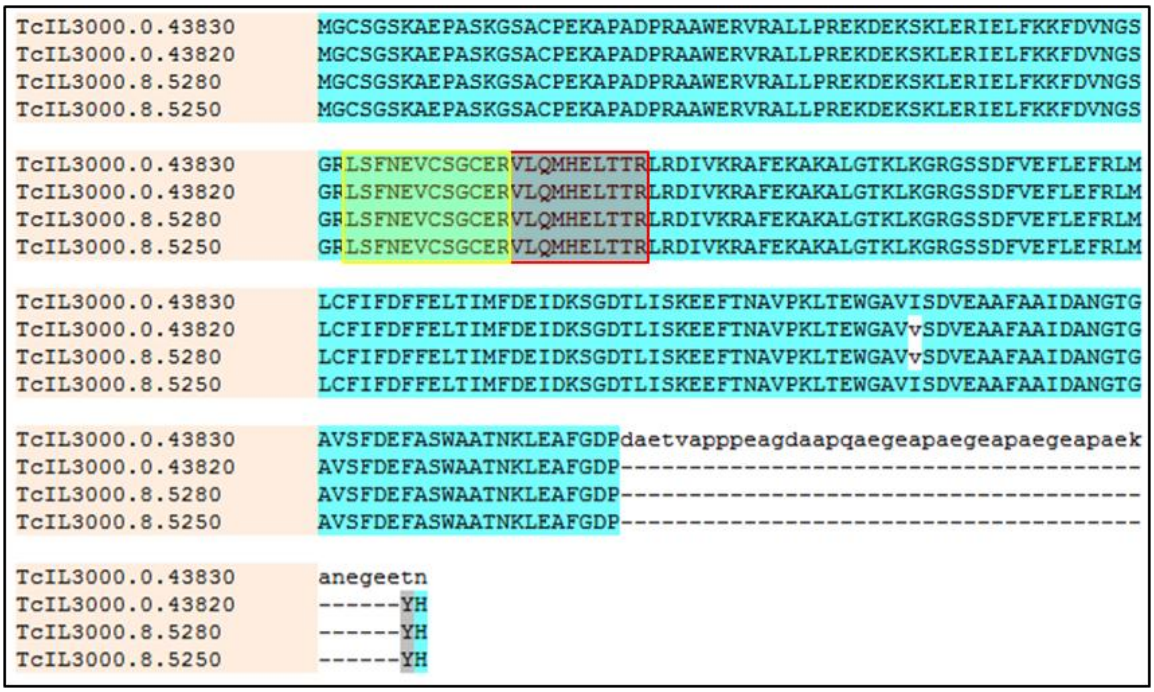


Figure 22. Multiple sequence alignment of the *T. congolense* calflagins. The positions of the two MS-identified tryptic peptides are highlighted.

Figure 23. Multiple sequence alignment of *T. congolense, T. brucei and T. cruzi* calflagins. Accession numbers: *T. brucei* (AAB40004.1), *T. congolense* (TcIL3000.0.43820) and *T. cruzi* (EAN83724.12).

4.3.3. Cloning and expression of recombinant *T. congolense* calflagin

To confirm that the MS identified protein was truly the antigen recognized by mAb Tc6/42.6.4, full length *T. congolense* calflagin (TcIL3000.0.43820/TcIL3000.05280) was expressed as a recombinant protein in *E. coli*. An immunoblot was performed on the recombinant calflagin using mAb Tc6/42.6.4 as the primary antibody (Figure 24). A colloidal Coomassie Brilliant Blue stained band of approximately 23 kDa was expressed by IPTG induced *E. coli* (Figure 24, lane 6). This band was clearly reactive with mAb Tc6/42.6.4 (Figure 24, lane 3) indicating that calflagin was in fact the antigen recognized by mAb Tc6/42.6.4.

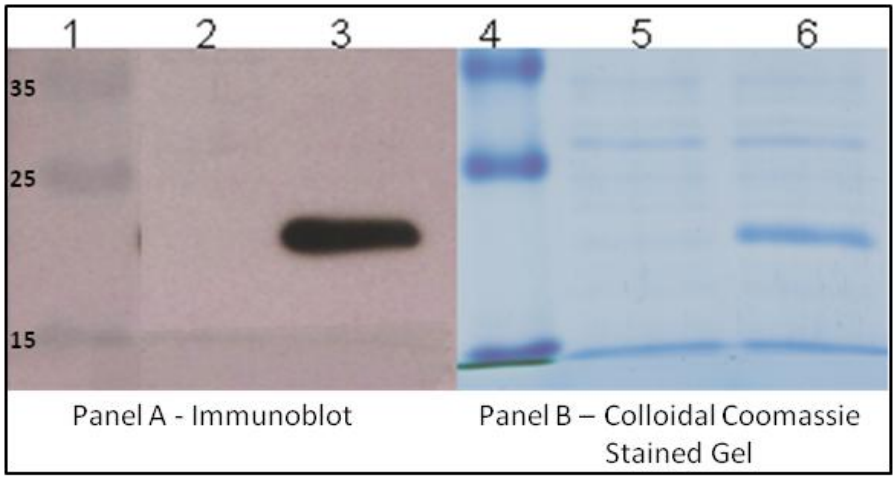


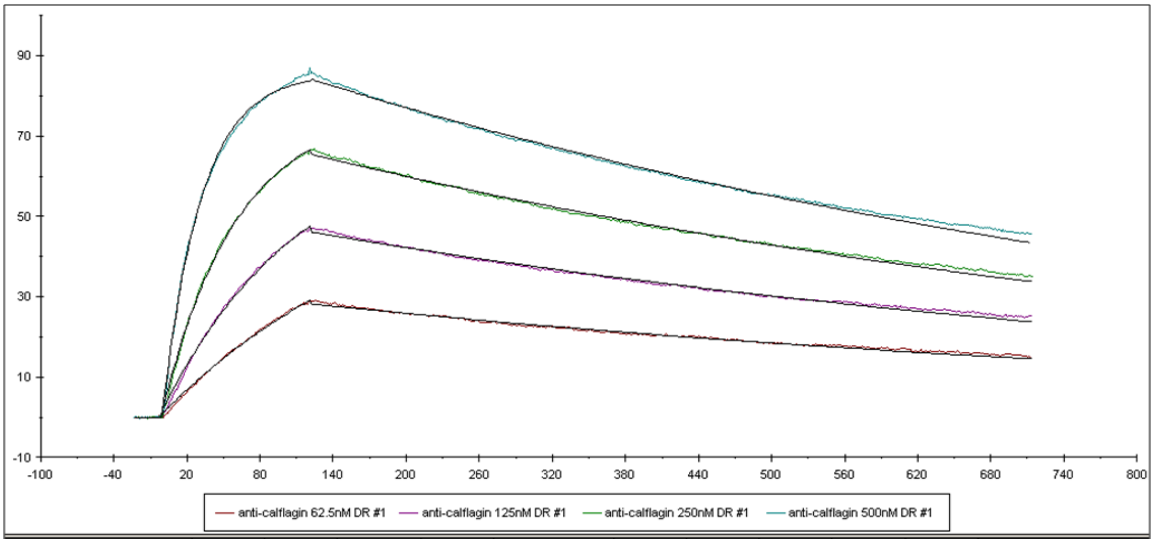
Figure 24. Expression of recombinant *T. congolense* calflagin.
Panel A. Immunoblot of recombinant calflagin using mAb Tc6/42.6.4.
Lane 1: Molecular weight standards.
Lane 2: *E. coli* transformed with calflagin in pET-24a, not induced
Lane 3: *E. coli* transformed with calflagin in pET-24a, induced with IPTG.
Panel B. Colloidal Coomassie Brilliant Blue staining of *E. coli* lysates.
Lane 4: Molecular weight standards.
Lane 5: *E. coli* transformed with calflagin in pET-24a, not induced
Lane 6: *E. coli* transformed with calflagin in pET-24a, induced with IPTG.

A truncated (AA12–202), His tagged recombinant calflagin was also expressed and purified by nickel-chelate chromatography. This purified calflagin was used to study mAb binding kinetics by SPR (section 4.2.12), to immunize mice for generation of new

mAbs (section 4.2.9), to measure anti-calflagin antibody responses from *T. congolense* infected mice (section 4.2.15) and to attempt protein crystallization (in progress in the laboratory of Dr. Marty Boulanger).

4.3.4. MAb Tc6/42.6.4 – calflagin binding kinetics

The purified recombinant calflagin (section 4.2.8) was used as analyte in SPR assays for determining the kinetics of binding to mAb Tc6/42.6.4. The results are shown in Figure 25. The mAb bound strongly to the target antigen, with a dissociation constant (KD) of 18 nM. The antibody showed an on-rate (ka) of $6.08e^{4}$ M^{-1} s^{-1} and an off-rate (kd) of $1.12e^{-3}$ s^{-1}.



	ka (1/Ms)	kd (1/s)	Rmax (RU)	RI (RU)	Conc of analyte	KA (1/M)	KD (M)	Req (RU)	kobs (1/s)	Chi2
	6.08e4	1.12e-3				5.44e7	1.84e-8			0.213
anti-calflagin 62.5nM DR #1			81.3	0.83	62.5n			62.8	4.92e-3	
anti-calflagin 125nM DR #1			80.8	1.56	125n			70.4	8.72e-3	
anti-calflagin 250nM DR #1			81.4	1.17	250n			75.8	0.0163	
anti-calflagin 500nM DR #1			89.1	-0.315	500n			85.9	0.0315	

Figure 25. Multi-concentration analysis of binding kinetics of mAb Tc6/42.6.4 and calflagin by SPR.
Calflagin was injected over captured mAb Tc6/42.6.4 at concentrations of 62.5 nM, 125 nM, 250 nM and 500 nM (bottom to top curves respectively). Data were double referenced and fit globally using a 1:1 Langmuir binding model.

4.3.5. Sub-cellular localization of calflagin

In *T. cruzi* and *T. brucei*, calflagin is thought to act as a calcium sensor that oscillates between the plasma membrane and the cytoplasm based on intracellular calcium concentration [141]. To test whether or not calcium induced variable localization occurs in *T. congolense,* confocal immunofluorescence microscopy was performed using two different fixative protocols, one under normal cellular conditions and one in the presence of a calcium chelator. The results are shown in Figure 26. In red is shown the para-flagellar rod protein representing the flagellum and the blue DAPI staining shows the presence of nuclear and kDNA. In the presence of calcium, calflagin (in green) is distributed throughout the cells but showed a marked increased intensity along the flagellum and the cell membrane (Figure 26 A-C). When incubated with a fixative containing EGTA, the cells showed only diffuse fluorescence in the cytoplasm and a complete lack of fluorescence along the flagellum and cell periphery (Figure 26 D-F).

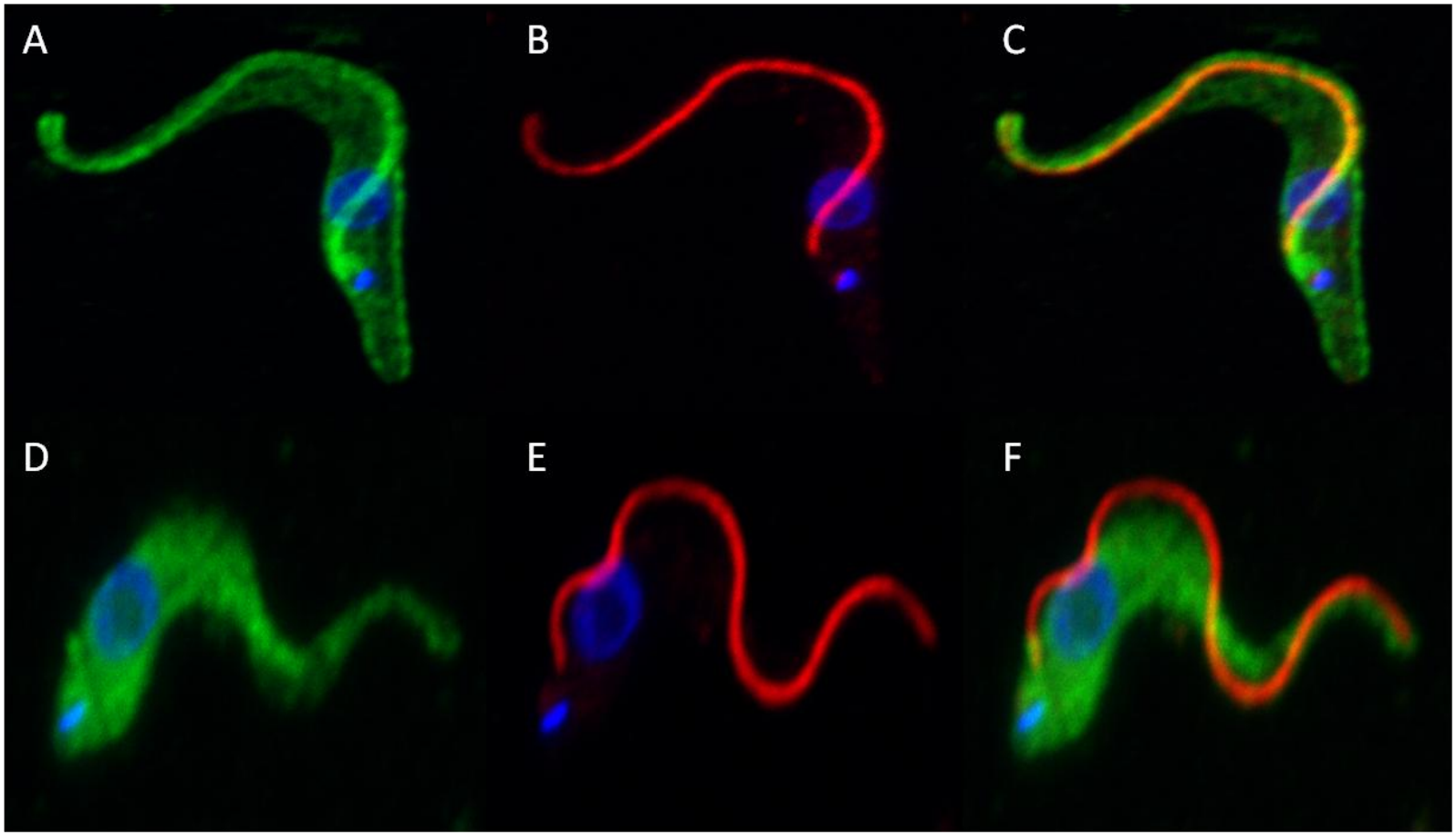


Figure 26. Confocal immunofluorescence microscopy showing localization of calflagin.
Fixed *T. congolense* PCF were probed with mAb Tc6/42.6.4 and an antiserum against para-flagellar rod protein in the presence (A, B, and C) and absence (D, E, and F) of calcium. Green: calflagin detected by mAb Tc6/42.6.4; Red: para-flagellar rod protein detected by rat anti-para-flagellar rod protein antiserum; Blue: DAPI staining of nuclear and kDNA.

4.3.6. Serodiagnosis of *T. congolense* by detection of mouse anti-calflagin antibodies

Mice were infected experimentally with *T. congolense* by collaborators at the University of Manitoba and serum was collected periodically. I tested these sera by indirect ELISA for the presence of both IgG and IgM antibodies against calflagin. All mice sampled throughout this experiment produced anti-calflagin antibodies as a result of infection by *T. congolense* (Figure 27). Even at early time points (days 4 and 8 post infection), all 9 mice produced anti-calflagin antibodies with IgM predominating, as expected for primary immune responses. In these mice, low-level IgG responses were observed and only at day 15 were the levels significantly above background. After longer intervals (30 days post infection), increased IgG responses were seen. Interestingly, both IgM and IgG responses were seen in all of the mice with late-stage infections.

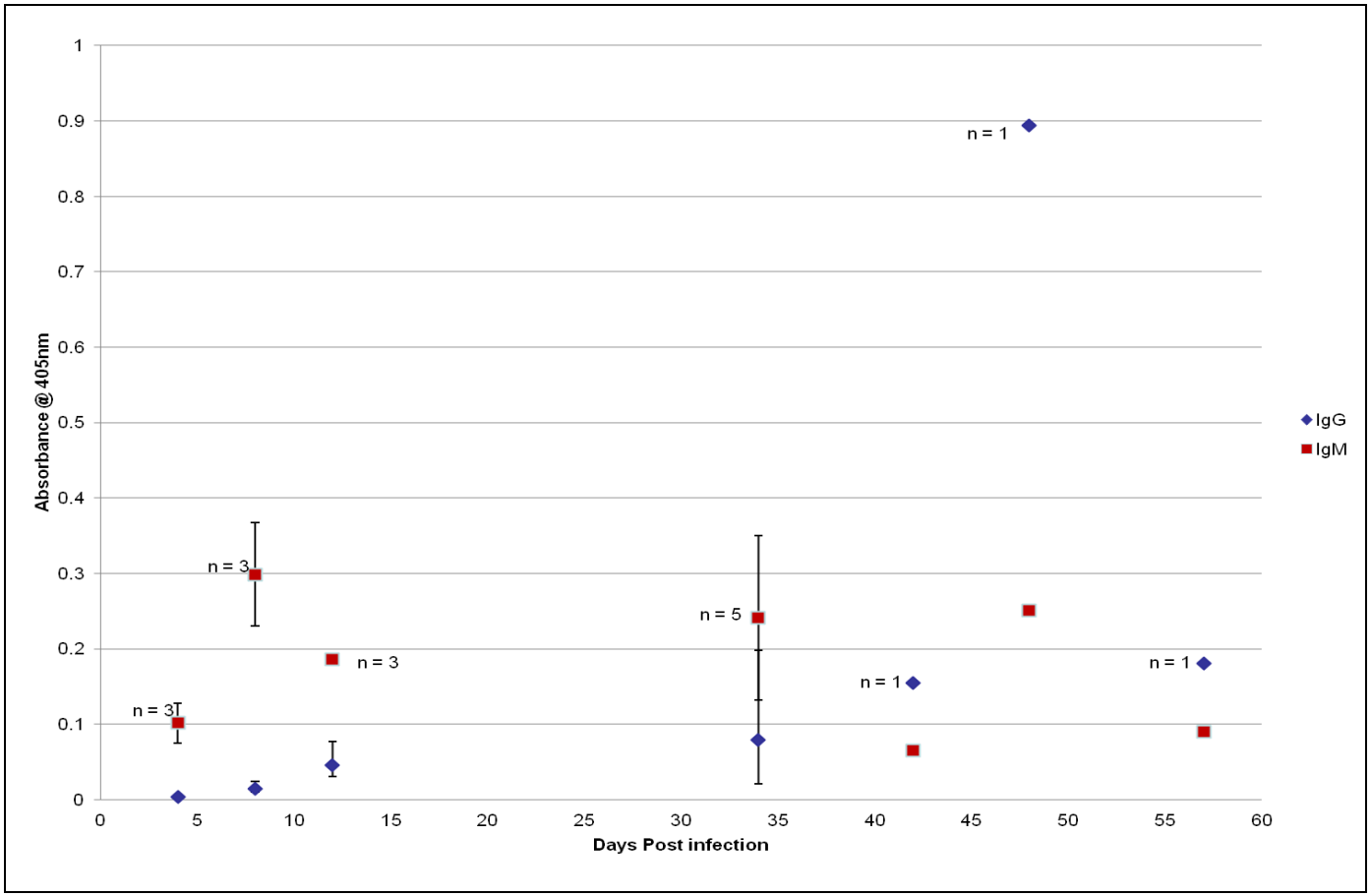


Figure 27. Indirect ELISA measurement of anti-calflagin antibodies in sera of *T. congolense*-infected mice.
At intervals after infection with *T. congolense* BSF, sera were tested by indirect ELISA using recombinant, purified *T. congolense* calflagin as antigen. All data were averaged (when n>1) and normalized by subtracting the signal obtained from prebleed serum. Error bars represent 1 standard deviation.

4.3.7. Development of an antigen capture ELISA for calflagin

Monoclonal antibodies specific for calflagin were derived by immunizing mice with purified recombinant calflagin. These antibodies were created for use as calflagin capture reagents in a sandwich ELISA in conjunction with the species specific detection mAb, Tc6/42.6.4 which had been labelled with alkaline phosphatase. A total of 7 new anti-calflagin mAbs were generated. All 7 were able to bind to recombinant calflagin and also showed reactivity with *T. congolense* lysate by ELISA, presumably binding to the native calflagin. Two of the new mAbs (1F9 and 2C7) were also able to recognize an antigen, presumably calflagin, from *T. b. brucei*. This indicates that these mAbs recognize a shared epitope between the *T. b. brucei* and *T. congolense* calflagins, which is likely not the same epitope recognized by Tc6/42.6.4. For this reason, mAbs 1F9 and 2C7 were produced at high concentrations in ascites fluids for testing of their ability to capture calflagin out of solution with subsequent detection by AP labelled mAb Tc6/42.6.4. Neither of the selected capture antibodies appeared to bind the recombinant or native calflagin strongly enough or long enough to be detected by mAb Tc6/42.6.4 (data not shown). Perhaps the fact that both mAbs were of the IgM isotype contributed to this poor result as IgM antibodies are notorious for having low affinity.

4.4. Discussion

Trypanosome-specific antibodies are useful for basic molecular research and for their potential application, for example in monitoring the presence of parasites in tsetse for epidemiological studies and for possible diagnosis of infections. In this regard, the monoclonal antibody Tc6/42.6.4 is of interest since it is one of the few trypanosome species-specific mAbs described to date.

Many attempts were made over more than two decades to identify the antigen recognized by the *T. congolense*-specific mAb Tc6/42.6.4. The antigen appeared to be moderately abundant as judged by immunofluorescence observations of fixed, permeabilized whole parasites yet only a relatively faint protein band corresponding to immunoreactive material was visualized by SDS-PAGE. The protein appears to be highly immunogenic, since antibodies were produced against this molecule after immunization of mice with whole parasites or parasite lysates. Previous attempts at identifying the antigen generally involved immunoenrichment followed by gas-phase Edman N-terminal sequencing from PVDF membranes and all failed, perhaps due to inadequate amounts of purified antigen or to blocking of N-terminal amino acid residues.

In the work reported here, I used more sensitive mass spectrometric methods to identify the antigen after enrichment using mAb Tc6/42.6.4. Based on the sequence of two contiguous peptides, the antigen was determined to be a trypanosome flagellar calcium-binding protein (also known as FCaBP or calflagin).

Homologues of calflagin are found throughout the order Trypanosomatidae [129, 141-143]. Calflagin is di-acylated by palmitoylation and myristoylation of its N terminus and is thus capable of close association with the plasma membrane, with the highest

concentration located along the flagellum [130]. Triton X-114 extraction showed that the *T. congolense* calflagin exists in both aqueous- and detergent-soluble forms (data not shown), implying that it has the capacity to become membrane bound. Again, with *T. cruzi*, calcium binding (through EF hand domains) is required for membrane localization of calflagin and when calcium chelators are added, these proteins become cytoplasmic [144, 145]. This phenomenon was also recapitulated through the immunofluorescence experiments described in this chapter and the results indicated that trypanosome calflagins behave similarly in different parasite species. The calcium dependent localization of calflagin to the flagellar membrane implies that its effect is directed at the flagellum, possibly involving the regulation of flagellar function although its function as a calcium sensor in the cytoplasm cannot be discounted.

It is known that depletion of calflagin (by RNA interference) in *T. brucei* does not inhibit growth or motility *in vitro,* indicating that calflagin is not an essential component for flagellar activity [146]. However, these same calflagin depleted BSF parasites were markedly attenuated when used to infect mice. The mice experienced a reduced parasite burden and consequently survived longer than those infected with wild-type parasites. The reason for this attenuation remains unknown but provides tantalizing possibilities for interference with calflagin as a new drug target.

It was long believed that the antigen recognized by mAb Tc6/42.6.4 was a protein unique to *T. congolense*. However, after the identification of the antigen as calflagin, it became clear that there must be a unique epitope found in the *T. congolense* calflagin but absent in other trypanosome species. Indeed several regions of primary sequence diverge between *T. congolense*, *T. brucei* and *T. cruzi*. The epitope could also be due to a unique

secondary structure adopted by *T. congolense* calflagin. It is likely that the epitope is not formed by topographical folding because SDS-PAGE denaturation did not abolish antibody binding.

I made several attempts to define the epitope of Tc6/42.6.4 but none were successful (data not shown). Tryptic or Glu-C digestion of recombinant calflagin seems to destroy the epitope recognized by mAb Tc6/42.6.4 since reactivity in ELISA or immunoblots was abolished (data not shown). It should be noted that both of these immunoassays require the adsorption of peptides to a solid surface (polystyrene plate for ELISA and PVDF for immunoblotting). Inefficient peptide adherence or physical inaccessibility of peptide to the mAb could not be ruled out as the cause for the observed negative results. Therefore, an in solution epitope mapping approach was attempted. This technique uses MS to analyze the tryptic peptides of calflagin to establish a standard peptide profile. If an antibody is added to the tryptic peptide mixture and is able to bind to a peptide, the antibody-peptide complex becomes too large to be observed by MALDI-MS. This causes the absence or reduction in intensity of the corresponding peptide peak from the mass spectrum when compared to the established peptide profile. When this experiment was performed, peptide peak reduction corresponding to antibody binding was not evident, adding credence to the theory that tryptic digestion destroys the epitope recognized by mAb Tc6/42.6.4.

In an attempt to circumvent the apparent epitope destruction by protease digestion, a method based on incomplete digestion of calflagin (with trypsin or Glu-C) coupled with protection of the epitope by antibody binding was attempted (section 4.2.13). The results of these experiments were inconclusive (data not shown). However, at the first time

point sampled in the independent protease digestions, peaks were observed at masses which would correspond to a calflagin tryptic peptide (AA 14-26) and a Glu-C peptide (AA 20-30). These peaks were not observed at later time points and were detected inconsistently when the experiments were replicated. Based on this inconclusive evidence, several overlapping peptides (AA 12-21, 14-30, and 18-35) were synthesized and assessed for binding to Tc6/42.6.4 by ELISA and MiSCREEN [140] but no binding was detected (data not shown). From these failed experiments using synthetic peptides, and with the knowledge that mAb Tc6/42.6.4 can bind to the truncated calflagin construct (AA12-202), it can be concluded that the epitope does not include the first 35 or last 2 amino acids of calflagin.

There are two ways in which mAb Tc6/42.6.4 and calflagin might be employed as diagnostic tools. The first is by serodiagnosis (the detection of host antibodies made against calflagin) and the second is by antigen capture (the detection of calflagin itself in the bloodstream of infected animals). Both options are being pursued but owing to calflagin's strong immunogenicity yet low abundance, the serodiagnostic approach seems more likely to succeed, although this assay format has its limitations (discussed in Chapter 5). *T. cruzi* calflagin has been used since the 1990s as an antigen for serodiagnosis of Chagas' disease [147, 148] and for monitoring treatment efficacy following chemotherapy [149]. *T. cruzi* calflagin is highly immunogenic in humans. Similarly, the results reported in this dissertation showed that *T. congolense* calflagin is highly immunogenic in mice, resulting in easily detected anti-calflagin antibodies from the sera of all mice infected with *T. congolense*, even early in infection. It will be necessary to test infected livestock for anti-calflagin antibodies to learn whether or not

the protein will be worth pursuing further as a serodiagnostic antigen in cattle and other domestic animals. This work will be pursued with colleagues in Africa as we cannot ship the required sera into Canada. Currently, this work is beyond the scope of this dissertation.

Attempts were made to develop an antigen capture ELISA for measurement of calflagin. MAb Tc6/42.6.4 has already been shown to be an excellent species specific detection reagent. To complete an antigen capture ELISA, a second capture antibody is needed. To this end, new anti-calflagin mAbs were generated. Unfortunately the two most promising mAbs did not function well in an antigen capture ELISA assay.

4.4.1. Future work

Experiments to identify the calflagin epitope recognized by mAb Tc6/42.6.4 using various protease digestion experiments were unsuccessful. It should be possible to synthesize an array of overlapping peptides from the entire calflagin sequence and find the peptide(s) bound by mAb Tc6/42.6.4. Having so many peptides synthesized would be expensive so instead a cloning method is currently being employed. The calflagin epitope of mAb Tc6/42.6.4 likely lies between residue 35 and 202. PCR primers recognizing internal regions of the calflagin have been designed and used in conjunction with the primers described section 4.2.8 which were designed amplify PCR the full length calflagin. Together, three calflagin gene fragments corresponding to AA 1-159, 78-159 and 78-204 have been amplified by PCR. These gene fragments are being cloned into pET-24a and will be expressed in *E. coli* and tested for binding to mAb Tc6/42.6.4 in order to define the location of the epitope.

The calflagin serodiagnostic results need to be replicated (section 4.3.6). Gaining ethical approval for a new set of mouse infections is underway. The next step for serodiagnosis of AAT would be to test sera from infected African animals. This work cannot be performed in Canada due to import restrictions on animal products. Therefore, an African collaborator must be found to perform these experiments.

Finally, for the antigen detection ELISA for calflagin, the two potential capture mAbs mentioned in section 4.3.7 were selected because they clearly did not share the same epitope as mAb Tc6/42.6.4 (based on their ability to cross react with *T. b. brucei*). Additional anti-calflagin mAbs should be tested, perhaps by SPR rather than ELISA. SPR will provide more quantifiable and sensitive data by measuring mAb-antigen binding directly rather than through a surrogate enzymatic reaction which is susceptible to interference from off target or non-specific binding. SPR will also provide information about the strength of binding which will be useful for ranking the candidate capture mAbs.

Chapter 5. Identification of trypanosome proteins in plasma of patients with late-stage African sleeping sickness

All experiments presented in this chapter were conceived by Brett Eyford and Terry Pearson in consultation with Steve Carr and Rushdy Ahmad of the Broad Institute, Cambridge MA. HAT plasma sample collection from Central Uganda was directed by Brett Eyford and Terry Pearson with the help of Dr. John Enyaru of the Makerere University, Kampala, Uganda. Plasma processing and top-down proteomic analyses were conducted by Brett Eyford, with mass spectrometry performed by Darryl Hardie at the UVic-Genome BC Proteomics Centre. Trypanosome antigen detection in human plasma was performed by Brett Eyford and "deep-mining" mass spectrometry using extended LC with a Q-Exactive instrtument was conducted by Rushdy Ahmad from the Broad Institute.

Some of the work presented in this dissertation has been prepared as a manuscript for submission:

Eyford BA, Ahmad R, Enyaru J, Carr SA and Pearson TW. Identification of trypanosome proteins in the plasma of humans infected with *T. b. rhodesiense*. PLoS Pathogens

5.1. Introduction

5.1.1. Current diagnosis of human African trypanosomiasis

Several methods are currently used to diagnose human African trypanosomiasis (HAT) also known as African sleeping sickness. These include detection of anti-trypanosome antibodies [150, 151], amplification of DNA sequences [152, 153] and direct observation of parasites by microscopic examination of patient blood or cerebrospinal fluid (CSF), usually preceded by parasite enrichment techniques [154]. Although each of these methods has problems that hinder reliable, high throughput and cost effective disease diagnosis, together, they do help disease control efforts [75].

As mentioned in Chapter 1, section 1.3, the symptoms of the early stage of HAT (headaches, fever, fatigue and joint pain) give no definitive indicators of trypanosome infection, thus HAT is often mistakenly identified as other diseases such as influenza or malaria. Many HAT patients will not seek medical aid until the parasites invade the central nervous system and the more alarming symptoms associated with late stage disease arise. Due to the risks associated with the drug melarsoprol used to treat late stage *T. b. rhodesiense* HAT patients and the cost and time required for eflornithine treatment of late stage *T. b. gambiense* HAT, it is preferable to treat early stage infections. Currently, the only way to definitively diagnose HAT is to microscopically observe the parasites in the blood (early stage disease) or in the CSF (late stage disease). Using parasite enrichment techniques, the current limit of microscopic detection is ~100 parasites/mL of blood [154]. However, between parasitemic waves, low numbers of parasites make microscopic detection unreliable. Due to low sensitivity and low

throughput, microscopic diagnosis is only used to confirm suspected infections and is not an effective tool for mass screening campaigns.

The card agglutination test for trypanosomiasis (CATT), although not registered with any regulatory agency, was heralded as a major breakthrough for diagnosis of HAT [150]. The CATT is the only serological assay widely used to detect infections with African trypanosomes. The CATT operates under the premise that one subspecies of human infective African trypanosome, *T. b. gambiense* that causes chronic infections in western Africa, often expresses a particular VSG type (LiTat 1.3) early in an infection. Many patients will generate an antibody response against LiTat 1.3 VSG and these antibodies can persist long after these parasites have been cleared. The CATT is performed by mixing lyophilized, stained *T. b. gambiense* LiTat 1.3 bloodstream form parasites with a blood sample on a plasticized card. When the blood of an infected person is exposed to the LiTat 1.3 parasites, anti-LiTat 1.3 antibodies present will cause visible agglutination. The CATT was a step forward in trypanosome diagnostics because it is a simple test that can be performed and interpreted with minimal training. It is also very quick, yielding immediate results thus facilitating the screening of large numbers of people. However, the CATT is not perfect, since the LiTat 1.3 VSG does not appear in every *T. b gambiense* infection. As with all antibody detection assays, false positive tests are possible when a person has circulating antibodies despite being cured of infection. Thus the CATT and other serological tests are not necessarily indicative of active infection. For this reason the CATT is only used as a first line screening tool to establish suspicion of an infection and a positive result must be followed up with microscopic demonstration of parasites in the blood or CSF. Of great importance is the fact that LiTat

1.3 is not expressed by *T. b. rhodesiense* parasites, thus making the CATT useless for diagnosis of the acute form of HAT found in eastern and southern Africa.

5.1.2. HAT diagnostic tools

The CATT has been the most successful HAT diagnostic tool for field application. However, due to its limitations, much effort has been put into finding other trypanosome proteins that may have use as serodiagnostic antigens for detection of anti-parasite antibodies [155]. Unfortunately, such tests based on antibody detection are hindered by variable antibody responses made by patients and the persistence of antibodies in the bloodstream even after the parasites have been eliminated.

Another avenue being explored for diagnosis and staging of HAT is to measure changes in the levels of normal human proteins in response to trypanosome infections [156-158]. Biomarkers identified by this approach are likely to be general stress response proteins and not specific to HAT. Even if a human protein is found to be specifically up- or down-regulated in response to trypanosome infection, using that protein as the basis for a diagnostic test is problematic since accurate quantitation is required to differentiate between healthy and disease associated levels. Quantitation of proteins in a complex mixture is a complicated issue but can be achieved by sandwich ELISAs or MS-based technologies [159]. However, quantitative antigen capture ELISAs are, so far, not practical for rural point-of-care diagnosis of HAT and clearly MS-based assays are also not easily adapted to field use.

Much recent attention has focused on the technique called loop-mediated isothermal amplification (LAMP) for diagnosis of HAT [152, 153]. This technique uses

oligonucleotide primers to amplify parasite DNA (encoding rRNA) in a manner similar to PCR but under isothermal conditions. In most DNA amplification techniques, the amplified products are visualized by staining of DNA in agarose gels, not easily performed in field situations. Detection can now be achieved by visual observation of turbidity resulting from magnesium pyrophosphate precipitation [160]. Isothermal heating can now be produced by the controlled exothermic reaction of calcium oxide with water to maintain the required 60 - 65 ºC for at least 1 hour [161]. The improvements in both product detection and heat supply have helped with LAMP's feasibility as a viable point-of-care diagnostic tool. Additionally, the limit of detection has been improved from 1000 parasites/ mL to as low as 1 parasite/ mL [162]. It remains to be determined if this level of sensitivity is adequate to detect trypanosome infections between parasitemic waves. It is also not known if trypanosome DNA persists in human plasma for any appreciable amount of time or if the LAMP test requires intact cells to be present in the input sample. Regardless, LAMP is a very promising diagnostic tool and will likely be applied to HAT screening campaigns in the near future.

Diagnostic assays for trypanosomiasis based on the direct detection of parasite-derived proteins theoretically offer several advantages over antibody-based and DNA-based assays. First, detection of parasite proteins released into the bloodstream of infected patients may be more indicative of active infections. Secondly, soluble protein antigens may persist at detectable levels despite fluctuating parasitemias and may become rapidly undetectable after parasite clearance, either through natural immunity or after drug treatment. Third, the presence or absence of the trypanosome protein(s) may be used for diagnosis. Quantification is not required. According to the dominant model of

trypanosome infection, parasites in sequential waves are cleared from the bloodstream by specific anti-VSG antibodies, either through opsonization and phagocytosis, agglutination and glomerular filtration or by complement–mediated lysis. Lysed parasites will theoretically release their cellular contents into the bloodstream where some of the soluble proteins might persist and possibly accumulate with subsequent parasitemias and lysis cycles. Indeed, previous work has shown that trypanosome antigens are immunologically detectable in the sera of trypanosome infected cattle [126-128], rodents [163], vervet monkeys [164] and humans [165, 166] although none of the antigens were identified. Several interesting observations arose from these studies. First, in those animals where they were examined, parasite antigens remained relatively constant and easily detectable throughout an infection, regardless of parasite cell concentration, making these antigens more reliable targets than the direct observation of the parasites. Second, in humans, the parasite antigens decreased to undetectable levels within weeks of patients being drug cured of their infections, suggesting that the antigens are better tools for monitoring active infections than the antibody surrogates detected by the CATT. Third, antigens were detected in animals and humans infected with different species of trypanosomes, suggesting that at least some of the antigens are likely to be conserved between species. This is unlike the LiTat 1.3 VSG used in the CATT that is limited to some strains of *T. b. gambiense*. Fourth, in at least two human patients, the levels of parasite antigens in the plasma began to rise several weeks after drug treatment (and after the corresponding reduction of soluble antigen concentration), prior to remission of these patients [166]. In these particular cases, monitoring of circulating parasite proteins detected re-infection (or possibly incomplete cure) before symptoms or parasites were

observed, offering the possibility that an assay based on antigen detection could be used to follow treatment efficacy.

Several antigen detection assays for human and animal trypanosomiasis have been developed [127, 167, 168] although only one (for AAT) has reported the identification of the molecule detected [128]. No reliable antigen detection tests have been developed and implemented for wide-scale use in the field. The collective data suggest that antigen detection assays have great potential for diagnosis of African sleeping sickness although it is clear that more effort is required to identify parasite antigens of greatest utility. To this end, several strategies towards antigen identification have been put forward [155, 169] although these are based on examination of the parasites themselves and are not aimed at identification of the most relevant molecules, those that are found circulating in a patient's bloodstream. Identification of soluble circulating parasite proteins in the blood of an infected animal is perhaps the most promising approach for discovery of candidate biomarkers for diagnosis of trypanosomiasis. However, this strategy is made technically difficult by the fact that it trypanosome proteins are likely to be of relatively low abundance compared to the abundant normal human plasma proteins. Indeed the detection of low levels of proteins in plasma samples has proven to be difficult due to the "large dynamic range problem" which essentially describes the fact that the most abundant protein in plasma (albumin) compared to the least (some of the cytokines) covers a dynamic range of 10^{10} [170, 171].

It is hypothesized that trypanosome proteins are released into the bloodstream of infected animals, including humans, and that detection and identification of such

molecules in plasma or serum will be indicative of active infection, thus useful for diagnosis and monitoring of trypanosomiasis.

To test this hypothesis, I chose to design and execute an "infection proteomics" strategy to identify parasite proteins in the plasma of HAT patients. To do this, an approach involving immunological detection of trypanosome antigens was coupled with mass spectrometry for protein identification.

Because of the expected low abundance of parasite proteins in plasma, plasmas from parasitologically confirmed late stage *T. b. rhodesiense* HAT patients were chosen. Plasma from late stage patients, who have potentially experienced several waves of parasitemia, was used in order to maximize the chances that they would contain higher levels of parasite antigens than plasma from patients with early stage disease. I chose to work with humans as the source of plasma since this is a "real life" situation where methods for parasite monitoring and staging of trypanosome infection are available and routinely applied. Also, the results of this work should more easily be applied to improving human health compared to using an animal model. Patients from central Uganda were chosen as this is a known focus of *T. b. rhodesiense* infections. These parasites cause an acute form of HAT where parasite numbers can reach high levels, thus potentially releasing relatively large amounts of antigens into the patients' bloodstreams and maximizing chances for their detection.

Plasmas containing trypanosome antigens (detected by ELISA) were selected for analysis by two distinct mass spectrometric approaches (Figure 28). The first method (top-down proteomics) relied on anti-trypanosome antibodies to enrich parasite proteins from the human samples. This was followed by trypsin digestion and LC-MS/MS

analysis for peptide and protein identification. The second method (bottom-up proteomics) began with Lys-C cleavage and trypsin digestion of the plasma sample. The resulting peptide mixture was then fractionated by extensive LC followed by MS/MS analysis to sequence the vast number of separated peptides. This latter approach, "deep-mining", was made possible by initial immunodepletion of the high and medium abundance plasma proteins (accounting for >99% of the total plasma protein content). In this way, the large dynamic range of proteins in plasma [170] was reduced to allow the separation and sequencing of peptides derived from lower abundance proteins, including those from trypanosomes. Nevertheless, long LC separation times and highly sensitive MS analysis were required.

Trypanosome proteins identified by either of these immunoproteomics techniques may be candidate biomarkers for either initial diagnosis or for monitoring trypanosomiasis. In addition, identification of parasite proteins found in plasma throughout the course of trypanosome infections may provide new insights into host-parasite interactions and trypanosome biology.

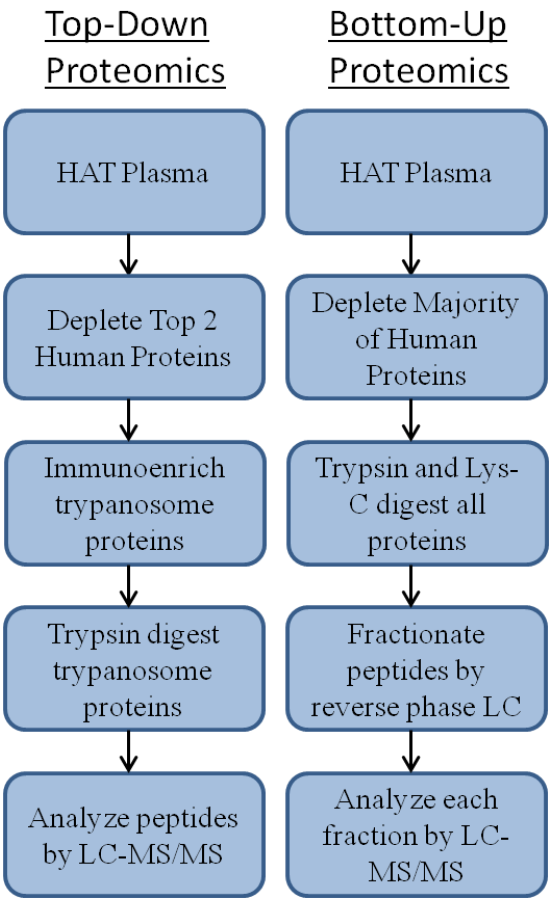


Figure 28. Schematic outline of the two immunoproteomic workflows used to identify trypanosome proteins in human plasma.

5.2. Methods

5.2.1. Ethics

Ethical clearance for the project "Biomarker discovery for staging human African trypanosomiasis patients" involving human plasma and CSF from sleeping sickness patients and control subjects was approved by the Ethical Review Board of the Vector Control Division of the Ministry of Health, Kampala, Uganda. All sleeping sickness patients gave written informed consent before enrolment and all had the option of withdrawing from the studies at any time. The Human Research Ethics Board of the University of Victoria (protocol number: 11-531) and the Massachusetts Institute of Technology's Committee on the Use of Humans as Experimental Subjects, one of the Broad Institute's institutional review boards of record, also approved the studies.

5.2.2. HAT plasma and CSF collection and preparation

Plasma and CSF were collected from HAT patients received at the Lwala Hospital, Kaberamaido District, Uganda. Patients were first identified by passive surveillance by the clinician on duty who referred them for parasitological analysis in the hospital laboratory. First, wet blood smears were examined for trypanosomes by light microscopy, followed by the hematocrit centrifugation technique for detection of low level parasitemias [172]. Patients with detectable bloodstream parasitemias were staged by double centrifugation of CSF and microscopic counting of trypanosomes and white blood cells. Blood samples from confirmed trypanosome infected patients were collected and prepared with the BD P100 Blood Collection System (Cat. No. 366448, Becton Dickinson). The resulting plasmas were stored in liquid nitrogen. Similarly, CSF

samples were collected from confirmed late-stage sleeping sickness cases and were centrifuged then stored in liquid nitrogen prior to transportation to Makerere University, Kampala followed by International shipment on dry ice to the University of Victoria. A descriptive summary of the HAT samples is shown in Figure 29.

5.2.3. Trypanosomes and cell culture

T. congolense IL3000 [86], *T. b. brucei* MiTat 1.2 [173], *T. b. brucei* 427.01 [174], *T. b. rhodesiense* ViTat 1.1 [175] and *T. b. gambiense* U2 [175] were originally obtained as BSF stabilates from the International Livestock Research Institute (ILRI), Nairobi, Kenya. PCF trypanosomes were transformed from BSF and cultured as described in Chapter 2, section 2.2.1.

5.2.4. Enzyme linked immunosorbent assay

Indirect ELISAs were performed as described in section 3.2.3. Specifically, primary antibodies used were either anti-*T. b. rhodesiense* PCF antiserum (section 5.2.5) or protein A-purified polyclonal antibodies from the same antiserum (section 5.2.6). Solid phase adsorbed antigens used in ELISAs were trypanosome PCF lysates (5×10^5 cells per well of *T. b. rhodesiense* ViTat 1.1, *T. b. gambiense* U2, *T. congolense* IL3000 and *T. b. brucei* 427.01), human CSF and plasma (diluted 1/100 in dH_2O) or fractions collected from HPLC separated trypanosome proteins immunoenriched from human plasma (section 5.2.8). The secondary antibody used in all cases was a 1:4000 dilution of goat anti-rabbit IgG (H+L) – AP (Cat. No. 075-1516, KPL, Gaithersburg MD).

5.2.5. Anti-trypanosome antibodies

Rabbit antisera were made by immunizing two New Zealand White rabbits with *T. b. rhodesiense* ViTat1.1 PCF lysate (contracted to Epitomics Inc., Burlingame CA). Rabbits were bled to obtain pre-immunization sera and after one day of rest were immunized with a mixture of trypanosome lysate prepared by sonicating 2 x 10^7 parasites/mL in PBS containing 1x protease inhibitor cocktail V (Cat. No. 539137, Calbiochem). An extended immunization procedure over a total of 78 days, involving 4 injections (1 priming dose of immunogen in FCA and 3 boosts with immunogen in FIA) of immunogen preparation was performed. Antisera were collected and prepared one week after each immunization and stored at -20 °C until used. All test bleeds reacted by ELISA (section 5.2.4) with all species of trypanosome mentioned in section 5.2.4, indicating that conserved proteins were being detected. All test bleeds were pooled at this point and were stored at -20 °C.

5.2.6. Antibody purification

Five mL of Protein A Sepharose bead slurry (Cat. No. ABP-01-5, EY Laboratories, San Mateo, CA) were packed into a column (bed volume of ~3 mL) and washed with 25 mL of PBS. The beads were mixed overnight at 4 °C with 15 mL of the pooled rabbit anti-*T. b. rhodesiense* PCF antiserum. Beads were washed with 75 mL of PBS prior to antibody elution with 20 mL of 50 mM Glycine, pH 2.5. Eluate was immediately neutralized with 3 mL of 1M Tris, pH 9 and dialyzed against PBS overnight at 4 °C. Aliquots were stored at -20 °C with 0.1% sodium azide as preservative.

5.2.7. Gel electrophoresis and immunoblotting

Electrophoresis and immunoblotting methods were performed as described in Chapter 2, section 2.2.7. The primary antibody was either a 1:1,000 dilution of rabbit anti-*T. b. rhodesiense* antiserum or a 1:200 dilution of protein A purified rabbit anti-*T. b. rhodesiense* antibodies. The secondary antibody was a 1:20,000 dilution of HRPO-conjugated goat anti-rabbit IgG (Cat. No. L42007 Caltag Laboratories, Burlingame, CA).

5.2.8. Top-down proteomics: Immunoenrichment of trypanosome proteins

Preparation of an immunoadsorbent HPLC column was performed as described by Neubert, Gale & Muirhead, 2010 [176]. A 30 mm x 2.1 mm PEEK column (Cat. No. 33.021.0030-01, Applied Research Europe) was packed with SelfPack POROS 20 G beads (Cat. No. 1-5128-10, ABSciex, Foster City, CA). Protein A purified rabbit anti-*T. b. rhodesiense* antibodies (0.6 mg/mL, 4 mL) were passed through the column in a closed loop at 1 mL/min for 4 h. The column was washed with PBS for 5 minutes at 1 mL/min and the antibody was cross-linked to the protein A with 7.78 mg/mL dimethyl pimelimidate in 100 mM triethanolamine, pH 8.2 for 30 minutes at 1 mL/min. The reaction was quenched with 100 mM ethanolamine buffer pH 8.2 for 15 minutes at 1 mL/min before washing the column with 200 mM Tris HCl pH 8.3 for 30 minutes at 1 mL/min. When not in use, the column was stored at 4°C in PBS plus 0.1% sodium azide.

For immunoenrichment of trypanosome antigens from human plasma the column was attached to an HPLC instrument comprised of a Beckmann HPLC 421 controller, a 160 absorbance detector and a 114M solvent delivery module. PBS was used as a running buffer at a flow rate of 0.25 mL/min. Column flow through was monitored by measuring

OD 280 nm. Normal human plasma, selected HAT plasma (from patients LWO150A, LWO155A and LWO156A), CSF (from patient LWO155A) and post-treatment cured plasma (from patient LWO150B) were depleted of albumin and IgG by a ProteoPrep Immunoaffinity depletion kit (Cat. No. Protia-1kt, Sigma-Aldrich, St. Louis, USA), diluted 2/5 in PBS. For each sample, 400 µL were injected. After sample injection the column was washed with PBS until flow through measured at OD 280 nm returned to baseline, at which time 0.5 mL of elution buffer (1% TFA, 3% ethanol) was injected. Additional elution injections were performed as required until the OD 280 nm reading indicated that no further protein was being eluted. All column flow through material (unbound plasma proteins, acid-eluted proteins and washes) was collected in 1 mL fractions. The presence of trypanosome antigens in each fraction was assessed by indirect ELISA as described in section 5.2.4. Fractions containing the highest levels of trypanosome antigens were pooled, the proteins digested with trypsin in solution and the peptides desalted and concentrated (described in first paragraph of Chapter 4, section 4.2.13). Peptides were identified by tandem mass spectrometry (described below).

5.2.9. Top-down proteomics: LC-MS/MS identification of peptides

(UVic-Genome BC Proteomics Centre protocol)

Ten µL of the 25 µL peptide mixtures obtained by trypsin digestion of the immunoenriched proteins were separated by on-line chromatography using a Thermo Scientific EASY-nLC II system with a reversed-phase pre-column ReproSil –Pur C18-AQC18 A1 EASY pre-column (Thermo Fisher Scientific) and an in-house prepared reverse phase nano-analytical column Magic C-18AQ (Michrom BioResources Inc.) at a

flow rate of 300 nL/min. The chromatography system was coupled on-line with an LTQ Orbitrap Velos mass spectrometer equipped with a Nanospray II source (Thermo Fisher Scientific). Solvents were A: 2% ACN/ 0.1% FA and B: 90% ACN/ 0.1% FA. Samples were separated by a 24 minute gradient (0 min: 5% B; 20 min: 40% B; 2 min: 80% B; 2 min: 80% B). Samples were ionized by electro-spray into a LTQ Orbitrap Velos (Thermo Fisher Scientific). The eight most intense ions of charge state 2-4 exceeding 5000 counts were selected for CID MS/MS fragmentation. The peak lists were submitted to an in-house Mascot 2.2 server and tested against the Uniprot-Swissprot 20110104 (523151 sequences; 184678199 residues) and T_brucei Proteins_May08_v4_20090416 (9192 sequences; 4592610 residues).

5.2.10. Bottom-up proteomics: Immunodepletion and protease digestion of plasma from HAT patients

Plasma samples from three Ugandan patients (LWO24A, LWO25A and LWO31A) infected with *T. b. rhodesiense* were pooled to achieve a total volume of 500 µl. The protein concentration in the pool plasma sample was determined by BCA assay to be 59 mg/ml. The pooled plasma sample was depleted of high and medium abundance human proteins using two immunoaffinity columns in a serial configuration: an LC20 Human IgY14 column (for removal of the top 14 most abundant proteins) followed by an LC10 Human Supermix column (Sigma-Aldrich, Saint Louis, MO). The latter resin has been reported to remove at least 155 proteins, 38% of the plasma proteome in protein number and 94% of plasma protein in mass [177]. The depleted plasma was concentrated to 400 µl with an Amicon 3 kDa concentrator and was then buffer exchanged and denatured with a 6 M urea and 50 mM ammonium bicarbonate. The protein yield of the depleted,

concentrated plasma was 100 µg as determined using a Bradford protein assay (Pierce, Rockford, IL), indicating that, approximately 99.6 % of the proteins were removed by the immunodepletion process.

The denatured sample was digested with Lys-C (Wako, Richmond, VA) and trypsin (sequencing grade modified, Promega, Madison, WI) according to an in-solution digestion protocol. The sample was first incubated for 30 minutes at 37 °C in 20 mM DTT, followed by addition of iodoacetamide to a final concentration of 50 mM followed by incubation for an additional 30 minutes at room temperature in the dark. The urea was diluted to 2 M with water prior to a 4 hour digestion with Lys-C at 1:50 (w/w) enzyme to substrate ratio at 37°C. The urea was further diluted with water to 0.6 M and the pH was adjusted to 8.0 with 1 M Tris base prior to trypsin addition (1:50 enzyme to substrate ratio) and incubation overnight at 37 °C. The reaction was stopped with FA to a final concentration of 1% and the solution was desalted using an Oasis HLB 1 cc Extraction Cartridge (Waters, Milford, MA) conditioned with 3 x 500 µL ACN, followed by 4 x 500 µL 0.1% FA. Samples were loaded onto the cartridges and washed with 3 x 500 µL 0.1% FA. Desalted peptides were eluted by 2 applications of 500 µL of 80% ACN/0.1% FA. Eluates were frozen, dried via vacuum centrifugation and stored at -80°C prior to MS analysis.

5.2.11. Bottom-up proteomics: LC-MS/MS identification of peptides from HAT plasma

(Protocol from Dr. Steven Carr's lab, Broad Institute, Cambridge MA)

Basic (pH 10) reverse phase HPLC was used to fractionate peptides from all 100 µg of the original digested protein sample using a narrow-bore 2.1 × 150 mm capillary reverse

phase column (Agilent: ZORBAX) packed with 3.5 µm beads and coupled to an Agilent 1100 HPLC system. The mobile phases were as follows: mobile phase A (20 mM ammonium formate in water, 2% ACN, pH 10) and mobile phase B (90% ACN/ 10% 20 mM ammonium formate, pH 10). The gradient was 0-5 min 0% B, 5-55 min 50% B, 55-61 min 100% B, 61-80 min 0% B at a constant flow rate of 0.2 mL/min. A total of 90 fractions were collected from 0-80 min and then combined to generate a total of 30 fractions. The fractions were frozen at -80 °C and dried before MS analysis.

The peptides samples in each other the 30 fractions were resuspended in 8 µL of 3% ACN/ 0.5% FA before LC-MS/MS analysis using a Q Exactive mass spectrometer coupled to an EASY-nLC 1000 UHPLC: Thermo Scientific). A PicoFrit column (New Objective, Woburn, MA), with an inner diameter of 75 µm packed with 20 cm of ReproSil-Pur C18 1.9 µm particles, was directly interfaced to the Q Exactive instrument equipped with a custom nano-electrospray ionization source. Two µg of the peptide mixture from each of the 30 fractions were injected and separated by a 180 minute gradient from 5 - 60% solvent B. MS analysis settings for protein identification were as follows: one precursor MS scan at 70,000 resolution in profile mode was followed by data-dependent scans of the top 12 most abundant ions at low-resolution (17,500) in profile mode. Dynamic exclusion was enabled for a duration of 20 seconds. MS/MS spectra were collected with a normalized collision energy of 28 and an isolation width of 2.5 m/z. The total peptide separation and mass spec instrument time was 90 hours, allowing detection and identification of very low levels of peptides.

5.2.12. Bottom-up proteomics: Data analysis and quantitation

(Dr. Steven Carr lab protocol)

All MS data were processed using Agilent Spectrum Mill MS Proteomics Workbench (Agilent Technologies, Palo Alto, CA) Rev B.04.00.120. MS/MS spectra were searched against a concatenated human (UNIPROT) and a *T. brucei* protein database (http://phenyx.proteomics.washington.edu/FASTAcreator/index.cgi) with parent mass tolerance of 20 ppm, fragment mass tolerance 30 ppm, a maximum of two missed cleavages and carbamidomethylation and oxidized methionine/pyroglutamic acid as fixed and variable modifications, respectively. Database matches for individual spectra were auto-validated according to user-defined scoring thresholds for both peptides (false discovery rate $\leq 1.2\%$) and proteins (minimum protein score of 20) in a two-step process. No manual inspection of spectra or validation was performed.

The sum of the precursor-ion signal intensities of all peptides derived from each protein was used to calculate total protein ion intensity. The peak area for the extracted ion chromatogram of each precursor ion was calculated automatically by the Spectrum Mill software using narrow windows around each individual member of the isotope cluster. Peak widths in both the time and m/z domains were dynamically determined based on MS scan resolution. For a given protein the ratio of the total protein intensity to total unique peptides can be calculated and is called the "molar intensity" for that protein. Although subject to some error, molar intensity can be used to rank, within an order of magnitude, the abundance of the proteins within the given sample.

5.3. Results

5.3.1. HAT plasma and CSF collection

Plasma was collected from six, parasitologically confirmed, late stage, *T. b. rhodesiense* infected patients at the Lwala Hospital in Uganda between 2008 and 2010 (Figure 29). CSF was also collected from the same six patients and 2 additional patients. Plasma was also available from some patients who had completed drug therapy and who were apparently cured. All samples were assigned a patient code (e.g. LWO24). An "A" or "B" following the patient code indicates that the sample was collected while the patient was infected or after drug cure, respectively. Although specific durations of infection prior to hospitalization for these patients are unknown, it has been established that near Lwala, Uganda, the late stage of the disease is reached, on average, 4 weeks after infection with *T. b. rhodesiense* [178]. Thus presumably, all patients had experienced several waves of parasitemia, potentially allowing trypanosome proteins to accumulate in the bloodstream after secretion from live parasites or released by parasite destruction mediated by the host immune system.

Plasmas from three patients (LWO 150, 155 and 156) were used for the top-down proteomics analysis and samples from the other three patients (LWO24, 25 and 31) were used for the bottom-up proteomics analysis (outlined in Figure 28).

No	Sample code	Date of isolation	Sex	Age	WBC	Stage	Form	CSF	Plsama
1	LWO 23A	10/01/2008	M	14	4.0	I	*T.b.rhodesiense*	√	No
2	LWO 24 A	14/01/2008	F	55	7.0	II	*T.b.rhodesiense*	√	√
3	LWO25A	17/01/2008	M	12	13.0	II	*T.b.rhodesiense*	√	√
4	LWO 30A	9/02/2008	M	72	12.0	II	*T.b.rhodesiense*	√	No
5	LWO 31A	20/02/2008	M	54	86.0	II	*T.b.rhodesiense*	√	√
6	LWO 150A	12/02/2010	M	45	25.0	II	*T.b.rhodesiense*	√	√
7	LWO 155A	11/03/2010	M	14	204.0	II	*T.b.rhodesiense*	√	√
8	LWO 156A	12/03/2010	M	27	86.0	II	*T.b.rhodesiense*	√	√
Subtotal of samples								**8**	**6**
1	LWO 23B	13/04/2008	M	14			Matched Control	No	√
2	LWO 24B	26/03/2008	F	55			Matched Control	No	√
3	LWO 25B	10/03/2008	M	12			Matched control	No	√
4	LWO 30B	13/03/2008	M	71			Matched control	No	√
5	LWO 31B	13/04/2008	M	52			Matched control	No	√
6	LWO 150B	3/03/2010	M	45			Matched control	No	√
Subtotal of samples									**6**

Figure 29. Summary of HAT patient information.
Plasma and CSF were collected from Ugandan patients at Lwala Hospital in central Uganda.

5.3.2. Top-down proteomics for discovery of trypanosome proteins in human plasma

As probes for parasite antigens, anti-trypanosome antibodies were generated by immunizing rabbits with lysates of *T. b. rhodesiense* PCF. The PCF life cycle stage was chosen as the immunogen in order to generate antibodies against conserved cellular proteins and to avoid reactivity with the immunodominat VSG molecules found on BSF trypanosomes. In immunoblotting experiments, the resulting antisera (and antibodies purified by protein A) reacted with antigens in lysates of the immunizing species of trypanosome (Figure 30). The antisera and purified antibodies also reacted in ELISA with antigens from lysates of PCF from *T. congolense*, *T. b. brucei*, and *T. b. gambiense* (not shown). These results showed that the antibodies recognize conserved antigens from multiple species of trypanosomes.

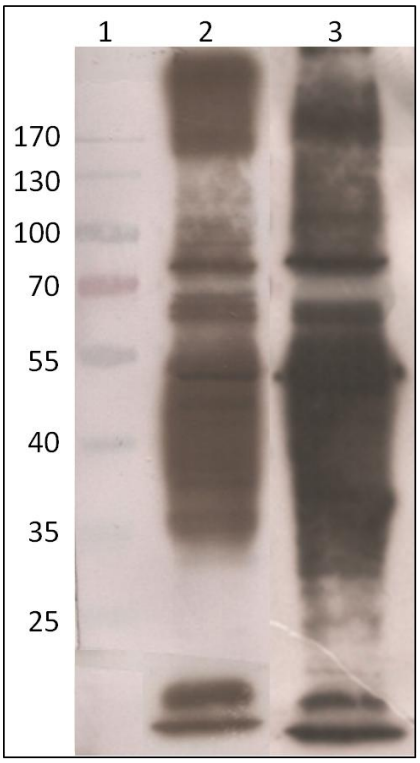


Figure 30. Immunoblot analysis of antigens in lysates of *T. b. rhodesiense.*
Lane 1: Molecular weight standards (kDa)
Lane 2: *T. b. rhodesiense* PCF lysate probed with anti-*T. b. rhodesiense* PCF antiserum
Lane 3: *T. b. rhodesiense* PCF lysate probed with protein A purified anti-*T. b. rhodesiense* antibodies

The protein A purified antibodies were covalently coupled to POROS G beads in an HPLC column. This column was used to enrich trypanosome proteins from albumin/IgG depleted HAT plasma (LWO150A, LWO155A, and LWO156A), CSF (LWO155A) and a matched plasma from a post-treatment cured patient (LWO150B). All fractions from each HPLC run were collected (initial column flow through, washes and elutions) and assessed by ELISA for the presence of trypanosome antigens. As an example, Figure 31 shows the ELISA results for the enrichment of proteins from plasma sample LWO150A. All HAT samples tested; “cured plasma”, “infected plasma” and “infected CSF”, showed similar profiles, with antigens present in the initial flow through and in as many as 3 elution fractions. A control HPLC experimental run, performed on normal (uninfected)

human plasma, showed baseline levels of background reactivity with all fractions (Figure 32). It is interesting that antigens were detected in the plasma of a "cured" patient (LWO150B) although this is not surprising. Plasma sample LWO150B was collected from the patient 20 days after the trypanosome infected sample LWO150A which also marked the start of drug therapy (Figure 29). Presumably, three weeks was enough time to cure the infection but not long enough for trypanosome proteins to be cleared from the patient's bloodstream. This time frame for antigen persistence is consistent with reports from previous literature [166].

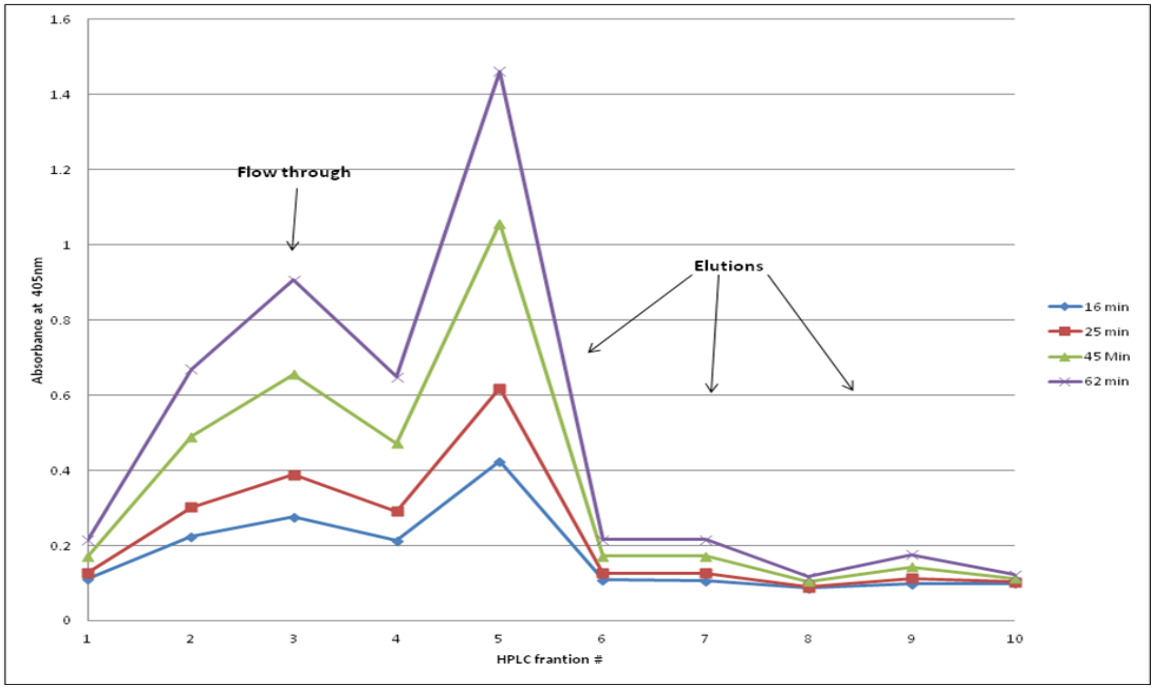


Figure 31. ELISA analysis of HPLC fractions of trypanosome proteins enriched from plasma from patient LWO150A.
The coloured lines indicate different incubation times after substrate addition.

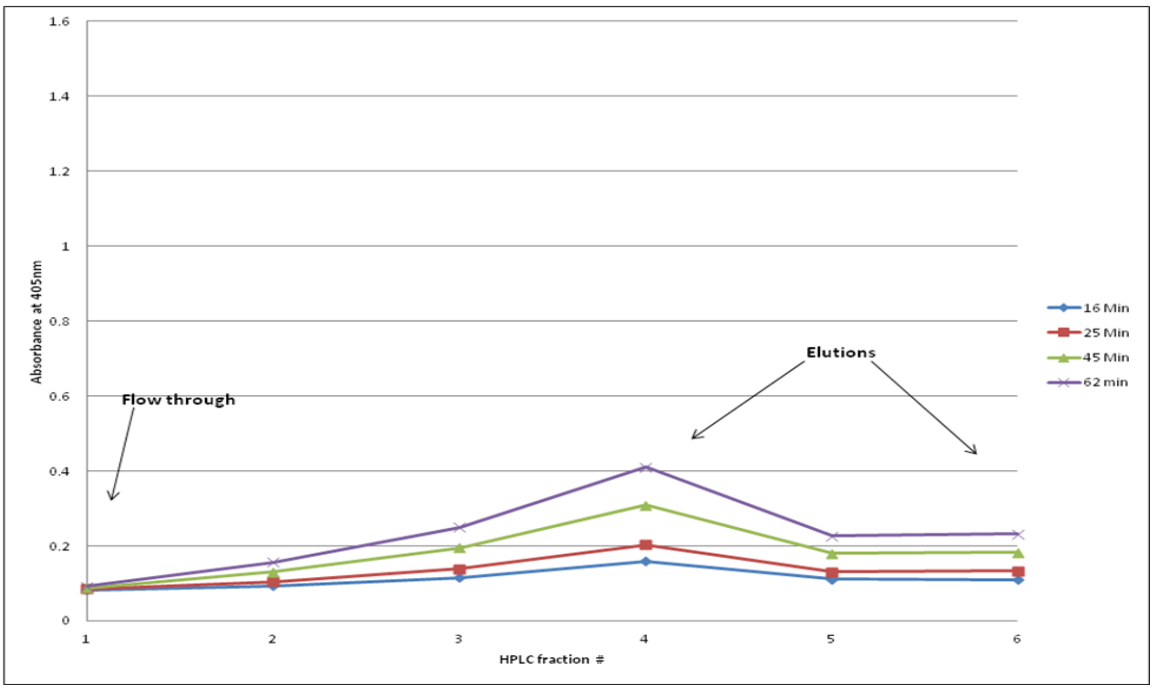


Figure 32. ELISA analysis of HPLC fractions from normal human plasma. The coloured lines indicate different incubation times after substrate addition.

The degree to which trypanosome proteins were enriched and isolated from human plasma proteins was unknown, thus to increase antigen concentration and reduce variability between samples, pools were created from selected elution fractions (Table 14).

Table 14. Pools of immunoenriched trypanosome proteins for MS/MS analysis

Pool name	Antigen source	Contributing HPLC fractions
056_1	LWO150A plasma LWO155A plasma LWO156A plasma	First elution fraction (represented by fraction 5 from Figure 31)
056_2-3	LWO150A plasma LWO155A plasma LWO156A plasma	Second and third elution fractions (represented by fractions 7 & 9 from Figure 31)
155CSF	LWO155A CSF	All three elution fractions
150B	LWO150B plasma	All three elution fractions

Enriched proteins in each plasma pool were digested with trypsin, analyzed by LC-MS/MS (section 5.2.9) and the resulting peptide data were used to search both the *T. brucei* and the UniProt non-redundant protein databases. No trypanosome proteins were identified when peptide searches were performed using the non-redundant database (data not shown). All of the proteins identified by these searches were either normal human proteins or trypsin. When the data were used to query the *T. brucei* database, the resulting hits had very weak scores. The best trypanosome protein MASCOT score was 48 from a hypothetical protein (Tb11.02.2030) with a single contributing peptide. For reference, the best human protein hit (constant region of the kappa antibody light chain, P01834|IGKC_HUMAN) had a score of 4900. When a decoy, randomized, *T. brucei* database was queried, the false discovery rates were found to range between 77 - 200% (a false discovery rate of < 1% is considered publishable). This gives further indication that no trypanosome proteins were reliably identified. These results are disappointing considering the fact that trypanosome antigens (presumably proteins) were known to be present in all samples (as detected by ELISA). Perhaps their identification by MS/MS was inhibited by ion suppression caused by an excess of peptides from human proteins in the samples. For the top down method to succeed, either further purification of trypanosome proteins would be required, or the LC-MS/MS protocol needs modification to achieve more extensive peptide fractionation and more sensitive MS/MS peptide sequencing. Unfortunately, by this time all three plasma samples had been exhausted. Further sample collection will be required before top-down immunoenrichment methods can be optimized.

5.3.3. Bottom-up proteomics for identification of trypanosome proteins in human plasma

Using the polyclonal anti-trypanosome serum, human plasma and CSF samples from late-stage sleeping sickness patients (LWO24, LWO25 and LWO31) were assessed by ELISA for the presence of detectable trypanosome antigens. The results are shown in Figure 33. Small amounts of patients' samples were used (1 µL diluted in 99 µL of water) to coat the ELISA plate wells in order to avoid overloading the ELISA wells with protein which could either mask low abundance parasite proteins or lead to erroneously high signals resulting from non-specific protein-protein interactions.

These three patients provided a good cross section of the patient population. One patient (LWO24) was a middle aged female. Of the two males, one was middle aged (LWO25) and the other was in his teens (LWO31). By ELISA, two of the patients (LWO24 and 31) had detectable trypanosome antigens in their plasma and one of these had detectable antigens in their CSF (LWO31). No trypanosome antigens were detected in the post-treatment plasmas from all 3 patients. Despite variations in apparent antigen levels, all three patients were parasitologically confirmed to have late stage HAT. The plasmas from these three patients were pooled and used as source material for MS analysis

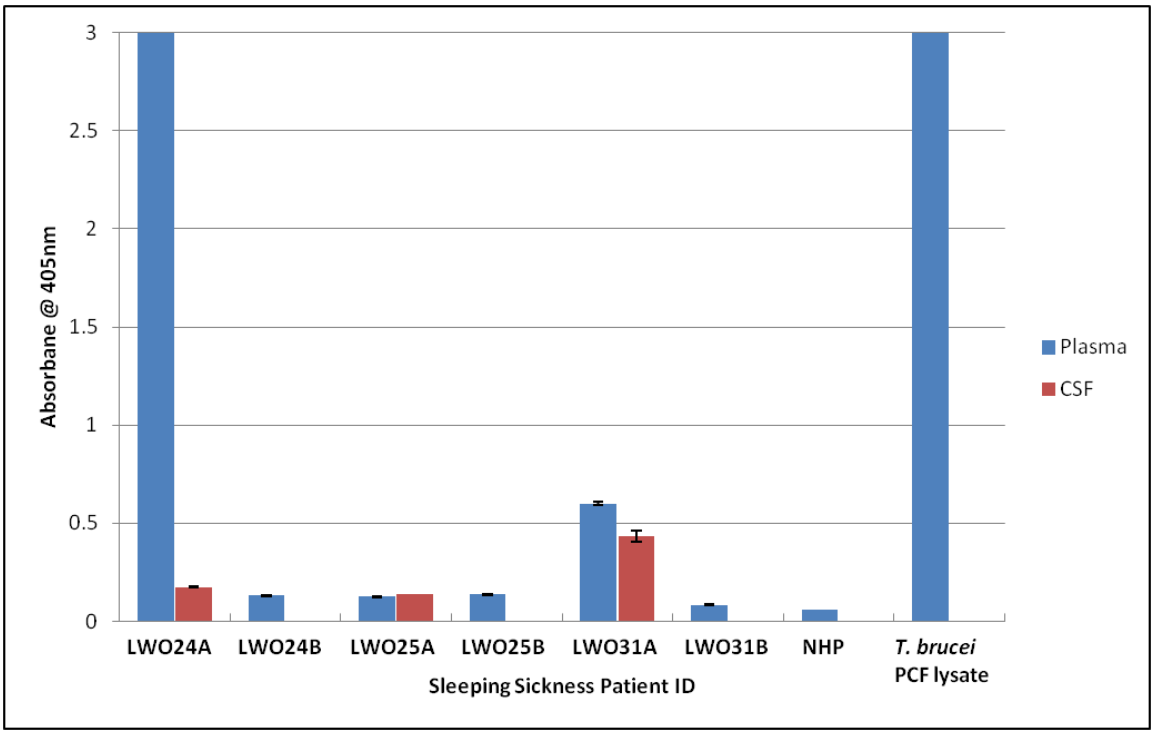


Figure 33. ELISA detection of trypanosome antigens in unmodified HAT plasma and CSF. All samples were tested in duplicate and the average response is shown. CSF samples were tested in a separate experiment from the plasmas, therefore the absolute signal intensities are not directly comparable. Error bars represent 1 standard deviation.
NHP = normal human plasma

For MS analysis, the pooled plasma sample was first immunodepleted of high-and medium-abundance human proteins to allow detection of the remaining low-abundance proteins, including those from trypanosomes. One hundred µg of total protein remained after protein depletion and concentration, representing approximately 0.4 % of the initial protein amount. The proteins were reduced, alkylated and digested sequentially with Lys-C and trypsin followed by extensive fractionation of the peptides by reverse phase LC and finally LC-MS/MS analysis.

The MS/MS peptide data generated by the bottom-up method were searched against an amalgamated NCBI human and *T. brucei* protein database. In total, 4326 human proteins and 254 trypanosome proteins were confidently identified (with at least 2 contributing peptides). An additional 1843 human proteins and 570 trypanosome proteins were

identified based on a single contributing peptide. Protein data resulting from the sequencing of a single peptide may still be relevant but is less reliable. These single peptide data have been included in the supplemental file (Appendix 2) but will not be discussed in detail in the text of this dissertation.

The number of trypanosome proteins discovered in the plasma of HAT patients was surprisingly high and can be attributed to the sensitivity of the method achieved by depletion of the highly and moderately abundant human proteins, the long reverse phase LC peptide fractionation (30 fractions collected over 80 min) and extensive MS analysis (3 hours of MS time per fraction for a total instrument time of 90 hours) using a highly sensitive Orbitrap Q Exactive mass spectrometer. As with Chapter 2, too much data was generated to describe in detail in the body of this dissertation, therefore only some of the interesting groups of trypanosome proteins will be discussed below. A complete list of all the proteins identified (both human and trypanosome) can be found in an Excel file in electronic Appendix 2.

Unlike the iTRAQ method described in Chapter 2, the bottom-up MS/MS protocol used here has no built-in quantitation capability. The proteins in the tables below are listed by molar intensity. This value is used as an approximate measure of protein abundance and is calculated by averaging the total ion intensity for a given protein by the number of peptides identified from that protein. This value represents ion intensity per peptide. If it is assumed that a mole of peptide (regardless of source or chemical nature) gives rise to the same level of signal intensity, then molar intensity can be used as a means to rank protein abundance. However, this approach is flawed, because it is known that the mass, charge and physio-chemical properties of peptides influence their

behaviour and thus signal strength in a mass spectrometer. Nevertheless, this method can be used to rank proteins with an accuracy of approximately one order of magnitude (personal communication, Dr. Rushdy Ahmad, Broad Institute).

Surprisingly, trypanosome VSGs immediately stood out as the most conspicuous group of proteins identified in the pooled HAT plasma. It was exciting yet puzzling to find twelve highly divergent trypanosome VSG sequences (Table 15 and Figure 34). The human antibody response is known to be directed against the VSG coat presented on the dominant antigenic types of trypanosomes of each parasitemic wave and it has long been believed that within each wave a single VSG type dominates. It has often been assumed that the predominant, abundant and highly immunogenic VSGs would be cleared from the body rapidly after parasite lysis. These results challenge this traditional model of antigenic variation. In the sample used here (pooled from three patients) and considering the text-book model of antigenic variation (Chapter 1, Figure 3), it would be expected that perhaps three VSG types would be seen, those from the most immediately cleared parasitemic wave. The identification of such a wide variety of VSGs in acute infections was unexpected and of interest for understanding the infection process and our current model of antigenic variation.

Table 15. VSGs discovered in HAT plasma

No. of Peptides	% Sequence Coverage	Molar Intensity	Accession No.	Protein Name
2	5.1	3.49E+09	Tb927.5.291b	variant surface glycoprotein, putative
3	5.4	9.80E+08	Tb09.244.0200	variant surface glycoprotein, putative
2	3.5	9.10E+08	Tb10.v4.0152	variant surface glycoprotein, putative
2	5.3	5.20E+08	Tb11.57.0012	variant surface glycoprotein pseudogene, putative
2	6.6	4.97E+08	Tb927.5.5210	variant surface glycoprotein, putative
2	4	4.46E+08	Tb11.30.0008	variant surface glycoprotein pseudogene, putative
5	14.4	2.58E+08	Tb927.5.4670	variant surface glycoprotein, putative
2	5.3	2.27E+08	Tb927.4.5410	variant surface glycoprotein, putative
2	5.2	2.11E+08	Tb10.v4.0058	variant surface glycoprotein, putative
2	5.6	1.71E+08	Tb05.5K5.530	variant surface glycoprotein, putative
2	3.8	1.37E+08	Tb05.5K5.320	variant surface glycoprotein, putative
2	4.8	6.85E+07	Tb10.v4.0134	variant surface glycoprotein, putative

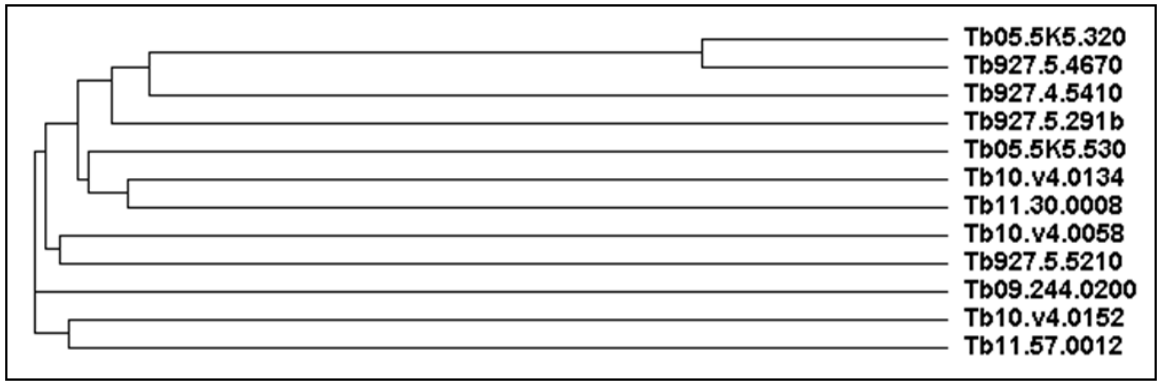


Figure 34. Cladogram of the twelve VSGs discovered in pooled HAT plasma.

The largest groups of annotated protein hits were those involving protein folding (chaperones and isomerases, Table 16) and protein degradation (proteases, peptidases and ubiquitin related proteins, Table 17). This can possibly be explained by the high rate of cell division and VSG synthesis or recycling. Beyond the known functions of the folding proteins and proteases inside the trypanosome cell, it is possible that some (especially the proteases) have a function once released from the cell. It is easy to envision a trypanosome derived protease acting as a virulence factor when released into the bloodstream. This is purely speculation but trypanosome proteases have previously been hypothesized to act as virulence factors [179]. The data generated here may offer supporting evidence for future research into the topic.

All of the trypanosome proteins in HAT plasma are interesting for their potential significance for understanding this disease. However, the abundant, conserved proteins are perhaps the most interesting from an applied point of view. The most abundant trypanosome proteins in plasma of sleeping sickness patients are mostly likely to be useful as biomarkers for diagnosis and monitoring of HAT. Table 18 shows a list of the 15 most abundant trypanosome proteins discovered in pooled HAT plasma. The proteins in this table have a broad spectrum of functional roles but all appear to be conserved soluble proteins. To estimate of their concentrations, the molar intensities of the most abundant trypanosome proteins can be compared to the molar intensities from normal human proteins in the HAT plasma (Table 19). Using this comparison, it appears that the most abundant trypanosome proteins were present in the human bloodstream in the range of 5-40 µg/mL.

Table 16. Chaperones and protein isomerases discovered in HAT plasma

No. of Peptides	% Sequence Coverage	Molar Intensity	Accession No.	Protein Name
4	6.7	3.08E+10	Tb927.7.710	heat shock 70 kDa protein, putative
4	8.6	2.65E+10	Tb927.6.3740	heat shock protein 70, mitochondrial precursor, putative
2	2.2	1.73E+10	Tb10.61.1940	chaperone protein DNAJ, putative
2	6.7	1.52E+10	Tb09.211.1350	peptidyl-prolyl cis-trans isomerase, putative
33	47.9	8.33E+09	Tb11.01.3110	heat shock protein 70
2	2.1	6.10E+09	Tb927.7.4590	chaperone protein DNAJ, putative
8	64.7	4.96E+09	Tb927.7.4770	peptidyl-prolyl cis-trans isomerase, putative
29	40.3	4.79E+09	Tb10.26.1080	heat shock protein 83
6	77	4.22E+09	Tb927.7.1320	1010 kDa heat shock protein, putative
10	74	3.89E+09	Tb11.03.0250	cyclophilin a
4	25.1	2.05E+09	Tb927.7.5790	protein disulfide isomerase, putative
11	23.5	1.81E+09	Tb11.02.5450	glucose-regulated protein 78, putative
10	22.5	1.69E+09	Tb10.6k15.2290	protein disulfide isomerise
17	50.6	1.49E+09	Tb927.8.7410	calreticulin, putative
16	45	1.49E+09	Tb927.4.5010	calreticulin, putative
9	38.5	1.12E+09	Tb10.61.0180	peptidylprolyl isomerase-like protein, putative
13	39.5	2.86E+08	Tb927.7.1300	protein disulfide isomerase, putative
10	26.3	2.47E+08	Tb10.70.0280	heat shock protein 60 chaperonin, mitochondrial precursor
2	26.9	2.07E+08	Tb927.8.690	peptidyl-prolyl cis-trans isomerase, putative
10	20.9	1.39E+08	Tb927.5.2940	stress-induced protein sti1, putative
9	17	9.19E+07	Tb10.389.0880	heat shock protein, putative
2	6.2	8.05E+07	Tb927.3.5340	heat shock cognate 70 - interacting protein, putative
8	12.2	6.35E+07	Tb11.02.0250	heat shock protein 84, putative

Table 17. Proteases and ubiquitin proteins discovered in HAT plasma

No. of Peptides	% Sequence Coverage	Molar Intensity	Accession No.	Protein Name
12	9.4	5.43E+10	Tb11.01.1680	polyubiquitin, putative
5	3.9	2.90E+09	Tb09.211.3610	ubiquitin-activating enzyme E1, putative
5	23.3	2.70E+09	Tb09.211.0050	ubiquitin-conjugating enzyme E2, putative
5	68.9	2.08E+09	Tb927.5.1000	ubiquitin-conjugating enzyme E2, putative
2	1.8	1.68E+09	Tb927.2.3030	ATP-dependent Clp protease subunit, putative
2	0.4	1.48E+09	Tb11.47.0035	calpain-like cysteine peptidase, putative
3	9.7	1.44E+09	Tb11.02.0070	aminopeptidase NPEPL1
24	70.8	9.67E+08	Tb11.02.4440	aminopeptidase, putative
6	60	7.17E+08	Tb927.7.4060	calpain-like cysteine peptidase, putative
2	18.1	6.30E+08	Tb11.02.0815	ubiquitin-conjugating enzyme, putative
34	56.3	3.38E+08	Tb10.6k15.2520	prolyl oligopeptidase, putative
2	4.6	1.25E+08	Tb10.6k15.0580	RPT6 proteasome regulatory ATPase subunit 6
3	3.5	9.93E+07	Tb10.6k15.3800	dipeptidyl-peptidase 8-like serine peptidase
2	10.3	9.80E+07	Tb10.70.7080	serine carboxypeptidase III precursor, putative
2	4	8.25E+07	Tb927.6.2150	cell division cycle protein 16, putative
4	6.4	7.73E+07	Tb927.7.190	oligopeptidase A, putative
3	8.1	5.77E+07	Tb11.02.0100	carboxypeptidase, putative
2	6.2	4.66E+07	Tb10.61.1870	aminopeptidase, putative

Table 18. The 15 highest intensity trypanosome proteins discovered in HAT plasma

No. of Peptides	% Sequence Coverage	Molar Intensity	Accession No.	Protein Name
12	9.4	5.43E+10	Tb11.01.1680	polyubiquitin, putative
4	6.7	3.08E+10	Tb927.7.710	heat shock 70 kDa protein, putative
4	8.6	2.65E+10	Tb927.6.3740	heat shock 70 kDa protein, mitochondrial precursor, putative
2	2.2	2.13E+10	Tb927.2.1170	retrotransposon hot spot protein, putative
2	2.2	1.73E+10	Tb10.61.1940	chaperone protein DNAJ, putative
3	16.2	1.63E+10	Tb927.8.5600	transaldolase, putative
2	6.7	1.52E+10	Tb09.211.1350	peptidyl-prolyl cis-trans isomerase, putative
11	57.2	1.42E+10	Tb11.02.3210	triosephosphate isomerase
13	61.7	1.37E+10	Tb11.01.4621	calmodulin
2	1.6	1.11E+10	Tb10.6k15.3410	pre-mRNA splicing factor ATP-dependent RNA helicase, putative
2	2.2	9.30E+09	Tb927.4.5340	hypothetical protein, conserved
3	0.5	9.27E+09	Tb927.8.3250	dynein heavy chain, putative
2	3.4	9.20E+09	Tb10.70.4880	eukaryotic translation initiation factor 5, putative
3	1.7	8.73E+09	Tb927.5.2090	kinesin, putative
33	47.9	8.33E+09	Tb11.01.3110	heat shock protein 70

Table 19. List of normal human plasma proteins with molar intensities similar to those from the most abundant trypanosome proteins found in HAT plasma

No. of Peptides	% Sequence Coverage	Molar Intensity	Concentration (µg/mL)	Accession No.	Protein Name
15	58.8	5.01E+10	19.0	P22352	glutathione peroxidase 3
19	38	4.17E+10	5.0	P18428	lipopolysaccharide-binding protein
28	71.4	3.93E+10	45.0	P25311	zinc-alpha-2-glycoprotein
34	64.8	3.56E+10	5.0	P36955	pigment epithelium-derived factor
33	51.7	2.48E+10	25.0	P07225	vitamin K-dependent protein S
31	67.6	1.69E+10	14.0	P29622	kallistatin
25	80	1.53E+10	6.5	P04278	sex hormone-binding globulin
40	73.4	1.38E+10	5.0	P03951	coagulation factor XI
22	57.4	1.30E+10	9.5	Q96IY4	carboxypeptidase B2
29	51.3	1.18E+10	40.0	P23142	fibulin-1
17	58.4	9.18E+09	5.0	P17936	insulin-like growth factor-binding protein 3
142	61	9.01E+09	15.0	P04275	von Willebrand factor

Values for protein concentration were retrieved from reference [171]. Human proteins reported to be depleted by the LC20 IgY14 and Supermix LC10 columns [177] have been omitted from this table.

5.4. Discussion

African trypanosomiasis may be ideally suited for this biomarker discovery methodology because of the high parasite burden experienced by patients and the fact that antigenic variation results in sequential and extended lysis of parasites, which should lead to a relatively high concentration of released parasite material when compared to other infectious diseases. To my knowledge, there is only one previously published report describing unbiased discovery of parasite proteins in patients (three proteins identified, with relatively poor data, in the saliva of malaria patients [180]), although targeted assays have been created and used to detect preselected parasite proteins from plasmodia and trypanosomes in mammalian fluids [126-128, 181]. By using an unbiased discovery method, it is more likely that biomarkers with the most diagnostic utility, those that are abundant, conserved, soluble and persistent, will be identified.

The discovery of twelve distinct VSG proteins in the plasma of HAT patients is a good example of the value of data driven, discovery based scientific methods. These results challenge the traditional model of antigenic variation (displayed in Figure 3). By using an unbiased survey to "look and see", information has been acquired which may expand our understanding of these parasites. The high number of different VSGs detected in plasma of patients with acute trypanosome infection was clearly at odds with what would have been expected from current models of antigenic variation. Either the VSGs accumulate in the bloodstream with successive parasitemic waves, or the waves of parasitemia have become unsynchronized with parasites expressing multiple different VSG types. All the patients used here had late stage infections so either explanation is possible but each scenario offers a starting point for further study into the infection

process and host-parasite interactions. Unfortunately, it is not known which VSGs came from which patient's plasma. Is this presence of multiple VSG types a common feature among HAT patients or is it the result of one patient's unique physiology? The ability to detect individual VSG types in plasma in a temporal "snapshot" bodes well for use of this sensitive proteomics method for studying the trypanosome infection process in infected humans and other animals. The general approach and method is likely to be useful for studying the "infection proteomics" of many different microbial pathogens, so far a relatively unexplored area of research.

Chaperones and proteins involved in protein degradation were also well represented in the protein discovery results. Although these proteins have less of a direct apparent impact on the disease process (compared to the well known VSGs), the identification of so many of this group (23 chaperones and 18 proteases/ubiquitin related proteins) raises questions as to why they persist in the bloodstream and whether or not they play a role in the disease process. Indeed, based on molar intensity, six of the fifteen most abundant trypanosome proteins are chaperones or proteases/ubiquitin related proteins. The rest of the top fifteen were a collection of enzymes, motor proteins, signaling proteins and two proteins of unknown function (a hypothetical protein and a retrotransposon hot spot protein).

Although trypanosome alpha- and beta-tubulin were both identified in the pooled plasma, I initially found it odd that they were not among the most abundant proteins. They are thought to be the most abundant intracellular proteins but, by molar intensity, were ranked at positions 66 (alpha tubulin) and 223 (beta tubulin). Tubulins are known to be easily detectable by MS. They were the most confidently identified proteins in the

iTRAQ experiments reported in Chapter 2. Furthermore, 8 different human tubulins were found in the HAT plasma. The reason for the relative low abundance of the trypanosome tubulins might be due to their tendency to polymerize, resulting in precipitation and increases rate of clearance from the bloodstream. Although the actual reason for their under representation is unknown, this finding further highlights the strength of using an unbiased discovery approach. Tubulins were thought to be the most abundant protein released from dying parasites and thus, a likely candidate biomarker but the experimental data suggests otherwise.

In addition to the protein families mentioned above, many metabolic, cytoskeletal, and signaling proteins were identified in human plasma. Similar to the iTRAQ results presented in Chapter 2, the largest group of trypanosome proteins identified (92 proteins) were annotated as "hypothetical proteins" which highlights the understudied nature of these parasites.

5.4.1. Future Work

While some of the trypanosome proteins discovered in plasma may be diagnostically useful, their identification here is based on a single experiment. This type of "deep mining" protein identification experiment must at least be replicated before the results enable selection of diagnostically useful (abundant and persistent) proteins. Nevertheless, the data obtained provide strong proof of principal for the discovery of trypanosome proteins in human fluids. The next step would be to perform more bottom-up MS proteomics experiments using plasma from patients with early stage HAT to identify proteins that appear early in the infection process. In this way, protein antigens

might be identified for both early diagnosis and for determination of early and late stage HAT. Identification of trypanosome proteins in CSF may also prove to be useful for biomarker discovery. This approach can be further developed for identification of biomarkers useful for monitoring clearance of infection after drug cure and for detection of relapses.

Any biomarker that will be used for development of diagnostic tests will ideally be a conserved trypanosome protein that is found in both human infective subspecies of *T. brucei* (*T. b. rhodesiense* and *T. b. gambiense*), allowing detection of HAT across Africa. Depending on the degree of protein conservation and the specificity of the mAbs used, it is possible that such tests may also allow diagnosis of AAT caused by other species of African trypanosomes such as *T. b. brucei*, *T. congolense* and *T. vivax*.

One strategy for validation of potential biomarkers is to choose several candidates from the most abundant circulating protein antigens and to use them as immunogens to produce polyclonal antibodies for use in ELISA assays. To do this, the selected proteins can be expressed as recombinant proteins and used as immunogens to generate polyclonal antisera. Affinity purification of specific antibodies followed by enzyme labeling will allow separate aliquots of the same antibodies to be used as capture and detection reagents in sandwich ELISAs. Such ELISAs can be used to screen sera or plasma from a recently established HAT biobank of HAT [182] or from fresh samples collected specifically for this kind of project. Results from this preliminary screening will indicate the most useful, reliable biomarkers (single proteins or panels) and will direct the creation of monoclonal antibody pairs for use in a standardized ELISA which can be adapted to a

point-of-care format (e.g. a lateral flow assay) for further validation in the field and ultimately for diagnosis and monitoring of trypanosome infections.

The only appreciable costs associated with this type of field test would be reagent production, kit assembly and distribution. Compared to the reagents needed for DNA based assays (primers, polymerases, nucleotides, heat source) mAbs are, once derived, relatively inexpensive to produce. This would lower production costs and ultimately the cost of the test itself. Health care agencies will be able to purchase more test kits, which will facilitate the more frequent screening of the at-risk population. In the short term, the test created using biomarkers discovered by the work presented in this dissertation will help to manage HAT by identifying active infections and, in the long term, may aid in the control of human African trypanosomiasis or ideally, to its eradication.

References

1. Hamilton, P.B., Gibson, W.C., Stevens, J.R., *Patterns of co-evolution between trypanosomes and their hosts deduced from ribosomal RNA and protein-coding gene phylogenies.* Molecular Phylogenetics and Evolution, 2007. **44**(1): p. 15-25.

2. Martin, D.S., Wright, A.D.G., Barta, J.R., Desser, S.S., *Phylogenetic position of the giant anuran trypanosomes Trypanosoma chattoni, Trypanosoma fallisi, Trypanosoma mega, Trypanosoma neveulemairei, and Trypanosoma ranarum inferred from 18S rRNA gene sequences.* Journal of Parasitology, 2002. **88**(3): p. 566-571.

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

4. Valentin, G., *Über ein Entozoon im Blute von Salmo fario.* Archgenpsychiatry, 1841: p. 435-436.

5. Roditi, I., *The surface coat of african trypanosomes*2001: Stiftung Professor Dr. Max Cloëtta.

6. de Raadt, P. *The history of sleeping sickness*. in *Fourth International Cours on African Trypanosomoses*. 2005. Tunis.

7. Steverding, D., *The history of African trypanosomiasis.* Parasites & Vectors, 2008. **1**(3).

8. Poinar, G.J., *Triatoma dominicana sp. n. (Hemiptera: Reduviidae: Triatominae), and Trypanosoma antiquus sp. n. (Stercoraria: Trypanosomatidae), the First Fossil Evidence of a Triatomine-Trypanosomatid Vector Association.* Vector-Borne and Zoonotic Diseases, 2005. **5**(1): p. 72-81.

9. Centers for Disease Control. *Parasites - American Trypanosomiasis (also known as Chagas Disease).* [cited 2012 August 30th]; Available from: http://www.cdc.gov/parasites/chagas/disease.html.

10. Bruce, D., *Preliminary report on the tsetse fly disease or nagana in Zululand Durban.* Bennett and Davis, 1895.

11. McNamara, J.J., Mohammed, G., Gibson, W.C., *Trypanosoma (Nannomonas) godfreyi sp. nov. from tsetse flies in The Gambia: biological and biochemical characterization.* Parasitology, 1994. **109**(Pt 4): p. 497-509.

12. Bruce, D., Harvey, D., Hamerton, A.E., Davey, J.B., Bruce, M., *The morphology of T. simiae, sp. nov.* Proceedings of the Royal Society B, 1912. **85**: p. 477-81.

13. Ziemann, H., *Beitrag zur Trypanosomen frage.* Parasitenk Infektionskrankh, Abt I,, 1905. **38**: p. 307-14.

14. Benne, R., *RNA-editing in trypanosome mitochondria.* Biochimica et Biophysica Acta - Gene Structure and Expression, 1989. **1007**(2): p. 131-139.

15. Robinson, D.R., Gull, K., *Basal body movement as a mechanism for mitochondrial genome segregation in the trypanosome cell cycle.* Nature, 1991. **352**: p. 731-3.

16. Wiser, M.F. *Kinetoplastid Biology and Human Disease*. 1999 [cited 2012 August 31st]; Available from: http://www.tulane.edu/~wiser/protozoology/notes/kinet.html.

17. Sutton, R.E., Boothroyd, J.C., *Evidence for Trans splicing in trypanosomes.* Cell, 1986. **47**(4): p. 527-535.

18. Mair, G., et al., *A new twist in trypanosome RNA metabolism: cis-splicing of pre-mRNA.* RNA, 2000. **6**(2): p. 163-169.

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

20. Van Den Abbeele, J., Claes, Y., van Bockstaele, D., Le Ray, D., Coosemans, M., *Trypanosoma brucei spp. development in the tsetse fly: characterization of the post-mesocyclic stages in the foregut and proboscis.* Parasitology, 1999. **118**(5): p. 469-78.

21. Seed, J., Wenck, M., *Role of the long slender to short stumpy transition in the life cycle of the african trypanosomes.* Kinetoplastid Biology and Disease, 2003. **2**(1): p. 3.

22. El-Sayed, N.M., Hegde, P., Quackenbush, J., Melville, S.E., Donelson, J.E., *The African trypanosome genome.* International Journal for Parasitology, 2000. **30**(4): p. 329-345.

23. Opperdoes, F.R., *Compartmentation of carbohyrdrate metabolism in trypanosomes.* Annual Review of Microbiology, 1987. **41**: p. 127-51.

24. Chaudhuri, M., Ajayi, W., Temple, S., Hill, G.C., *Identification and partial purification of a stage-specific 33 kDa mitochondrial protein as the alternative oxidase of the Trypanosoma brucei brucei bloodstream trypomastigotes.* Journal of Eukaryotic Microbiology, 1995. **42**(5): p. 467-72.

25. Chaudhuri, M., Hill, G.C., *Cloning, sequencing, and functional activity of the Trypanosoma brucei brucei alternative oxidase.* Molecular and Biochemical Parasitology, 1996. **83**(1): p. 125-9.

26. Bringaud, F., Rivière, L., Coustou, V., *Energy metabolism of trypanosomatids: Adaptation to available carbon sources.* Molecular and Biochemical Parasitology, 2006. **149**(1): p. 1-9.

27. Turner, M.J., *The Biochemistry of the Surface Antigens of the African Trypanosomes.* British Medical Journal, 1985. **41**(2): p. 137-143.

28. Engstler, M., Pfohl, T., Herminghaus, S., Boshart, M., Wiegertjes, G., Heddergott, N., Overath, P., *Hydrodynamic Flow-Mediated Protein Sorting on the Cell Surface of Trypanosomes.* Cell, 2007. **131**(3): p. 505-515.

29. Berriman, M., et al., *The Genome of the African Trypanosome Trypanosoma brucei.* Science, 2005. **309**(5733): p. 416-422.

30. International Laboratory for Research on Animal Diseases (ILRAD). *Annual Report of the International Laboratory for Research on Aniamal Diseases*. 1989 [cited 2012 October 3rd]; ILRAD is now part of the International Livestock Research Institute (ILRI)]. Available from: http://www.ilri.org/InfoServ/Webpub/fulldocs/ilrad89/Trypano.htm#P23_5072.

31. Vickerman, K., *Developmental Cycles and Biology of Pathogenic Trypanosomes.* British Medical Journal, 1985. **41**(2): p. 105-114.

32. Brun, R., Schönenberger M *Cultivation and in vitro cloning or procyclic culture forms of Trypanosoma brucei in a semi-defined medium.* Acta Tropica, 1979. **36**(3): p. 289-92.

33. Pearson, T.W., Moloo, S.K., Jenni, L. , *Culture form and tsetse fly midgut for procyclic Trypanosoma brucei express common proteins.* Molecular and Biochemical Parasitology, 1987. **25**(3): p. 273-8.

34. Roditi, I., Liniger, M., *Dressed for success: the surface coats of insect-borne protozoan parasites.* Trends in Microbiology, 2002. **10**(3): p. 128-34.

35. Vassella, E., et al., *Major Surface Glycoproteins of Insect Forms of Trypanosoma brucei Are Not Essential for Cyclical Transmission by Tsetse.* PLoS ONE, 2009. **4**(2): p. e4493.

36. Utz, S., Roditi, I., Kunz Renggli, C., Almeida, I.C., Acosta-Serrano, A., Bütikofer, P., *Trypanosoma congolense Procyclins: Unmasking Cryptic Major Surface Glycoproteins in Procyclic Forms.* Eukaryotic Cell, 2006. **5**(8): p. 1430-1440.

37. Bütikofer, P., Vassella, E., Boschung, M., Renggli, C.K., Brun, R., Pearson, T.W., Roditi, I., *Glycosylphosphatidylinositol-anchored surface molecules of Trypanosoma congolense insect forms are developmentally regulated in the tsetse fly*. Molecular and Biochemical Parasitology, 2002. **119**(1): p. 7-16.

38. Oberle, M., Balmer, O., Brun, R., Roditi, I., *Bottlenecks and the Maintenance of Minor Genotypes during the Life Cycle of Trypanosoma brucei.* PLoS Pathogens, 2010. **6**(7): p. e1001023.

39. Urwyler, S., Studer, E., Renggli, C.K., Roditi, I., *A family of stage-specific alanine-rich proteins on the surface of epimastigote forms of Trypanosoma brucei.* Molecular Microbiology, 2007. **63**(1): p. 218-228.

40. Sakurai, T., Sugimoto, C., Inoue, N., *Identification and molecular characterization of a novel stage-specific surface protein of Trypanosoma congolense epimastigotes.* Molecular and Biochemical Parasitology, 2008. **161**(1): p. 1-11.

41. World Health Organization. *Human African trypanosomiasis (sleeping sickness) Fact Sheet*. 2012 [cited 2012 August 8th]; Available from: http://www.who.int/mediacentre/factsheets/fs259/en/.

42. World Health Organization. *Weekly epidemiological record*. 2006 Feb 24, 2006 [cited 2012 August 9th]; 81:[69-80]. Available from: http://www.who.int/wer/2006/wer8108.pdf.

43. Louis, F.J., Simarro, P.P. , *Rough start for the fi ght against sleeping sickness in French equatorial Africa.* Medecine Tropicale (Marseilles), 2005. **65**(3): p. 251-57.

44. Pays, E., Vanhollebeke, B., Vanhamme, L., Paturiaux-Hanocq, F., Nolan, D.P., Pérez-Morga, D., *The trypanolytic factor of human serum.* Nature Reviews Microbiology, 2006. **4**(6): p. 477-486.

45. Vanhollebeke, B., De Muylder, G., Nielsen, M.J., Pays, A., Tebabi, P., Dieu, M., Raes, M., Moestrup, S.K., Pays, E., *A Haptoglobin-Hemoglobin Receptor Conveys Innate Immunity to Trypanosoma brucei in Humans.* Science, 2008. **320**(5876): p. 677-681.

46. Drain, J., Bishop, J.R., Hajduk, S.L., *Haptoglobin-related Protein Mediates Trypanosome Lytic Factor Binding to Trypanosomes.* Journal of Biological Chemistry, 2001. **276**(32): p. 30254-30260.

47. Wheeler, R.J., *The trypanolytic factor–mechanism, impacts and applications.* Trends in Parasitology, 2010. **26**(9): p. 457-464.

48. Vanhamme, L., Paturiaux-Hanocq, F., Poelvoorde, P., Nolan, D.P., Lins, L., Van Den Abbeele, J., Pays, A., Tebabi, P., Van Xong, H., Jacquet, A., Moguilevsky, N., Dieu, M., Kane, J.P., De Baetselier, P., Brasseur, R., Pays, E., *Apolipoprotein L-I is the trypanosome lytic factor of human serum.* Nature, 2003. **422**(6927): p. 83-87.

49. Molina Portela, M.P., Raper, J., Tomlinson, S., *An investigation into the mechanism of trypanosome lysis by human serum factors.* Molecular and Biochemical Parasitology, 2000. **110**(2): p. 273-282.

50. Van Xong, H., Vanhamme, L., Chamekh, M., Chimfwembe, C., Van Den Abbeele, J., Pays, A., Van Meirvenne, N., Hamers, R., De Baetselier, P., Pays, E., *A VSG Expression Site–Associated Gene Confers Resistance to Human Serum in Trypanosoma rhodesiense.* Cell, 1998. **95**(6): p. 839-846.

51. Milner, J.D., Hajduk, S.L., *Expression and localization of serum resistance associated protein in Trypanosoma brucei rhodesiense.* Molecular and Biochemical Parasitology, 1999. **104**(2): p. 271-283.

52. Oli, M.W., Cotlin, L.F., Shiflett, A.M., Hajduk, S.L., *Serum Resistance-Associated Protein Blocks Lysosomal Targeting of Trypanosome Lytic Factor in Trypanosoma brucei.* Eukaryotic Cell, 2006. **5**(1): p. 132-139.

53. Stephens, N.A., Hajduk, S.L., *Endosomal Localization of the Serum Resistance-Associated Protein in African Trypanosomes Confers Human Infectivity.* Eukaryotic Cell, 2011. **10**(8): p. 1023-1033.

54. Symula, R.E., Beadell, J.S., Sistrom, M., Agbebakun, K., Balmer, O., Gibson, W., Aksoy, S., Caccone, A., *Trypanosoma brucei gambiense Group 1 Is Distinguished by a Unique Amino Acid Substitution in the HpHb Receptor Implicated in Human Serum Resistance.* PLoS Neglected Tropical Diseases, 2012. **6**(7): p. e1728.

55. Kieft, R., Capewell, P., Turner, C.M.R., Veitch, N.J., MacLeod, A. Hajduk, S., *Mechanism of Trypanosoma brucei gambiense (group 1) resistance to human trypanosome lytic factor.* Proceedings of the National Academy of Sciences, 2010. **107**(37): p. 16137-16141.

56. Simarro, P.P., Jannin, J., Cattand, P., *Eliminating Human African Trypanosomiasis: Where Do We Stand and What Comes Next?* PLoS Medicine, 2008. **5**(2): p. e55.

57. World Health Organization. *WHO Report on Global Surveillance of Epidemic-prone Infectious Diseases - African trypanosomiasis*. 2012 [cited 2012 November 8th]; Available from: http://www.who.int/csr/resources/publications/CSR_ISR_2000_1tryps/en/index.html.

58. Abenga, J.N., Lawal, I.A., *Implicating roles of animal reservoir hosts in the resurgence of Gambian trypanosomosis (Sleeping Sickness).* African Journal of Biotechnology, 2005. **4**(2): p. 134-7.

59. Jamonneau, V., Ilboudo, H., Kaboré, J., Kaba, D., Koffi, M., Solano, P., Garcia, A., Courtin, D., Laveissière, C., Lingue, K., Büscher, P., Bucheton, B., *Untreated*

Human Infections by Trypanosoma brucei gambiense Are Not 100% Fatal. PLoS Neglected Tropical Diseases, 2012. **6**(6): p. e1691.

60. Urech, K., Neumayr, A., Blum, J., *Sleeping Sickness in Travelers - Do They Really Sleep?* PLoS Neglected Tropical Diseases, 2011. **5**(11): p. e1358.

61. Brun, R., Blum, J., Chappuis, F., Burri, C., *Human African trypanosomiasis.* Lancet, 2010. **375**(9709): p. 148-159.

62. Fairlamb, A.H., *Chemotherapy of human African trypanosomiasis: current and future prospects.* Trends in Parasitology, 2003. **19**(11): p. 488-494.

63. Burri, C., *Chemotherapy against human African trypanosomiasis: Is there a road to success?* Parasitology, 2010. **137**(Special Issue 14): p. 1987-1994.

64. Delespaux, V., de Koning, H.P., *Drugs and drug resistance in African trypanosomiasis.* Drug Resist Update, 2007. **10**(1–2): p. 30-50.

65. Yun, O., Priotto, G., Tong, J., Flevaud, L., Chappuis, F., *NECT Is Next: Implementing the New Drug Combination Therapy for Trypanosoma brucei gambiense Sleeping Sickness.* PLoS Neglected Tropical Diseases, 2010. **4**(5): p. e720.

66. Priotto, G., Kasparian, S., Mutombo, W., Ngouama, D., Ghorashian, S., Arnold, U., Ghabri, S., Baudin, E., Buard, V., Kazadi-Kyanza, S., Ilunga, M., Mutangala, W., Pohlig, G., Schmid, C., Karunakara, U., Torreele, E., Kande, V., *Nifurtimox-eflornithine combination therapy for second-stage African Trypanosoma brucei gambiense trypanosomiasis: a multicentre, randomised, phase III, non-inferiority trial.* The Lancet, 2009. **374**(9683): p. 56-64.

67. Checchi, F., Piola, P., Ayikoru, H., Thomas, F., Legros, D., Priotto, G., *Nifurtimox plus Eflornithine for Late-Stage Sleeping Sickness in Uganda: A Case Series.* PLoS Neglected Tropical Diseases, 2007. **1**(2): p. e64.

68. The RTS S Clinical Trials Partnership, *First Results of Phase 3 Trial of RTS,S/AS01 Malaria Vaccine in African Children.* New England Journal of Medicine, 2011. **365**(20): p. 1863-1875.

69. Regules, J.A., Cummings, J.F., Ockenhouse, C.F., *The RTS,S vaccine candidate for malaria.* Expert Review of Vaccines, 2011. **10**(5): p. 589-599.

70. Rasooly, R., Balaban, N., *Trypanosome microtubule-associated protein p15 as a vaccine for the prevention of African sleeping sickness.* Vaccine, 2004. **22**(8): p. 1007-1015.

71. Luhrs, K.A., Fouts, D.L., Manning, J.E., *Immunization with recombinant paraflagellar rod protein induces protective immunity against Trypanosoma cruzi infection.* Vaccine, 2003. **21**(21–22): p. 3058-3069.

72. Soltys, M.A., *Immunity in African Trypanosomiasis.* Bulletin of the World Health Organization, 1963. **28**: p. 753-61.

73. Maudlin, I., Holmes, P.H., Miles, M.A. , *The Trypanosomiases*2004, Wallingford, UK: CABI Publishing.

74. Swallow, B.M., *Impacts of Trypanosomiasis on African agriculture*, in *International Scientific Council for Trypanosomosis Research and Control*1999.

75. Programme Against African Trypanosomosis, *Tsetse and Trypanosomiasis Information Bulletin, Volume 30, Part 2*, 2007.

76. Programme Against African Trypanosomosis. *PAAT: The Disease* 2012 [cited 2012 August 13th]; Available from: http://www.fao.org/ag/againfo/programmes/en/paat/disease.html.

77. Gadelha, C., Holden, J.M., Allison, H.C., Field, M.C., *Specializations in a successful parasite: What makes the bloodstream-form African trypanosome so deadly?* Molecular and Biochemical Parasitology, 2011. **179**(2): p. 51-58.

78. Besteiro, S., Barrett, M.P., Rivière, L., Bringaud, F., *Energy generation in insect stages of Trypanosoma brucei: metabolism in flux.* Trends in Parasitology, 2005. **21**(4): p. 185-191.

79. Roditi, I., Lehane, M.J., *Interactions between trypanosomes and tsetse flies.* Current Opinions in Microbiology, 2008. **11**(4): p. 344-51.

80. Walshe, D.P., Ooi, C.P., Lehane, M.J., Haines, L.R., *Chapter 3 The Enemy Within: Interactions Between Tsetse, Trypanosomes and Symbionts*, in *Advances in Insect Physiology*, J.S. Stephen and C. Jeacuterocircme, Editors. 2009, Academic Press. p. 119-175.

81. Haines, L.R.L., S.M., Pearson, T.W., Lehane, M.J., *Tsetse EP Protein Protects the Fly Midgut from Trypanosome Establishment.* PLoS Pathog, 2010. **6**(3): p. e1000793.

82. Gibson, W., Bailey, M., *The development of Trypanosoma brucei within the tsetse fly midgut observed using green fluorescent trypanosomes.* Kinetoplastid Biology and Disease, 2003. **2**(1): p. 1.

83. Evans, D.A., Ellis, D.S., Stamford, S., *Ultrastructural Studies of Certain Aspects of the Development of Trypanosoma congolense in Glossina morsitans morsitans.* Journal of Eukaryotic Microbiology, 1979. **26**(4): p. 557-563.

84. Helm, J.R., Hertz-Fowler, C., Aslett, M., Berriman, M., Sanders, M., Quail, M.A., Soares, M.B., Bonaldo, M.F., Sakurai, T., Inoue, N., Donelson, J.E., *Analysis of expressed sequence tags from the four main developmental stages of*

Trypanosoma congolense. Molecular and Biochemical Parasitology, 2009. **168**(1): p. 34-42.

85. Bienen, E.J., Webster, P., Fish, W.R., *Trypanosoma (Nannomonas) congolense: changes in respiratory metabolism during the life cycle.* Experimental Parasitology, 1991. **73**(4): p. 403-12.

86. Fish, W.R., Muriuki CW, Muthiani AM, Grab DJ, Lonsdale-Eccles JD, *Disulfide bond involvement in the maintenance of the cryptic nature of the cross-reacting determinant of metacyclic forms of Trypanosoma congolense.* Biochemistry, 1989. **28**(13): p. 5415-21.

87. Sadowski, P.G., Dunkley, T.P.J., Shadforth, I.P., Dupree, P., Bessant, C., Griffin, J.L., Lilley, K.S., *Quantitative proteomic approach to study subcellular localization of membrane proteins.* Nat. Protocols, 2006. **1**(4): p. 1778-1789.

88. Zieske, L.R., *A perspective on the use of iTRAQ™ reagent technology for protein complex and profiling studies.* Journal of Experimental Botony, 2006. **57**(7): p. 1501-1508.

89. Lanham, S.M., Godfrey D.G., *Isolation of salivarian trypanosomes from man and other mammals using DEAE-cellulose.* Experimental Parasitology, 1970. **28**(3): p. 521-34.

90. Hirumi, H., Hirumi, K., *In vitro cultivation of Trypanosoma congolense bloodstream forms in the absence of feeder cell layers.* Parasitology, 1991. **102**(Pt2): p. 225–36.

91. Wellcome Trust Sanger Institute. *Trypanosoma congolense*. 2012; Available from: http://www.sanger.ac.uk/Projects/T_congolense/.

92. Seymour, S.L. *ProteinPilot Software for Protein Identification and Quantitation*. 2006 [cited 2012 October 1st]; Available from: http://appliedbiosystems.cnpg.com/lsca/webinar/proteinpilot/20060516/.

93. Kuzyk, M.A., Ohlund, L.B., Elliott, M.H., Smith, D., Qian, H., Delaney, A., Hunter, C.L., Borchers, C.H., *A comparison of MS/MS-based, stable-isotope-labeled, quantitation performance on ESI-quadrupole TOF and MALDI-TOF/TOF mass spectrometers.* Proteomics, 2009. **9**(12): p. 3328-3340.

94. Beecroft, R.P., Roditi, I., Pearson, T.W., *Identification and characterization of an acidic major surface glycoprotein from procyclic stage Trypanosoma congolense.* Molecular and Biochemical Parasitology, 1993. **61**: p. 285-294.

95. Kelley, R.J., Brickman, M.J., Balber, A.E., *Processing and transport of a lysosomal membrane glycoprotein is developmentally regulated in African trypanosomes.* Molecular and Biochemical Parasitology, 1995. **74**(2): p. 167-178.

96. Stebeck, C.E., Frevert, U., Mommsen, T.P., Vassella, E., Roditi, I., Pearson, T.W., *Molecular characterization of glycosomal NAD(+)-dependent glycerol 3-phosphate dehydrogenase from Trypanosoma brucei rhodesiense.* Molecular and Biochemical Parasitology, 1996. **76**(1-2): p. 145-58.

97. Loveless, B., *Studies on the expression of the major cell surface molecules of insect forms of Trypanosoma congolense, a major parasite of cattle in Africa* in *Department of Biochemistry & Microbiology*2010, University of Victoria: Victoria.

98. Parish, N.M., Morrison, W.I., Pearson, T.W., *Identification of an antigen specific to Trypanosoma congolense by using monoclonal antibodies.* Journal of Immunology, 1985. **134**(1): p. 593-597.

99. Bayne, R.A., Kilbride, E.A., Lainson, F.A., Tetley, L., Barry, J.D. , *A major surface antigen of procyclic stage Trypanosoma congolense* Molecular and Biochemical Parasitology, 1993. **61**(2): p. 295-310.

100. Crowe, J.S., Barry, J.D., Luckins, A.G., Ross, C.A., Vickerman, K. , *All metacyclic variable antigen types of Trypanosoma congolense identified using monoclonal antibodies.* Nature, 1983. **306**(5941): p. 389-91.

101. Lenardo, M.J., Rice-Ficht, A.C., Kelly, G., Esser, K.M., Donelson, J.E., *Characterization of the genes specifying two metacyclic variable antigen types in Trypanosoma brucei rhodesiense.* Proceedings of the National Academy of Sciences, 1984. **81**(21): p. 6642-6646.

102. Donelson, J.E., *Antigenic variation and the African trypanosome genome.* Acta Tropica, 2003. **85**(3): p. 391-404.

103. Ziegelbauer, K., Multhaup, G., Overath, P., *Molecular characterization of two invariant surface glycoproteins specific for the bloodstream stage of Trypanosoma brucei.* Journal of Biological Chemistry, 1992. **267**(15): p. 10797-10803.

104. Overath, P., Chaudhri, M., Steverding, D., Ziegelbauer, K., *Invariant surface proteins in bloodstream forms of Trypanosoma brucei.* Parasitology Today, 1994. **10**(2): p. 53-8.

105. Jackson, D.G., Smith, D.K., Luo, C., Elliott, J.F., *Cloning of a novel surface antigen from the insect stages of Trypanosoma brucei by expression in COS cells.* Journal of Biological Chemistry, 1993. **268**(3): p. 1894-900.

106. Marcoux, V., Wei, G., Tabel, H., Bull, H.J. , *Characterization of major surface protease homologues of Trypanosoma congolense* Journal of Biomedicine and Biotechnology, 2010.

107. El-Sayed, N.M.A., Donelson, J.E., *African Trypanosomes Have Differentially Expressed Genes Encoding Homologues of the Leishmania GP63 Surface Protease.* Journal of Biological Chemistry, 1997. **272**(42): p. 26742-26748.

108. LaCount, D.J., Gruszynski, A.E., Grandgenett, P.M., Bangs, J.D., Donelson, J.E., *Expression and Function of the Trypanosoma brucei Major Surface Protease (GP63) Genes.* Journal of Biological Chemistry, 2003. **278**(27): p. 24658-24664.

109. Grandgenett, P.M., Otsu, K., Wilson, H.R., Wilson, M.E., Donelson, J.E., *A Function for a Specific Zinc Metalloprotease of African Trypanosomes.* PLoS Pathogens, 2007. **3**(10): p. e150.

110. Michels, P.A.M., Bringaud, F., Herman, M., Hannaert, V., *Metabolic functions of glycosomes in trypanosomatids.* Biochimica et Biophysica Acta (BBA) - Molecular Cell Research, 2006. **1763**(12): p. 1463-1477.

111. Richardson, J.P., Beecroft, R.P., Tolson, D.L., Liu, M.K., Pearson, T.W., *Procyclin: an unusual immunodominant glycoprotein surface antigen from the procyclic stage of African trypanosomes.* Molecular and Biochemical Parasitology, 1988. **31**(3): p. 203-216.

112. Loveless, B.C., Mason, J.W., Sakurai, T., Inoue, N., Razavi, M., Pearson, T.W., Boulanger, M.J., *Structural Characterization and Epitope Mapping of the Glutamic Acid/Alanine-rich Protein from Trypanosoma congolense: Defining Assembly of the Parasite Cell Surface.* Journal of Biological Chemistry, 2011. **286**(23): p. 20658-65.

113. Quevillon, E., Silventoinen, V., Pillai, S., Harte, N., Mulder, N., Apweiler, R., Lopez, R., *InterProScan: protein domains identifier.* Nucleic Acids Research, 2005. **33**(suppl 2): p. W116-W120.

114. Fragoso, C.M., Schumann Burkard, G., Oberle, M., Renggli, C.K., Hilzinger, K., Roditi, I., *PSSA-2, a Membrane-Spanning Phosphoprotein of Trypanosoma brucei, Is Required for Efficient Maturation of Infection.* PLoS ONE, 2009. **4**(9): p. e7074.

115. Tolson, D.L., Turco SJ, Beecroft RP, Pearson TW, *The immunochemical structure and surface arrangement of Leishmania donovani lipophosphoglycan determined using monoclonal antibodies.* Molecular and Biochemical Parasitology, 1989. **35**(2): p. 109-118.

116. Battye, T.G., et al., *iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM.* Acta Crystallogr D Biol Crystallogr, 2011. **67**(Pt 4): p. 271-81.

117. Winn, M.D., et al., *Overview of the CCP4 suite and current developments.* Acta Crystallogr D Biol Crystallogr, 2011. **67**(Pt 4): p. 235-42.

118. Matthews, B.W., *Solvent content of protein crystals.* J Mol Biol, 1968. **33**(2): p. 491-7.

119. Desquesnes, M., Bengaly, Z., Millogo, L., Meme, Y., Sakande, H., *The analysis of the cross-reactions occurring in antibody-ELISA for the detection of trypanosomes can improve identification of the parasite species involved.* Annals of Tropical Medicine and Parasitology, 2001. **95**(2): p. 141-155.

120. Greiner, M., Kumar, S., Kyeswa, C., *Evaluation and comparison of antibody ELISAs for serodiagnosis of bovine trypanosomosis.* Veterinary Parasitology, 1997. **73**(3–4): p. 197-205.

121. Njiru, Z.K., Ouma, J.O., Enyaru, J.C., Dargantes, A.P., *Loop-mediated Isothermal Amplification (LAMP) test for detection of Trypanosoma evansi strain B.* Experimental Parasitology, 2010. **125**(3): p. 196-201.

122. Laohasinnarong, D., Thekisoe, O., Malele, I., Namangala, B., Ishii, A., Goto, Y., Kawazu, S., Sugimoto, C., Inoue, N., *Prevalence of Trypanosoma sp. in cattle from Tanzania estimated by conventional PCR and loop-mediated isothermal amplification (LAMP).* Parasitology Research, 2011. **109**(6): p. 1735-1739.

123. Bajyana, S.E., Hamers R., *A card agglutination test (CATT) for veterinary use based on an early VAT RoTat 1/2 of Trypanosoma evansi.* Annales de la Société Belge de Médecine Tropicale, 1988. **68**(3): p. 233-40.

124. Boulangé, A., Katende, J., Authié, E., *Trypanosoma congolense: Expression of a Heat Shock Protein 70 and Initial Evaluation as a Diagnostic Antigen for Bovine Trypanosomosis.* Experimental Parasitology, 2002. **100**(1): p. 6-11.

125. Bossard, G., Boulange, A., Holzmuller, P., Thévenon, S., Patrel, D., Authie, E., *Serodiagnosis of bovine trypanosomosis based on HSP70/BiP inhibition ELISA.* Veterinary Parasitology, 2010. **173**(1-2): p. 39-47.

126. Masake, R.A., Nantulya VM, *Sensitivity of an antigen detection enzyme immunoassay for diagnosis of Trypanosoma congolense infections in goats and cattle.* Journal of Parasitology, 1991. **77**(2): p. 231-6.

127. Nantulya, V.M., Lindqvist KJ, Stevenson P, Mwangi EK, *Application of a monoclonal antibody-based antigen detection enzyme-linked immunosorbent assay (antigen ELISA) for field diagnosis of bovine trypanosomiasis at Nguruman, Kenya.* Annals of Tropical Medicine and Parasitolology, 1992. **86**(3): p. 225-30.

128. Kashiwazaki, Y., Snowden K, Smith DH, Hommel M, *A multiple antigen detection dipstick colloidal dye immunoassay for the field diagnosis of trypanosome infections in cattle.* Veterinary Parasitology, 1994. **55**(1-2): p. 57-69.

129. Maldonado, R.A., Linss, J., Thomaz, N., Olson, C.L., Engman, D.M., Goldenberg, S., *Homologues of the 24-kDa Flagellar Ca2+-Binding Protein Gene of Trypanosoma cruzi are Present in Other Members of the Trypanosomatidae Family.* Experimental Parasitology, 1997. **86**(3): p. 200-205.

130. Godsel, L.M., Engman, D.M., *Flagellar protein localization mediated by a calcium-myristoyl/palmitoyl switch mechanism.* EMBO Journal, 1999. **18**(8): p. 2057-2065.

131. Paling, R.W., Moloo, S.K., Jenni, L., *Trypanosoma congolense: host responses following tsetse transmitted infection of Kilifi isolates in goats.* Experimental Parasitology, 1987. **63**(3): p. 279-87.

132. Zweygarth, E., Röttcher, D., *The occurrence of Trypanosoma (Nannomonas) simiae in the cerebrospinal fluid (CS) of domestic pigs.* Parasitology Research, 1987. **73**(5): p. 479-80.

133. Okwor, I., Muleme, H., Jia, P., Uzonna, J.E., *Altered Proinflammatory Cytokine Production and Enhanced Resistance to Trypanosoma congolense Infection in Lymphotoxin β-Deficient Mice.* Journal of Infectious Diseases, 2009. **200**(3): p. 361-369.

134. Hu, C., Rio, R.V.M., Medlock, J., Haines, L.R., Nayduch, D., Savage, A.F., Guz, N., Attardo, G.M., Pearson, T.W., Galvani, A.P. Aksoy, S., *Infections with Immunogenic Trypanosomes Reduce Tsetse Reproductive Fitness: Potential Impact of Different Parasite Strains on Vector Population Structure.* PLoS Neglected Tropical Diseases, 2008. **2**(3): p. e192.

135. Bordier, C., *Phase separation of integral membrane proteins in Triton X-114 solution.* Journal of Biological Chemistry, 1981. **256**(4): p. 1604-1607.

136. Bains, J., Leon, R., Temke, K.G., Boulanger, M.J., *Elucidating the reaction mechanism of the benzoate oxidation pathway encoded aldehyde dehydrogenase from Burkholderia xenovorans LB400.* Protein Science, 2011. **20**(6): p. 1048-1059.

137. Schlaeppi, K., Deflorin, J., Seebeck, T, *The major component of the paraflagellar rod of Trypanosoma brucei is a helical protein that is encoded by two identical, tandemly linked genes.* Journal of Cell Biology, 1989. **109**(4): p. 1695-1709.

138. Pope, M.E., Soste, M.V., Eyford, B.A., Anderson, N.L., Pearson, T.W., *Anti-peptide antibody screening: Selection of high affinity monoclonal reagents by a refined surface plasmon resonance technique.* Journal of Immunological Methods, 2009. **341**(1–2): p. 86-96.

139. Kiselar, J., Downard, K.M., *Direct Identification of Protein Epitopes by Mass Spectrometry without Immobilization of Antibody and Isolation of Antibody−Peptide Complexes.* Analytical Chemistry, 1999. **71**(9): p. 1792-1801.

140. Razavi, M., Pope, M.E., Soste, M.V., Eyford, B.A., Jackson, A.M., Anderson, N.L., Pearson, T.W., *MALDI Immunoscreening (MiSCREEN): A method for selection of anti-peptide monoclonal antibodies for use in immunoproteomics.* Journal of Immunological Methods, 2011. **364**(1–2): p. 50-64.

141. Buchanan, K.T., Ames, J.B., Asfaw, S.H., Wingard, J.N., Olson, C.L., Campana, P.T., Araujo, A.P., Engman, D.M., *A flagellum-specific calcium sensor.* Journal of Biological Chemistry, 2005. **280**(48): p. 40104-11.

142. Engman, D.M., Krause, K.H., Blumin, J.H., Kim, K.S., Kirchhoffd, L.V., Donelson, J.E., *A Novel Flagellar Ca^{2+}-binding protein in Trypanosomes.* Journal of Biological Chemistry, 1989. **264**(31): p. 18627-31.

143. Wu, Y.M., Deford, J., Benjamin, R., Lee, M.G.S., Ruben, L. , *The gene family of EF-hand calcium binding proteins from the flagellum of Trypanosoma brucei.* Biochemical Journal, 1994. **304**(3): p. 833-41.

144. Maldonado, R.A., Mirzoeva, S., Godsel, L.M., Lukas, T.J., Goldenberg, S., Watterson, D.M., Engman, D.M., *Identification of calcium binding sites in the trypanosome flagellar calcium-acyl switch protein.* Molecular and Biochemical Parasitology, 1999. **101**(1–2): p. 61-70.

145. Pinto, A.P.A., Campana, P.T., Beltramini, L.M., Silber, A.M., Araújo, A.P.U., *Structural characterization of a recombinant flagellar calcium-binding protein from Trypanosoma cruzi.* Biochimica et Biophysica Acta, 2003. **1652**(2): p. 107-114.

146. Emmer, B.T., Daniels, M.D., Taylor, J.M., Epting, C.L., Engman, D.M., *Calflagin Inhibition Prolongs Host Survival and Suppresses Parasitemia in Trypanosoma brucei Infection.* Eukaryotic Cell, 2010. **9**(6): p. 934-942.

147. Godsel, L.M., Tibbetts, R.S., Olson, C.L., Chaudoir, B.M., Engman, D.M., *Utility of Recombinant Flagellar Calcium-Binding Protein for Serodiagnosis of Trypanosoma cruzi Infection.* Journal of Clinical Microbiology, 1995. **33**(8): p. 2082-5.

148. Marcipar, I.S., Roodveldt, C., Corradi, G., Cabeza, M.L., Brito, M.E.F., Winter, L.M.F., Marcipar, A.J., Silber, A.M., *Use of Full-Length Recombinant Calflagin and Its C Fragment for Improvement of Diagnosis of Trypanosoma cruzi Infection.* Journal of Clinical Microbiology, 2005. **43**(11): p. 5498-503.

149. Sosa E.S., S., E.L., Ruiz, A.M., Velazquez, E., Porcel, B.M., Yampotis, C., *Efficacy of chemotherapy with benznidazole in children in the indeterminate phase of Chagas' disease.* American Journal of Tropical Medicine and Hygiene, 1998. **59**(4): p. 526-9.

150. Magnus, E., Vervoort T, Van Meirvenne N, *A card-agglutination test with stained trypanosomes (C.A.T.T.) for the serological diagnosis of T. b. gambiense*

trypanosomiasis. Annales de la Société Belge de Médecine Tropicale, 1978. **58**(3): p. 169-76.

151. Lejon, V., Büscher P, Magnus E, Moons A, Wouters I, Van Meirvenne N, *A semi-quantitative ELISA for detection of Trypanosoma brucei gambiense specific antibodies in serum and cerebrospinal fluid of sleeping sickness patients.* Acta Tropica, 1998. **69**(2): p. 151-164.

152. Njiru, Z.K., Mikosza AS, Armstrong T, Enyaru JC, Ndung'u JM, Thompson AR, *Loop-mediated isothermal amplification (LAMP) method for rapid detection of Trypanosoma brucei rhodesiense.* PLoS Neglected Tropical Diseases, 2008. **2**(1): p. e147.

153. Njiru, Z.K., Traub R, Ouma JO, Enyaru JC, Matovu E, *Detection of Group 1 Trypanosoma brucei gambiense by Loop-Mediated Isothermal Amplification.* Journal of Clinical Microbiology, 2011. **49**(4): p. 1530-1536.

154. Chappuis, F., Loutan L, Simarro P, Lejon V, Buscher P, *Options for field diagnosis of human african trypanosomiasis.* Clinical Microbiology Reviews, 2005. **18**(1): p. 133-46.

155. Manful, T., Mulindwa J, Frank FM, Clayton CE, Matovu E, *A search for Trypanosoma brucei rhodesiense diagnostic antigens by proteomic screening and targeted cloning.* PLoS ONE, 2010. **5**(3): p. e9630.

156. Papadopoulos, M.C., et al., *A novel and accurate diagnostic test for human African trypanosomiasis.* The Lancet, 2004. **363**(9418): p. 1358-1363.

157. Agranoff, D., Stich, A., Abel, P., Krishna, S., *Proteomic fingerprinting for the diagnosis of human African trypanosomiasis.* Trends in Parasitology, 2005. **21**(4): p. 154-157.

158. Tiberti, N., et al., *Cerebrospinal Fluid Neopterin as Marker of the Meningo-Encephalitic Stage of Trypanosoma brucei gambiense Sleeping Sickness.* PLoS ONE, 2012. **7**(7): p. e40909.

159. Anderson, N.L., Anderson, N.G., Haines, L.R., Hardie, D.B., Olafson, R.W., Pearson, T.W., *Mass Spectrometric Quantitation of Peptides and Proteins Using Stable Isotope Standards and Capture by Anti-Peptide Antibodies (SISCAPA).* Journal of Proteome Research, 2004. **3**(2): p. 235-244.

160. Mori, Y., Nagamine, K., Tomita, N., Notomi, T., *Detection of Loop-Mediated Isothermal Amplification Reaction by Turbidity Derived from Magnesium Pyrophosphate Formation.* Biochemical and Biophysical Research Communications, 2001. **289**(1): p. 150-154.

161. Curtis, K.A., Rudolph, D.L., Nejad, I., Singleton, J., Beddoe, A., Weigl, B., LaBarre, P., Owen, S.M., *Isothermal Amplification Using a Chemical Heating Device for Point-of-Care Detection of HIV-1.* PLoS ONE, 2012. **7**(2): p. e31432.

162. Grab, D.J., Nikolskaia, O.V., Inoue, N., Thekisoe, O.M.M., Morrison, L.J., Gibson, W., Dumler, J.S., *Using Detergent to Enhance Detection Sensitivity of African Trypanosomes in Human CSF and Blood by Loop-Mediated Isothermal Amplification (LAMP).* PLoS Neglected Tropical Diseases, 2011. **5**(8): p. e1249.

163. Liu, M.K., Pearson TW, *Detection of circulating trypanosomal antigens by double antibody ELISA using antibodies to procyclic trypanosomes.* Parasitology, 1987. **95**(2): p. 277-90.

164. Liu, M.K., Pearson TW, Sayer PD, Gould SS, Waitumbi JN, Njogu AR, *Serodiagnosis of African sleeping sickness in vervet monkeys by detection of parasite antigens.* Acta Tropica, 1988. **45**(4): p. 321-20.

165. Liu, M.K., Cattand P, Gardiner IC, Pearson TW, *Immunodiagnosis of sleeping sickness due to Trypanosoma brucei gambiense by detection of antiprocyclic antibodies and trypanosome antigens in patients' sera.* Acta Tropica, 1989. **46**(4): p. 257-66.

166. Liu, M.K., Wellde BT, Pearson TW, *Trypanosoma brucei rhodesiense: Immunodiagnosis of African sleeping sickness by detection of antigen and antibodies in documented sera from Lambwe Valley, Western Kenya. M. Liu, PhD thesis*, 1989, University of Victoria.

167. Rae, P.F., Luckins, A.G., *Detection of circulating trypanosomal antigens by enzyme immunoassay.* Annals of Tropical Medicine and Parasitology, 1984. **78**(6): p. 587-596.

168. Nantulya, V.M., *An antigen detection enzyme immunoassay for the diagnosis of rhodesiense sleeping sickness.* Parasite Immunology, 1989. **11**(1): p. 69-75.

169. Hutchinson, O.C., Webb H, Picozzi K, Welburn S, Carrington M, *Candidate protein selection for diagnostic markers of African trypanosomiasis.* Trends in Parasitology, 2004. **20**(11): p. 519-523.

170. Anderson, N.L., Anderson NG, *The Human Plasma Proteome: history, character, and diagnostic prospects.* Molecular & Cellular Proteomics, 2002. **1**(11): p. 845-867.

171. Hortin, G.L., Sviridov D, Anderson NL, *High-Abundance Polypeptides of the Human Plasma Proteome Comprising the Top 4 Logs of Polypeptide Abundance.* Clinical Chemistry, 2008. **54**(10): p. 1608-1616.

172. Woo, P.T.K., *The haematocrit centrifuge technique for the diagnosis of African trypanosomiasis.* Atca Tropica, 1970. **27**(4): p. 384-6.

173. Cross, G., *Immunochemical Aspects of Antigenic Variation in Trypanosomes* Journal of General Microbiology, 1979. **113**(1): p. 1-11.

174. Cross, G.A., Manning JC, *Cultivation of Trypanosoma brucei spp. in semi-defined and defined media.* Parasitology, 1973. **67**: p. 315–31.

175. Anderson, N.L., Parish NM, Richardson JP, Pearson TW, *Comparison of African trypanosomes of different antigenic phenotypes, subspecies and life cycle stages by two-dimensional gel electrophoresis.* Molecular and Biochemical Parasitology, 1985. **16**(3): p. 299-314.

176. Neubert, H., Gale, J., Muirhead, D., *Online High-Flow Peptide Immunoaffinity Enrichment and Nanoflow LC-MS/MS: Assay Development for Total Salivary Pepsin/Pepsinogen.* Clinical Chemistry, 2010. **56**(9): p. 1413-1423.

177. Patel, B.B., Barrero CA, Braverman A, Kim PD, Jones KA, Chen DE, Bowler RP, Merali S, Kelsen SG, Yeung AT, *Assessment of Two Immunodepletion Methods: Off-Target Effects and Variations in Immunodepletion Efficiency May Confound Plasma Proteomics.* Journal of Proteome Research, 2012. **11**(12): p. 5947-5958.

178. Kuepfer, I., Hhary EP, Allan M, Edielu A, Burri C, Blum JA, *Clinical Presentation of T. b. rhodesiense Sleeping Sickness in Second Stage Patients from Tanzania and Uganda.* PLoS Neglected Tropical Diseases, 2011. **5**(3): p. e968.

179. Authie, E., *Trypanosomiasis and trypanotolerance in cattle: a role for congopain?* Parasitology Today, 1994. **10**(9): p. 360-364.

180. Huang, H., Mackeen M, Cook M, Oriero E, Locke E, Thezenas M, Kessler B, Nwakanma D, Casals-Pascual C, *Proteomic identification of host and parasite biomarkers in saliva from patients with uncomplicated Plasmodium falciparum malaria.* Malaria Journal, 2012. **11**(1): p. 178.

181. Murray, C.K., Bennett JW, *Rapid Diagnosis of Malaria.* Interdisciplinary Perspectives on Infectious Diseases, 2009.

182. Franco, J.R., Simarro, P.P., Diarra, A., Ruiz-Postigo, J.A., Jannin, J.G., *The Human African Trypanosomiasis Specimen Biobank: A Necessary Tool to Support Research of New Diagnostics.* PLoS Neglected Tropical Diseases, 2012. **6**(6): p. e1571.

Appendices

Appendix 1. Excel sheet 1 contains all proteins identified by iTRAQ and sorted by replicate; Excel sheet 2 shows only those proteins present in 2 or 3 replicates, sorted by accession number. Expression data have been averaged; Excel sheet 3 contains the same data as Excel sheet 2 although the proteins are organized by their putative function. All proteins that showed positive or negative changes in expression (with high confidence, regardless of intensity) are highlighted in red and blue, respectively. All uncertain expression data (P-value > 0.05) has been replaced with the label "data not reliable (DNR)".

Appendix 2. Excel sheet 1 contains a list of all proteins (human and trypanosomes) identified in HAT plasma and is ordered by number of contributing peptides. Excel sheet 2 shows only the trypanosome proteins discovered in HAT plasma. Molar intensity has been calculated for those proteins with at least two contributing peptides and the proteins are listed by this value. Excel sheet 3 shows only the trypanosome proteins discovered in HAT plasma (with at least 2 contributing peptides) and is organized by putative function.