Investigating the molecular basis of tsetse-trypanosome interactions

by

Sarah Goomeshi Nobary
M.Sc., University of Alberta, 2011

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

MASTER OF SCIENCE

in the Department of Biochemistry and Microbiology

Supervisory Committee

Investigating the molecular basis of tsetse-trypanosome interaction

by

Sarah Goomeshi Nobary
M.Sc., University of Alberta, 2011

Supervisory Committee

Dr. Martin Boulanger, Department of Biochemistry and Microbiology
Supervisor

Dr. Perry Howard, Department of Biochemistry and Microbiology
Departmental Member

Dr. Leigh Anne Swayne, Division of Medical Sciences
Outside Member

Abstract

Supervisory Committee
Dr. Martin Boulanger, Department of Biochemistry and Microbiology
Supervisor
Dr. Perry Howard, Department of Biochemistry and Microbiology
Departmental Member
Dr. Leigh Anne Swayne, Division of Medical Sciences
Outside Member

The parasitic pathogens of genus *Trypanosoma* cause significant morbidity and mortality worldwide. The most well studied *Trypanosoma* related diseases are African sleeping sickness (*Trypanosoma brucei*) and African Animal Trypanosomiasis (*Trypanosoma congolense*). Despite more than 100 years of research these diseases continue to have a devastating impact on the socioeconomic development of Africa. A major impediment to controlling outbreaks is the lack of an effective vaccine due, in part, to the parasite's ability to continually alter its protein coat while in the host, which results in effectively evading the host immune system. Recent studies have identified *Trypanosoma congolense* proteins that are selectively expressed during transmission in the tsetse arthropod vector where the parasite's protein coat is not constantly recycled. Of these proteins, Congolense Insect Stage Specific Antigen (*Tc*CISSA) and Congolense Epimastigote Specific Protein (*Tc*CESP) were selected for characterization based on cellular localization, expression levels and predicted roles in facilitating transmission by the tsetse fly.

The goal of the present study is to understand the crosstalk between *T. congolense* and its vector, the tsetse fly. Revealing the structure of proteins is a crucial step in determining their functions. In order to gain insight into the molecular basis of structure and function of *Tc*CESP and *Tc*CISSA we took various biophysical and biochemical approaches. *Tc*CISSA was recombinantly produced in *E. coli*, crystallized and diffraction quality data collected to 2.5 Å resolution. Structure determination, however, has been problematic due to the absence of homologous models and the inability to take advantage of SelMet phasing due to the presence of only a single methionine in the sequence. Structure determination efforts are ongoing using multiple approaches including NMR. In

contrast to *Tc*CISSA, the size and complexity of *Tc*CESP required insect cells for efficient recombinant production. While crystallization trials have yet to yield diffraction quality crystals, a combination of homology modeling validated by chemical crosslinking and mass spectrometry, and circular dichroism spectroscopy have yielded intriguing insight into the architecture of CESP. Characterizing the function of these proteins offers the potential for rare insight into the molecular crosstalk between the parasite and vector and may support the development of novel transmission blocking vaccines.

Table of Contents

List of Tables

List of Figures

List of abbreviations

AAT	African animal trypanosomiasis
AEBSF	4- benzenesulfonyl fluoride hydrochloride
BARP	Brucei alanine rich protein
BLAST	Basic local alignment search tool
BSF	Bloodstream
CBDPS	Cyanurbiotindimercaptopropionylsuccinimide
CESP	Congolense epimastigote stage protein
CID	Collision induced dissociation
CISSA	Congolense insect stage specific antigen
DDT	Dichlorodiphenyltrichloroethane
DNA	Deoxyribonucleic acid
E. coli	*Escherichia coli*
EMF	Epimastigote form
GARP	Glutamate/alanine rich protein
GPI	Glycophosphatydilinositol
HEPES	2-[4-(2-hydroxyethyl)piperazin-1-yl] ethanesulfonic acid
His	Histidine
HpHbR	Haptoglobin haemoglobin receptor
HPLC	High pressure liquid chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
I-TASSER	Iterative threading assembly refinement
iTRAQ	Isobaric tags for relative and absolute quantitation
kDa	Kilo Dalton
kDNA	Kinetoplast deoxyribonucleic acid
LB	Luria broth
LC-MS	Liquid chromatography-Mass spectrometry
Leu	Leucine
Met	Methionine
MF	Metacyclic Form
MHz	Mega Hertz
MS	Mass spectrometry
MWCO	Molecular weight cut off
Ni	Nickel
NMR	Nuclear magnetic resonance
PDB	Protein data bank
PEG	Poly-ethylene glycol
PF	Procyclic form
Phyre	Protein homology/analogy recognition engine
ProSa	Protein structure analysis
PSSA	Procyclic stage specific antigen
RNA	Ribonucleic acid

rpm	Revolutions per minute
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SelMet	Selenomethionine
Sf9	*Spodoptera frugiperda* (insect cell line)
TEV	Tobacco etch virus
VSG	Variable surface glycoprotein

Acknowledgments

I would like to thank my supervisor, Dr. Boulanger for his help and guidance and providing me the opportunity to study in the field of Structural Biology. I am also thankful to my committee members Dr. Perry Howard and Dr. Leigh Anne Swayne for their help and guidance during my studies and thesis defence. I would also like to thank our collaborators from Dr. Evgenye Petrochenko's group from the Borcher lab for the MS experiments on CESP and the Smith lab at Queen's University for carrying out the NMR experiments. I am grateful to Dr. Ausio for his assistance with the CD spectrometer. I am very grateful to the University of Victoria for providing funding for this research and my studies.

I am very grateful to my family for their support from the other side of the world and to all the friends and colleagues who helped me during my studies in UVic.

Chapter 1 - Introduction

1.1. Vector borne diseases

Vector borne diseases are illnesses that require an organism, a vector, in order to be transmitted between hosts. Examples of these diseases include malaria, lyme disease, dengue fever and African trypanosomiasis. Vector borne diseases cause significant morbidity and mortality on a global scale (Figure 1.1). The disease burden resulting from infections of vector borne pathogens is dictated both by the acute pathology of the infection and by the efficiency by which the pathogen is transmitted by the vector. While many different vectors exist to enable dissemination and transmission of pathogens, the most widely studied are the arthropod vectors such as mosquitoes, flies and ticks.

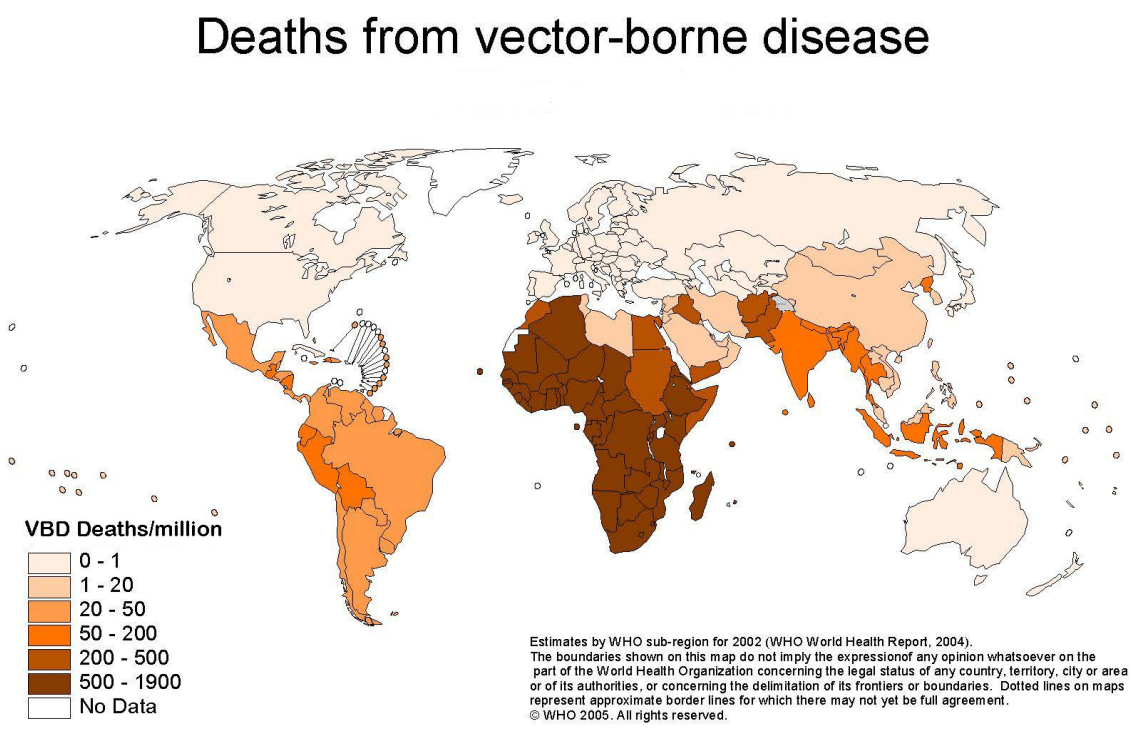

Figure 1.1. Number of deaths resulting from vector borne diseases. While every country in the world suffers from vector borne diseases, the continent of Africa pays the heaviest toll. This map has been reproduced with permission from (1).

Historically, the majority of vector-borne diseases have been largely restricted to tropical countries. However, the impact of climate change on vector habitat and the increased mobility of the population at large has significantly increased the distribution of vector borne pathogens targeting humans, and agricultural animals and plants on a global scale (2). The mechanisms by which microorganisms are transmitted by insects range from relatively simple mechanical events, as in the cases of many vector-borne viruses, through complex biological processes that enable microbial growth within the vector. Examples of this latter case are the tsetse fly transmitted kinetoplastid parasites of the genus *Trypanosoma* infamously known for their role in causing sleeping sickness in humans. The *Trypanosoma*-tsetse fly pairing is an interesting model of a complex vector-pathogen interaction and it is the focus of my thesis work.

1.2. African trypanosomiasis

1.2.1. Etiologic agent

Protozoan species from the genus *Trypanosoma* are responsible for diseases known as African trypanosomiasis. In humans, the disease is better known as African sleeping sickness while in animals it is referred to as African animal trypanosomiasis (AAT) or chronic wasting disease. The Zulu word for AAT, N'gana, means useless or to be in low or depressed spirit (3). Trypanosomes are extracellular parasites, which reside in the blood or lymph of their host. Figure 1.2 shows an early drawing of trypanosomes among the blood cells.

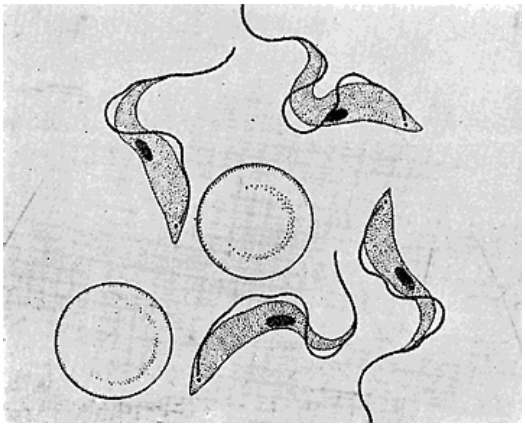

Figure 1.2. An early drawing of trypanosomes among blood cells. The trypanosomes are extracellular parasites and do not enter the host blood cells. This drawing was reproduced from Reiley, 1915 (4).

1.2.2. Historical perspective and Societal Impact

Trypanosoma infections have shaped the development of the African continent. With much of the sub-Saharan African economy supported by agriculture and livestock production (5,6), the effect of AAT is far reaching (5,7) with direct impacts on local economies and ultimately on human health. It is noteworthy that AAT affects domesticated animals more seriously than wild animals. The infection in wild animals is usually mild, compared to the fatal disease observed in domesticated animals. While numerous different domesticated animal species can acquire AAT, it is of major importance in cattle, as cattle are the main animals reared in sub-Saharan Africa. N'gana kills three million cattle every year and causes general chronic ill-health, abortion and reduced productivity of cattle herds. The economic loss in cattle production is estimated to be 4 billion USD every year. The greatest loss is in human food production and it is shown that the absence of oxen for cultivation can result in an 80% decrease in crop production (8).

In ancient times the north coast of the continent of Africa had more vegetation compared to the present (9) and records show that the distribution of AAT was different. During 3000 BC – 2000 BC (Old Kingdom), the flora and fauna of the Nile Valley were different and it was probably similar to the current region of the Gazelle River, one of the major tributaries of the Nile (3), in the Sudan. It is reasonable to suggest, therefore, that the distribution of the vector for trypanosomiasis and therefore the disease was extended northwards and ranged into the Nile delta. There is evidence that Egyptians of the Old Kingdom raised and kept their cattle together with game animals in hope of rearing trypanotolerant animals (3). In addition, a veterinary papyrus (Figure 1.3) dating from 2nd millennium BC describes a disease that is similar to N'gana. An ointment made from the fat of particular birds was used to treat the bite of the vector. This shows that although the Egyptians did not know the basis of the disease, they associated the disease with the bite of a fly (3).

Figure 1.3. Part of the veterinary papyrus describing the disease, ushau. The descriptions translated from this papyrus are very similar to symptoms of N'gana. On this papyrus, a treatment for the bite of the fly that was believed to cause the disease has also been described. Reproduced with permission from Stevering, 2008 (3).

1.2.3. Therapeutic treatments

There are three medications for N'gana (AAT): isometamidium, homidium, and diminazene and they have been reviewed (10,11). Isometamidium is mainly used as a prophylactic and provides up to 6 months protection against the tsetse challenge. Homidium is used as a therapeutic agent and does not have much prophylactic property and diminazene is only a therapeutic agent. Although there is constant demand for these medications, the market value does not seem enough to justify the investments by the pharmaceutical companies. In addition, resistance to drugs has been reported and is on the rise (12). It is reasonable to assume that some of the areas affected by AAT are remote and there are no formal reports about the drug resistance from those parts of the continent (13).

1.2.4. Evolutionary classification of *Trypanosoma*

The genus *Trypanosoma* belongs to the order Kinetoplastida due to the presence of the kinetoplast, a disc shaped organelle, which contains many copies of circular DNA. The kinetoplast DNA (kDNA) is considered the most structurally complex DNA network and

there are two types of circular DNA in the organelle: maxicircles, which contain genes for expressing mitochondrial proteins; and minicircles, which encode RNA editing genes. These are required for deciphering the encrypted genes in the maxicircles (14-17). Figure 1.4.A shows the organelle located in the *Trypanosoma* cell and 1.4.B shows an electron micrograph showing the circular DNA from the kinetoplast.

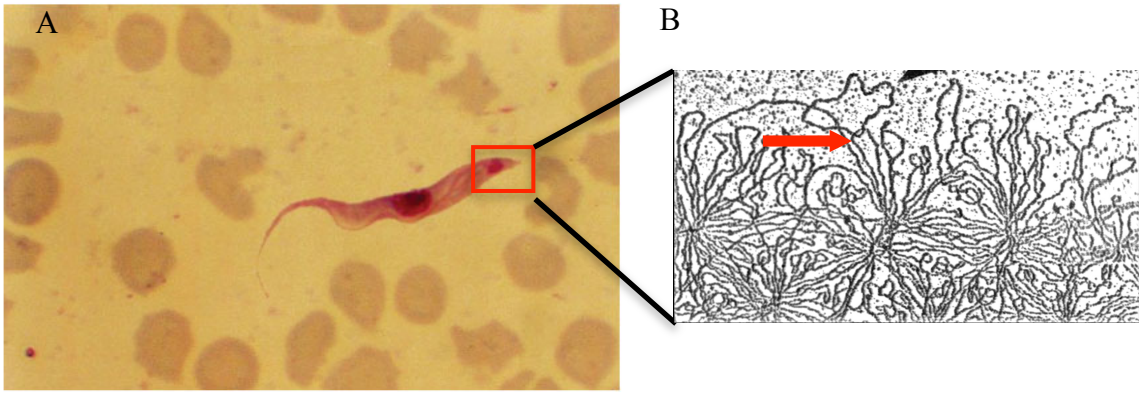

Figure 1.4. The kinetoplast shown in the parasite (A) and an electron micrograph (B) showing the maxicircles and minicircles. A: The size of the kinetoplast is different in the different species of *Trypanosoma.* The biggest kinetoplast belongs to *T. vivax.* **B:** as seen in the figure minicircles are closely associated with each other and in some cases are interlocked (red arrow). Panel A is reproduced from (18) and Panel B is modified with permission from (14).

The *Trypanosoma* species are divided into two main groups based on their mode of transmission (19): Stercoraria and Salivaria. The Stercoraria migrate to the posterior (hindgut) of their vector and are transmitted by the feces of the vector (by means of contamination). An example of these species is *T. cruzi,* the cause of Chaga's disease, a disease that is found in the South American continent and in the southern United States. The other group, Salivaria is transmitted via inoculation and the members are spread through the salivary gland or mouthparts of their vectors. The trypanosomes of the Salivaria group are the only trypanosomes that are capable of antigenic variation in their mammalian host. The species that are in this group are the causes of African trypanosomiasis. They can change their antigens to evade the immune response of the host and have, therefore, posed major problems in attempts to develop a vaccine.

There are three subgenera in the Salivaria group (19). *Duttonella* is one of these subgenera and the best-known species from this subgenus is *Trypanosoma (Duttonella) vivax*. This species is transmitted by the tsetse fly and its final development occurs in the proboscis of tsetse. *T. (D.) vivax* infects domestic animals. The second subgenus from the Salivarian trypanosomes is the *Nannomonas*. This subgenus includes the infamous *Trypanosoma (Nannomonas) congolense,* the cause of more than 80% of the N'gana cases in Africa (20-23). *T. congolense* infects a broad range of domestic animals. This pathogen is transmitted through the proboscis (mouthparts) of the tsetse fly (19). *T. congolense* is further divided into three genetically distinct subgroups: Savannah, or the 'Dry' division, which includes isolates from East and West African savannah; Riverine/forest or the 'Humid' division; and finally, the Kenya coast, which is also named 'Kilifi' (24). The last subgenus from the Salivarian trypanosomes is called *Trypanozoon* and it includes *Trypanosoma (Trypanozoon) brucei,* and its subspecies *T. b. brucei, T. b. rhodesiens* and *T. b. gambiense.* While *T. b. brucei* infects animals, the other two sub-species cause human trypanosomiasis or sleeping sickness (25).

1.3. Transmission and dissemination

1.3.1. The tsetse fly: A scourge of sub-Saharan Africa

Species of the genus *Trypanosoma* are transmitted by the bite of the tsetse fly (Figure 1.5. A). There are 33 species and sub-species of tsetse flies (genus *Glossina*) (26,27), but the most important species responsible for parasite transmission is *Glossina morsitans*. Tsetse flies are found only in Africa and they are limited to the sub-Saharan region. The tsetse belt represents an area as of approximately 8 million km^2 (Figure 1.5. B), nearly equivalent to the United States (26,28-30). N'gana was associated with the tsetse fly for a long time and the people living close to the tsetse infested areas have always tried to graze their herds in safer places, which they determined by trial and error. The first report describing a direct relationship with tsetse flies was by the missionary David Livingston in 1857. He described tsetse flies as "poisonous insects to ox, horse and dog". It was David Bruce, a Scottish pathologist and microbiologist who proved that trypanosomes are transmitted by tsetse flies (31,32). Tsetse flies are characterized by a distinct proboscis, an antenna with branched arista hairs and wings that fold at rest and that have a characteristic "hatchet" cell when opened (33). Tsetse flies feed exclusively on blood and

they require blood meals every 3-4 days. The life span of a tsetse fly is approximately 3-4 months. A female tsetse fly gives birth to only one offspring at a time and nurses the offspring for a few weeks (26,34).

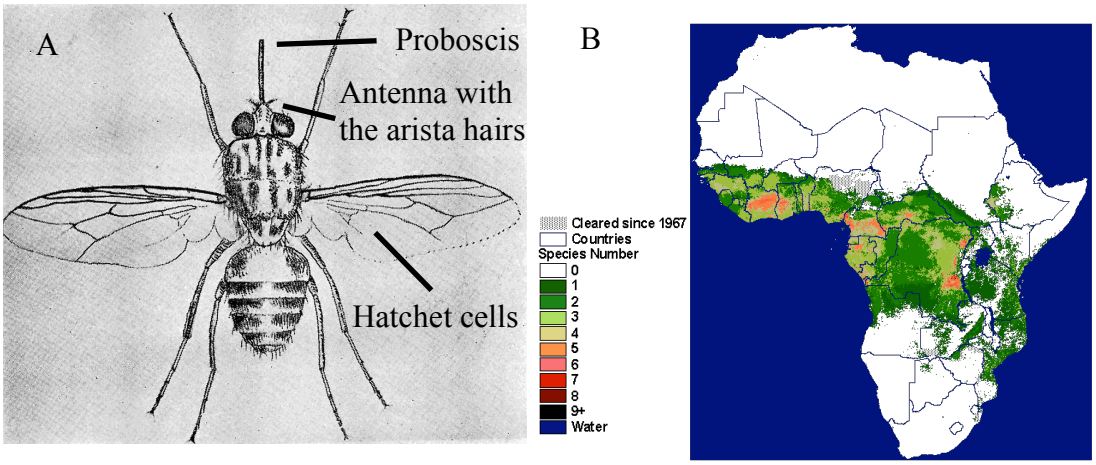

Figure 1.5. A. A general diagram of *Glossina morsitans.* The proboscis, antenna and the hatchet cells on the wings are shown in the figure above. Figure reproduced from (4). **B. Distribution of different species of tsetse flies.** As the guide at the left hand side shows, the number of species of *Glossina* found in each region is shown in different colours from green (one species) to black (more than nine species). The map has been taken from (35)

1.3.2. Controlling tsetse populations: evolving approaches

Traditional measures to control African Trypanosomiases relied on strategies to control the tsetse fly. One of the approaches taken towards vector control has been the use of insecticides. Two chemical groups of insecticides have proven successful in controlling the tsetse flies. One group, the organochlorines, includes DDT (dichlorodiphenyltrichloroethane), dieldrin, and endosulfan. The second group of chemicals is the group of synthetic pyrethroids. For nearly 40 years, the organochlorines were the basis for tsetse control using various delivery methods, however their persistence and negative environmental impacts eventually led to restrictions and ultimately a ban on their use in most developed countries (36-40).

Use of baits and traps (41,42) and sterile insect techniques (43-46) have also been pursued as more strategic measures to control tsetse populations. However, with all of these approaches, the areas that were cleared from tsetse flies were re-infested soon after the clearing. In addition, the environmental impacts of decreasing the population of tsetse

flies and the possible effects on the lives of humans have not been clearly understood. Therefore, a control strategy with an emphasis on deactivating, killing or reducing the infectivity of the parasites inside the vector is widely considered as a significantly more effective and long-term solution. Such a strategy will be specific and the environmental impacts can be minimized. Strategically disrupting transmission, however, will require a better understanding of the molecular interactions and relationship between trypanosomes and tsetse fly. Such an understanding can be accomplished through a detailed investigation of the surface proteins of trypanosomes that are expressed during transmission by the tsetse vector.

1.4. Tsetse flies play a essential role in the life cycle of Salivarian trypanosomes

1.4.1. Morphological changes during the life cycle

Salivarian trypanosomes go through a complex life cycle as they alternate between their vector and their mammalian host. Trypanosomes enter the tsetse fly during a blood meal and establish an infection that spans the life of the vector. As the trypanosomes go through different stages in their life cycle, they undergo dramatic morphological and metabolic changes. For the sake of convenience we will refer to these different forms by their acronyms: bloodstream forms (BSF) reside in the bloodstream of their mammalian host, procyclic forms (PF) are found in the midgut, epimastigote forms (EMF) and the metacyclic forms (MF) are both found in the mouthparts of tsetse (Figure 1.5). The BSF of *T. congolense* is a short, stumpy form with swollen mitochondria. At this stage the trypanosomes use glycolysis and excrete the resulting pyruvate instead of oxidizing it (47,48). The BSF are covered by $\sim 10^7$ copies of variable surface glycoproteins (VSG), which are attached to the surface of trypanosomes by glycophosphatydilinositol (GPI) anchors. These proteins shed periodically and this helps the trypanosomes evade the host immune system. In addition to immune evasion, the tight packing of the VSG molecules on the membrane shields the surface of the parasite from the host antibodies (49-51).

In the BSF, the kinetoplast is located towards the posterior end of the cell and the nucleus is in the center, closer to the posterior than the anterior of the cell (47). After the blood is ingested, the parasite enters the digestive tract of the tsetse. The first three days are of crucial importance in establishing a successful infection, because the parasites have

to overcome the innate immune responses and digestive enzymes of the vector in order to survive and eventually penetrate the peritrophic matrix that lines up the gut epithelium. The peritrophic matrix of the tsetse fly is composed of chitin microfibrils, glycoproteins and proteins that are in a proteoglycan matrix produced by the cardia, a region in the anterior midgut specialized for producing this matrix (33).

In the midgut of the tsetse fly, *T.congolense* cells start proliferating and they eventually differentiate into the procyclic forms. The differentiation begins when the cells start to elongate and their kinetoplast moves further from the posterior and closer to the nucleus. Over time the procyclic forms elongate and become more slender, until their length doubles and their width halves (on day 17). Cell division in the procyclic forms of *T. congolense* is symmetrical and results in two cells that are equal in size. The appearance of trypanosomes in the proventriculus starts six days post infection and they look longer and thinner than the procyclic forms in the midgut. At this point the procyclics are unable to divide and their population is morphologically homogeneous (47).

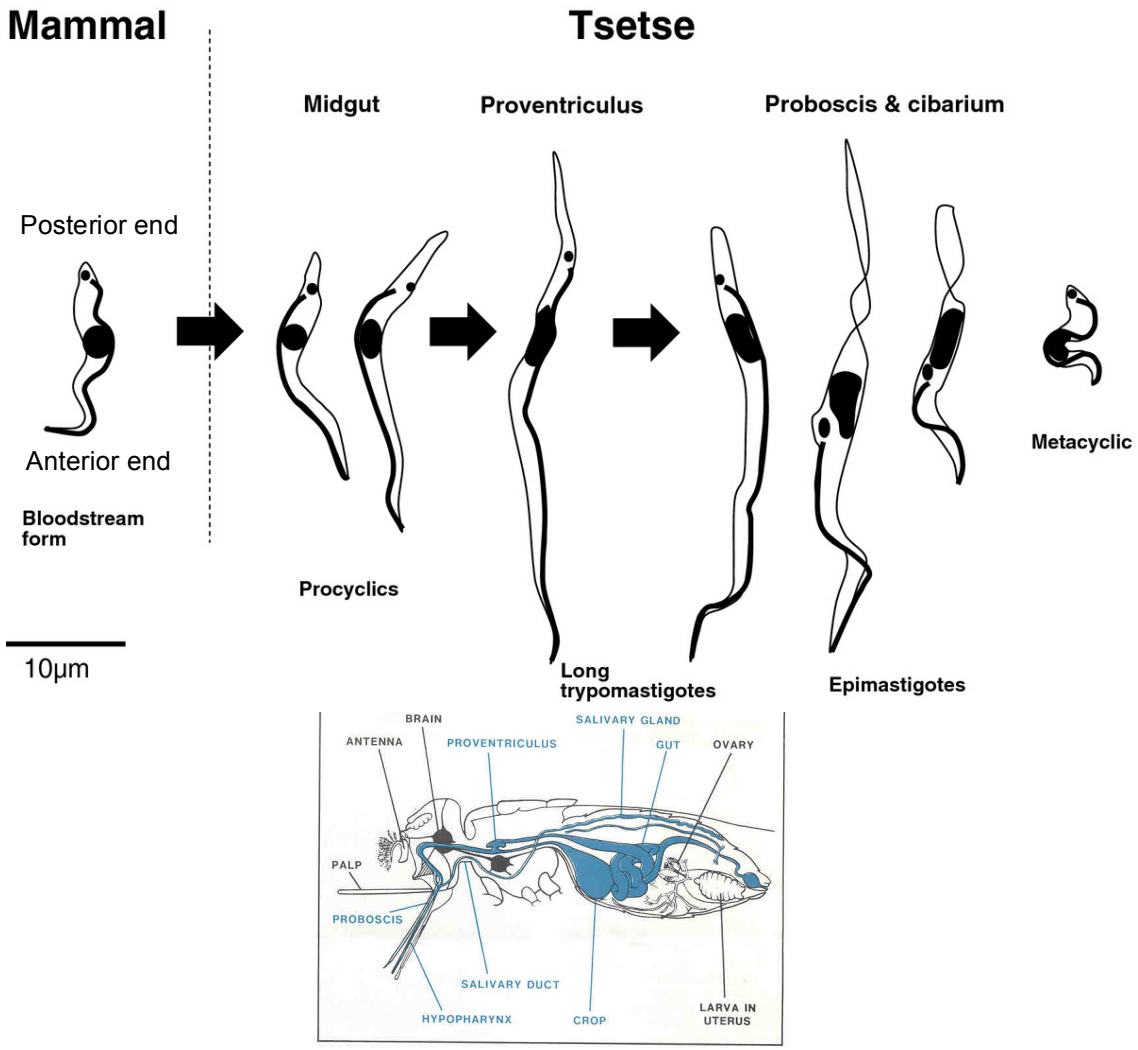

Figure 1.6. Top: Morphological changes of *T. congolense* during its life cycle stages. Bottom: digestive tract of a tsetse fly. The anterior and posterior ends of the parasite are only shown on the BSF, but apply to the other forms. As the BSF enter the midgut they differentiate to procyclics become longer. Procyclics stop dividing in the proventriculus. These forms then migrate to the cibarium and proboscis, and differentiate to epimastigotes. Some epimastigotes possess extremely long posterior ends and some have their posterior ends truncated. The metacyclics are unable to divide and are very small. Metacyclics are the infective form and are able to infect a new mammalian host. The top panel is adapted from (47) and the bottom panel is reproduced with permission from (52).

The proboscis infection can be detected even before the saliva of the fly contains epimastigotes (47). Previous studies have also reported the detection of proboscis infection prior to the emergence of epimastigotes (53,54). The number of epimastigotes in the saliva is initially very few and they are variable in length. During the transition into the epimastigotes the position of the kinetoplast changes relative to the nucleus. First, the distance between the two organelles decreases and then they pass by each other. Eventually the kinetoplast will be positioned towards the anterior and the nucleus is located towards the posterior. In some epimastigotes the posterior end appears truncated, as if it has been cut off or twisted onto itself. In addition epimastigotes with truncated posteriors were previously observed in cells grown *in vitro* (55), so it is very unlikely that they are artifacts resulting from the fixation process. Epimastigotes adhere to the chitinous lining of the proboscis and divide and eventually develop into metacyclics, which are the infective forms of the parasite to the next mammalian host (56,57). The first metacyclics, identified by their characteristic S-shape (55) and very short length (much shorter and thinner than the BSF), can be observed in the saliva on day 21(47).

1.4.2. The life cycle of *T. congolense* and other Salivarian trypanosomes

The life cycle of *T. congolense* is summarized in Figure 1.7. They both go through the midgut and migrate to the anterior of the vector. Although this species is classified with *T. vivax* in the Salivarian group, the life cycle of the latter species is significantly different from the *T. congolense*. *T. vivax* only develops in the proboscis of tsetse fly and never enters the midgut. There they develop into epimastigotes and later into metacyclic forms. The BSF of *T. vivax* exhibits dimorphism. One of these two forms of BSF is a slender form with a rounded posterior. The other form is club-shaped and has a swollen posterior and then abruptly tapers toward the anterior. The subgenus *Dutttonella* has a characteristically large kinetoplast, which is very useful for diagnosis (19).

T. vivax has been isolated outside of the tsetse belt in Africa and even in South America (58). This is because *T. vivax* can be transmitted by other biting insects. The transmission of *T vivax* by other insect vectors is strictly mechanical and it gives *T. vivax* the benefit of a more efficient transmission and a more widespread infection of the infection. The tsetse fly is the only vector in which *T. vivax* can multiply and develop into

epimastigotes and metacyclic forms (19,59,60). Although *T. vivax* is more widespread than *T. congolense,* the latter is still the major species of concern in African trypanosomiasis and the focus of my thesis work.

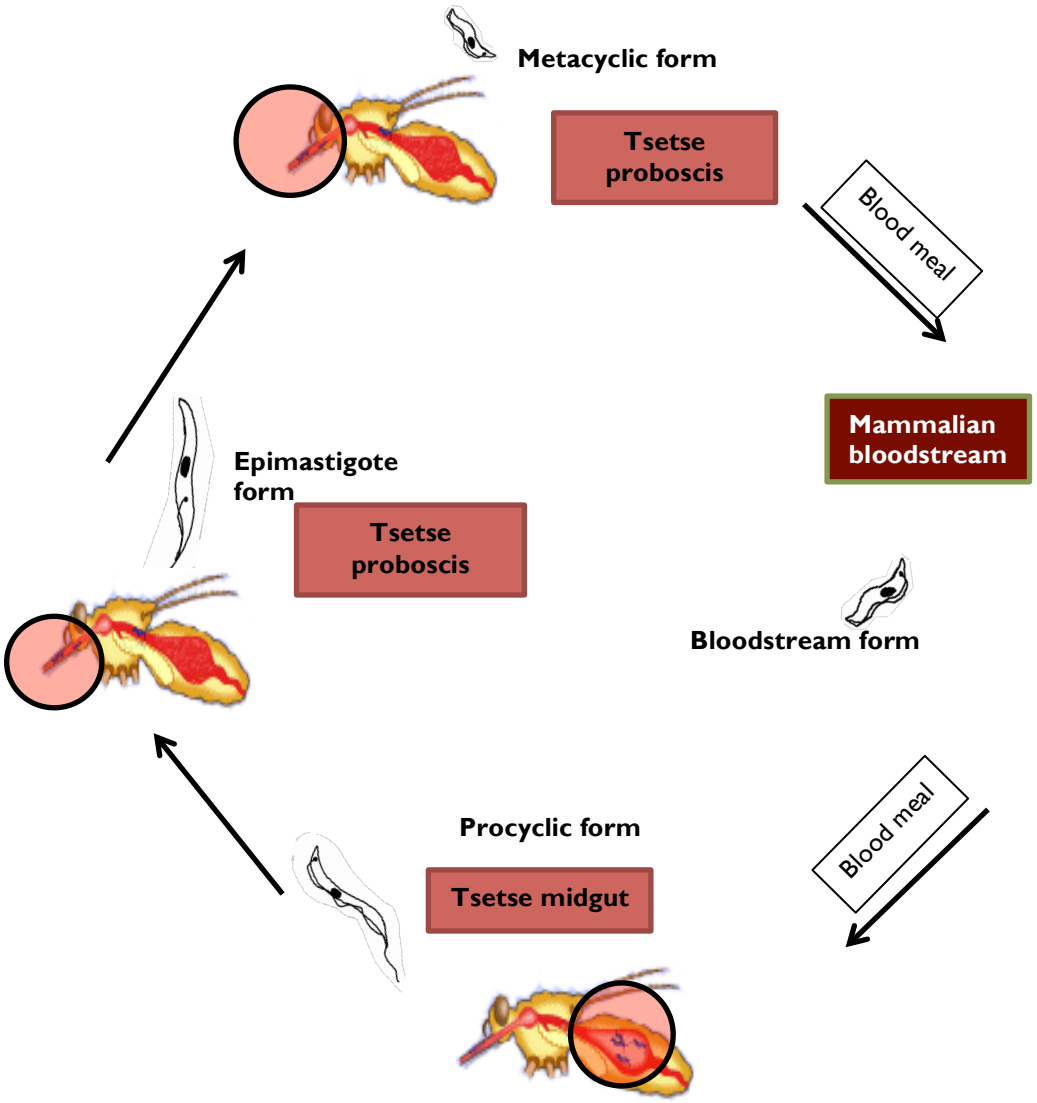

Figure 1.7. Lifecycle of *T. congolense.* The BSF enter the midgut of tsetse fly after a blood meal. The BSF develops into PF are shown as blue flagellates in the midgut, (red circle). The PF migrate to the proboscis and develops into EMF, which are able to divide and adhere to the lining of the proboscis. The EMF then develops into MF, which are not able divide or adhere. The MF is the infective form and start the cycle again in a new host. The image of the tsetse fly has been adapted from: http://homepages.ed.ac.uk/nsavill/tryps.htmL.

1.5. Interface between *T. congolense* and the tsetse fly

1.5.1. Background

Infection of tsetse flies by trypanosomes is not an easy process and there are many bottlenecks and roadblocks for the establishment of the infection. Laboratory infection experiments for *Glossina* involve multiple colonization failures and problems with parasite survival in the fly. In addition, even in the hot spots for Human African trypanosomiasis, the percentage of tsetse flies infected with trypanosomes is very low (61,62). These suggest that the process of tsetse colonization is much contested and it is a balance between the tsetse immunity, physical barriers and the evading power of the infecting trypanosome (63). One of the factors that affect the susceptibility of tsetse fly is the symbionts that reside in this vector. There is evidence that one of these symbionts, *Sodalis glossinidius,* increases the susceptibility of tsetse flies to Trypanosoma infection. However the exact mechanism is unknown (64-66). Another factor is the infectivity of different strains of *T. congolense.* Some strains have a higher infectivity than others (63). Levels of sialic acids present in the blood of the mammalian host may also affect the survival and colonization of *T. congolense* in the midgut. Trans-sialidase is an enzyme expressed only in the procyclic forms and is the only means by which trypanosomes are able to acquire sialic acid (67). It has been shown that trypanosome mutants with no trans-sialidases are unable to colonize tsetse flies (68). Although these data suggest a possible correlation between sialic acids and the colonization of trypanosomes in the midgut of tsetse fly, the exact mechanism is yet to be elucidated (63).

Tsetse flies possess physical and immune barriers against trypanosomes and they are important defence mechanisms for the vector. The physical barriers of tsetse flies against pathogens include the peritrophic matrix (PM), cuticle, epithelial layers and basal membranes. The PM is a barrier between the midgut epithelium and the ingested blood, and it protects the cells from infection and other damaging or toxic substances that are present in the blood meal. It also allows for compartmentalized digestion and metabolism of the blood meal. The PM is mainly composed of chitin, glycosamineglycans, peritrophin proteins and mucin-like glycoproteins (69-71).

Similar to the vertebrate immune system, the immune system of tsetse fly is comprised of humoral and cellular arms. Since trypanosomes do not enter the hemocoel of tsetse fly,

the involvement of the cellular immunity of tsetse has not been documented (33). The two main humoral immune response pathways of tsetse fly are the Toll and immune deficiency (Imd) pathways. The Imd pathway is stimulated by trypanosomes and acts in systemic and epithelial immunity, where effector molecules are released into the hemolymph or secreted by an epithelium, respectively (72). Many genes are upregulated in tseste flies after trypanosoma invasion, but it is yet not clear how many of these genes are related to tsetse immunity (63).

1.5.2. Surface molecules of *T. congolense*

There is a significant difference between the surface coats of the bloodstream forms and the metacyclic forms and the coats of the insect stage forms (EMF and PF). As discussed earlier, the trypanosomes in the bloodstream of mammals are covered with a dense layer of VSG. The metacyclic forms are also covered by VSG. Unlike the VSG coat, the surface proteins of the insect stage are not constantly recycled and thus are better suited as molecular targets for developing inhibitory therapeutics that impair the transmission of parasite through tsetse fly. Therefore, gaining structural insight into these proteins will help us understand the function of these proteins and to propose possible binding partners or biologically active surfaces.

A major technological advance in identifying surface proteins displayed on the parasite life cycle stages during passage through the tsetse fly was the ability to grow each life cycle stage of the strain *T. congolense* IL3000 *in vitro* (73,74). Prior to this, the study of their surface proteins was limited due to the small number of parasites that can be harvested from tsetse flies. Recently, Eyford *et al.* performed an extensive iTRAQ based proteomics study of the different life cycles of *T. congolense* IL3000. Most of their results were confirmed using Western blots (75). A significant outcome of this work was the identification of several novel proteins that are selectively expressed on the procyclic and epimastigote insect stages of *T. congolense.* Of the proteins identified by mass spectrometry, the major surface glycoprotein glutamate/alanine rich (GARP) had been previously shown by transcriptome analysis to be highly expressed in epimastigotes (76-78). The structure of GARP (79) was published in the same year by the Boulanger lab and it resembled the structure of a monomer of VSG (80) and a more recently identified structure, the haptoglobin/hemoglobin receptor (HpHbR). Despite being of different

sizes, a common architectural feature displayed by each of these three proteins is a core triple helix bundle motif (Figure 1.8).

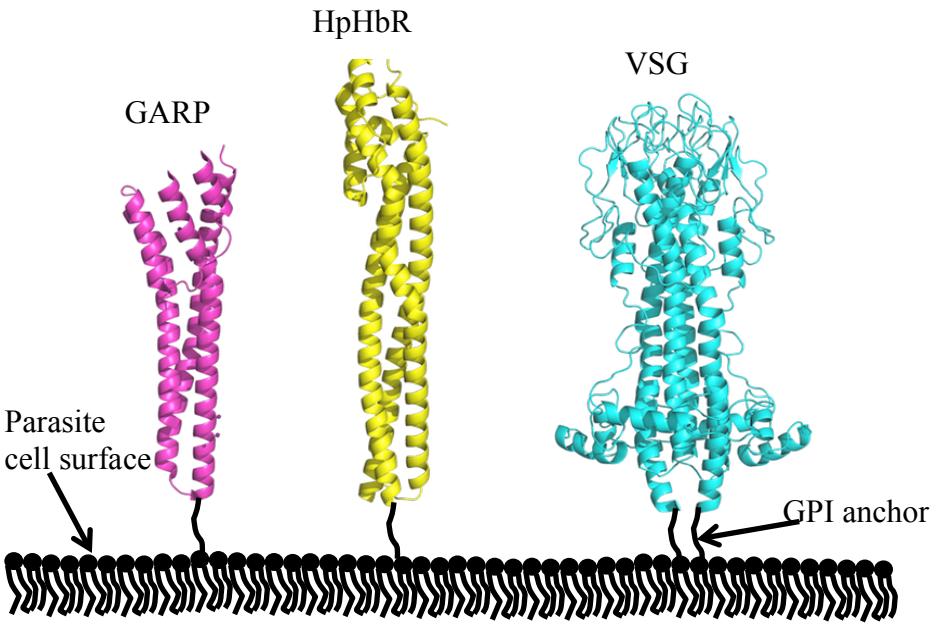

Figure 1.8. Structures of three surface proteins of trypanosomes. The structures of GARP shown in pink (79), HpHbR shown in yellow (81), and VSG shown in cyan (82), share a common triple helix bundle motif. These proteins are GPI anchored and their elongated structures allows their packing on the cell, creating an impenetrable coat to protect the plasma membrane of the parasite from the immune system of the host (79).

1.5.3. *Tc*CESP and *Tc*CISSA

Analysis of the Eyford *et al.* data revealed two intriguing parasite proteins expressed during tsetse transmission: Congolense Insect Stage Specific Antigen (CISSA – Chapter 2) and Congolense Epimastigote Specific Protein (CESP – Chapter 3). *Tc*CISSA (*T. congolense* CISSA) was shown to be expressed in both procyclic and epimastigote forms with an up-regulation in the epimastigotes (75) and the observed expression profile of *Tc*CESP (*T. congolense* CESP) confirmed a previously reported observation by Sakurai et al (75,83). *Tc*CISSA shared a high amino acid sequence identity with a previously identified a transmembrane surface glycoprotein from *T. brucei* known as Procyclic Stage Specific Antigen (PSSA-2). This protein was shown to be necessary for the maturation of

trypanosome infection in the tsetse fly (84). Intriguingly, both CISSA and PSSA-2 incorporate several unusual amino acid repeats (RR, EE, VV, TT, FF, CC, AA, GG, DD, LL, SS, QQ, RRR, VVV, SSS, GGG) and proline rich repeats that render tertiary structure prediction programs unusable. Ultimately, the information obtained from the study by Eyford *et al.* combined with previous studies about the trypanosome surface proteins provide a valuable platform, from which to identify target proteins such as *Tc*CISSA and *Tc*CESP that are likely to have significant biological roles in promoting parasite transmission.

1.6. Objectives

T. congolense is the major cause of African animal trypanosomiasis in sub-Saharan Africa. In contrast to the mechanisms of host infection, however, the molecular strategies that enable the transmission of this disease remain largely enigmatic. This is due, in part, to the technical challenges associated with studying the vectors themselves. The goal of my research is to describe the molecular basis of transmission. Structural information about proteins offers valuable insight into the function, I chose biophysical approaches to gain insight into the structures of *Tc*CISSA and *Tc*CISSA. The approach I have taken in my research is to build on recent studies of describing the surface protein coat of *Trypanosoma congolense* as the parasite is transmitted by the tsetse fly. Based on existing data, we postulate that both *Tc*CISSA (Chapter 2) and *Tc*CESP (Chapter 3) play important roles in mediating cross talk between parasite and insect vector. To establish the molecular mechanism through which this occurs, the goal of my thesis work is to elucidate the structural and biochemical characteristics of *Tc*CISSA and *Tc*CESP.

Chapter 2 – *T. congolense* Insect Stage Specific Antigen (*Tc*CISSA)

Contribution

The initial molecular biology steps were performed by Drew Bowie and Sean Workman. The preliminary native crystals of *Tc*CISSA were obtained by Sean Workman and I was responsible for growing the native crystal with higher diffraction quality. I was also responsible for the engineering and the cloning and the expression trials of the mutants and crystallization of the selenomethionine (SelMet) derivative. In addition, I purified labelled *Tc*CISSA for NMR studies.

2.1. Introduction

According to the research done by Eyford et al., *Tc*CISSA is only expressed in the procyclic and epimastigote forms of *T. congolense* (75). It is therefore logical to assume that *Tc*CISSA plays a role in facilitating transmission. The potential for *Tc*CISSA to play an important biological role is also reflected in its shared 64% sequence identity (Figure 2.1) with the functionally characterized *T. brucei* homolog, (*Tb*PSSA-2). Fragoso *et al.* have previously shown that *Tb*PSSA-2 is required for both maturation of *T. brucei* infection in the tsetse fly and migration of the parasite to the salivary glands. More specifically, deletion of *Tb*PSSA-2 resulted in a decrease in the infection of salivary glands by *T. brucei,* without altering the level and intensity of the midgut infections. Complementing the null-mutants with full-length *Tb*PSSA-2 restored the ability of *T. brucei* to infect the salivary glands. Despite the functional observations associated with *Tb*PSSA-2, there are no reports describing the mechanism by which the protein promotes survival of the parasite in the fly. Structural and biophysical characterization of *Tc*CISSA, therefore, has the potential to provide substantial insight into how both congolense and brucei species of trypanosomes engage the tsetse vector at the molecular level.

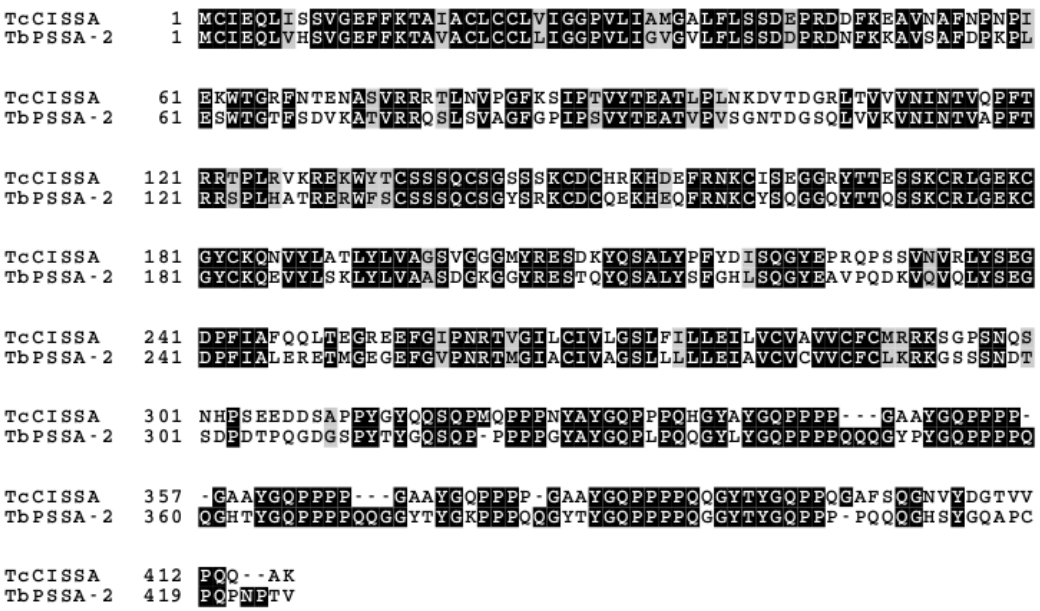

Figure 2.1. Amino acid sequence alignment between *Tb*PSSA-2 and *Tc*CISSA. The two proteins share about 64% sequence identity. The identical amino acids are shaded black and the amino acids that have similar chemical properties are shaded gray. Alignment has been done using BLAST (85) and the shading was done using BoxShade.

2.2. Materials and methods

2.2.1. Sequence alignment of different homologs of *Tc*CISSA

Using protein BLAST (85,86) and the amino acid sequence of *Tc*CISSA as the query sequence, homologues of *Tc*CISSA from different species of *Trypanosoma.* Since my research focused on the N-terminal of *Tc*CISSA, I aligned the first 311 amino acids of this protein with the first 311 amino acids of the homologues. The multiple alignment was carried out using ClustalW (87), using the slow pairwise alignment. All the different sequence comparison matrices, BLOSUM, ID, PAM and Gonnet were examined for the first ~300 amino acids of these proteins (containing their N-terminal).

2.2.2. Construct design, cloning and mutagenesis

The designed gene construct for recombinant *Tc*CISSA (r*tc*cissa) starts after the predicted signal peptide cleavage site (88) and ends before the predicted transmembrane coil. Figure 2.2 shows the amino acid sequence of *Tc*CISSA that was included in the construct.

The gene was codon optimized for expression in *Escherichia coli* and was synthesized by GenScript. The synthesized gene was flanked by NcoI (5' end) and NotI sites (3' end). The synthetic gene was digested for 30 minutes at 37°C with 1 µL of Not1 (Fermentas, cat. no. FD0594) and 1 µL of Nco1 (Fermentas, cat. no. ER0572) in 1X FastDigest buffer (Fermentas). The DNA fragment corresponding to the r*Tc*CISSA gene was separated from the vector on a 1% agarose gel by agarose gel electrophoresis. The DNA band corresponding to *rtccissa* was excised from the gel and purified using a QIAquick Gel Extraction Kit (Qiagen, cat. No. 28704). Two mutants, *Tc*CISSA L103M and L154M were designed, each with an additional methionine and were codon optimized and synthesized by GenScript. The basis for choosing these leucines residues was to have them about 50 amino acids apart from the methionine residue.

The DNA concentration was measured using a NanoDrop. It was then ligated in a 2:1 ratio of insert to vector into the 5' NcoI and 3' NotI endonuclease sites of pET32amod, which was previously treated with alkaline phosphatase, using T4 DNA ligase and 1X ligase buffer according to manufacturer's instruction (NEB cat. No. M0202S). The N-terminal of this *E. coli* specific vector contains a thioredoxin tag to support the formation of disulphide bonds and a hexa-histidine (His) tag, and these tags are separated from the

rTccissa construct by a thrombin cleavage site. The resistance marker for this vector was ampicillin. Figure 2.3 shows the predicted amino acid sequence of the resulting clone.

A tube containing 30 µl of chemically competent *E. coli* DH5α was transformed with *rtccissa/*pET32amod DNA and grown on Luria Bertani (LB) agar plates supplemented with ampicillin to a final concentration of 50 µg/mL. The plate was incubated at 37°C overnight. A single colony was grown in 5 mL LB broth supplemented with ampicillin (50 µg/mL). *rTccissa/*pET32amod DNA was then isolated from the cells grown in the 5 mL LB using a Qiagen Plasmid Miniprep Kit (Qiagen, cat. no. 27104).

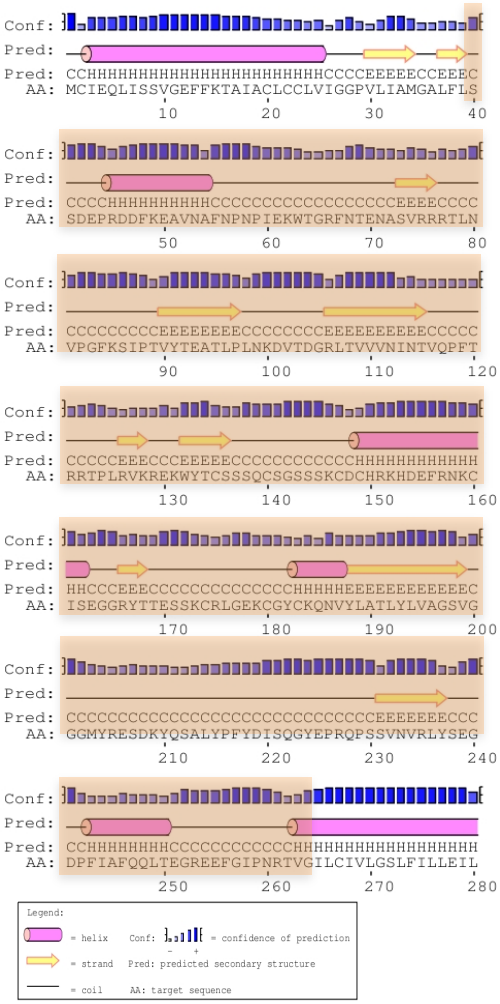

Figure 2.2. Amino acid sequence of extracellular region of *Tc*CISSA. The area highlighted in orange shows the secondary structure prediction of the construct designed for expression in *E. coli* and crystallization trials. The secondary structure prediction (pink bars represent helices and yellow arrows show the beta-strands) was used to ensure that the construct does not contain too far from the start of a predicted secondary structure.

MSDKIIHLTDDSFDTDVLKADGAILVDFWAEWCGPCKMIAPILDEIADEYQGKLTVAKL
NIDQNPGTAPKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAGSGSGHMHH
HHHHSSGLVPR/GSGMKETAAAKFERQHMDSPDLLVPR/GSAMGSSDEPRDDFKEAVNA
FNPNPIEKWTGRFNTENASVRRRTLNVPGFKSIPTVYTEATLPLNKDVTDGRLTVVVNI
NTVQPFTRRTPLRVKREKWYTCSSSQCSGSSSKCDCHRKHDEFRNKCISEGGRYTTESS
KCRLGEKCGYCKQNVYLATLYLVAGSVGGGMYRESDKYQSALYPFYDISQGYEPRQPSS
VNVRLYEGDPFIAFQQLTEGREEFGIPNRTVGAAA

Trx tag His tag Thrombin cut site NcoI NotI C = 8 cysteines

Figure 2.3. Expected amino acid of r*Tc*CISSA. The thioredoxin tag (red bold text) NcoI (yellow highlight) and NotI (green) sites, the cysteines (gray highlight), thrombin cleavage sites (blue highlight) and the His-tag (blue text) have been shown in different colours in the sequence.

2.2.3. Small scale protein expression

A tube of 30 µL Rosetta-gami 2 (DE3) cells (Novagen) was thawed on ice and the cells were transformed with 1 µL miniprepped *rtccissa/*pET32amod DNA (10 µL competent cells), and plated onto LB plates supplemented with ampicillin following an overnight incubation at 37°C. One colony was used to inoculate 5 mL 2x YT medium for starter culture. The medium contained 16 g/L tryptone, 10 g/L yeast extract and 5 g/L NaCl, supplemented with 50 µg/mL ampicillin and 0.5% glucose to avoid undesired expression. The cells from the starter culture were transferred to 50 mL of ZYP-5052 autoinduction medium supplemented with 50 µg/mL ampicillin. The ingredients of this medium are listed in Tables 2.1 and 2.2.

Table 2.1. Ingredients of ZYP 5052 medium.

Component	Volume
ZY (5 g/L tryptone, 10 g/L yeast extracts)	928 mL
1 M $MgSO_4$	1 mL
50x 5052	20 mL
20x NPS	50 mL

Table 2.2. Ingredients of 20x NPS and 50x 5052.

20x NPS		50x 5052	
dd H_2O	90 mL	Glycerol	25 g
$(NH_4)_2SO_4$	6.6 g	H_2O	73 mL
KH_2PO_4	13.6 g	Glucose	2.5 g
Na_2HPO_4	14.2 g	α-lactose	10 g

The flask was incubated at 37°C for 4 hours followed by a 44 hour incubation at 16°C. In order to take a sample for negative control, a 2 mL aliquot was taken right before the temperature was switched to 16°C and it was centrifuged at 10000 rpm for 2 minutes at 4°C (Eppendorf). After removing the supernatant, the cells were stored at -20°C. In order to assess the protein expression levels, a 2 mL

sample was taken at various time points, centrifuged at 10000 rpm for 1 min at 4°C (Eppendorf). Cells from these time points and the cells taken as negative control were lysed using 150 µL of BugBuster Protein Extraction Reagent (Novagen) and 0.1 µL of 2.5 mg/mL DNAse (Sigma) were added to each tube and they were incubated at room temperature for 10 minutes. The lysates were centrifuged at 13000 rpm for 10 minutes at 4°C. The supernatant was transferred to a fresh tube and 20 µL was taken to represent the soluble fractions. The insoluble fraction of the lysate was re-suspended in 130 µL of dH_2O.

In order to harvest r*Tc*CISSA from the soluble fraction, Nickel purification was performed using a bed-volume of 30 µL of 50 % Ni-chelated Sepharose beads (Amersham Pharmacia Biotech AB) in 1X binding buffer (10 mM Tris pH 8.5, 1 M NaCl, 30 mM imidazole). The tube was incubated for 1 hour at 4°C on a Mini LabRoller (LabNet International). After the incubation, the samples were centrifuged at 2300 rpm for 5 minutes at 4°C (Eppendorf) and the supernatants were saved and represented the flow thru fraction (proteins un-bounded to Ni^{2+} beads). Ni^{2+} beads were re-suspended in 300 µL of binding buffer and the tubes were gently mixed to keep the beads re-suspended, then centrifuged, and the supernatant was saved as the wash fraction. Protein bound to the Ni^{2+} beads was eluted by adding 30 µL of stripping buffer (10 mM Tris pH 8.5, 1 M NaCl, 500 mM imidazole) and incubated for 3 minutes on ice and 20 µL was taken as the purified fraction. The aliquots from different fractions were analyzed on a 15% SDS poly acrylamide gel run at 185 V for 1 hour and 30 minutes.

2.2.4. Large scale protein expression and purification

The transformation, culture media and the growth conditions were as described for the small scales. The volume of the starter culture was 100 mL and for the ZYP 5052 it was 2 L. Following the 48 hours incubation in ZYP 5052, the cells were re-suspended in 40 mL

1X binding buffer supplemented with 60 µL protease inhibitor cocktail (Calbiochem), and 100 µL DNase. The cell suspension was homogenized and de-gassed, and the cells were lysed using a French press (SLM Aminco, SLM Instrument, Inc.). The lysate was centrifuged at 14000 rpm for 20 minutes at 4°C (JA-20 rotor, Beckman Coulter Avanti J-E centrifuge) and the insoluble fraction was removed following centrifugation. The supernatant was taken and diluted by adding 100 mL binding buffer. A volume of 60 µL of protease inhibitor cocktail and a 2.5 mL bed volume of Ni-chelated Sepharose beads (equilibrated in binding buffer) were added to the lysate. The batch bind was stirred at 4°C for 30 minutes. Ni^{2+} beads were separated using vacuum filtration through a coarse frit (Chemglass Life Sciences). The beads were then washed with 60 mL binding buffer and were transferred to a Bio-Rad Poly-Prep disposable column (731-1550EDU). The r*Tc*CISSA bound to beads was eluted in a series of three 6-8 mL elutions with elution buffer (10 mM Tris pH 8.5, 1 M NaCl, 250 mM imidazole). Following the elutions, the remaining protein bound to the beads was stripped using 5 mL of stripping buffer (10 mM Tris pH 8.5, 1 M NaCl, 500 mM imidazole). The absorbance of the protein solution was measured at 280nm using spectrophotometer (DU 730 Life Science UV/Vis Spectrophotometer, Beckman Coulter) and the concentration was calculated based on the extinction coefficient of 29380 M^{-1} cm^{-1}. Samples were analyzed by SDS-gel, fractions pooled based on purity, concentrated and buffer exchanged into size exclusion buffer (20 mM HEPES pH 7.0, 150 mM NaCl, 1% glycerol) using a 10 kDa molecular weight cut-off (MWCO) Amicon Ultra spin concentrator (Millipore). The thioredoxin tag was removed by an overnight thrombin cleavage (1 µL thrombin/ 20 mg protein) at 18°C.

The harvested r*Tc*CISSA was purified by size exclusion chromatography using a Superdex 75 16/60 HiLoad column (GE Healthcare) equilibrated in size exclusion buffer. The expected elution volume of r*Tc*CISSA monomer was approximately 70 mL and the thioredoxin tag approximately 80 mL. Fractions from the size exclusion were run on a 15% SDS-gel pooled, concentrated and buffer exchanged into A1 buffer (20 mM HEPES pH 6.8, 10 mM NaCl, 1% glycerol) for cation exchange using a 3 kDa MWCO Amicon spin concentrator (Millipore). After buffer exchanging, cation-exchange chromatography was performed using a HETP source 30S column (GE Healthcare) and *Tc*CISSA bound to the column was eluted with an increasing gradient of NaCl using of buffer B (20 mM

HEPES pH 6.8, 0.5 M NaCl, 1% glycerol). Fractions were run on a 15% SDS gel, pooled based on purity and concentrated to 22 mg/mL using a 3 kDa MWCO Amicon spin concentrator (Millipore).

2.2.5. Protein expression in SelMet and M9 media

In order to obtain SelMet derivatized crystals from r*Tc*CISSA, we had to use a culture medium to support the growth of *E. coli* Rosetta gami cells. The incubation conditions were identical to native *Tc*CISSA. The ingredients are listed in Tables 2.3 and 2.4. r*Tc*CISSA labelled with ^{13}C and ^{15}N was expressed in M9 medium in order to prepare a sample for NMR experiments. After growing a single colony of Rosetta gami transformed with *rTccissa* DNA in 500 mL 2x YT medium, cells were transferred to 4 L of M9 medium. Cultures in both media were induced with 0.7 mM IPTG for M9 and SelMet media and the recipe for the M9 medium is given in Table 2.5. The micronutrient solution for both media was the same. All the media were supplemented with ampicillin as described.

Table 2.3. Ingredients of the SelMet medium

SelMet medium (per liter)	
40% glucose	7.5 mL
Amino acid solution	30 mL
SelMet (Molecular dimensions 10 g/L)	4 mL
$MgSO_4$ (1M)	1 mL
Thiamine and Biotine (10 mg/mL)	1 mL
Micronutrient solution	1 mL
20x NPS	50 mL

Table 2.4. Ingredients of M9 medium with ^{13}C and ^{15}N.

M9 medium (per liter)	
^{13}C glucose	2 g
$^{15}NH_4Cl$	1 g
20x NPS	50 mL
$CaCl_2$ (0.1 M)	1 mL
$MgSO_4$ (1M)	1 mL
Thiamine and Biotine (10 mg/mL)	1 mL
Micronutrient solution	1 mL

Table 2.5. Ingredients for the micronutrient solution

Micronutrient solution (100 mL)	
$FeSO_4.7H_2O$	0.6 g
$CaCl_2.2H_2O$	0.6 g
$MnCl_2.6H_2O$	0.1 g
$CoCl_2.6\ H_2O$	0.08 g
$ZnSO_4.7\ H_2O$	0.07 g
$CuCl_2.2H2O$	0.03 g
H_2BO_3	0.002 g
$(NH_4)_6Mo_7O_{24}.4H_2O$	0.025
EDTA	0.5 g

2.2.6. Crystallization and data collection

Initial crystallization trials for r*Tc*CISSA were performed using the Gryphon crystallization robot (Art Robbins Instruments). The commercial screens and the protein were set in 96-well sitting drop Intelli-Plates (Hampton Research) with 55 μL reservoir solution and different drop ratios in the wells. The screens that were used for initial trials were Index screen (Hampton Research), Wizard I and II screens (Emerald BioSystems), and MCSG-1 screen (Emerald BioSystems) and PEG/Ion (Hampton Research). The resulting crystals were soaked in NaI, KI, NaBr, KBr, LiBr, NH_3Br, and heavy atoms.

Native and SelMet derivatized crystals were looped and then soaked into 25% glycerol as the cryoprotectant in a stepwise manner. The crystals were flash-frozen in liquid nitrogen (77 K or -196°C). Diffraction data from the native *Tc*CISSA crystals was collected on beamline 7-1 at the Stanford Synchrotron Radiation Laboratory (SSRL) using an ADSC Quantum 315r CCD detector. In total, 320 images were collected with 0.5 degree oscillations.

2.3. Results

2.3.1. *Tc*CISSA Homologs

To investigate the presence of *Tc*CISSA homologs in other *Trypanosoma* species, a BLAST search was performed. The homolog of *Tc*CISSA that shares the highest amino acid sequence identity (64%) is *Tb*PSSA-2 from *T. brucei*. *Tb*PSSA-2 is required for maturation of *T. brucei* infection in tsetse fly (84). Interestingly, there were genes present in other species of *Trypanosoma* that had at least 40% amino acid sequence identity to *Tc*CISSA. The most identical amino acid sequence after *Tb*PSSA-2 was observed a putative PSSA (Procyclic Stage Specific Antigen) from *T. vivax.* This species infects cattle and is a major cause of N'gana. The interesting feature of this parasite is that it can be transmitted mechanically as well as cyclically and is therefore more widespread than *T. congolense* (89).

It is intriguing that *T. cruzi* and *T. rangeli,* which are from the Stercoraria with completely different life cycles and vectors, also harbour genes encoding potential PSSA homologs. These putative PSSA homologs share ~40% sequence identity with *Tc*CISSA in their N-terminal region (the first 311 amino acids). While *T. cruzi* is a well-known human pathogen in the South American continent, *T. rangeli* is non-pathogenic (90). The number of *Trypanosoma* species that express a *Tc*CISSA and have different life cycles suggests the possibility that the protein is a key multifunctional protein, or perhaps plays a more structural role such as promoting membrane integrity. An alignment of these amino acid sequences is shown in Figure 2.4.

```
CISSA         MCIEQLISSVGEFFKTAIACLCCLVIGGPVLIAMGALFLSSDEPRDDFKEAVNAFNPNPI 60
TbPSSA        MCIEQLVHSVGEFFKTAVACLCCLLIGGPVLIGVGVLFLSSDDPRDNFKKAVSAFDPKPL 60
TvPSSA-2      MCLMDLVESIDNCFKSAVACLCFLVIGGPVLIAIGVLFLRHDDGENEFKEALKAFNPESV 60
TcrPSSA       MCLFGIVDAVAGFFKSVVACACCLIIGGPALIIAGSMLLNQQDLRKAFSDDVKEFNPTPL 60
TrPSSA        MC-FDILGAVEDCYKFVVSFACCLIIGGPALIIAGGILLNKGDGKKAFTEAVKEFDPTFI 59
              **   :: ::   :* .::  * *:****.**  * ::*   : .. *.. :. *:*  :

CISSA         EKWTGRFNTENASVRRRTLNVPGFKSIPTVYTEATLPLNKDVTDGRLTVVVNINTVQPFT 120
TbPSSA        ESWTGTFSDVKATVRRQSLSVAGFGPIPSVYTEATVPVSGNTDGSQLVVKVNINTVAPFT 120
TvPSSA-2      SHWSGFINGHNASLRRGSLNVPGYSGVPTRYVTASVVYP-PVGAKNLRVEVDIRSVERFS 119
TcrPSSA       NAWTGTINDVPITVRRESLNVQGVDGAISVFAEAVVSVPQ-RSSSRFPVSVNVNTVASFV 119
TrPSSA        NAWTGKINDAPITVRRESLNVQGVTGATSVFAEAVIPVESSRSLGVISISVNVFNVTPFS 119
              . *:* :.    ::** :*.* *     : :. * :           : : *:: .*  *

CISSA         RRTPLRVKREKWYTCSSSQCSGSSS-KCDCHRKHDEFRNKCISEGGRYTTESSKCRLGEK 179
TbPSSA-2      RRSPLHATRERWFSCSSSQCSGYSR-KCDCQEKHEQFRNKCYSQGGQYTTQSSKCRLGEK 179
TvPSSA        RSVPLAVRSEKTYSCYSSDCDKYSSSKCRCTKENDKFRRKCEANGGKYSPTVTDCSLGHQ 179
TcrPSSA       RKAPFRAIKKTSYTCTSSDCRSRLN--CRCNELLNSFMDQCMASGGRFVRTSGMCVLDRT 177
TrPSSA        RRAALQTTRKSVFDCSSTSCRAGKS--CRCDREAREFQDKCIAMKGVFDAYPSWCRSGHK 177
              *  .: .  :  : * *:.*       * * .   .*  :* :  * :      *  ..

CISSA         CGYCKQNVYLATLYLVAGSVGGGMYRESDKYQSALYPFYDISQGYEPRQPSSVNVRLYSE 239
TbPSSA-2      CGYCKQEVYLSKLYLVAASDGKGGYRESTQYQSALYSFGHLSQGYEAVPQDKVQVQLYSE 239
TvPSSA        CSTCTQRVYLNTVYLVAEDIGNGKFRESSRYASATHPMGSRS-GYSSRRGDTIEVRLYSD 238
TcrPSSA       CGTCERTVYLRRLYLVVSEVGNGKYVEDTKLRSAMYPFGDLDNDYQPGIPSTVTVRLYSS 237
TrPSSA        CGQCTTAIYLSRLYLVVQEVSKGHYAEDTALQSAKYAFGAMDNDYQPTMPSEVAVRLYSN 237
              *. *   :**  :***. . . * : *.    ** :.:   . .*..   . : *:***.

CISSA         GDPFIAFQQLTEGREEFGIPNRTVGILCIVLGSLFILLEILVCVAVVCFCMRRKSGPSNQ 299
TbPSSA-2      GDPFIALERETMGEGEFGVPNRTMGIACIVAGSLLLLLEIAVCVCVVCFCLKRKGSSSND 299
TvPSSA        KDPLVALEHLTHGRGEFGLGNRTIGIAAIVFGSLLILLEICACAVMICFCMKRKESAGQA 298
TcrPSSA       KDPYIALQRLTRGSNDLGVNPRTVGIVLIVLGCLFLLLEIGVCTALICYWTRRKKTSSGA 297
TrPSSA        KDPYIALQRETSGTGEFGPNPRTVGIILIVLGVLFLLLEVCVATAIICYCTRRNKTTSDA 297
               ** :*::: * *  ::*   **:**  ** * *::***: ... ::*:  :*:  ..

CISSA         SNHPSEEDDSAP 311
TbPSSA-2      TSDPDTPQGDGS 311
TvPSSA        SGEDNDEMGYSP 310
TcrPSSA       TPYLAPASYGIS 309
TrPSSA        SPNFSADNYQTA 309
              :          .
```

Figure 2.4. Alignment of the first 311 amino acids of *Tc*CISSA and *Tb*PSSA-2 with the other potential PSSA proteins from other trypanosome species. about 40% of the amino acids are identical in all sequences. Considering how different the species are from each other, this level of amino acid sequence identity is very high. Tv, Tcr, and Tr represent *T. vivax, T. cruzi* and *T. rangeli,* respectively.

2.3.2. Recombinant protein production

The N-terminal of *Tc*CISSA contains eight cysteine residues. Since this part of the protein is located outside the cell in an oxidizing environment compared to the cytoplasm, it is reasonable to expect that these cysteines are in the form of disulfide bonds. Because of the number of potential disulfide bonds in our r*Tc*CISSA construct, we chose the Rosetta-gami strain of *E. coli* for expressing r*Tc*CISSA. This strain of *E. coli* harbours mutations in *trxB* (thioredoxin reductase) and *gor* (glutathione reductase) genes. These genes express enzymes that reduce the disulfide bonds; therefore the Rosetta-gami strain is ideal for expressing proteins where the cysteines are in disulfide bonds. Rosetta-gami also contains a plasmid that encodes rare tRNAs to facilitate expression of proteins with substantial codon bias (91-94). Expression from the pET32a vector augments the protein folding capabilities of Rosetta-gami in that pET32a contains a thioredoxin fusion, an N-terminal chaperon that promotes disulfide bonding in the recombinant protein (95-97). Using this engineered strain of *E. coli* in conjunction with the optimal expression vector, I was able to successfully produce recombinant *Tc*CISSA that showed a clear protein band on an SDS-PAGE gel following Nickel affinity purification (Figure 2.5 A).

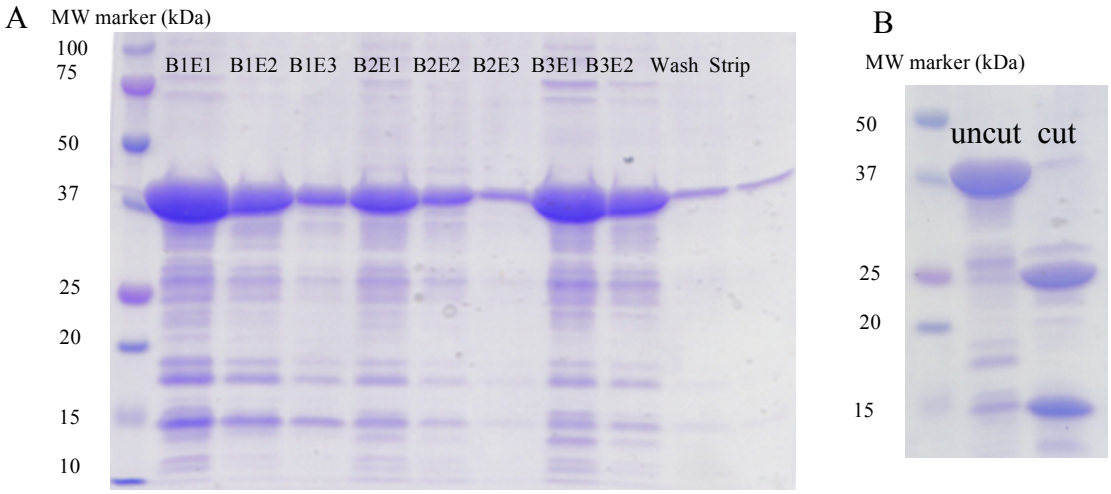

Figure 2.5. SDS-gel from r*Tc*CISSA batch-bind elutions. A: Following the cell lysis and after each batch-bind, the beads were separated and washed (W) with binding buffer. Beads were eluted 3 times, each time with about 7 mL elution buffer. Beads were then stripped (S) of all the proteins bound to them using stripping buffer. Three batch binds were performed. The letter B stands for Batch and E for Elution. B1E1: Batch1, elution 1 etc. **B: Verification of the thrombin cleavage.** The expected size of r*Tc*CISSA is about 26 kDa and the thioredoxin tag is about 13kDa. The gel demonstrates that no un-cut r*Tc*CISSA is left. The othe bands observed on the gel show the contaminants that are present after the affinity purification.

Fractions were pooled based on purity and thrombin removal of the thioredoxin tag was tested in small scale (Figure 2.5 B). Following the thrombin cleavage step, the protein was further concentrated to a final volume of 2 mL and centrifuged to remove the insoluble debris. The protein sample was run on a Superdex 75 size exclusion chromatography column (Figure 2.6). Three largely distinct peaks were observed, representing soluble aggregate, monomeric *Tc*CISSA and thioredoxin tag. The protein purity requirement for structural studies is very high and since the sample still showed residual contaminating thioredoxin tag following size exclusion chromatography, a cation exchange chromatography (Figure 2.7) step was included, resulting in a highly pure sample. As the following two figures show, the SDS gels and the single peaks from the chromatogram suggest that the protein is highly purified and consists of a homogeneous population.

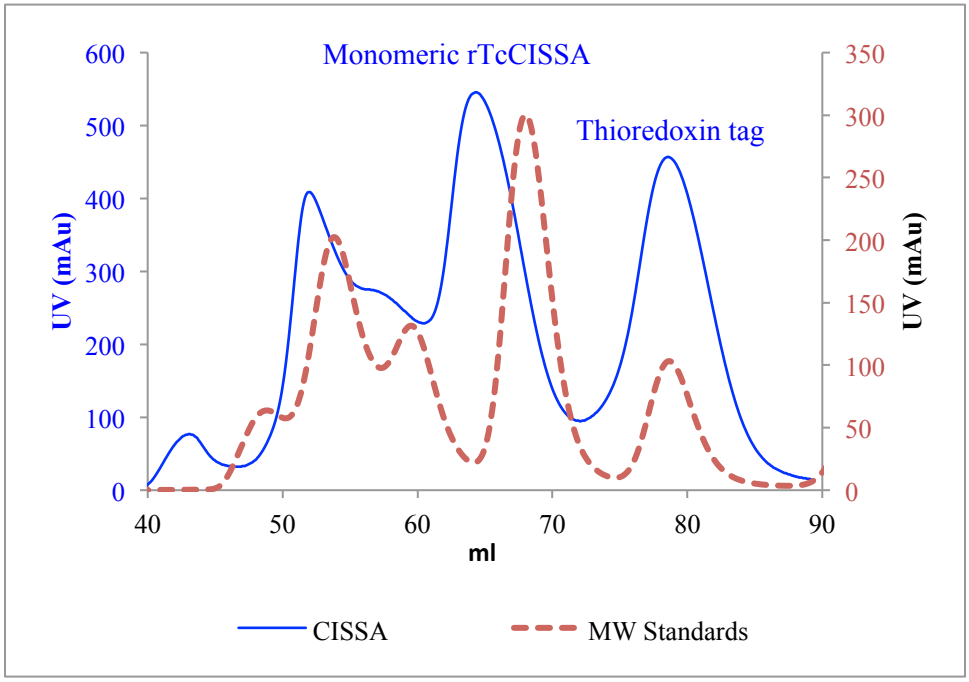

Figure 2.6. Size exclusion chromatography of r*Tc*CISSA. The dotted red and the secondary y axis (red) curve represent the molecular weight standards for Superdex 75 column. r*Tc*CISSA is plotted in blue. The gel picture at the right hand side shows the pooled fractions from monomeric r*Tc*CISSA, showing a trace amount of thioredoxin tag remaining in the sample.

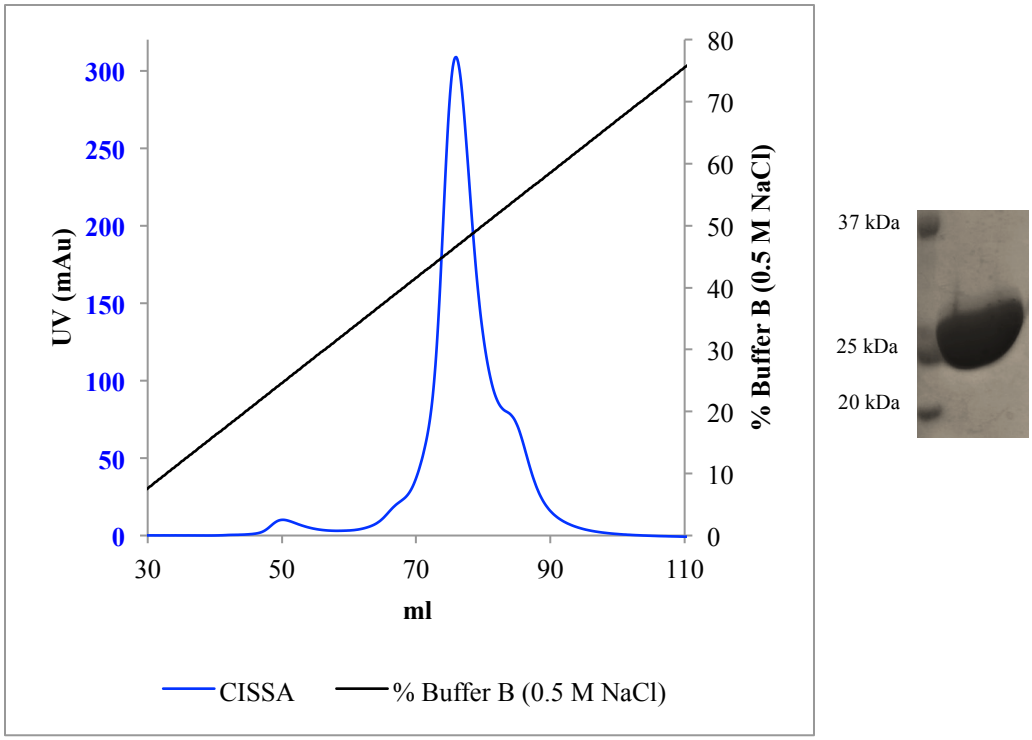

Figure 2.7. Cation exchange chromatography of r*Tc*CISSA. The black line and the secondary y axis show the percentage of buffer B (0.5 M NaCl). r*Tc*CISSA is plotted in blue. The protein bound to the column was eluted with ~ 300 mM NaCl. The gel at the right hand side confirms the removal of the thioredoxin tag and confirms the expected size.

2.3.3. Crystallization

Initial crystals were observed in polyethylene glycol (20%) based conditions supplemented with either ammonium sulphate or lithium sulphate. The resulting crystals diffracted to 2.5 Å resolution. The previous dataset collected from *Tc*CISSA crystals had a resolution of 2.8 Å (98). Despite early crystallization success, reproducibility of the crystals was a major problem that may have arisen due to the instability of PEG based solutions. It is well documented that aging of the PEG can result in a reduction of pH and increased ionic strength and elevated levels of aldehydes, carboxylates and peroxides, as well as increased metal binding (99,100). However, through crystal seeding experiments I was able to generate subsequent rounds of diffraction quality crystals, which allowed us to explore derivatization experiments to properly phase the structure.

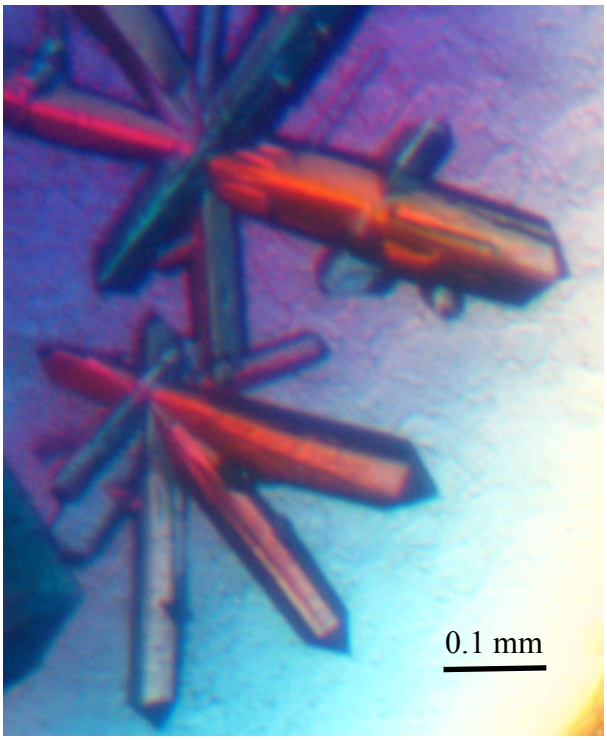

Figure 2.8. r*Tc*CISSA crystals. These crystals formed in two different conditions and grew overnight. Both conditions yielded the same crystal morphology.

2.3.4. SelMet derivatization

Since we were unable to identify appropriate models to enable structure determination through molecular replacement, the phase problem was approached through SelMet derivatization. Using this strategy, methionine residues in a protein are replaced with SelMet that provide an anomalous signal that can be used to "phase" or determine the structure. Efforts to express SelMet derivatized r*Tc*CISSA using the B834 strain of *E. coli,* a strain incapable of synthesizing methionine, were unsuccessful. Therefore, a protocol was developed to express SelMet derivatized r*Tc*CISSA using the Rosetta gami *E. coli.* SelMet was incorporated in r*Tc*CISSA by growing the Rosetta gami *E. coli* in a homemade defined culture medium supplemented with SelMet. The yield of protein harvested from this medium was approximately half the yield of native r*Tc*CISSA from ZYP-5052. The protein harvest and purification and the size exclusion and cation exchange traces were similar to native r*Tc*CISSA. The crystallization condition for SelMet r*Tc*CISSA was the lithium sulfate condition.

X-ray diffraction studies revealed that the SelMet crystals were also orthorhombic and had the same general characteristics as the native r*Tc*CISSA crystals. However, the diffraction resolution was limited to 3.3 Å and the anomalous signal from the single methionine was not strong enough to solve the structure. To increase the anomalous signal, we therefore pursued an engineering strategy to replace leucine residues in r*Tc*CISSA with methionine residues.

2.3.5. Engineering r*Tc*CISSA mutants to increase the number of methionines

Leucine and methionine are structurally similar, therefore it was expected that replacing leucines with methionines would not disrupt the structure of r*Tc*CISSA. Since there was only one methionine (excluding the one at the N-terminal of the construct) in the construct, which was more than 250 amino acids long, I selected two leucines that were 50-100 amino acids apart from the methionine. The aim was to replace each by a methionine and create two mutants with two methionines each. This would create a stronger anomalous signal from SelMet (Figure 2.9). While the double mutant isoforms expressed in regular media, no expression was observed in minimal media supplemented with SelMet. This result suggests that while the protein can tolerate substitutions of Leu to Met at these positions, it cannot tolerate substitutions of Leu to SelMet.

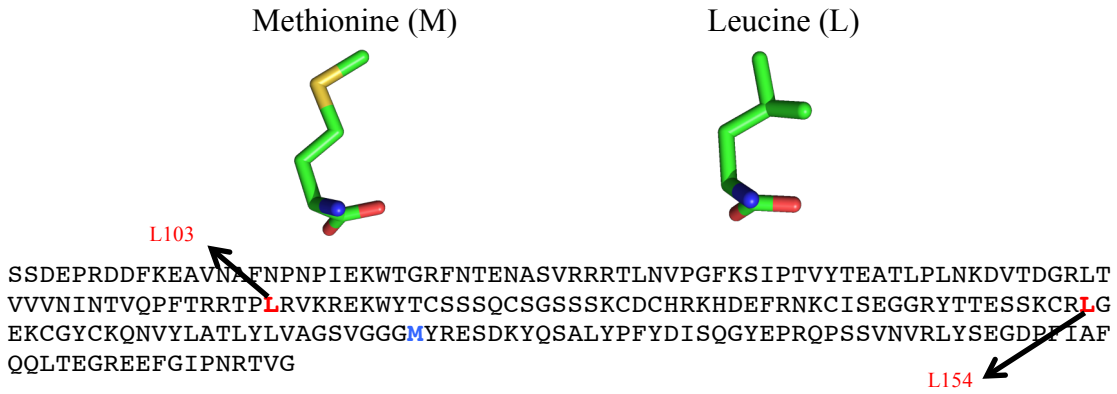

Figure 2.9. Structures of leucine and methionine and the amino acid sequence of r*Tc*CISSA. Since these two amino acids are close in structure, replacing one with the other usually does not disrupt the fold of the protein, as long as the residue is on the surface of the protein. The two chosen leucines are showed in red and the methionine residue in blue. Each of the leucines was replaced by a methionine to generate a mutant with two methionines.

2.3.6. ^{13}C /^{15}N labelled r*Tc*CISSA

As we were unable to solve the phase problem using SelMet derivatization and X-ray crystallography, we initiated a collaboration to pursue structural studies using Nuclear Magnetic Resonance (NMR) with Dr. Steve Smith's Lab at Queen's University. In support of this approach, I prepared r*Tc*CISSA labelled with ^{13}C and ^{15}N. At the time of writing, data had been collected on the 600 MHz NMR at Queen's University and the 800 MHz NMR at the University of British Columbia. As Figure 2.10 shows, the peaks have a homogeneous intensity and they are well separated from each other. These demonstrate that the quality of the NMR data was good and the sample was pure and monomeric. However, due to the size of r*Tc*CISSA, a more efficient deuteration step is required for solving the structure of this protein via NMR.

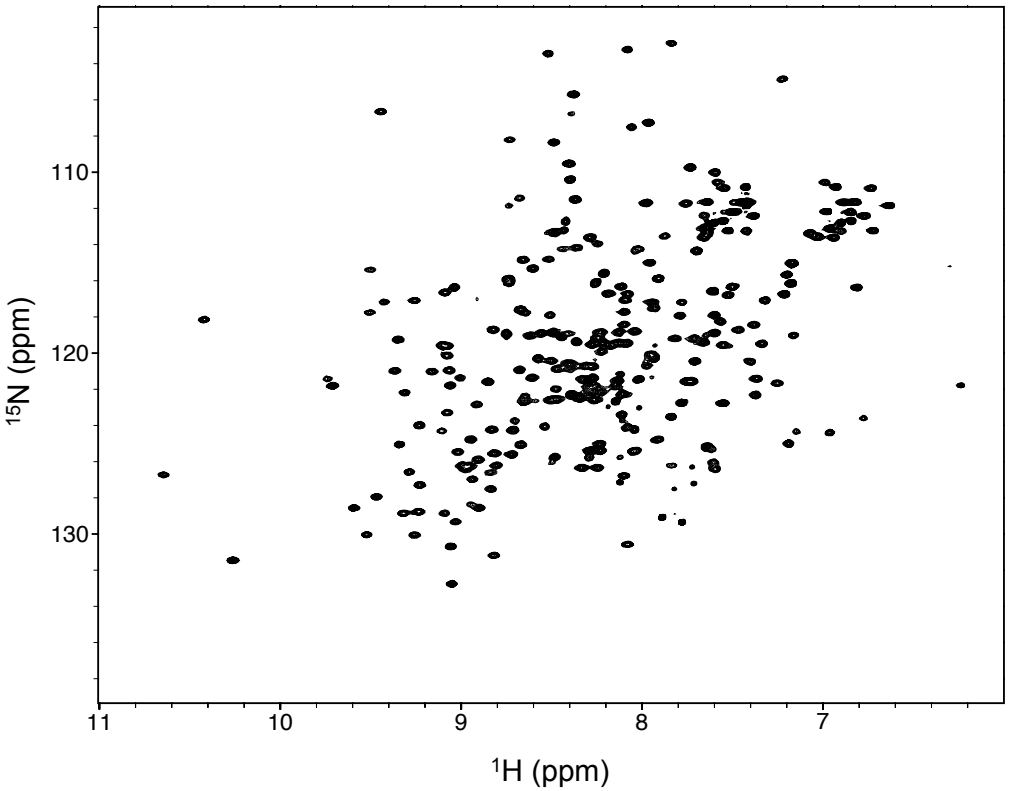

Figure 2.10. Two dimensional ^{1}H-^{15}N heteronuclear single quantum coherence spectrum of 1 mM uniformLy $^{l3}C/^{15}N$-labeled CISSA at pH 6.9 recorded at 14.1 Telsa and 308 Kelvin. The homogeneity of the spots and the fact that the spots are not clustered in one part indicate that the protein is folded properly and that the sample is monomeric and pure.

2.4. Discussion

2.4.1. *Tc*CISSA an important protein in the insect stages of *T. congolense*

As discussed earlier, *Tc*CISSA is a homologue of *Tb*PSSA-2, a protein required for maturation of *T. brucei* infection in the tsetse fly (84). The amino acid sequences of these two proteins are highly identical (64%), they have the same cell localization and they are also expressed in the same life cycle stage. We therefore expect that they carry out a similar function. In addition, there are homologues of *Tc*CISSA in other species of *Trypanosoma,* which could mean that this is a protein needed for maturation of infection in other vectors as well.

2.4.2. Other procyclic stage specific antigens (PSSA) in other trypanosomes

As figure 2.4 shows, the PSSA protein is a conserved protein in many different trypanosomes. The conserved N-terminal is extracellular and interacts with the vector. Different models of alignment were examined for the N-terminal and they all showed the same model of alignment and guide tree. I therefore presented the default model (Gonnet) It is possible that the N-terminus of these proteins is interacting with a universal signal common to all the arthropod vectors.

Fragoso *et al.* suggested that *Tb*PSSA2 may form a dimer and they reported that C_{289} could be involved in this dimerization (84). This cysteine residue is conserved in the other species except *T. cruzi (Tcr*PSSA). The *Tcr*PSSA is shorter though and there is a cysteine residue in position 272 of this protein. The difference in length could explain the difference in the position of this cysteine residue.

Fragoso *et al.* showed that the phosphorylation of T^{305} is needed for the localization of *Tb*PSSA-2. They showed that a T305D mutant has the same general properties as the wild type. Replacing a threonine with an aspartate mimics the phosphorylation of Threonine (84). Interestingly, in the sequence of *Tv*PSSA there is an aspartate instead of the threonine in *Tb*PSSA2. As discussed earlier, *T. vivax* has a different life cycle from the other Salivarian trypanosomes (19,101), yet it still shares a significant amino acid identity with *Tc*CISSA and *Tb*PSSA-2. This may reflect a general function for the protein

such as stabilizing membranes or acting as chaperones to promote assembly of protein complexes.

While *T. vivax* is from the Salivaria group, *T. rangeli* and *T. cruzi* belong to the Stercoraria. *T. rangeli* is an unusual species. The transmission of *T. rangeli* occurs both through the saliva and the feces of the vector, with the anterior transmission being more efficient. This unusual behaviour could mean that *T. rangeli* is an intermediate between the Stercoraria and Salivaria or that it belongs to a different subgenus, however small-subunit RNA sequences have shown that this species belongs to the Stercoraria (90). This unusual position of this *T. rangeli* could explain why it lacks the C-terminal. The species mentioned above are the only species of trypanosomatids that contain homologs of CISSA. It is possible that this protein is specific to the genus *Trypanosoma.*

2.4.3. The phase problem and SelMet r*Tc*CISSA

During a diffraction experiment, the intensity of the overall diffracted X-rays is measured, but the phases are unknown. In order to solve the structure of a molecule using X-ray diffraction data, one needs to find the relative phases to mathematically describe the X-ray wave function. Using SelMet derivatized crystals is a common method for solving the phase problem in crystallography (102,103). Selenium and sulphur are in the same chemical group and have the same chemical properties. However selenium has 18 electrons more than sulphur and the advantage of this method over soaking crystals in anomalous scatterers or heavy atoms is that soaking usually damages the crystals, like the case for r*Tc*CISSA crystals. The classical way of incorporating SelMet in the protein is by expressing the protein in a defined, methionine deficient medium supplemented with SelMet as the sole source of methionine. A methionine auxotroph *E. coli* strain is used to ensure the 100% incorporation of SelMet in the target protein (103). The challenge we had for *Tc*CISSA was that our attempts to express the SelMet protein using the auxotroph strain, B834, were unsuccessful and most of the expressed protein was insoluble. We hypothesized that the possible reason for this is that the B834 strain does not have the folding capacity of Rosetta-gami. Therefore, we tried a defined medium supplemented with SelMet, with the aim of suppressing methionine biosynthesis in Rosetta-gami, so that the cells would utilize the SelMet from the medium and incorporate it in the medium.

We were able to harvest a reasonable amount of SelMet protein and get SelMet crystals with clear anomalous signals.

There was a problem with the number of methionines in the r*Tc*CISSA construct. In order to solve the structure of a protein using SelMet, there should be one methionine present per 50-100 amino acids in a protein. Our construct only had one methionine in >200 amino acids. In addition, the resolution of the SelMet crystals was low (3.3 Å). So despite our success in getting the SelMet derivatized crystals, we were still unable to solve the structure. We therefore created the L103M and L154M mutants. We chose these leucines because they were about 50 amino acids apart from the methionine residue and hydrophilic residues surrounded them. Unfortunately, these mutants were not stable when expressed in the SelMet medium. It is possible that replacing the leucine residue with a SelMet is sufficient to disturb the fold of r*Tc*CISSA and thereby promote its degradation by the proteases present in the Rosetta-gami strain (104). A future direction would be to choose other leucines for mutation or choose isoleucine instead of leucines.

2.4.4. The NMR sample

As figure 2.10 shows, The ^{1}H-^{15}N HSQC spectrum of CISSA displays good chemical shift dispersion of ^{1}H-^{15}N correlation resonances, indicative of a folded protein sample. The resonances are of a uniform intensity, consistent with a monomeric, homogeneous protein sample. And approximately the number of resonances expected for non-proline amino acid residues comprising CISSA (Fig. 2.10). A subset of well resolved resonances with backbone HN chemical shifts of 8.5 ppm or greater is indicative of extended or β-structure within CISSA while the large number of resonances located in the central region of the spectrum (7.0–8.5 ppm) also suggests the presence of random coil and α-helical structure. The size of CISSA necessitates that the protein be deuterated prior to collecting additional data on the 800 kHz instrument.

Once the structure of the N-terminal of *Tc*CISSA is determined, the architecture of the protein will be revealed. This will enable us to we can gain insight into the possible functions or potential ligands for *Tc*CISSA. Structure determination of *Tc*CISSA will also provide us with additional, valuable information as to the overall shape of the surface protein coat of the parasite during passage through the tsetse fly.

Chapter 3 – Congolense Epimastigote Specific Protein (*Tc*CESP)

Contributions

Drew Bowie and Daniel Moller performed the initial cloning and DNA amplification of *Tc*CESP. Bianca Loveless was responsible for the virus amplification and insect cell expression. I performed the protein purification steps and the CD experiments. CBDPS crosslinking analysis was done by the Petrochenko group from Dr. Borcher's lab and the homology model was generated collaboratively by myself and Swapna Seshadri (the Parkinson lab, University of Toronto). I was also responsible for the analysis of crosslinked lysines and disulfide bonding on the proposed model.

3.1 Introduction

There are very few proteins identified on the surface of *Trypanosoma* parasites that are even loosely characterized as adhesins. Not only has *Tc*CESP been reported to display adhesive properties (83), it is only expressed in the epimastigote stage of *T. congolense* (75,83) (Figure 3.1). Of note, only the epimastigotes are able to adhere to the lining of the mouthparts of tsetse fly (105,106). In their study, Sakurai et al. reported that adding the epimastigote lysate (E-sup) to the procyclic form causes them to adhere to the flask. After analysing the different saturated ammonium sulfate fractions of E-sup they found the fraction containing the protein responsible for adhesion. They initially called this protein "flask-adhesion factor" and carried out further experiments and after further characterization they changed the name to CESP.

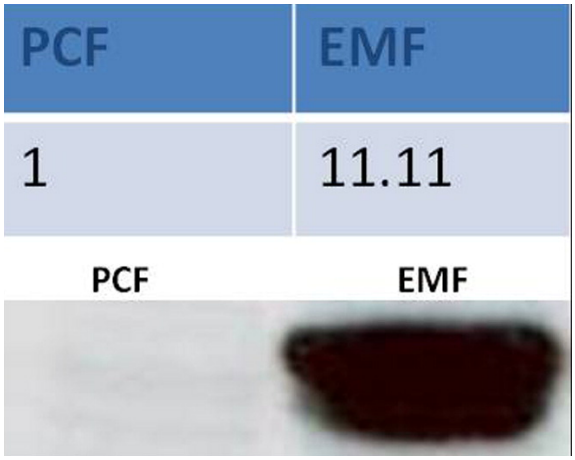

Figure 3.1. Immunoblot on cell lysate from *T. congolense* using a monoclonal antibody generated against *Tc*CESP. This figure shows that the expression of *Tc*CESP is restricted to the epimastigotes and there is no expression in the procyclic forms. Reproduced with permission from (75,107).

*Tc*CESP is 689 amino acids long and contains a 22 amino acid long signal peptide and a glycophosphatidylinositol (GPI) binding sequence (19 amino acids). Sakurai *et al.* predicted that the overall secondary structure of *Tc*CESP is mainly α-helical and they also showed that it is a GPI anchored protein. Southern analyses also showed that there are multiple copies of the *Tc*CESP gene in the *T. congolense* chromosome (83).

As stated in the introduction, the structures of HpHbR, VSG and GARP, all of which are GPI anchored proteins from trypanosomes, share the same overall predominantly α-helical structural motif (79,81,82,107). These proteins are from different life cycle stages and although they have structural similarity, they carry out different functions. One of these proteins is *T. congolense* glutamate/alanine rich protein (*Tc*GARP), which covers the surface of *T. congolense* procyclic and epimastigote forms (76,77). The *T. brucei* counterpart of GARP is brucei alanine rich protein (*Tb*BARP) and it covers the insect forms of *T. brucei.* The structures of both *Tc*GARP and *Tb*BARP have been described. Although these two proteins share around 30% amino acid sequence identity, their structures are almost identical (79,107). Interestingly, as Figure 3.2.B shows, the phylogenetic studies by Sakurai et al. showed that *Tc*GARP and *Tc*CESP are related (83). Our bioinformatics analyses have revealed that there is about 30% sequence identity

between *Tc*GARP and *Tc*CESP. It is therefore possible that the structures of these two proteins are similar.

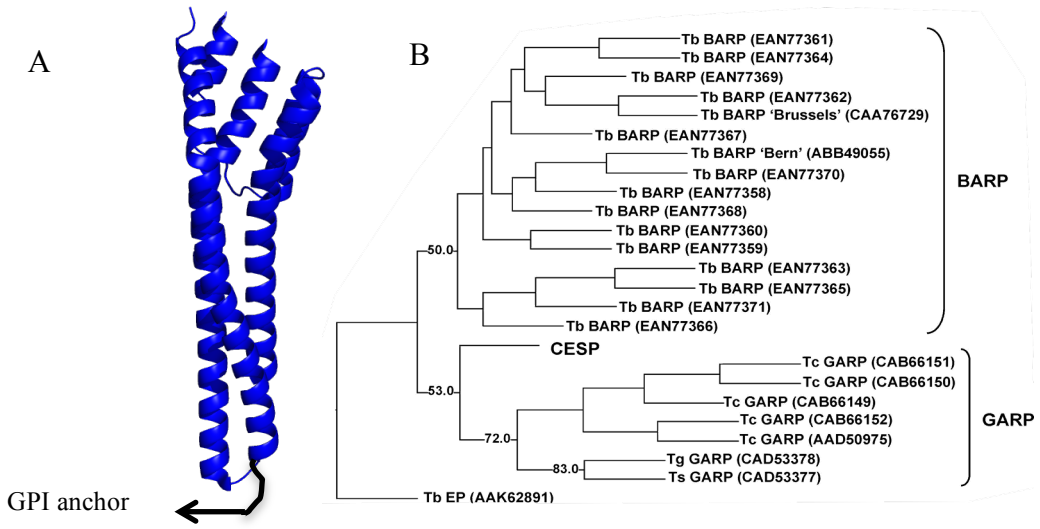

Figure 3.2. A. Structure of *Tc*GARP (79). B. Phylogenetic tree of the GARP, BARP and CESP proteins. Ts and Tg stand for *Trypanosoma simiae* and *Trypanosoma godfreyi*, respectively. This tree suggests an evolutionary relation between GARP and CESP. The phylogenetic tree has been reproduced with permission from (83).

The adhesive properties of *Tc*CESP are important, because if the epimastigotes do not adhere to the mouthparts, they would not be able to continue their growth and life cycle and will never differentiate into the infective forms, the metacyclic. Further biophysical and biochemical characterization of *Tc*CESP offers the potential for detailed insight into the molecular cross talk between parasite and insect vector. The studies presented here represent the first stage in ultimately defining the binding mechanism in the tsetse fly.

3.2 Materials and Methods

3.2.1 Construct design and cloning

The r*Tc*CESP (recombinant *Tc*CESP) gene construct starts after the predicted signal peptide cleavage site (88) and ends after the GPI anchor addition site (108-111). The gene was codon optimized for expression in insect cells and was synthesized by GenScript. The synthesized gene was flanked by NcoI (5' end) and NotI sites (3' end). The DNA was digested for 30 minutes at 37°C with 1 µL of Not1 (Fermentas, cat. no. FD0594) and 1 µL of Nco1 (Fermentas, cat. no. ER0572) in 1X FastDigest buffer (Fermentas). The DNA fragment corresponding to the r*Tc*CESP gene was separated from the vector on a 1% agarose gel. The DNA was excised from the gel and purified using a QIAquick Gel Extraction Kit (Qiagen, cat. No. 28704).

The DNA concentration was measured using a NanoDrop. It was then ligated in a 2:1 ratio of insert to vector into pAcGP67 His TEV N-term, which was previously treated with alkaline phosphatase, using T4 DNA ligase and 1X ligase buffer (NEB). The N-terminal of this insect cell specific vector contains a GP67 secretion signal and a hexa-histidine (His) tag, which is separated from the *Tc*CESP construct by a TEV-protease cleavage site. The resistance marker for this vector was ampicillin.

A tube of chemically competent *E. coli* DH5α was transformed with *rTccesp*/pAcGP67 His TEV N-term DNA construct and grown on Luria Bertani (LB) agar plates supplemented with ampicillin to a final concentration of 50 µg/mL. The plate was grown at 37°C overnight and a single colony was grown in 5 mL LB broth supplemented with ampicillin (50 µg/mL). This culture was used as a starter for a 50 mL LB broth culture. *rTccesp*/pAcGP67 His TEV N-term DNA was isolated using a Qiagen Plasmid Midi Kit (Qiagen, cat. no. 12143).

MLLVNQSHQGFNKEHTSKMVSAIVLYVLLAAAAHSAFAADLGSHHHHHHSSGRENLYFQ
/GSMGDEFSLDHVNAFCELTKQFRHLPIAVGQQLDEAIDEEAKAAKARKDSQKAVNRAD
AAAKKSEDAKKHAQDAKEAAKEAFEAADGAEVHLQAASELATELNGLVETHLSRLEEYL
KDAETKGEDERAQEAAKECTGNATNVTSGTLLESKEKLEKAAGKEDYEELRYDTERAGS
LLDDLKGAQLEVSSLQIKATNAEERATEAAAKAKAATPDLYDLTVGQAKAFCKLTEQLR
GLLDTVSRHNVTVTEEVAKTASSKNRSDEAVKQAESAAGRNSDAAPHAKKAKEAGEEVS
AAADEAQKAHEDSSRVVQDLRLLAKDKLSSYDRFLKNVGERASGKDAQNAAKECTDSAV
DVKSEVVEKLNAAFEESTPGKPFQPVKDLFKKFSDNLKELEGAADVASKSRAKAEAAEK
IVNESAAEAEIAAFDVKKLTLKQVNAFCGLAEQFRGLLNSVKKLEGQATHCATVAAAVK
ARSTETVKRVVAAAGKNASAKELAERAQAVEAEVTAAADKAQEAYAGVSGFVGNLTASR
DEHLLVLEDVITGAKTSASCSCGIRAAKACTAAAEDVTAESLKEAKDTFRQDIPGDHYK
ELSDGAENVSGRLEELRSTAQRVSASRAAAQAAEERLNAIAEEAAA*

Secretion signal His tag TEV cut site NcoI NotI
N-X-ST predicted N-glycosylation
Cysteines = 9

Figure 3.3. Expected amino acid sequence of the r*Tc*CESP construct. The NcoI (yellow highlight) and NotI (green) sites, the cysteines (gray highlight), TEV cleavage sites (cyan highlight), secretion signal (pink highlight) and the His-tag (blue text) have been shown in different colours in the sequence. There are eight predicted N-linked glycosylation sites.

3.2.2 Transfection and amplification of viral titre

The transfection, amplification and small-scale protein expressions and infection of cells for large-scale protein production were carried out by Bianca Loveless in the Boulanger lab. First 2.5 mL Sf9 of insect cells at $2x10^6$ cells/mL were added to a single well of a 6- well tissue culture plate and then incubated at 27°C without shaking for 1 hour to facilitate adherence. The Sapphire™ baculovirus (Orbigen) was used for the transfection of the Sf9 cells by adding *rTccesp/*pAcGP67 His TEV N-term DNA to Sapphire Baculovirus DNA in a 5:1 ratio, with Fugene HD (Fugene, Fisher Scientific cat. no. PRE2311) as a transfection reagent. After seven hours of transfection the supernatant was contained and the cells were discarded. This supernatant contained the P1 virus. Since the viral titer is usually very low for large-scale infections, an amplification step is

necessary. In order to amplify the virus and obtain a higher viral titer, 70 µL of P1 virus was added to 40 mL of Sf9 cells split at $1.5x10^6$ cells/mL, and then incubated at 27°C with and 140 rpm for 10 days. The cells were discarded and the supernatant was then harvested. This supernatant contained the P2 virus, which were the next generation after amplification.

3.2.3. Small scale protein expression and purification

In order to establish the optimum amount of virus needed for maximum protein expression, different volumes of P2 were added to 2.5 mL of High Five™ cells (Invitrogen, cat. no. BTI-TN-5B1-4) at $2.0x10^6$ cells/mL in a 6-well tissue culture plate. A final concentration of 300 ng/mL tunicamycin was added to prevent N-linked glycosylation. Glycosylation The plate was incubated for 65 hours at 27°C and 140 rpm. After the incubation period, 10 µL of cells was taken and stained using trypan blue to visualize and count the cells. The media were harvested from the wells of the plate and were centrifuged (4000 rpm, 4 min, 4°C). Cells were discarded and 1.8 mL of the supernatant was taken. 200 µL of 10x binding buffer (200 mM HEPES pH 8.0, 1.5 M NaCl, 200 mM imidazole) and 30 µL bed volume of nickel chelated sepharose beads were added to the supernatant. The tubes were incubated on a Lab-rocker for one hour at 4°C and then the beads were harvested by centrifugation, and washed with binding buffer. A sample from the wash fraction was taken and run on an SDS gel. The bound r*Tc*CESP was eluted by adding stripping buffer (20 mM HEPES pH 8.0, 150 mM NaCl, 500 mM imidazole) to the beads and incubating the tube on ice for 3 min. This represented the purified fraction. Some of the supernatant from the beads was run on the gel as the soluble fraction. All the different fractions were analyzed by PAGE and were run on a 15% SDS-gel.

3.2.4. Large scale protein expression and purification

Protein expression was carried out by adding 5 µL of P2 virus per 2.5 mL of High Five™ insect cells at a density of $2.6x10^6$ cells/mL. The cells were grown in Sf900III media and supplemented with gentamycin and tunicamycin was added to a final concentration of 300 ng/mL. The flasks were incubated for 65 hours at 27°C (140 rpm).

After the incubation, cells were visualized and stained using trypan blue as described above. In order to harvest the protein, the culture was centrifuged (JLA 9.1 rotor, 4500 rpm, 15 min, 4°C) and the cells were discarded. 60 µL of protease inhibitor cocktail (Calbiochem, cat. no. 539134) was added to the supernatant and then the supernatant was filtered on ice using 1 and 0.45 µm glass fibre filter (Pall Corporation, cat. no. 66215 and HVLP04700, respectively). Concentration and buffer exchange of the supernatant was done using a diafilteration system equipped with membrane filters that had a 10 kDa molecular weight cut off (Pall Corporation, cat. no. 32037044R). The supernatant was buffer exchanged into binding buffer and was concentrated to a volume of approximately 400 mL. This buffer exchanged supernatant was then centrifuged (JA 14 rotor, 13000 rpm, 30 minutes, 4°C) and the debris was removed. A 2 mL bed volume of nickel chelated sepharose beads was added and the mixture was incubated with a magnetic stirrer at 4°C for 30 minutes. The subsequent steps are as described for r*Tc*CISSA in the previous chapter.

The elutions were pooled and concentrated to 2 mL and the sample was run on a Superdex 200 size exclusion column (HiLoad 16/60 Superdex 200; GE Healthcare, cat. no. 17-1069-01). The selected fractions from the size exclusion were analyzed by SDS-PAGE. In order to further purify the protein sample, the fractions from the size exclusion step were pooled and concentrated and buffer exchanged into the A1 buffer for anion exchange (20 mM HEPES pH 8.0, 10 mM NaCl). The sample was concentrated to 5 mL and was run on an anion exchange column (Source 30Q; GE Healthcare, cat. no. 17-1275-10). The bound r*Tc*CESP was eluted off the column using an increasing B1 buffer (20 mM HEPES pH 8.0, 1 M NaCl). Fractions were analyzed by SDS-PAGE.

3.2.5. Crystallization trials

Initial crystallization trials were carried out using similar protocols as described for *Tc*CISSA (Chapter 2). However, crystallization was performed using the following screens: Index, MCSG-1 and Crystal I and II.

3.2.6. Circular dichroism

We performed a circular dichroism experiment to validate the secondary structure of r*Tc*CESP biochemically. CD spectra were taken using a Jasco-810 spectropolarimeter (Ausio lab, University of Victoria). CESP was dissolved in a low salt buffer (10 mM

NaCl, 20 mM HEPES pH 8.00). Three measurements were taken between 185-240 nm and the values were averaged and plotted using Microsoft Excel.

3.2.7. CBDPS crosslinking

Crosslinking analysis was performed in collaboration with Jason Serpa from the Borcher lab (University of Victoria) using CBDPS-H8/D8 (Cyanurbiotindimercaptopropionylsuccinimide Creative Molecules Inc.), which is an isotopically coded CID-cleavable affinity-purifiable crosslinker (112). CESP in HBS was crosslinked using 10 μM CBDPS-H8/D8. Samples were incubated at 25°C for 30 minutes. Crosslinking was quenched by adding 25 μM ammonium bicarbonate and incubated for 30 minutes at 25°C. The crosslinked samples were then digested with trypsin and proteinase K at 1:25 and 1:10 (w:w) enzyme:substrate ratios, respectively. Both digests were overnight at 37°C, with shaking at 300 rpm. Trypsin and proteinase K were inhibited by adding AEBSF at a final concentration of 10 mM. The crosslinked peptides were enriched on monomeric avidin beads (Thermo Scientific, Rockford, IL) and were eluted using 0.1% TFA 50% acetonitrile. The peptides were then concentrated by lyophilization. A nano-HPLC system (Easy-nLCII, ThermoFisher Scientific, Bremen, Germany) was used for the Mass Spectrometric analysis. This system was coupled to the ESI-source of an LTQ Orbitrap Velos mass spectrometer (ThermoFisher Scientific, Bremen, Germany). MS data were acquired with Mass Tags and Dynamic Exclusion enabled in global data dependent settings ((mass deltas: 4.02511, 2.68340, 2.01255; ratio range (%): 50-100). MS scans (*m/z* 400-2000 range) and MS/MS scans were acquired at 60000 and 30000 resolution respectively. MSMS fragmentation was performed by CID activation at normalized collision energy of 35%. Data analysis was performed using DXMSMS Match of ICC-CLASS.

3.2.8. Homology modelling

*Tc*CESP consists of 3 internally repeated domains (residues 26-228 (d1), 238-438 (d2), 450-650(d3)), each sharing about 30% sequence identity (d1-d2 ~31%, d2-d3 ~34.7%, d1-d3 ~27.2%). A PSI-BLAST (86) search against PDB (113) was performed to identify closest structural homologue. Since the search did not reveal any structural template that

could be used to model full sequence, the three domains were modelled individually using I-TASSER (114). I-TASSER used *Tc*GARP from *Trypanosoma congolense* (PDB: 2Y44) (79), as the template for modelling all the 3 domains. 2Y44 spans almost the entire length of the individual domains and shares 25.5%, 25.5%, and 22.4% sequence identity with d1, d2 and d3 domains, respectively. The fit of the sequence to the modelled structure was verified using ProSa (115).

3.3. Results

3.3.1. *Tc*CESP is predicted to adopt three helical domains

Previous studies in our lab showed that the amino acid sequence of *Tc*CESP contains three sequence repeats. Each of these repeats share approximately 30% sequence identity with *Tc*GARP. When the whole sequence of *Tc*CESP was submitted to the one-to-one threading tool from the Phyre server (116) for tertiary structure prediction, no results were obtained. However, individually submitting the three repeats of *Tc*CESP to the server yielded three structures similar to *Tc*GARP. Figure 3.4 shows the predicted tertiary structure of each of the three repeats of *Tc*CESP.

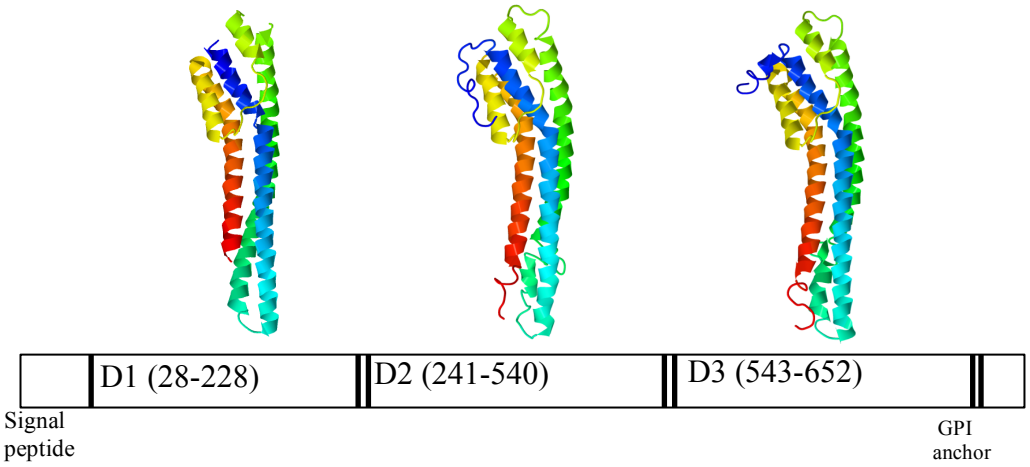

Figure 3.4. Prediction of Tertiary structure of the three repeats of *Tc*CESP using Phyre (116). The box shows the sequence of CESP, the location of the signal peptide, the repeat domains (D1-D3), and the GPI anchor. After modeling each domain by Phyre2, using the GARP structure as template, each domain adapted the same fold as GARP. The modeled structure of each repeat domain is shown above each repeat, coloured blue to red from the N to C-terminus.

3.3.2. Recombinant protein production

The complexity of r*Tc*CESP necessitated the use of a more advanced expression platform than *E. coli*. Consequently, r*Tc*CESP expression studies were carried out in High Five™ insect cells and the protein was secreted in the culture medium. In order to prevent N-linked glycosylation (117) of r*Tc*CESP, the protein was expressed in the presence of tunicamycin, which is a potent inhibitor of N-linked glycosylation in eukaryotes. Glycosylation adds a degree of heterogeneity to the protein sample and prevents crystallization. A volume of 2.8 L of insect cells was harvested and after concentrating and buffer exchanging the culture medium the protein was harvested using Ni-chelated beads and the elutions were pooled and concentrated down to 2 mL. The protein was purified on the size exclusion column and it eluted almost exclusively as a monomer (Figure 3.5). After pooling the Sx200 fractions and buffer exchanging, the protein was applied to an Anion exchange (Figure 3.6).

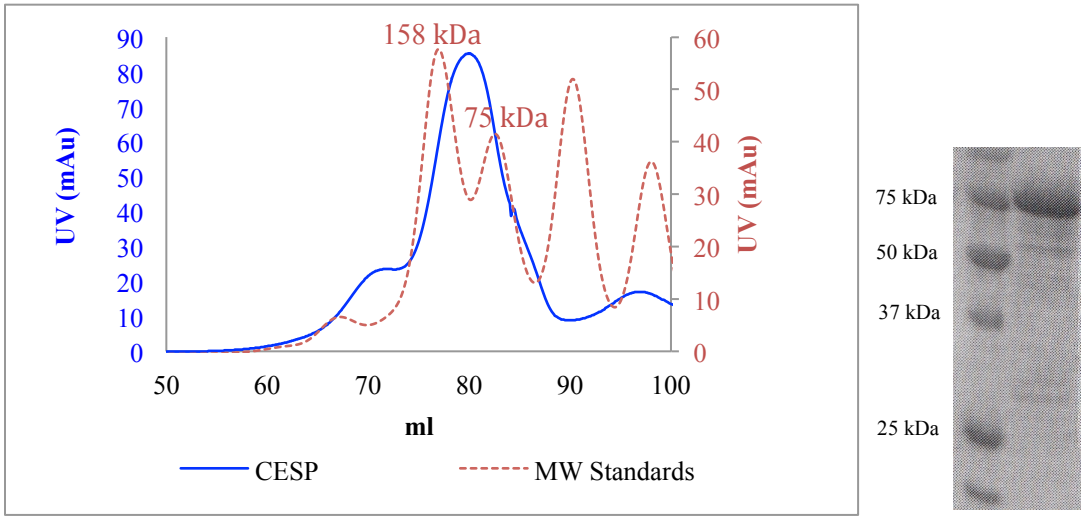

Figure 3.5. Size exclusion chromatography of r*Tc*CESP. The dotted red and the secondary y axis (red) curve represent the molecular weight standards for Superdex 200 column. r*Tc*CESP is plotted in blue. The gel picture at the right hand side shows the pooled fractions from monomeric r*Tc*CESP and the contaminants.

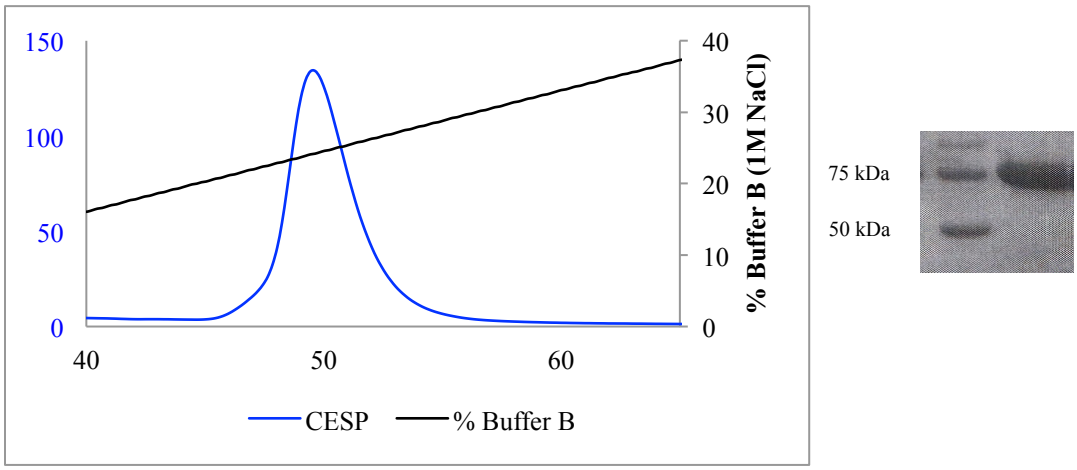

Figure 3.6. Anion exchange chromatography of r*Tc*CESP. The black line and the secondary y axis show the percentage of buffer B (1M NaCl). r*Tc*CESP is plotted in blue. The protein bound to the column was eluted with ~ 300 mM NaCl. The gel at the right hand side confirms the removal of the contaminants and confirms the expected size.

3.3.3. Building a model for *Tc*CESP

Despite setting numerous crystallization trials with the highly purified r*Tc*CESP, no crystals of diffraction quality were obtained. Thus characterizing the overall architecture of *Tc*CESP required the alternative approach of combining circular dichroism spectroscopy, advanced homology modelling and chemical cross linking with mass spectrometry.

The first step was to obtain experimental data confirming that the secondary structure of *Tc*CESP is α-helical. An accurate method for determining the secondary structure of proteins is the spectroscopic technique of circular dichroism (CD), which takes advantage of the unique spectroscopic signatures of α-helices, β-sheets and random coils. As observed in Figure 3.7. A, the CD spectra for *Tc*CESP are consistent with a protein that is predominantly α-helical.

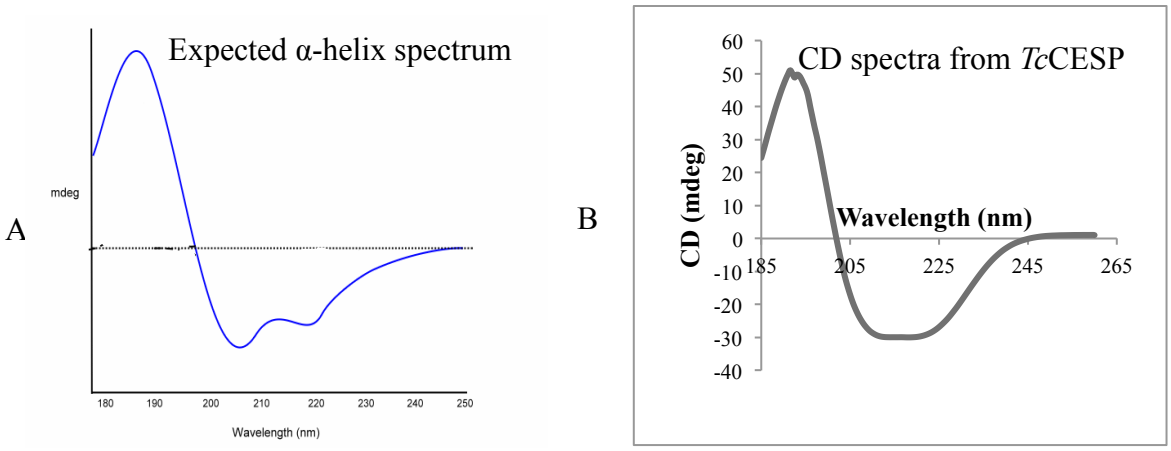

Figure 3.7. A. Expected CD spectrum from an α-helical protein. B: the CD spectrum from r*Tc*CESP. A: The α-helix spectrum contains two valleys around 220 nm and 208 nm and a peak around 190 nm. Figure adapted from http://www.proteinchemist.com/cd/cdspec.htmL **B:** This spectrum was an average of three readings of a 2 mg/mL solution of *Tc*CESP in 10 mM NaCl in pH 8.00 (20 mM HEPES). As the figure shows, the peak obtained from r*Tc*CESP clearly shows that it is an α-helical protein.

The obtain experimental data describing the spatial relationships within the protein, we pursued a CBDPS crosslinking approach. When combined with mass spectrometry, this approach provides relative spatial information between the primary amines of lysine residues and allows for the mapping of disulphide bonds. After the ideal concentration of CBDPS for crosslinking was determined, r*Tc*CESP was treated with appropriate amount of the crosslinker and digested and the peptides were identified using LC-MS/MS. Figure 3.7 shows the CBDPS titration gel and Table 3.1 shows the confirmed crosslinks from r*Tc*CESP. Thirty one out of 63 lysines from the construct were identified in the crosslinks. The cross linking data provided a platform from which to assemble the larger spatial relationships between the three helical bundle domains. Figure 3.8 shows how the distances between the primary amines of the crosslinked lysines were measured using PyMol.

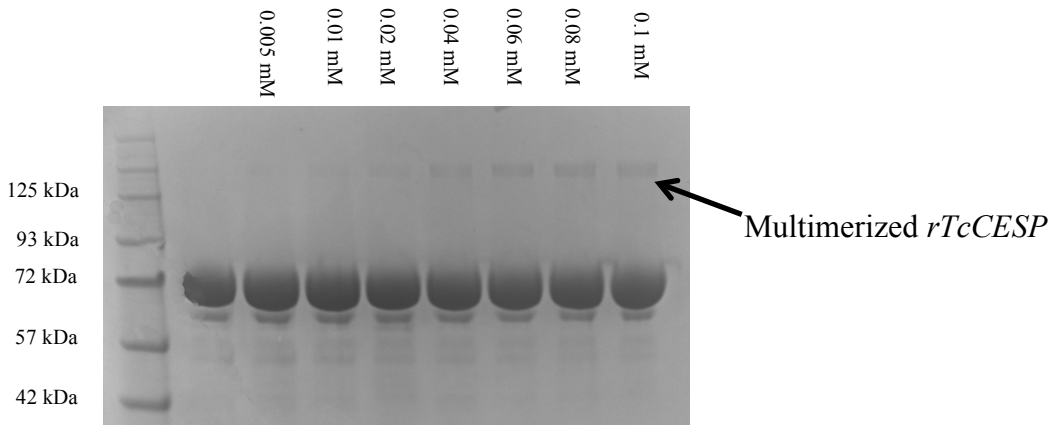

Figure 3.8. CBDPS titration. The concentration of CBDPS added to the purified r*Tc*CESP is written on each lane. The first lane has no CBDPS added. As the concentration increases a population of multimerized *Tc*CESP appears. The ideal concentration of CBDPS is the concentration in which there is no multimer or higher

Table 3.1. List of the crosslinked lysines in r*Tc*CESP crosslinked by CBDPS. The domains represent the three repeats in the TcCESP sequence. Most of the crosslinked lysines were between domains 1 and 2. Only three were between domains 2 and 3 and none between domains 1 and 3.

lysine 1	lysine 2	Crosslinked domain	lysine 1	lysine 2	Crosslinked domain
60	87	1-1	87	402	1-2
63	70	1-1	87	403	1-2
66	70	1-1	93	337	1-2
66	228	1-1	93	339	1-2
66	348	1-2	93	403	1-2
70	81	1-1	175	302	1-2
70	228	1-1	228	348	1-2
80	93	1-1	272	430	2-2
81	87	1-1	282	303	2-2
81	228	1-1	285	448	2-2
86	87	1-1	285	576	2-3
86	93	1-1	303	430	2-2
86	226	1-1	320	424	2-2
86	337	1-2	339	402	2-2
86	339	1-2	448	576	2-3
86	403	1-2	452	576	2-3
87	339	1-2	497	505	2-2
87	392	1-2			

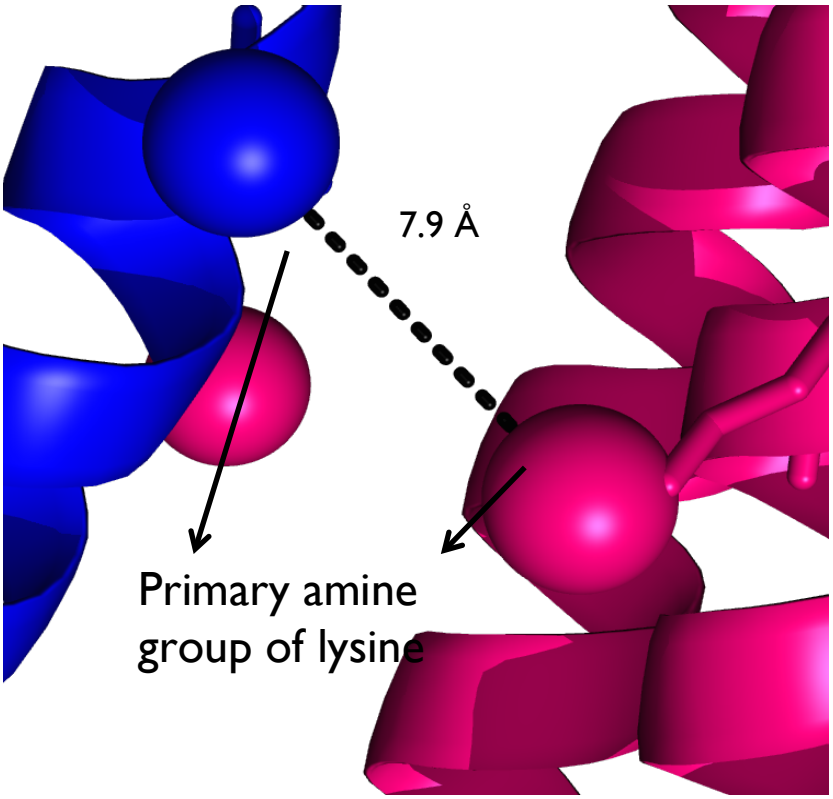

Figure 3.9. Measuring the distance between the terminal amines of lysines. The terminal amine groups are shown as spheres and the distance between the amine groups of these two lysines has been shown (7.9 Å). This measurement was carried out using PyMol.

3.4. Discussion

3.4.1. Selecting an optimal expression system

Previously, efforts were made to express *Tc*CESP in *E. coli.* Since *Tc*CESP is a large eukaryotic protein (80 kDa), there was minimal success with the expression trials with the majority of protein being in the insoluble fraction. The ability to produce r*Tc*CESP using the insect cell expression system provided the opportunity to study this intriguing protein. *Tc*CESP is a GPI-anchored protein therefore the chosen vector for this construct was one with a secretion signal enabling the protein to be secreted into the medium.

There are eight predicted glycosylation sites on *Tc*CESP. Protein glycosylation is not desirable for crystallization, because it adds chemical and conformational heterogeneity to the protein and a heterogeneous population of proteins would either not crystallize, or produce crystals with low diffraction quality. Tunicamycin is a potent inhibitor of N-linked-glycosylation (118,119), and it inhibits cell growth. Therefore the ideal amount to be added should be tested in small-scale expression experiments. To inhibit glycosylation, tunicamycin was added to the growth medium of the insect cells.

The reason why the His-tag of r*Tc*CESP was not removed was that buffer for TEV protease requires a reductant to ensure the optimum catalytic activity of the enzyme. In order to keep the disulfide bonds intact, the protein was left un-cut. r*Tc*CESP has more than 600 amino acids and the His-tag is very short (40 amino acids). Thus, it was not expected that this short tag would affect the overall potential for the protein to crystallize.

3.4.2. Triplication of the CESP domains and the avidity

The sequence identity between GARP and each of the CESP domains suggests that these proteins have a common ancestor. The phylogenetic tree reported by Sakurai et al. (Figure 3.2) suggests that GARP and CESP are related, although there is very low bootstrap support. In addition, there is no scale to show how divergent these genes actually are.

Our biochemical and biophysical studies together with the sequence analyses suggests the same evolutionary origin for CESP and GARP. It is possible that the triplication event

in CESP helps create a binding surface and increase the avidity of this protein. This supports the proposed role of CESP as an adhesin. We propose two possible models for CESP, which are described below.

3.4.3. Homology models for *Tc*CESP

Despite the pure protein sample that was properly folded and homogeneous (according to the CD results) we were unable to obtain X-ray diffraction quality crystals for r*Tc*CESP. In addition, due to the size of the protein, we could not pursue NMR experiments for structure determination. We therefore opted for an alternative approach. Preliminary analysis of *Tc*CESP showed the presence of three repeats in its amino acid sequence, each with ~30% sequence identity to GARP. The result from circular dichroism (CD) confirmed it's α-helical nature. Using a variety of chemical crosslinkers, our collaborators (Borchers lab, University of Victoria) were able to provide spatial information for *Tc*CESP that could be used to assemble the subdomains and validate a homology model from the Parkinson Lab (University of Toronto).

Based on the crosslinking data, one plausible model for *Tc*CESP would be the final CESP model 1 shown in figure 3.9. In this model domains 1 and 2 are located antiparallel to each other and domain three, which is attached to the GPI-anchor, is attached to the membrane. Domain 3 is masked in the GARP coat and the domain1-domain2 bundle is completely out side the coat. Another model, model 2, supports the idea that the gene triplication event in *Tc*CESP is responsible for increased avidity and facilitating adhesion. Figure 3.9 shows this model together with the other triple bundle motif structures of GARP, HpHbR and VSG.

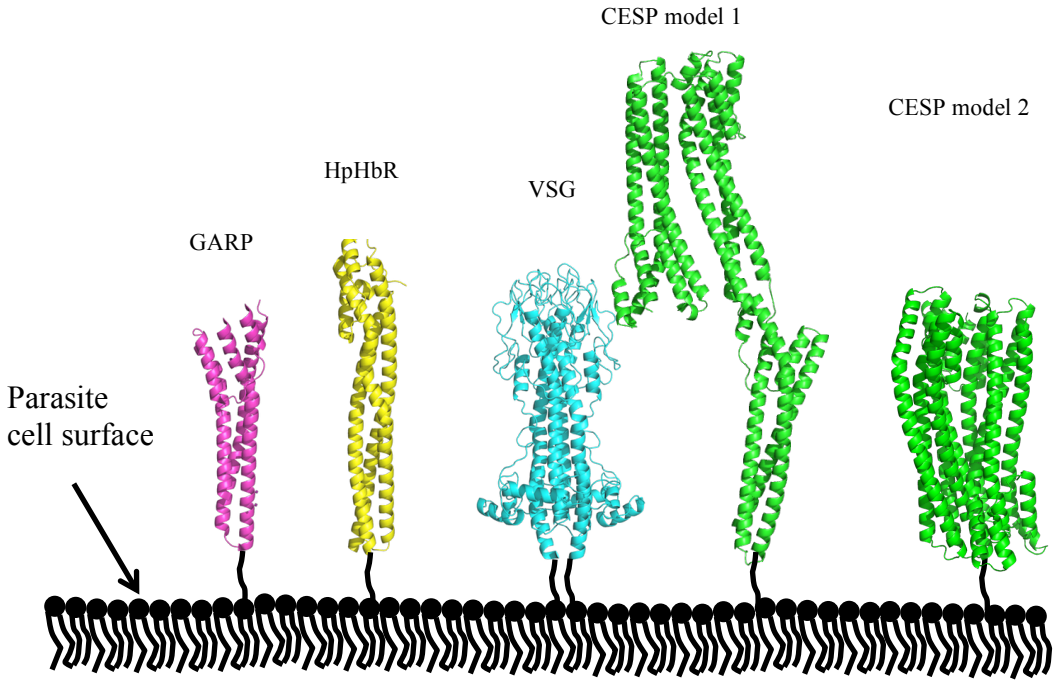

Figure 3.10. Structures of GARP, HpHbR, VSG and the proposed homology models for CESP. All these proteins have the conserved triple helix bundle motif. Domain three of *Tc*CESP is the C-terminal domain and is attached to the membrane by a GPI anchor. Although model 1 corresponds to the crosslinks obtained and identified, model 2 suggests a more stable and more compact structure. It also supports our hypothesis about increased avidity by a gene triplication event.

Model 1 could explain the adhesive properties of *Tc*CESP in two ways. Either the domains interact chemically with the lining of the proboscis, or it might be possible that the elongated *Tc*CESP molecules act as an anchor and cause the attachment, which is a mere mechanical interaction. The linker between domains 2 and 3 can allow the movement of the bundle of domains 1 and 2. These domains are outside of the GARP coat, so they can easily interact with their substrate.

In the case of CESP model 2 there are no crosslinks identified to support this model. However this model is a more stable and compact model compared to model 1. In addition, this model suggests an efficient binding interface at the membrane distal region of the protein. As explained in the results, CESP contains three repeats that could be a

result of a triplication event. This could enhance the avidity of *Tc*CESP by enhancing the number of the binding sites.

Attachment of the epimastigote forms from different species of *Trypanosoma* has been described based on ultrastructure studies (101,105,106,120). The epimastigotes of *T. congolense* and *T. brucei* have the ability to adhere. A BLAST search with the amino acid sequence of CESP as query yielded one protein from *T. brucei.* This protein was about 400 amino acids long and had 28% sequence identity to *Tc*CESP. Interestingly, among the other identical amino acids, there were four conserved cysteines between these proteins. These cysteines covered four of the cysteines that make disulfide bonds in *Tc*CESP. So if this 400 amino acid long putative protein from *T. brucei* is really a counterpart of *Tc*CESP, the difference in size might be related to differences between their functions and between the environments they are expressed.

The counterpart of CESP in *T. brucei* was the only protein that was shown in the results from BLAST and there was no other protein from other species of *Trypanosomes,* or other trypanosomatids. This is not surprising, since *T. brucei* and *T. congolense* are the two species that have the same life cycle and their epimastigote forms show adhesion to the tsetse mouthparts. To our knowledge, *T. vivax* is the only other species with epimastigotes adhering to the mouthparts. It is possible that the CESP counterpart in *T. vivax* has a completely different amino acid from the *T. congolense* and *T. brucei* CESP counterparts, and for this reason it cannot be identified by a BLAST search. The other species of *Trypanosoma* have different vectors and life cycles and it is not very likely for them to have CESP counterparts.

Chapter 4 – Conclusion and future directions

AAT is a disease with devastating socio-economic impact on the sub-Saharan region of Africa. It is caused by species of *Trypanosoma* and is transmitted by the bite of *Glossina ssp* (tsetse fly). *T. congolense* is responsible for most of the diagnosed cases of AAT. Antigenic variation of the bloodstream forms makes vaccine generation impossible and, combined with increasing resistance to the therapeutics, it is necessary to develop alternative strategies for fighting the disease. Vector control is an attractive option as long as it can be strategically targeted and delivered. A detailed understanding of the molecular cross talk between parasite and vector may offer the potential for such targeted control.

The goal of this research is to gain this understanding by studying the surface proteins that express in the insect stage of *T. congolense.* From the surface proteins identified so far, we chose the ones that had a putative function according to previous studies on these proteins. One of these proteins, *Tc*CISSA (congolense insect stage specific antigen), has more than 60% sequence identity to *Tb*PSSA-2 that it is necessary for maturation of parasite infection in the tsetse fly. A construct was designed that contained the N-terminal of *Tc*CISSA, the recombinant protein was crystallized and good quality X-ray diffraction data was obtained from native CISSA crystals. In order to solve the phase problem, many efforts were made and finally a $^{15}N/^{13}C$ labelled sample of r*Tc*CISSA was produced and sent to the Smith Lab at Queen's University. Because of the size of r*Tc*CISSA, the protein needs to be expressed in D_2O to ensure full deuteration of the protein. Since this protein is conserved among other species, the information from this structure might help understand the *Trypanosoma* interactions with its vectors in a more global perspective.

The other protein that was a target of this study was *Tc*CESP (congolense epimastigote specific protein). This protein is a putative adhesin (83) and contains three repeats that are predicted to resemble GARP. A model for *Tc*CESP is proposed based on the structure of GARP and validated by disulfide binds and crosslinking data. This model explains how the domains 1 and 2 are located outside the GARP coat and it provides an overall shape and positioning of this protein on the surface of the parasite. *Tc*CESP is only

expressed in the epimastigotes (75,83) and it is proposed to be responsible for the adhesion in this lifecycle stage. It is expressed in high copy numbers (83). Further experiments to identify the binding partner of *Tc*CESP are needed to confirm the interaction between *Tc*CESP and its substrate.

In conclusion this research represents a step forward towards characterizing two important proteins from the insect stage of *T. congolense.* The structural information from these targets will help us understand the tsetse-*Trypanosoma* interactions.

We have a good dataset from *Tc*CISSA crystals, but have been unable to solve the phase problem. The NMR experiments did not solve the structure because of the size of r*Tc*CISSA and inefficient hydrogen-deuterium exchange. Recently we have been able to produce the SelMet crystals in a different condition by seeding. Further experiments are needed to solve the phase problem for *Tc*CISSA. In addition, a sample of *Tc*CISSA expressed with $^{15}N/^{13}C$ and D_2O can be used for NMR structure determination. Further crystallography attempts including designing new constructs and different mutants with leucines/isoleucines replaced by methionines are also ongoing.

We have presented a homology model for *Tc*CESP that explains a possible function and geometry on the cell membrane. There is no binding partner identified for this protein. We have expressed *Tc*CESP and *Tc*CISSA and other insect stage proteins from trypanosomes and have sent them to the Liverpool School of Tropical Medicine to identify binding partners for them. Combined with the structural information on *Tc*CISSA and *Tc*CESP, finding the binding partner will shed more light on the functions of these proteins.

References

1. Vector borne diseases. in *Health and Environment Linkages Initiative*, World Health Organization, Geneva, Switzerland, retrieved on April 1st, 2014 http://www.who.int/heli/risks/vectors/vector/en/

2. Githeko, A. K., Lindsay, S. W., Confalonieri, U. E., and Patz, J. A. (2000) Climate change and vector-borne diseases: a regional analysis. *Bulletin of the World Health Organization* **78**, 1136-1147

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

4. Riley, W. A. a. J. O. A. (1915) *Handbook of medical entomology*, Comstock publishing company, Ithaca, N. Y.

5. Mukhebi, A., Perry, BD (1992). Economic implications of the control of East Coast fever in eastern, central and southern Africa.

6. Swallow, B. M. (1998) Impacts of African Animal Trypanosomiasis on Human Migration, Livestock and Crop Production. in *24th meeting of the International Scientific Council for Trypanosomosis Research and Control* (Swallow, B. M. ed., OAU/ISCTC, Maputo, Mozambique

7. Simuunza, M., Weir, W., Courcier, E., Tait, A., and Shiels, B. (2011) Epidemiological analysis of tick-borne diseases in Zambia. *Vet Parasitol* **175**, 331-342

8. (SOS), S. o. S. S. Nagana, Animal African trypanosomiasis - a threat to economic development and food security. http://www.stampoutsleepingsickness.com/about-sos-/sleeping-sickness-and-nagana/nagana.aspx

9. Jolly, D., Prentice, I. C., Bonnefille, R., Ballouche, A., Bengo, M., Brenac, P., Buchet, G., Burney, D., Cazet, J.-P., Cheddadi, R., Edorh, T., Elenga, H., Elmoutaki, S., Guiot, J., Laarif, F., Lamb, H., Lezine, A.-M., Maley, J., Mbenza, M., Peyron, O., Reille, M., Reynaud-Farrera, I., Riollet, G., Ritchie, J. C., Roche, E., Scott, L., Ssemmanda, I., Straka, H., Umer, M., Van Campo, E., Vilimumbalo, S., Vincens, A. and Waller, M. (1998) Biome reconstruction from pollen and plant macrofossil data for Africa and the Arabian peninsula at 0 and 6000 years. *Journal of Biogeography* **25**, 21

10. Leach, T. M., and Roberts, C. J. (1981) Present status of chemotherapy and chemoprophylaxis of animal trypanosomiasis in the Eastern hemisphere. *Pharmacol Ther* **13**, 91-147

11. Mulligan, H. W. (1970) *The African Trypanosomiases*, George Allen & Unwin, London

12. Geerts, S. a. H., P.H. (1998) Drug Management and Parasite Resistance in Bovine Trypanosomiasis in Africa. in *PAAT Technical and Scientific Series No. 1.*, FAO, Rome

13. Holmes P.H., E., M.C. and Geerts, S. (2004) Current Chemotherapy of Animal Trypanosomiasis. in *The Trypanosomiases* (Maudlin, I., Holmes, P.H., Miles, M.A. ed.), CABI Publication, Oxford, UK.

14. Lukes, J., Guilbride, D. L., Votypka, J., Zikova, A., Benne, R., and Englund, P. T. (2002) Kinetoplast DNA network: evolution of an improbable structure. *Eukaryot Cell* **1**, 495-502

15. Simpson, L. (1999) RNA editing-an evolutionary perpective. in *The RNA world* (Gesteland, F., Cech, T. R. and Atkins, J. F. ed.), Cold Spring Harbor Laboratory Press, Cold Spring Harbr, N. Y.

16. Gott, J. M., and Emeson, R. B. (2000) Functions and mechanisms of RNA editing. *Annu Rev Genet* **34**, 499-531

17. Estevez, A. M., and Simpson, L. (1999) Uridine insertion/deletion RNA editing in trypanosome mitochondria--a review. *Gene* **240**, 247-260

18. Osorio, A. L., Madruga, C. R., Desquesnes, M., Soares, C. O., Ribeiro, L. R., and Costa, S. C. (2008) Trypanosoma (Duttonella) vivax: its biology, epidemiology, pathogenesis, and introduction in the New World--a review. *Mem Inst Oswaldo Cruz* **103**, 1-13

19. Hoare, C. A. (1972) *The trypanosomes of mammals. A zoological monograph*, Blackwell Scientific Publications, Oxford, UK

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

21. Simukoko, H., Marcotty, T., Phiri, I., Geysen, D., Vercruysse, J., and Van den Bossche, P. (2007) The comparative role of cattle, goats and pigs in the epidemiology of livestock trypanosomiasis on the plateau of eastern Zambia. *Vet Parasitol* **147**, 231-238

22. Tafese, W., Melaku, A., and Fentahun, T. (2012) Prevalence of bovine trypanosomosis and its vectors in two districts of East Wollega Zone, Ethiopia. *Onderstepoort J Vet Res* **79**, E1-4

23. van den Bossche, P., Akoda, K., Kubi, C., and Marcotty, T. (2006) The transmissibility of Trypanosoma congolense seems to be associated with its level of resistance to isometamidium chloride. *Vet Parasitol* **135**, 365-367

24. Gashumba, J. K., Baker, R. D., and Godfrey, D. G. (1988) Trypanosoma congolense: the distribution of enzymic variants in east and west Africa. *Parasitology* **96**, 475-486

25. Shaw, A. P. M. (2004) Economics of African Trypanosomiasis. in *The Trypanosomiases* (Maudlin, I., Holmes, P.H., Miles, M.A. ed.), CABI Publications, Oxford, UK. pp 369-402

26. Leak, S. G. A. (1999) *Tsetse biology and ecology: their role in the epidemiology and control of trypanosomosis.*, Cabi, Wallingford, UK

27. Elsen, P., Amoudi, M. A., and Leclercq, M. (1990) First record of Glossina fuscipes fuscipes Newstead, 1910 and Glossina morsitans submorsitans Newstead, 1910 in southwestern Saudi Arabia. *Ann Soc Belg Med Trop* **70**, 281-287

28. Molyneux, D. H., Pentreath, V., Doua, F. (1996) African trypanosomiasis in man. in *Manson's Tropical Diseases* (C., C. G. ed.), 20 Ed., W. B. Saunders Company Ltd, London. pp 1171-1196

29. Krafsur, E. S. (2009) Tsetse flies: genetics, evolution, and role as vectors. *Infect Genet Evol* **9**, 124-141

30. Despommier, G., Hotez, K. (2006) *Parasitic diseases*, 5 ed., Apple trees production, New York, NY

31. de Raadt, P. (2005) The history of sleeping sickness. World Health Organization, http://www.who.int/trypanosomiasis_african/country/history/en/print.html

32. Cox, F. E. (2004) History of sleeping sickness (African trypanosomiasis). *Infect Dis Clin North Am* **18**, 231-245

33. Caljon, G., De Vooght, L., and Van Den Abbeele, J. (2014) The Biology of Tsetse–Trypanosome Interactions. in *Trypanosomes and Trypanosomiasis* (Magez S., R. M. ed.), Springer-Verlag, Wien.

34. Lehane, M. J. (1999) *The biology of blood-sucking in insects.*, Cambridge University Press, Cambridge

35. Wint W., R. D. Predicted tsetse distribution. http://ergodd.zoo.ox.ac.uk/tseweb/all_species.htm

36. Allsopp, R. a. H., B. H. (2004) The Trypanosomiases. (Maudlin, I., Holmes, P.H., Miles, M.A. ed.), CABI Publication, Oxford, UK.

37. Ormerod, W. E. (1976) Ecological effect of control of African trypanosomiasis. *Science* **191**, 815-821

38. Linear, M. (1982) Gift of poison – the unacceptable face of development aid. *Ambio* **11**, 2-8

39. Linear, M. (1985) Billion-dollar war on tsetse fly could be disastrous for Africa. Sunday times

40. Carson, R. (1962) *Silent Spring*, Houghton Mifflin, Cambridge, Massachusetts

41. Whiteside, E. F. (1949) An experiment in control of tsetse with DDT-treated oxen. *Bull Entomol Res* **40**, 123-134

42. Harris, R. H. T. P. (1932) Some facts and figures regarding the attempted control of Glossina pallidipes in Zululand. . *South African Journal of Science* **29**, 495-507

43. Williamson, D. L., Baumgartner, H.H., Mtuya, A.G., Warner, P.V., Tarimo, S.A. and Dame, D.A. (1983) Integration of insect sterility and insecticides for control of Glossina morsitans morsitans Westwood (Diptera: Glossinidae) in Tanzania. I. Production of tsetse flies for release. *Bulletin of Entomological Research* **73**, 259-265

44. Williamson, D. L., Baumgartner, H.H., Mtuya, A.G., Warner, P.V., Tarimo, S.A. and Dame, D.A. (1983) Integration of insect sterility and insecticides for control of Glossina morsitans morsitans Westwood (Diptera: Glossinidae) in Tanzania. II. Methods of sterilisation, transportation and release of sterilised males. *Bulletin of Entomological Research*

45. Cuisance, D., Politzar, H., Merot, P., Février Tamboura, I., Bauer, B., Kabore, I. and Fidelier, J. . (1986) La campagne de lutte integrée contre les glossines dans la zone pastorale d'accueil de Sideradougou (Republique de Burkina Faso). in *18th Meeting of the International Scientific Council for Trypanosomiasis Research and Control (ISCTRC)*, OAU/STRC, Harare, Zimbabwe

46. Dame, D. A. a. S., C.H. (1970) The sterile male technique against tsetse flies, Glossina spp. *Bulletin of the Entomological Society of America* **16**, 24-30

47. Peacock, L., Cook, S., Ferris, V., Bailey, M., and Gibson, W. (2012) The life cycle of Trypanosoma (Nannomonas) congolense in the tsetse fly. *Parasites & Vectors* **5**

48. Vickerman, K. (1965) Polymorphism and mitochondrial activity in sleeping sickness trypanosomes. *Nature* **208**, 762-766

49. Baral, T. N. (2010) Immunobiology of African trypanosomes: need of alternative interventions. *J Biomed Biotechnol* **2010**, 389153

50. Schwede, A., Jones, N., Engstler, M., and Carrington, M. (2011) The VSG C-terminal domain is inaccessible to antibodies on live trypanosomes. *Mol Biochem Parasitol* **175**, 201-204

51. Vickerman, K., and Luckins, A. G. (1969) Localization of variable antigens in the surface coat of Trypanosoma brucei using ferritin conjugated antibody. *Nature* **224**, 1125-1126

52. ILRAD. (1984) Tsetse flies—vectors of trypanosomiasis. Nairobi, Kenya

53. Lloyd, L., Johnson, W.B. (1924) The trypanosome infections of tsetse flies in northern Nigeria and a new method of estimation. *Bull Entomol Res* **14**, 265-288

54. Robertson, M. (1913) Notes on the life-history of Trypanosoma gambiense, with a brief reference to the cycles of Trypanosoma nanum and Trypanosoma pecorum in Glossina palpalis. *Phil Trans Roy Soc London* **203**, 161-184

55. Prain, C. J., and Ross, C. A. (1990) Trypanosoma congolense: appearance and distribution of variable antigen types during metacyclic differentiation in vitro. *Parasitology* **100**, 107-113

56. Jefferies, D., Helfrich, M. P., and Molyneux, D. H. (1987) Cibarial infections of Trypanosoma vivax and T. congolense in Glossina. *Parasitol Res* **73**, 289-292

57. Thevenaz, P., and Hecker, H. (1980) Distribution and Attachment of "Trypanosoma-(Nannomonas)-Congolense in the Proximal Part of the Proboscis of Glossina-Morsitans-Morsitans. *Acta Tropica* **37**, 163-175

58. Shaw, J. J., and Lainson, R. (1972) Trypanosoma vivax in Brazil. *Ann Trop Med Parasitol* **66**, 25-32

59. Gardiner, P. R. (1989) Recent studies of the biology of Trypanosoma vivax. *Adv Parasitol* **28**, 229-317

60. Molyneux, D. H., and Ashford, R. W. (1975) Observations on a trypanosomatid flagellate in a flea, Peromyscopsylla silvatica spectabilis. *Ann Parasitol Hum Comp* **50**, 265-274

61. Auty, H. K., Picozzi, K., Malele, I., Torr, S. J., Cleaveland, S., and Welburn, S. (2012) Using molecular data for epidemiological inference: assessing the prevalence of Trypanosoma brucei rhodesiense in tsetse in Serengeti, Tanzania. *PLoS Negl Trop Dis* **6**, e1501

62. Jamonneau, V., Ravel, S., Koffi, M., Kaba, D., Zeze, D. G., Ndri, L., Sane, B., Coulibaly, B., Cuny, G., and Solano, P. (2004) Mixed infections of trypanosomes in tsetse and pigs and their epidemiological significance in a sleeping sickness focus of Cote d'Ivoire. *Parasitology* **129**, 693-702

63. Dyer, N. A., Rose, C., Ejeh, N. O., and Acosta-Serrano, A. (2013) Flying tryps: survival and maturation of trypanosomes in tsetse flies. *Trends in parasitology* **29**, 188-196

64. Weiss, B., and Aksoy, S. (2011) Microbiome influences on insect host vector competence. *Trends in parasitology* **27**, 514-522

65. Dale, C., and Welburn, S. C. (2001) The endosymbionts of tsetse flies: manipulating host-parasite interactions. *International journal for parasitology* **31**, 628-631

66. Farikou, O., Njiokou, F., Mbida Mbida, J. A., Njitchouang, G. R., Djeunga, H. N., Asonganyi, T., Simarro, P. P., Cuny, G., and Geiger, A. (2010) Tripartite interactions between tsetse flies, Sodalis glossinidius and trypanosomes--an epidemiological approach in two historical human African trypanosomiasis foci in Cameroon. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases* **10**, 115-121

67. Schauer, R. (2001) The occurrence and significance of sialic acids in insects. *Trends Glycosci. Glycotechnol.* **13**, 507–517

68. Nagamune, K., Acosta-Serrano, A., Uemura, H., Brun, R., Kunz-Renggli, C., Maeda, Y., Ferguson, M. A., and Kinoshita, T. (2004) Surface sialic acids taken from the host allow trypanosome survival in tsetse fly vectors. *The Journal of experimental medicine* **199**, 1445-1450

69. Weiss, B. L., Wang, J., Maltz, M. A., Wu, Y. and S. Aksoy, . (2013) Trypanosome infection establishment in the tsetse fly gut is influenced by microbiome-regulated host immune barriers. . *PLoS pathogens* **9**, e1003318

70. Rose, C., Belmonte, R., Armstrong S. D., Molyneux, G., Haines, L. R., Lehane, M. J., Wastling, J., and A. Acosta-Serrano. (2014) An investigation into the protein composition of the teneral Glossina morsitans morsitans peritrophic matrix. . *PloS Neglected Tropical Diseases* **8**, e2691

71. Lehane, M. J. (1997) Peritrophic matrix structure and function. *Annual review of entomology* **42**, 525-550

72. Lehane, M. J., Aksoy, S., Gibson, W., Kerhornou, A., Berriman, M., Hamilton, J., Soares, M. B., Bonaldo, M. F., Lehane, S., and Hall, N. (2003) Adult midgut expressed sequence tags from the tsetse fly Glossina morsitans morsitans and expression analysis of putative immune response genes. *Genome biology* **4**, R63

73. Hirumi, H., and Hirumi, K. (1991) In vitro cultivation of Trypanosoma congolense bloodstream forms in the absence of feeder cell layers. *Parasitology* **102 Pt 2**, 225-236

74. Urakawa, T., Eshita, Y., Fukuma, T., Hirumi, H., Hirumi, K., and Majiwa, P. A. (1995) Expression of Trypanosoma congolense antigens in Spodoptera frugiperda insect cells. *Exp Parasitol* **80**, 633-644

75. Eyford, B. A., Sakurai, T., Smith, D., Loveless, B., Hertz-Fowler, C., Donelson, J. E., Inoue, N., and Pearson, T. W. (2011) Differential protein expression throughout the life cycle of Trypanosoma congolense, a major parasite of cattle in Africa. *Mol Biochem Parasitol* **177**, 116-125

76. Bayne, R. A., Kilbride, E. A., Lainson, F. A., Tetley, L., and Barry, J. D. (1993) A major surface antigen of procyclic stage Trypanosoma congolense. *Mol Biochem Parasitol* **61**, 295-310

77. Beecroft, R. P., Roditi, I., and Pearson, T. W. (1993) Identification and characterization of an acidic major surface glycoprotein from procyclic stage Trypanosoma congolense. *Mol Biochem Parasitol* **61**, 285-294

78. 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., and Donelson, J. E. (2009) Analysis of expressed sequence tags from the four main developmental stages of Trypanosoma congolense. *Mol Biochem Parasitol* **168**, 34-42

79. Loveless, B. C., Mason, J. W., Sakurai, T., Inoue, N., Razavi, M., Pearson, T. W., and Boulanger, M. J. (2011) Structural characterization and epitope mapping of the glutamic acid/alanine-rich protein from Trypanosoma congolense: defining assembly on the parasite cell surface. *J Biol Chem* **286**, 20658-20665

80. Freymann, D., Down, J., Carrington, M., Roditi, I., Turner, M., and Wiley, D. (1990) 2.9 A resolution structure of the N-terminal domain of a variant surface glycoprotein from Trypanosoma brucei. *J Mol Biol* **216**, 141-160

81. Higgins, M. K., Tkachenko, O., Brown, A., Reed, J., Raper, J., and Carrington, M. (2013) Structure of the trypanosome haptoglobin-hemoglobin receptor and implications for nutrient uptake and innate immunity. *Proc Natl Acad Sci U S A* **110**, 1905-1910

82. Blum, M. L., Down, J. A., Gurnett, A. M., Carrington, M., Turner, M. J., and Wiley, D. C. (1993) A structural motif in the variant surface glycoproteins of Trypanosoma brucei. *Nature* **362**, 603-609

83. Sakurai, T., Sugimoto, C., and Inoue, N. (2008) Identification and molecular characterization of a novel stage-specific surface protein of Trypanosoma congolense epimastigotes. *Mol Biochem Parasitol* **161**, 1-11

84. Fragoso, C. M., Schumann Burkard, G., Oberle, M., Renggli, C. K., Hilzinger, K., and Roditi, I. (2009) PSSA-2, a membrane-spanning phosphoprotein of Trypanosoma brucei, is required for efficient maturation of infection. *PLoS One* **4**, e7074

85. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) Basic local alignment search tool. *J Mol Biol* **215**, 403-410

86. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389-3402

87. Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., and Higgins, D. G. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947-2948

88. Petersen, T. N., Brunak, S., von Heijne, G., and Nielsen, H. (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods* **8**, 785-786

89. Gibson, W. (2012) The origins of the trypanosome genome strains Trypanosoma brucei brucei TREU 927, T. b. gambiense DAL 972, T. vivax Y486 and T. congolense IL3000. *Parasites & vectors* **5**, 71

90. D'Alessandro, A. a. S., N. (1992) Trypanosoma rangeli. in *Parasitic Protozoa* (Kreier, J. P. a. B. J. R. ed.), Academic Press, Sandiego, California.

91. Brinkmann, U., Mattes, R. E., and Buckel, P. (1989) High-level expression of recombinant genes in Escherichia coli is dependent on the availability of the dnaY gene product. *Gene* **85**, 109-114

92. Del Tito, B. J., Jr., Ward, J. M., Hodgson, J., Gershater, C. J., Edwards, H., Wysocki, L. A., Watson, F. A., Sathe, G., and Kane, J. F. (1995) Effects of a minor isoleucyl tRNA on heterologous protein translation in Escherichia coli. *J Bacteriol* **177**, 7086-7091

93. Rosenberg, A., Griffin, K., Studier, W.S., McCormick, M., Berg, J., Novy, R., Mierendorf, R. (1996). *inNovations* **6**, 6

94. Seidel, H. M., Pompliano, D. L., and Knowles, J. R. (1992) Phosphonate biosynthesis: molecular cloning of the gene for phosphoenolpyruvate mutase from

Tetrahymena pyriformis and overexpression of the gene product in Escherichia coli. *Biochemistry* **31**, 2598-2608

95. Derman, A. I., Prinz, W. A., Belin, D., and Beckwith, J. (1993) Mutations that allow disulfide bond formation in the cytoplasm of Escherichia coli. *Science* **262**, 1744-1747

96. Stewart, E. J., Aslund, F., and Beckwith, J. (1998) Disulfide bond formation in the Escherichia coli cytoplasm: an in vivo role reversal for the thioredoxins. *EMBO J* **17**, 5543-5550

97. Prinz, W. A., Aslund, F., Holmgren, A., and Beckwith, J. (1997) The role of the thioredoxin and glutaredoxin pathways in reducing protein disulfide bonds in the Escherichia coli cytoplasm. *J Biol Chem* **272**, 15661-15667

98. Tonkin, M. L., Workman, S. D., Eyford, B. A., Loveless, B. C., Fudge, J. L., Pearson, T. W., and Boulanger, M. J. (2012) Purification, crystallization and X-ray diffraction analysis of Trypanosoma congolense insect-stage surface antigen (TcCISSA). *Acta Crystallogr Sect F Struct Biol Cryst Commun* **68**, 1503-1506

99. Jurnak, F. (1986) Effect of chemical impurities in polyethylene glycol on macromolecular crystallization. *Journal of Crystal Growth* **76**, 6

100. Ray, W. J., Jr., and Puvathingal, J. M. (1985) A simple procedure for removing contaminating aldehydes and peroxides from aqueous solutions of polyethylene glycols and of nonionic detergents that are based on the polyoxyethylene linkage. *Anal Biochem* **146**, 307-312

101. Vickerman, K. (1973) The mode of attachment of Trypanosoma vivax in the proboscis of the tsetse fly Glossina fuscipes: an ultrastructural study of the epimastigote stage of the trypanosome. *J Protozool* **20**, 394-404

102. Hendrickson, W. A., Horton, J. R., and LeMaster, D. M. (1990) Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. *EMBO J* **9**, 1665-1672

103. Cowie, D. B., and Cohen, G. N. (1957) Biosynthesis by Escherichia coli of active altered proteins containing selenium instead of sulfur. *Biochim Biophys Acta* **26**, 252-261

104. Phillips, T. A., VanBogelen, R. A., and Neidhardt, F. C. (1984) lon gene product of Escherichia coli is a heat-shock protein. *J Bacteriol* **159**, 283-287

105. Evans, D. A., Ellis, D. S., and Stamford, S. (1979) Ultrastructural studies of certain aspects of the development of Trypanosoma congolense in Glossina morsitans morsitans. *J Protozool* **26**, 557-563

106. Vickerman, K. (1969) On the surface coat and flagellar adhesion in trypanosomes. *J Cell Sci* **5**, 163-193

107. Evancio, J. (2013) *Trypanosoma brucei BARP crystallography research.* BSc Honours, University of Victoria

108. Eisenhaber, B., Bork, P., Yuan, Y., Loffler, G., and Eisenhaber, F. (2000) Automated annotation of GPI anchor sites: case study C. elegans. *Trends Biochem Sci* **25**, 340-341

109. Eisenhaber, B., Bork, P., and Eisenhaber, F. (1999) Prediction of potential GPI-modification sites in proprotein sequences. *J Mol Biol* **292**, 741-758

110. Sunyaev, S. R., Eisenhaber, F., Rodchenkov, I. V., Eisenhaber, B., Tumanyan, V. G., and Kuznetsov, E. N. (1999) PSIC: profile extraction from sequence alignments with position-specific counts of independent observations. *Protein Eng* **12**, 387-394

111. Eisenhaber, B., Bork, P., and Eisenhaber, F. (1998) Sequence properties of GPI-anchored proteins near the omega-site: constraints for the polypeptide binding site of the putative transamidase. *Protein Eng* **11**, 1155-1161

112. Petrotchenko, E. V., Serpa, J. J., and Borchers, C. H. (2010) Use of a Combination of Isotopically Coded Cross-Linkers and Isotopically Coded N-Terminal Modification Reagents for Selective Identification of Inter-peptide Crosslinks. *Analytical Chemistry* **82**, 817-823

113. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000) The Protein Data Bank. *Nucleic Acids Res* **28**, 235-242

114. Zhang, Y. (2008) I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* **9**, 40

115. Wiederstein, M., Sippl, M. J. (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res* **35**, W407-410

116. Kelley, L. A., and Sternberg, M. J. (2009) Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc* **4**, 363-371

117. Gupta, R., and Brunak, S. (2002) Prediction of glycosylation across the human proteome and the correlation to protein function. *Pac Symp Biocomput*, 310-322

118. Heifetz, A., Keenan, R. W., and Elbein, A. D. (1979) Mechanism of action of tunicamycin on the UDP-GlcNAc:dolichyl-phosphate Glc-NAc-1-phosphate transferase. *Biochemistry* **18**, 2186-2192

119. Takatsuki, A., , Kawamura, K., Okina, M., Kodama, Y., Ito, T., & Tamura, G. (1977) The Structure of Tunicamycin. *Agricultural and Biological Chemistry* **41**, 2307-2309

120. Tetley, L., and Vickerman, K. (1985) Differentiation in Trypanosoma brucei: host-parasite cell junctions and their persistence during acquisition of the variable antigen coat. *J Cell Sci* **74**, 1-19