Structural characterization of *Tb*Fam50, *Tb*PSSA2, and *Tc*CISSA, surface proteins expressed by the trypanosome inside the tsetse vector

by

RAGHAVENDRAN RAMASWAMY

MSc, Nottingham Trent University, 2012

A Thesis Submitted in Partial Fulfillment

of the Requirements for the Degree of

MASTER OF SCIENCE

in Biochemistry and Microbiology

University of Victoria

Supervisory Committee

Structural characterization of *Tb*Fam50, *Tb*PSSA2, and *Tc*CISSA, surface proteins expressed by the trypanosome inside the tsetse vector

by

RAGHAVENDRAN RAMASWAMY

MSc, Nottingham Trent University, 2012

Supervisory Committee

Dr. Martin J. Boulanger, (Department of Biochemistry and Microbiology)

Supervisor

Dr. Alisdair B. Boraston, (Department of Biochemistry and Microbiology)

Departmental Member

Dr. Steve J. Perlman, (Department of Biology)

Outside Member

Abstract

Supervisory Committee
Supervisor
Dr. Martin J. Boulanger, (Department of Biochemistry and Microbiology)
Departmental Member
Dr. Alisdair B. Boraston, (Department of Biochemistry and Microbiology)
Outside Member
Dr. Steve J. Perlman, (Department of Biology)

Vector-borne diseases such as malaria, leishmaniasis, and African trypanosomiasis are a major scourge to humans and animals in some of the most impoverished nations across the globe. Enabling the transmission of these disease-causing pathogens is a highly sophisticated molecular arsenal of surface proteins. My research focuses on biophysical characterization of these proteins with the ultimate goal of deciphering the molecular crosstalk between pathogen and vector. In support of this goal, I have selected the tsetse fly-transmitted parasites of the genus *Trypanosoma*, the etiological agent of African sleeping sickness, as a model system. Towards elucidating the molecular mechanism of transmission, I have attempted to characterize structurally three novel proteins; *Tb*Fam50.360, *Tb*PSSA2, and *Tc*CISSA and get insight into their functions. Before this study, GARP (Glutamic Acid Rich Protein from *T. congolense*), and VSG (Variant Surface Glycoprotein from *T. brucei*) were the only proteins to be structurally characterized in the vector stages of the parasite.

Our structural analysis revealed that while the N- terminal region of *Tb*Fam50.360 adopted a three-helical structure similar to previously characterized trypanosome surface proteins, ectodomains of both *Tb*PSSA2 and *Tc*CISSA adopted a previously uncharacterized bilobed architecture. The structural analysis further identified putative ligand binding regions in *Tb*Fam50.360 and *Tc*CISSA. However, in the absence of binding partners, the exact function of

these proteins could not be established. Our lab in conjunction with our collaborators is investigating the binding partners of these proteins within the tsetse.

The structures of *Tb*Fam50.360, *Tb*PSSA2, and *Tc*CISSA can be added to the repertoire of structurally characterized surface proteins expressed by trypanosomes. The information gained from these first structures of trypanosome surface proteins offer insight into their role in the trypanosome life cycle, and may, in the future, contribute to the control of African trypanosomiasis.

Table of Contents

Supervisory Committee ii

Abstract iii

Table of Contents v

List of Tables vii

List of Figures viii

List of Abbreviations ix

Acknowledgements xii

Chapter 1 : Introduction 1

1.1 Vector Borne Diseases-A global burden 1

1.2 General life cycle of VBDs 3

1.3 Treatment and Control of VBDs 4

1.4 Targeting the surface proteins at the vector-parasite interface 8

1.5 Tsetse-Trypanosome an ideal model system for understanding vector-pathogen interactions 9

1.6 Human and Animal African Trypanosomiasis 10

1.7 Life cycle of trypanosome in tsetse 12

1.8 *Tb*Fam50.360 17

1.9 *Tc*CISSA/*Tb*PSSA2 17

1.10 Research objectives 19

Chapter 2 : Structural characterization of *Tb*Fam50.360 20

2.1 Introduction 20

2.2 Materials and Methods 25

2.3 Results 29

2.4 Discussion 36

Chapter 3 : Structural characterization of *Tb*PSSA2 and *Tc*CISSA 41

3.1 Introduction 41

3.2 Materials and Methods 45

3.3 Results and Discussions 52

Chapter 4 : General discussion and future studies 62

Bibliography 64

List of Tables

Table 1.1: Major vector-borne diseases of humans, and associated aetiological agents and arthropod vectors 2

Table 2.1 Data collection and refinement statistics 27

Table 3.1: Data collection and refinement statistics 49

List of Figures

Figure 1.1: General transmission cycle for vector-borne diseases 4

Figure 1.2: Tsetse fly distribution in sub-Saharan Africa. 12

Figure 1.3: Life cycle of trypanosomes in the tsetse fly. 14

Figure 2.1: Predicted domain architecture of *Tb*Fam50.360. 22

Figure 2.2: Bayesian phylogeny and expression of genes in Fam50 family. 23

Figure 2.3: Structural and functional analysis of *Tb*Fam50.360. 31

Figure 2.4: *Tb*Fam50.360 closely resembles *Tc*GARP. 33

Figure 2.5: Comparison of the structures of *Tb*Fam50.360 and *Tc*GARP. 34

Figure 2.6: Model depicting the *Tb*Fam50.360 family of proteins in the context of the metacyclic stage of the trypanosome.. 38

Figure 3.1: Domain organization and sequence alignment of *Tb*PSSA2 and *Tc*CISSA .. 43

Figure 3.2: Both *Tb*PSSA2 and *Tc*CISSA behave as monomers in solution and adopt a unique bilobed architecture 54

Figure 3.3: *Tb*PSSA2 and *Tc*CISSA display conformational flexibility between their lobes. 56

Figure 3.4: Analysis of the hinge region in *Tc*CISSA 59

List of Abbreviations

1D	One dimensional
AAT	Animal African trypanosomiasis
AMA	Apical membrane antigen
ATP	Adenosine triphosphate
AU	Asymmetric unit
BARP	*Brucei* alanine rich protein
BLAST	Basic local alignment search tool
BSF	Bloodstream form
cDNA	Complementary DNA
CESP	Congolense epimastigote specific protein
CISSA	Congolense insect stage surface antigen
CLS	Canadian Light Source
Da	Daltons
DTT	Dithiothreitol
EDTA	Ethylene diamine tetra acetic acid
EMF	Epimastigote form
EST	Expressed sequence tags
GARP	Glutamic acid/alanine rich protein
GPI	Glycosylphosphatidylinositol
HA	Haemagglutinin
HAT	Human African trypanosomiasis

HDX	Hydrogen-Deuterium Exchange
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HPLC	High performance liquid chromatography
HRP	Heptatpeptide repeat protein
ISG	Invariant surface glycoprotein
iTRAQ	Isobaric tags for relative and absolute quantitation
kDa	KiloDalton
LB	Luria-Bertani
LC	Liquid chromatography
m/z	Mass to charge ratio
mAb	Monoclonal antibody
MALDI	Matrix assisted laser desorption ionization
MAP	Mitogen activated protein
MCF	Metacyclic form
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
PAGE	Polyacrylamide gel electrophoresis
Pb	*Plasmodium berghei*
PBS	Phosphate buffered saline
PCF	Procyclic culture form
PCR	Polymerase chain reaction
PDB	Protein data bank
PF	Procyclic form

PM	Peritrophic matrix
PRS	Protease resistant surface molecule
PSSA	Procyclic stage surface antigen
rmsd	Root means square deviation
SDS	Sodium dodecylsulphate
SSRL	Stanford Synchrotron Radiation Lightsource
Tb	*T. brucei*
Tc	*T. congolense*
Tcr	*Trypanosoma cruzii*
TM	Transmembrane
TRX	Thoredoxin
Tv	*Trypanosoma vivax*
VSG	Variant surface glycoprotein

Acknowledgements

I deem it a great pleasure to place on record my deep sense of gratitude and indebtedness to my research supervisor, Dr. Martin Boulanger, for his support, invaluable guidance and constant encouragement throughout the period of the research work. I am especially thankful to him for having confidence in me, which helped me to overcome many problems during my research work. I am indeed grateful to him for his constant support and full-fledged cooperation.

I am very grateful to my committee members, Dr. Alisdair Boraston and Dr. Steve Perlman for their support, encouragement and feedback during my research.

Very special thanks to Dr. Michelle Parker for her critical feedback which improved my critical thinking. Thanks to all friends and colleagues in Boulanger lab for their support.

Appreciation also goes out to Kaleigh, Crystal, Kevin, Kento, Claudia and Robert for the useful discussions over coffee. Special thanks to Mukundan, Nikhil, Onkar, Chakri, Rahul, Aditya, Anup, Jayaram and Karthik for keeping me motivated through these years.

At last, I would like to thank my parents. I want them to know that I am very grateful for their unreserved love and encouragement throughout my studies, to which has been a source of inspiration and moral support.

Chapter 1 : Introduction

1.1 Vector Borne Diseases-A global burden

Vector-borne diseases (VBD) caused by pathogens transmitted by blood-feeding insects have long impacted human affairs. The great king and conqueror, Alexander the Great, was defeated by the bite of a tiny mosquito vectoring pathogens for malaria. The Black Death that nearly decimated Europe and killed millions worldwide was the work of a tiny flea vectoring the pathogen responsible for bubonic plague from rats to humans. VBDs remain very influential even to this day, suppressing the economies of nations where they remain endemic.

Approximately one-sixth of the illness and disability caused worldwide is due to VBDs, with more than half the world's population currently estimated to be at risk (World Health Organization, 2014b). While there are many different VBDs worldwide (see Table 1.1), malaria, caused by *Plasmodium* and transmitted by Anopheline mosquitoes is predominant. Each year, approximately 225 million people are infected with the malaria parasite and in 2014 around 781,000 of these resulted in disease-induced mortality (World Health Organization, 2014b). Leishmaniasis, which is spread by female sandflies, is also quite deadly, with 12 million infected and 800,000 deaths each year (World Health Organization, 2013). However, because of the promiscuity of trypanosomes, to infect both humans and domestic animals through the bites of tsetse, trypanosomiasis arguably has a great economic impact. Approximately 70 million people and 50 million cattle were infected with the African variety over the last ten years, resulting in an economic loss of ~1-5 billion US dollars (Brun et al., 2010, Simarro et al., 2011, Simarro et al., 2012).

Table 1.1: Major vector-borne diseases of humans, and associated aetiological agents and arthropod vectors

Disease	Pathogen/parasite	Arthropod disease vector
Protozoan diseases		
Malaria	*Plasmodium falciparum, Plasmodium vivax, Plasmodium ovale, Plasmodium malariae*	*Anopheles spp.* (mosquitoes)
Leishmaniasis	*Leishmania spp.*	*Lutzomyia and Phlaebotomus spp.*
Trypanosomiasis	*Trypanosoma brucei* gambiense, *Trypanosoma brucei* rhodiense *Trypanosoma congolense*	*Glossina spp.*
Chagas disease	*Trypanosoma cruzi*	Triatomines
Viral diseases		
Dengue fever	DEN-1, DEN-2, DEN-3, DEN-4 flaviviruses	*Aedes aegypti*
Encephalitis	Flavi-, alpha- and bunyaviruses	Various mosquito and tick species
Yellow fever	Yellow fever *flavivirus*	*Aedes aegypti*
Filarial nematodes		
Lymphatic filariaisis	*Brugia malayi, Brugia timori, Wuchereria bancrofti*	*Anopheles, Culex, Aedes and Ochlerotatus* (mosquitoes)
Onchocerciasis	*Onchocerca volvulus*	*Simulum spp*.

VBDs have the greatest impact on poorly developed tropical countries, where a combination of optimal vector habitat and a lack of proper medical care can lead to large-scale epidemics. They affect rural, urban communities, but thrive predominantly among communities with poor living conditions such as lack of access to adequate housing, safe drinking water, and sanitation. Malnourished people and those with weakened immunity are especially vulnerable. These diseases also exacerbate poverty by preventing people from working and supporting themselves and their family, causing further hardship and impeding economic development (World Health Organization, 2014a, World Health Organization, 2014b). Moreover, globalization has made these diseases not just restricted to the tropics; they are spreading in the developed world and present a major global health and economic threat.

1.2 General life cycle of VBDs

Hosts can be infected either by biological or mechanical transmission of the pathogen. Mechanical transmission involves a simple transfer of pathogens from the mouth parts or other body parts of the vector to hosts. The vector, thus merely acts as a carrier of the pathogen and the pathogen does not undergo developmental changes in the vector (Greenberg, 1973). Though mechanical transmission is a public health concern, biological transmission is the most significant mode of transmission involving multiplication and/or development of the pathogen inside the vector. During a biological transmission, the vector takes a blood meal from an infected host and pathogens from this host enter the vector. Inside the vector, the pathogens develop and/or multiply. After an incubation period (varying for different pathogens), they are transmitted to a healthy host when the vector takes another blood meal (illustrated in Fig. 1.1). All major VBDs including malaria, African trypanosomiasis, and leishmaniasis are biologically transmitted. Therefore, due to the importance of biological transmission, this thesis will discuss biological transmission.

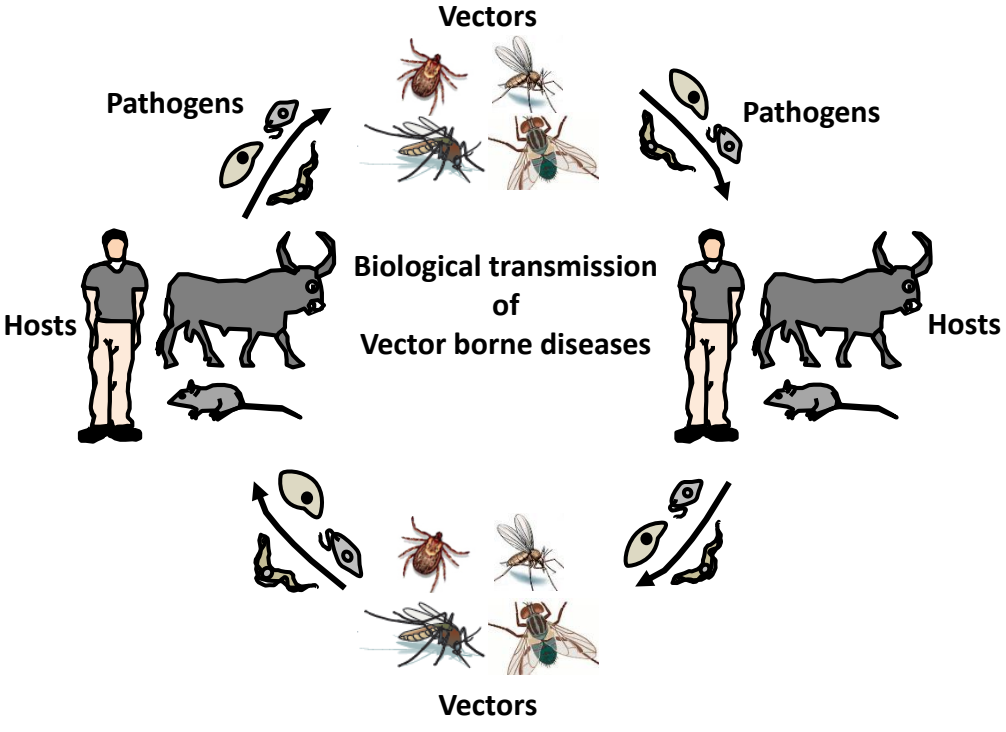

Figure 1.1: General transmission cycle for vector-borne diseases. The biological transmission of VBDs, involving insect vectors taking a blood meal from one host (infected host) and transmitting pathogens to other (healthy) hosts that become infected (Images of vectors copied from McGraw and Neill, 2013)

1.3 Treatment and Control of VBDs

While there are a number of methods for preventing and treating VBDs, there is still a major push to identify novel targets for vaccines and drugs. Current vaccines are generally not effective and drugs have severe side-effects or become virtually useless as the parasites develop resistance. Moreover, these drugs are too expensive for widespread use.

Treatment

To date, drugs remain the most effective treatment against VBDs in spite of causing severe side effects and challenges by contributing to the potential rise in pathogen resistance. Chloroquine derivatives (target the parasite's ability to detoxify heme) and artemisinin (thought to release free radicals causing disruption of parasite membrane) has been used successfully to combat malaria (De Vries and Dien, 1996; Mutabingwa, 2005; Nosten and White, 2007). Similarly, ivermectin

and Sodium stibogluconate have been effective against filariasis and leishmaniasis (Liu and Weller, 1996; Croft, 1997; Goodwin, 1984). Trypanicidal drugs such as eflornithine (inhibits ornithine decarboxylase and prevents spermidine synthesis) or benzidanazole (binds tubulin and prevents the uptake of glucose by parasites) can provide effective treatment against trypanosomiasis (Delespaux and de Koning, 2007, Burri, 2010). However, these drugs have severe side effects such as neurotoxicity, GI-tract disorders, and seizures that greatly limits their widespread use (White, 1985, Bouteille et al., 2003).

Control/Prevention of VBDs

Vaccination is an important strategy to control any infectious disease. However, they were a huge failure against VBDs with vaccines against yellow fever being the only exception. Current vaccines target pathogen-associated antigens expressed during host stage of their life cycle. Targeting the host specific antigens have proven difficult, owing to the complexities of different life cycle stages of the pathogen and our limited comprehension of the human immune response. One of the greatest challenges in developing a blood-stage vaccine is overcoming antigenic diversity. Most of the antigens presented by the pathogens show substantial polymorphism that facilitates its immune evasion (Volkman et al., 2002; Barry and McCulloch, 2001). Vaccine approaches not only need to account for this diversity, but also cover the majority of strains causing infection and diseases. Since an antigen has several allelic forms, it becomes impossible to incorporate all these forms into a single vaccine. Currently, the vaccines for malaria (RTS S/AS01) (Agnandji et al., 2011) and dengue (Sabchareon et al., 2012) have had mixed results. In the absence of a fully developed vaccine, the best way to control VBDs is by targeting the vector.

Vector control represents an important strategy for controlling VBDs, as the pathogens are powerless without their vectors. Currently, vector control is the only practical option for controlling dengue, Chagas disease, and plays a vital role in preventing malaria. Additionally, there is increasing evidence for vector control in preventing African animal trypanosomiasis and

lymphatic filariasis in several epidemiological settings (Allsopp, 2001). For decades, the primary method of vector control involved the use of organochlorine (DDT), organophosphate (Malathion), carbamate (Carbaryl), and pyrethroids (Deltamethrin) based insecticides. They reduced levels of transmission of dengue, leishmania, and filariasis in many parts of the world. Some countries such as Taiwan, are now celebrating 50 transmission free years of malaria (Yip, 2000). Recently, the use of insecticide-treated bed nets have been instrumental in reducing morbidity and mortality of VBDs. For example, in Africa, the use of insecticide-treated bed nets resulted in a dramatic decrease in parasitemia in young children by 62% and an increase in child survival by 27% (Schellenberg et al., 2001).

Although effective, insecticides have several limitations leading to their restricted use. In many countries of Asia, Africa, and South America, there is an emergence and spread of pyrethroid (insecticides targeting *Anopheles* mosquitoes) resistance (World Health Organization, 1992). There were also reports of organophosphate and carbamate resistance (broad-spectrum insecticides). Furthermore, the use of insecticides has adverse effects on human health. For example, DDT has been banned in many countries (United Nations, 1991) as exposure of low to moderate levels of DDT may cause nausea, increase liver enzyme activity, disrupt endocrine signaling and can be carcinogenic (Longnecker et al., 1997). Moreover, DDT also causes major effects in other organisms such as birds (thinning of eggshells and difficulties in egg hatching). The cost of insecticides has also been an important limiting factor. Therefore, financial burden, health risk and the evolution of resistance in insects have largely undermined the widespread application of insecticides.

Another way of targeting the vector is a method called sterile insect technique (SIT). This process involves mass rearing and release of male insects (mosquitoes or tsetse) that are sterilized by irradiation (Alphey et al., 2010). The released vectors mate with wild female vectors causing a

population fall. This technique has been used efficiently to eliminate the tsetse population in areas of Zanzibar and Nigeria (Vreysen et al., 2000; Politzar and Cuisance, 1984). Furthermore, this technique was used successfully to control *Anopheles* in El Salvador (Lofgren et al., 1974). Use of SIT is limited, despite its advantages. SIT requires the breeding of large numbers of insects before release, which is difficult. Vectors (males) often show negative fitness after irradiation making them incompetent for mating as compared to wild-type male vectors (Yakob et al., 2008). Furthermore, SIT does not provide a permanent elimination of vector population and creates a niche that can be re-colonized by immigrants (Thomé et al., 2010).

Emerging technologies for vector control

Genetic modification of the vector

Genetic modification, though still in its infancy, is aimed at making the vector recalcitrant to disease transmission. There are 2 main approaches. The first approach, also known as the release of insects with death lethal allele (RIDL), involves inducing mutations in such a way that the daughters of released males are either unable to fly or die as pupae (Phuc et al., 2007; Fu et al., 2010; Wise de Valdez et al., 2011). RIDL is currently trialled by Oxitec in Brazil and Malaysia (Lacroix et al., 2012). The second approach attempts to improve the defense system of vector. This approach uses RNA interference (RNAi) that recognizes and degrades invading pathogenic RNA. For example, the virus causing dengue carries DENV2 genomic RNA. This RNA was engineered into a genetic construct and transfected in male vectors. These males mate with wild type females and resulting females express DENV2 repeat RNA that activates RNAi and reduces vector competence (Franz et al., 2006; Gu et al., 2011).

Biological control using Wolbachia

The potential application of the symbiotic bacteria, Wolbachia, to control VBDs is a recent addition to the arsenal of weapons in the fight against VBDs. In this approach Wolbachia carrying

males or females mate with wild-type females or males, resulting in offsprings that either die at the embryonic stage or are resistant to pathogens (Moreira et al., 2009; O'Connor et al., 2012). Recently, this technique has been successfully used to reduce vector competence of Australian-Mosquito populations (Walker et al., 2011; Hoffmann et al., 2011).

Though genetic modification and biological control offer much promise to combat VBDs, their usage is restricted. The need to release a large number of vectors (>500-600), risk of evolution of resistance against the genetic modification in vectors and financial costs, partially offsets the success of these techniques (McGraw and Neill, 2013). Due to the ineffectiveness of current vector-control approaches, alternative therapeutic strategies against VBDs are highly warranted. One strategy which is receiving increasing attention from the past decade, is targeting the surface coat of the parasite when it traverses through the vector.

1.4 Targeting the surface proteins at the vector-parasite interface

The surface proteins at the vector-parasite interface play an important role in the successes of a pathogen's life cycle. These surface proteins protect the parasite from digestive enzymes and innate immune response of the insects (Roditi and Liniger, 2002). Also, surface proteins can act as tethering agents (adhesins), attaching the parasite to a specific area, and others facilitate the parasite to migrate from one life cycle stage to another (Roditi and Liniger, 2002). Most importantly, these vector-stage proteins are unlikely to undergo antigenic variation in the same manner in the host, and consequently, may be more suitable targets for the development of vaccines (Roditi and Liniger, 2002; Brun et al., 2010). Despite the contribution of these proteins to the life cycle of the pathogen, they are largely understudied. Therefore, understanding the molecular details of how these pathogens attach and migrate through their vectors is crucial in controlling their dissemination. Targeting these proteins may be one of the novel approaches to prevent/control VBDs.

Research on the surface proteins expressed in the vector-stages of the pathogen is currently undergoing a revival. Until a decade ago, morphology used to be the main criterion for distinguishing between different stages of a parasite life cycle, but it now transpires that differentiation from one stage to the next often accompanies alterations in the surface coat. Most of the surface molecules identified to date are peripheral membrane proteins usually anchored to the plasma membrane by a glycosylphosphatidylinositol (GPI) anchor (Utz et al., 2006; (Matthews et al., 2011). Besides GPI-anchor proteins, transmembrane proteins are also present on the pathogen's surface. They are likely to play an important role in multiple signaling pathways, such as communicating signals from the external environment to the cytoplasm or vice versa. However, due to the dominance of GPI anchor proteins, TM proteins are very hard to identify (Roditi and Liniger, 2002). Until now, very few TM proteins have been identified. Despite the identification of a number of surface proteins, our knowledge of these proteins at the molecular and functional level remains limited. Understanding the molecular mechanisms of these surface proteins is crucial in controlling pathogen dissemination. In this thesis, using tsetse-trypanosome as the model system, I will be discussing the surface proteins previously speculated to play an important role in the transmission of trypanosome inside the tsetse vector.

1.5 Tsetse-Trypanosome an ideal model system for understanding vector-pathogen interactions

The tsetse-trypanosome system shows a lot of promise as a highly tractable system to investigate vector-pathogen interaction. Culturing different life cycle stages of the trypanosome *in vitro*, and their genetic transformation has been achieved in *T. congolense,* thereby enabling the reproducibility of the entire life cycle of *T. congolense in vitro* (Coustou et al., 2010). Growing all the life cycle stages allows deciphering the mechanisms underlying the different differentiation steps of the parasite. Also, all the stages of the life cycle can be grown in sufficient quantities for

biochemical analysis, which gives a unique opportunity to study the differential protein expression throughout the infection cycle. The size of the tsetse (~6-15mm) is bigger than most of the vectors, such as mosquitoes (~2-12mm), and sandflies (~1.5-4mm) (Rozendaal, 1997) making dissection relatively easy. Therefore, it is possible to isolate the specific substructures including midgut, proboscis, and salivary glands and assess the molecular interactions between tsetse-trypanosome. Besides this, the biological architecture of tsetse is similar to other insects. Therefore, studying this tractable system will not only improve our understanding of trypanosome dissemination, but also provide valuable insight into other vector-pathogen systems.

1.6 Human and Animal African Trypanosomiasis

African trypanosomiasis are devastating diseases endemic to sub-Saharan Africa. These diseases are caused by kinetoplastid parasites of the genus *Trypanosoma* and threaten both humans and livestock; around 70 million people and 50 million cattle are under threat and 3 million are killed each year, causing a vast socio-economic damage in this region (Brun et al., 2010, Simarro et al., 2011, Simarro et al., 2012).

The causative agents for Human African trypanosomiasis (HAT) popularly known as sleeping sickness, are two sub-species of *T. brucei*: *T. brucei gambiense,* and *T. brucei rhodiense*. While the Rhodesian form of HAT is more rapid, progressing over the course of weeks or months, the Gambian form is characterized by a slow progression that can last for months or even years. The Gambian form is more severe, accounting for approximately 90% of cases reported to the WHO each year (Simarro et al., 2008). Nausea, lethargy, disruption of regular sleep cycles, loss of concentration and seizures are the most common HAT symptoms, which if left untreated, can cause death (Kennedy, 2008).

While trypanosome species causing HAT have a major impact on public health, other trypanosome species significantly affect livestock and cause African animal trypanosomiasis

(AAT). Among the species causing AAT, *T. congolense* is widely considered to be the most economically significant as it infects a broad range of livestock and domesticated animals, including cattle, sheep, pigs and dogs (Ilemobade, 2009). Symptoms include anemia, weight loss, and immunosuppression, and lack of treatment results in death (Vincendeau and Bouteille, 2006). Approximately 3 million cattle are killed by AAT every year, resulting in an economic loss in the range of ~1billion US dollars per annum. When including secondary losses such as reduced manure and draft power into consideration, the total GDP losses can be up to an amount of 5 billion per annum (Chappuis et al., 2005). Tsetse that transmits AAT has a territory covering almost a third of Africa (~9 million km^2), precluding much of the best-watered and most fertile land from cultivation that would otherwise be suitable for crop production or pastureland (Fig. 1.2) (Budd, 1999).

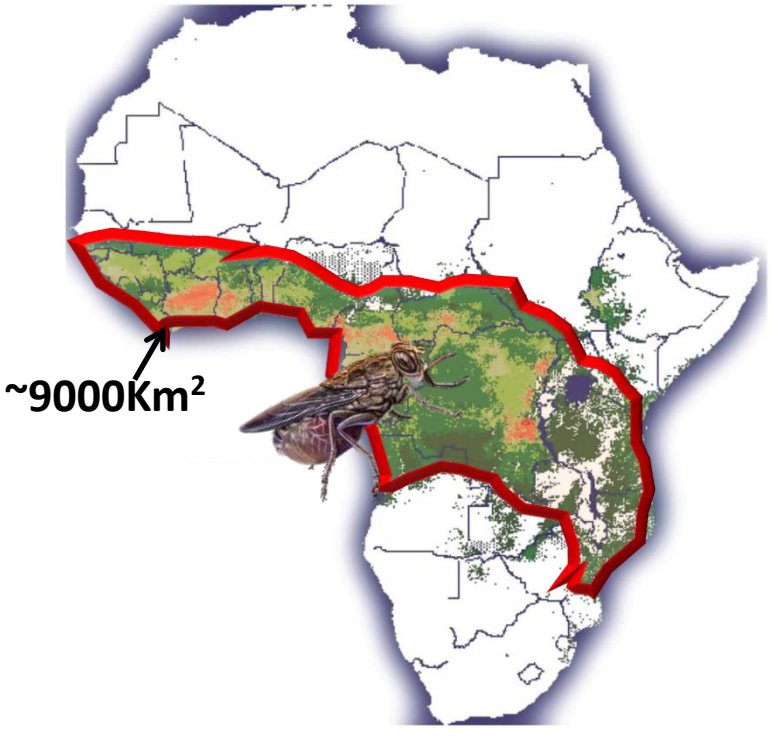

Figure 1.2: Tsetse fly distribution in sub-Saharan Africa. Map showing the distribution of different tsetse species in sub-Saharan African countries. The red outline indicates the area infested by the tsetse which corresponds to ~9000Km^2 (Map adapted from Kariithi et al., 2013.)

Currently, there are no vaccines available for HAT and AAT, and the few available treatments present severe side effects (Kennedy, 2008). Since these parasites have serious health and economic impacts, an understanding of their virulence mechanisms is essential to block the parasite and/or the associated pathogenesis. Interestingly, the most virulent form of the parasite develops in the tsetse (Matthews et al., 2011). Understanding the trypanosome developmental pathway in the tsetse will allow a better comprehension of the molecular crosstalk between parasite and vector, which is crucial for disrupting key interactions in preventing disease transmission.

1.7 Life cycle of trypanosome in tsetse

Trypanosomes (*T. brucei* and *T. congolense*) exhibit a complex life cycle alternating between the mammalian host and insect vector. As the trypanosome stages in both host and vector

are subject to dramatic environmental changes, they demonstrate an equally dramatic change in their metabolism and surface architecture. Trypanosomes assume four distinct life forms throughout their life cycle. These include the Blood stage forms (BSF), Procyclic forms (PF), Epimastigote forms (EMF) and Metacyclic forms (MF). While BSF and MF are common to both vector and host, the other stages are found exclusively in the vector. The majority of stage-specific surface proteins expressed by trypanosomes are anchored to the surface of parasite via a GPI anchor. The life cycle stages present in the tsetse express cell surface molecules that have been proposed to protect the parasite from proteolytic digestion (Acosta-Serrano et al., 2001) or to serve in parasite development and possible ligand-associated parasite-vector signaling (Richardson et al., 1988, Ruepp et al., 1997). It is important to note that at no stage is the surface naked, i.e., a continuous monolayer of glycoprotein or other glycoconjugates always covers trypanosomes. A dense coat of variable surface glycoprotein (VSG) covers the BSF and MF stages. The antigenic variation of VSG is responsible for the evasion of the immune response in the mammalian host (Barry and McCulloch, 2001). This antigenic variation is responsible for the waves of parasitemia characteristic of trypanosomiasis, and it presents as a huge roadblock to vaccine development (Barry and McCulloch, 2001).

BSF trypanosome appears in two forms; a long form that rapidly proliferates and after reaching a specific threshold, become non-proliferative short stumpy form (Fig. 1.3, a). This transformation appears to prepare the parasite in adapting to the life inside the tsetse. The short stumpy VSG covered BSF trypanosomes are ingested by the tsetse during a blood meal and move into the lumen of the tsetse midgut, where they differentiate irreversibly to the PF (Fig. 1.3, b).

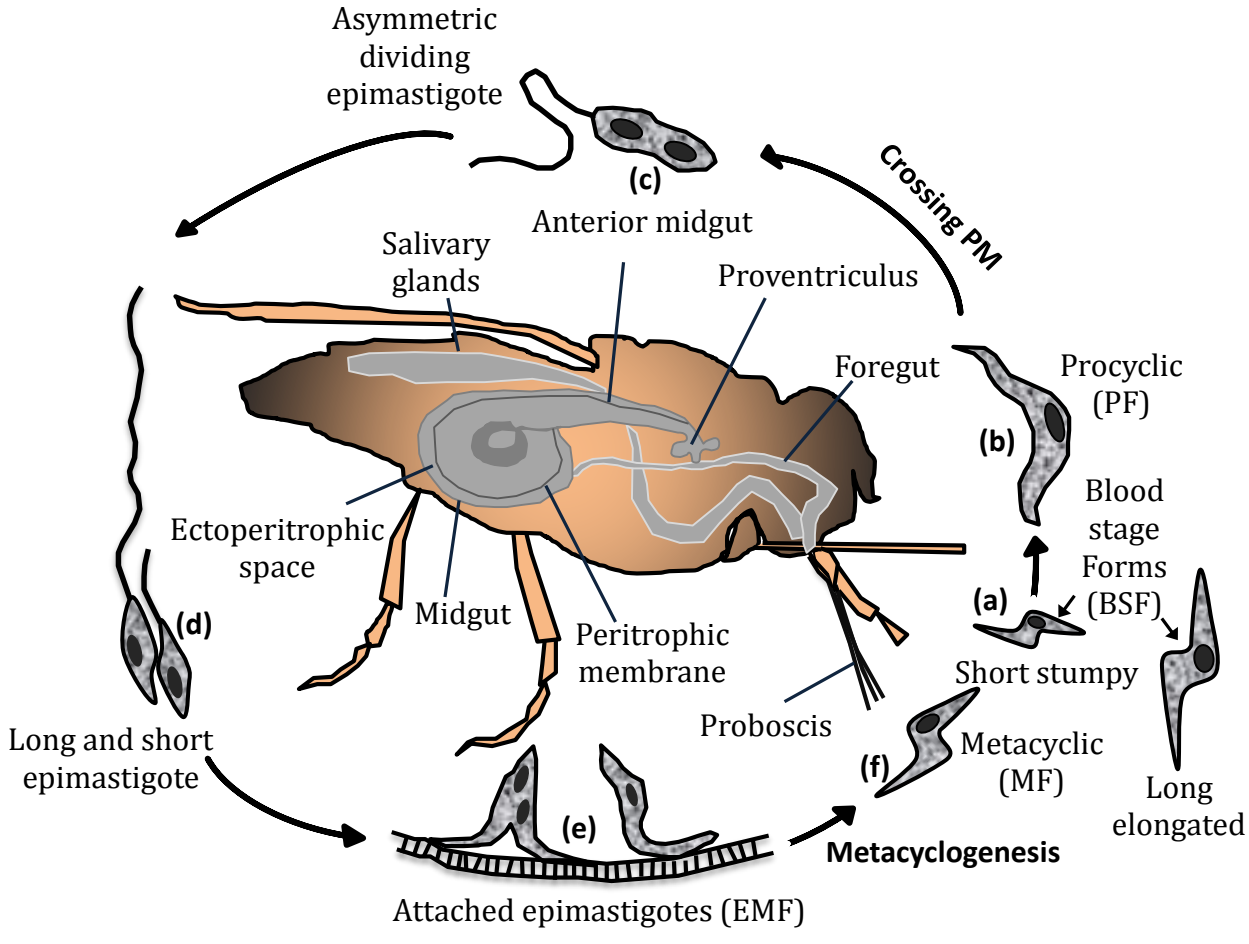

Figure 1.3: Life cycle of trypanosomes in the tsetse fly. Trypanosomes exist in 2 distinct forms, a long slender form and a short stumpy form **(a)**, in the mammalian bloodstream. Upon a successful blood meal, the short stumpy form differentiates into the PF within the lumen of the midgut **(b)**. To complete its development, the trypanosome must cross the PM, migrate to the foregut and differentiate further into long EMF. In the foregut, the long EMF are present that undergo asymmetrical cell division to produce a long and a short EMF **(c, d)**. The short EMF replicate and colonize the salivary gland epithelium **(e)**. These EMF further differentiate into the MF, thus completing the life cycle **(f)**. Both *T. brucei* and *T. congolense* share the same life cycle. However, the MF in *T. congolense* are found in the proboscis in contrast to *T. brucei* MF, which are found in the salivary glands.

This differentiation is characterized by shedding of the VSG coat and expression of invariant tsetse-specific glycoproteins, called procyclins (Vickerman et al., 1988). In *T. brucei*, the procyclin coat (PC) is comprised of two classes of procyclins EP procyclins that contain internal Glu-Pro dipeptide repeats and GPEET procyclins that contain the pentapeptide repeat Gly-Pro-Glu-Glu-Thr (Pays and Nolan, 1998, Roditi and Lehane, 2008, Roditi et al., 1998). A dense layer of

protease-resistance molecules (PRS), as well as glutamate- and alanine-rich protein (*Tc*GARP) (Beecroft et al., 1993, Bayne et al., 1993) and heavily glycosylated *T. congolense* heptapeptide repeat protein (*Tc*HRP) (Utz et al., 2006) covers the surface of *T. congolense* PCs. Following successful establishment within the tsetse midgut, the parasites must cross the peritrophic matrix (PM) to the foregut (Welburn and Maudlin, 1999). The PM acts as a physical and a biochemical barrier to toxins and pathogens. In the foregut, the trypanosome differentiates into long epimastigote form (Fig. 1.3, c), that undergoes an asymmetrical cell division to produce long and short EMFs (Fig. 1.3, d) (Van Den Abbeele et al., 1999), and differentiate further and migrate to the mouthparts. From here, the progression of short epimastigote form diverges in *T. brucei* and *T. congolense*. *T. brucei* short epimastigotes move to the salivary glands of the tsetse, attach to the microvilli with dendritic outgrowths from the flagellum, and differentiate into the non-motile, proliferative EMF (Fig. 1.3, e) (Vickerman, 1985). *T. congolense* short epimastigotes differentiate into the EMF in the proboscis of the fly, attaching to its chitinous interior via hemidesmosome-like structures (Vickerman, 1985). The EMF stage in *T. brucei* is characterized by the expression of glycoproteins such as BARP (*Brucei* Alanine rich protein), ISG65 (Invariant surface glycoprotein) and ISG75 whereas *T. congolense* expresses *Tc*GARP and *Tc*CESP (*Congolense* epimastigote specific protein) (Bütikofer et al., 2002, Sakurai et al., 2008). The EMF then differentiates into smaller non-dividing MF in a process called metacyclogenesis (Fig. 1.3, f). In the MF stage, procyclins are shed and VSG re-appears. MF trypanosomes are smaller than the EMF, are non-proliferative, and the virulent form of the parasite. MF trypanosomes are injected into new mammalian hosts by the tsetse during a blood meal where they differentiate into the BSF, thus completing the trypanosome life cycle. Therefore, the surface proteins expressed by the parasite play an important role in the differentiation of the parasite from one form to another. A

better understanding of the role played by the surface proteins could provide insights on novel ways to interrupt transmission.

Outstanding question

While a significant body of work has identified numerous surface molecules, the molecular mechanisms of vector-parasite interaction remain unknown as very few surface proteins have proposed functionality. Therefore, the key question in the field is "what is the exact functions of these surface proteins and how they facilitate parasite migration through tsetse". To address this question, my thesis focuses on understanding the structure-function relationship of proteins hypothesized to be involved in playing a critical role in the transmission of parasites inside the tsetse.

1.8 *Tb*Fam50.360

*Tb*Fam50.360 is a GPI-anchored protein belonging to the Fam50 family of proteins (Jackson et al., 2013) found in *T. brucei*. Transcriptomic data reveal a high expression in both EMF and MF, with MF having higher expression than EMF (Jackson et al., 2013). A high expression in the MF indicates a possible role played by *Tb*Fam50.360 in the insect to host transition since MF is the infective form. Recently, a proteomic study also demonstrated a high expression of this protein in MF, consistent with the genomic data (unpublished data). In this study, salivary proteins from both naïve and trypanosome-infected saliva were either subjected to an in-solution analysis or fractionated on 1D SDS-PAGE, followed by trypsin digestion. An LC-MS/MS analysis of the tryptic fragments from both preparations suggested a very high expression of *Tb*Fam50.360 in the saliva from *T. brucei* infected flies. Chapter 2 explains this protein in greater detail.

1.9 *Tc*CISSA/*Tb*PSSA2

A recent proteomic study taking advantage of the ability to culture *T. congolense* in all life cycle stages *in vitro* investigated the differential expression of proteins throughout its life cycle stages. This study revealed the existence of a novel protein overexpressed only in PCF and EMF. This protein was named *Tc*CISSA (Congolense Insect Stage Specific Antigen) (Eyford et al., 2011). Interestingly, *Tc*CISSA is a homolog of a previously identified surface protein in *T. brucei*, PSSA2 (Procyclic Stage Specific Antigen), and shares 60% identity. In contrast to the previously characterized GPI-anchored proteins from the surface of trypanosome, both *Tc*CISSA and *Tb*PSSA2 are transmembrane (TM) proteins containing a cytoplasmic domain. While the function of *Tc*CISSA is unknown, a recent study postulated the involvement of *Tb*PSSA2 in sensing and transmitting signals that contribute to the parasite's decision to divide, differentiate or migrate (Fragoso et al., 2009). In this study, the authors observed that *Tb*PSSA2 null mutants were

inefficient in establishing infections in salivary gland, despite successful infection in the midgut. Based on the sequence identity with *Tb*PSSA2, it is tempting to speculate that *Tc*CISSA would also play a similar role. Chapter 3 presents a detailed study on these proteins.

1.10 Research objectives

The African trypanosome *T. brucei* is a vector-borne parasite causing HAT in sub-Saharan Africa and, along with related species *T. congolense*, causes a similar disease in wild and domestic animals. Together, these parasites have a significant impact on socio-economic development in Africa. Trypanosomes express proteins on their surface that influence the host environment and allows for their transmission. Since the trypanosomes in the insect-stage undergo less antigenic variation in comparison with host-stage trypanosomes, the insect-stage offers an improved target for parasite control if strategies can be devised to disrupt their transmission to mammalian hosts. Though proteomic studies have identified a number of surface proteins, their functions remain unknown. Characterizing these currently unknown proteins will help facilitate the development of novel strategies to alter the host environment, thereby making it inhospitable for the parasite, and reducing disease transmission. Obtaining more information regarding parasite surface proteins and their interactions with their vector is critical to improving our understanding of parasite survival and transmission. The two main objectives of this dissertation project were as follows:

(i) Structurally characterize surface proteins hypothesized in facilitating the transmission of trypanosome from midgut and attachment with salivary glands.

(ii) Decipher the functions of the surface proteins from their structures.

Chapter 2：Structural characterization of *Tb*Fam50.360

Contributions:

Construct design, cloning, and initial crystallization trials were performed by Sean Workman (Materials and Methods section 2.2.1 and 2.2.2). Structure determination and refinement were performed alongside Dr. Marty Boulanger and I completed the final structural interpretation (Materials and Methods section 2.2.3-2.2.5)

2.1 Introduction

As discussed in Chapter I, HAT is a deadly parasitic disease caused by protozoan parasites of *T. brucei* species and transmitted by tsetse. A number of disease control strategies applied to date have not been very successful, and a long-term solution remains unidentified. Particularly relevant for disease control are surface proteins that play a significant role in interacting with the vector environment. They are predicted to play important roles in adhesion, signal transduction and membrane trafficking.

VSG and procyclin are two of the more well-studied surface proteins expressed by *T. brucei* in its PF and BSF stages. The VSG allows the parasite to evade the immune response in the host (Barry and McCulloch, 2001, Hajduk, 1984), whereas the role of procyclin is not entirely clear. Initially, procyclins were thought to protect the parasite against the digestive enzymes secreted by the tsetse midgut (Ruepp et al., 1997). However, a knockout study indicated that this is not necessarily true (Vassella et al., 2003). A third GPI-anchored family, BARP (*Brucei* alanine rich protein), expressed by immature salivary gland stages has been identified (Urwyler et al., 2007, Nolan et al., 2000) the function of which is unknown.

Identifying and characterizing more surface proteins is thus urgently needed. As a step in this direction, recent studies have applied *in silico* GPI-anchor attachment and signal sequence prediction approaches to identify genes predicted to encode products associated with the cell surface of trypanosome and evaluated their expression profile. Hypothetical protein, *Tb*927.7.360, was identified in these studies (Savage et al., 2012; Jackson et al., 2013).

*Tb*927.7.360 is a 360 amino acid polypeptide with a potential N-terminal leader sequence and a C-terminal hydrophobic sequence allowing GPI-anchor attachment (Fig. 2.1). *Tb*927.7.360 will henceforth be referred to as *Tb*Fam50.360.

*Tb*Fam50.360 belongs to clade iv of a larger family of proteins called the Fam50 family (Fig. 2.2) (Savage et al., 2012). Fam50 is one of the 79 gene families in the 'Fam' family group, which are known or predicted to be having surface localization. Besides *Tb*927.7.360, Fam50 also contains *Tc*GARP, *Tb*BARP, *Tc*CESP, and three other clades i-iii. Clade iv also contains four paralogs of *Tb*Fam50.360 which include *Tb*427.07.380, *Tb*427.07.440, *Tb*427.07.420, and *Tb*427.07.400. Since the genomic profile (transcriptomic data) of most of the members of Fam50 family reveals an upregulation in the midgut and salivary glands (Savage et al., 2012; Jackson et al., 2013), it is tempting to speculate their role in establishing infection in these stages. However, the exact functions of Fam50 family of proteins remain unknown.

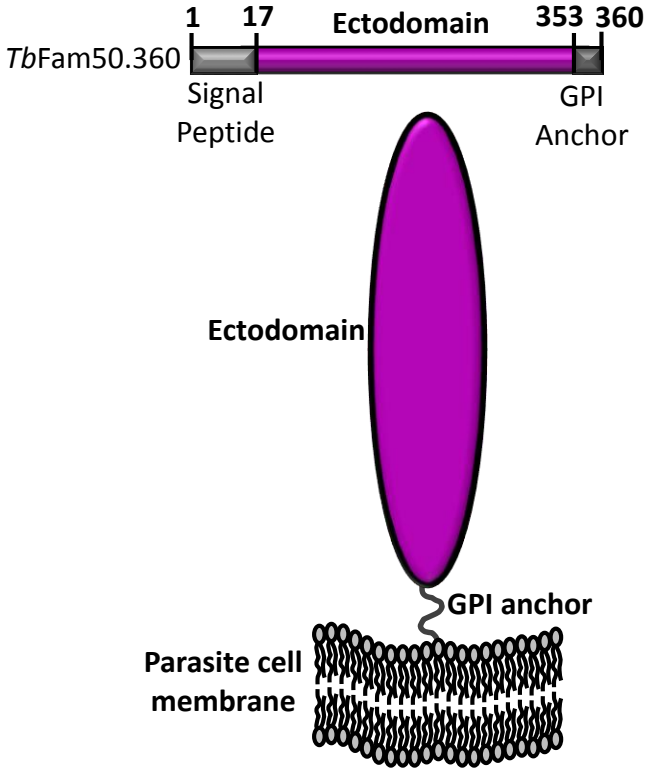

Figure 2.1: Predicted domain architecture of *Tb*Fam50.360. The figure shows the predicted organization of domains for *Tb*Fam50.360. *Light grey* indicates the predicted signal peptide, *dark grey*- predicted GPI anchor, and *deep purple*- ectodomain. (*Bottom*) shows the predicted localization of protein with respect to the plasma membrane. Note: The amino acids after the GPI anchor site are not shown as they get cleaved.

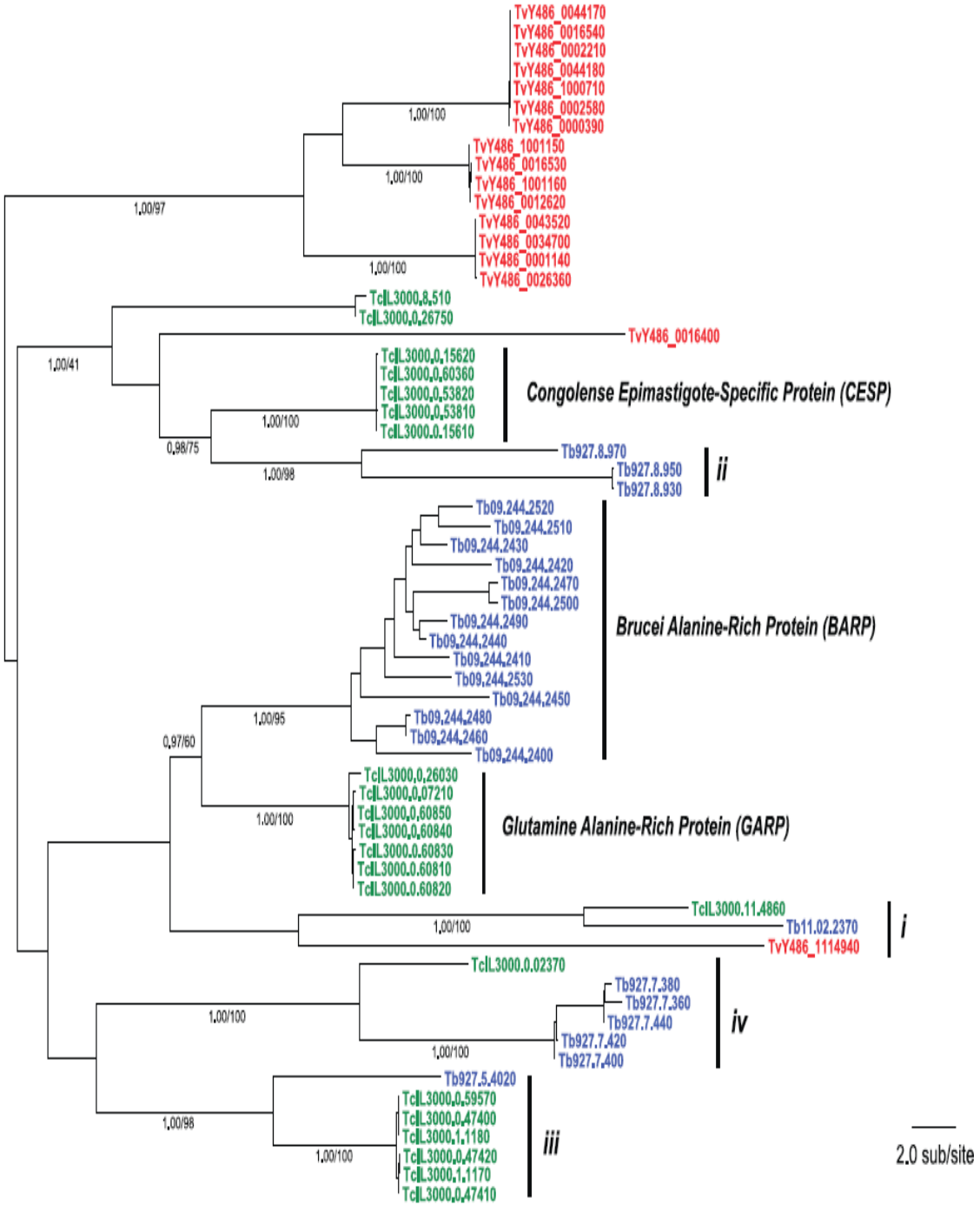

Figure 2.2: Bayesian phylogeny of Fam50 family. Phylogenetic analysis reveals different family members in distinct clades. The nodes are supported by posterior probability values and non-parametric bootstraps generated from a maximum likelihood analysis using an LG model with rate heterogeneity (Accession numbers of the proteins are indicated). The tree is midpoint rooted. Copied from (Jackson et al., 2013).

To date, *Tc*GARP (glutamic acid -alanine rich protein) remains the only surface protein from. The PF and EMF stages of *T. congolense* express *Tc*GARP. This protein is postulated to play a role in switching the trypanosome coat from VSG to other procyclins during the procyclic stage and protecting the parasite surface by shielding against digestive enzymes, proteins, and complement of the tsetse midgut (Loveless et al., 2011). But beyond this its functional roles are unknown. Besides *Tc*GARP, CESP has also been assessed previously and was shown to be a putative adhesin (Sakurai et al., 2008). Interestingly, the preliminary structure of BARP and a 3D model of CESP were also shown to adopt a three helical motif thereby showing a striking similarity to previously characterized trypanosome surface proteins including *Tc*GARP (Loveless et al., 2011), *Tb*VSG (Freymann et al., 1992) and *Tc*HpHbR (Higgins et al., 2013). Based on the structural similarities between *Tc*GARP, *Tb*BARP, *Tc*CESP, *Tb*VSG and *Tc*HpHbR, it is tempting to speculate that *Tb*Fam50.360 may also share the same architecture.

Among the members of the Fam50 family, clade iv is the only group having a predominant expression in MF at both genomic (Savage et al., 2012) and proteomic level (unpublished; from collaborators). Since MF is the infective form of the parasite that is transmitted to the vertebrate (host), it appears that clade iv may be important in insect-vertebrate transition. To gain insights into the structure and function of clade iv glycoproteins, we expressed the N-terminal region of one of the paralogs, *Tb*Fam50.360, in *E. coli* and determined its crystal structure.

2.2 Materials and Methods

2.2.1 Construct design and cloning of *Tb*Fam50.360

The sequence corresponding to *Tb*Fam50.360 from *Trypanosoma brucei* was obtained from TriTrypDB (Aslett et al., 2010); accession No. *Tb*427.07.360. A sequence encoding the mature *Tb*Fam50.360 from the end of signal peptide to the beginning of the predicted GPI anchor (Glu17 to Gly353, with numbering starting from the initiation methionine) was codon optimized for *E. coli* and synthesized by GenScript. A construct (Gly24 to Ala233) that aligns with the conserved portion of the *Tb*Fam50 clade iv was then sub-cloned into a modified pET32a vector (Novagen) containing N-terminal thioredoxin (Trx) and hexa-histidine tags separated from the gene of interest by a TEV protease site. Sequencing confirmed that no mutations were introduced during the amplification procedure.

2.2.2 Expression and Purification of *Tb*Fam50.360

*Tb*Fam50.360 was produced recombinantly in *E. coli* Rosetta-gami 2 (DE3) cells (Novagen) grown in autoinduction medium (Invitrogen) from a 5% inoculum. Following four hours of growth at 37 °C and 64 hours at 16 °C, the cells were harvested and the pellet re-suspended in 20 mM HEPES pH 8.3, 1 M NaCl, 30 mM imidazole. Cells in suspension were lysed using a French Press. The insoluble material was removed by centrifugation, and the soluble fraction allowed to batch-bind with Ni-agarose beads for 1 hour at 4°C. *Tb*Fam50.360 was eluted with buffer containing 250 mM imidazole, and fractions were analyzed by SDS-PAGE and pooled based on purity. The Trx-His_6 tag was removed by TEV cleavage overnight at 18 °C, and *Tb*Fam50.360 was further purified by size exclusion chromatography (Superdex Hi Load 16/60 75) in HEPES buffered saline (HBS). The protein concentration was determined by absorbance at 280 nm with a calculated extinction coefficient of 9970 M^{-1} cm^{-1} (The extinction coefficient was calculated using Protparam (Gasteiger et al., 2005)).

2.2.3 Crystallization and data collection

Crystals of purified recombinant *Tb*Fam50.360 were initially identified in the Index screen (Molecular Dimensions) using the sitting drop method at 18 °C. The final drops consisted of 0.3 µL of *Tb*Fam50.360 at 20 mg ml^{-1} with 0.3 µL of reservoir solution and were equilibrated against 60 µL of reservoir solution. Diffraction quality crystals grew in 48 hours in 25% PEG, 1500. A single crystal was looped, cryopreserved in 12.5% glycerol for 20 seconds, and flash-cooled to -173.15 °C directly in the cryostream. Diffraction data were collected at Canadian Light Source (CLS) at a wavelength of 0.9795 Å.

2.2.4 Data processing, structure solution and refinement

Diffraction data to 1.82 Å resolution were processed using Imosflm (Leslie, 1992) and Scala (Evans, 2006) in the CCP4 suite of programs (Dodson et al., 1997). Initial phases were obtained by molecular replacement using PHASER (McCoy et al., 2007) with one copy of the Glutamic Acid/Alanine-Rich Protein (*Tc*GARP) from *Trypanosoma congolense* (Loveless et al., 2011) (PDB 2y44; 15% identity over 210 residues). Solvent molecules were added using COOT (Emsley and Cowtan, 2004) and refinement carried out using Phenix Refine (Affonine et al., 2012). The overall structure of *Tb*Fam50.360 was refined to an R_{free} of 20.43%. Stereo-chemical analysis performed with PROCHECK and SFCHECK in CCP4 showed excellent stereochemistry with more than 99% of the residues in the favored conformations and no residues modeled in disallowed regions of the Ramachandran plot. Overall, 5% of the reflections were set aside for calculation of R_{free}. Data collection and refinement statistics are presented in Table 2.1

Table 2.1 Data collection and refinement statistics

	*Tb*Fam50.360
A. Data collection	
Synchrotron source	CLS
Space group	$P2_12_12_1$
a, *b*, *c* (Å)	24.81, 79.50, 108.17
$\alpha = \beta = \gamma$ (°)	90.00
Wavelength (Å)	0.9795
Temperature (K)	100
Resolution range (Å)	44.72-1.82 (1.92–1.82)
Measured reflections	136043
Unique reflections	19605 (2715)
Redundancy	6.9 (6.4)
Completeness (%)	98.2 (94.9)
$I/\sigma(I)$	20.8 (9.3)
R_{merge}[a] (%)	6.0 (15.9)
B. Refinement Statistics	
Resolution (Å)	37.31-1.82 (1.88–1.82)
R_{cryst}[b] / R_{free}[c] (%)	16.94(19.24)/20.28(29.20)
No. of atoms	
Overall	3,347
Protein	3,074
Solvent/Heterogen atoms	273
Mean temperature factor (Å^2)	
Overall	12.3
Protein	11.3
Solvent/Heterogen atoms	18.0
r.m.s. deviation from ideality	
Bond lengths (Å)	0.010
Bond angles (°)	1.14
Ramachandran statistics	
Most favored (%)	99.5
Allowed (%)	6.6
Generously allowed (%)	0.0
Disallowed (%)	0.0

Values in parentheses are for the highest resolution shell

[a] $R_{merge} = \sum_{hkl} \sum_i |I_{hkl,i} - [I_{hkl}]| / \sum_{hkl} \sum_i I_{hkl,i}$, where $[I_{hkl}]$ is the is the average of symmetry related observations of a unique reflection

[b] $R_{cryst} = \sum|F_{obs} - F_{calc}| / \sum F_{fobs}$, where F_{obs} and F_{calc} are the observed and the calculated structure factors, respectively.

[c] R_{free} is R using 5% of reflections randomly chosen and omitted from refinement

[d] Ramachandran statistics were determined using PROCHECK

2.2.5 Bioinformatics

Multiple sequences were aligned using Clustal Omega (Sievers et al., 2011) with BLOSUM62 matrix and pairwise alignment. The gap penalty at the beginning and the end was assigned a value of 1. The sequences were illustrated in ESPript 3.0. (Robert and Gouet, 2014) Accession numbers for the aligned sequences were obtained from TriTrypDB (Aslett et al., 2010) and are as follows; *Tb*Fam50.360 (*Tb*427.07.360), *Tb*Fam50.380 (*Tb*427.07.380), *Tb*Fam50.440 (*Tb*427.07.440), *Tb*Fam50.420 (*Tb*427.07.420), *Tb*Fam50.400 (*Tb*427.07.400). The models of the C-terminal domain of *Tb*Fam50.360, full-length *Tb*Fam50.380, *Tb*Fam50.440, *Tb*Fam50.400, and *Tb*Fam50.420 were generated using the IntFOLD algorithm (Roche et al., 2011). The template for modeling the N-terminal domains of *Tb*Fam50.360 homologs was the crystal structure of *Tb*Fam50.360. All the models of N-terminal domains had high confidence and P-values (P-value *Tb*Fam50.360- 6.21E-3; P-value *Tb*Fam50.380- 5.13E-3, P-value *Tb*Fam50.440- 6.11E-3; P-value *Tb*Fam50.400- 6.43E-3; P-value *Tb*Fam50.420- 6.57E-3). The signal peptide and GPI anchor were predicted using SignalP 4.0 (Petersen et al., 2011) and PredGPI (Pierleoni et al., 2008).

2.3 Results

2.3.1 *Tb*Fam50.360 adopts an extended helical architecture

The conserved, N-terminal domain of *Tb*Fam50.360 (Gly24 to Ala233) (Fig. 2.3A) was recombinantly produced in *E. coli* and purified to homogeneity using nickel affinity and size exclusion chromatography (SEC). Comparison of the SEC elution profile against a series of globular protein standards showed that *Tb*Fam50.360 eluted as a monomer of approximately 22 kDa (Fig. 2.3A). Crystals of purified *Tb*Fam50.360 were obtained using the sitting drop method and grew in space group $P2_12_12_1$ with a single molecule in the asymmetric unit. Molecular replacement using a truncated form of *T. congolense* Glutamic Acid/Alanine-Rich Protein (*Tc*GARP - PDB 2Y44) as the search model was used to determine the structure of *Tb*Fam50.360. Despite low sequence identity (15%), *Tc*GARP emerged as a suitable model based on secondary structure predictions. The overall structure of *Tb*Fam50.360 was refined to a resolution of 1.82 Å and is well defined with only two residues from the N-terminus remaining un-modelled.

The core of the *Tb*Fam50.360 N-terminal ectodomain adopts an elongated structure measuring approximately 83 Å in height and spanning approximately 25 Å in width (Fig. 2.3B – left panel). It has a well-ordered ectodomain with low *B*-factors throughout the structure (Fig 2.3B – middle panel). *Tb*Fam50.360 adopts an overall helical bundle structure composed of a core of extended twisted helices capped by a smaller helical bundle at the N-terminal end predicted to be distal from the parasite membrane (Fig. 2.3B). The helical bundle dominating the structure consists principally of three helices; Helix I (blue) comprises of residues V29-S83, helix II (green: E88-A129) and helix VI (red: F180-A233). The three helices adopt a bend of approximately 30° at G44 (helix I), A127 (helix II) and L194 (helix VI) and collectively, give rise to the helical bundle cap. In addition to the ends of the three major helices, this bundle includes three shorter helices: helix III (orange: D134-E142), which is connected by a five-residue loop to helix IV (cyan: S143-G156),

and helix V (pink: S166-F179) connected to helix IV by a nine residue loop. Helices IV and V lay on either side of helix I and form the broadest face of the helical bundle cap.

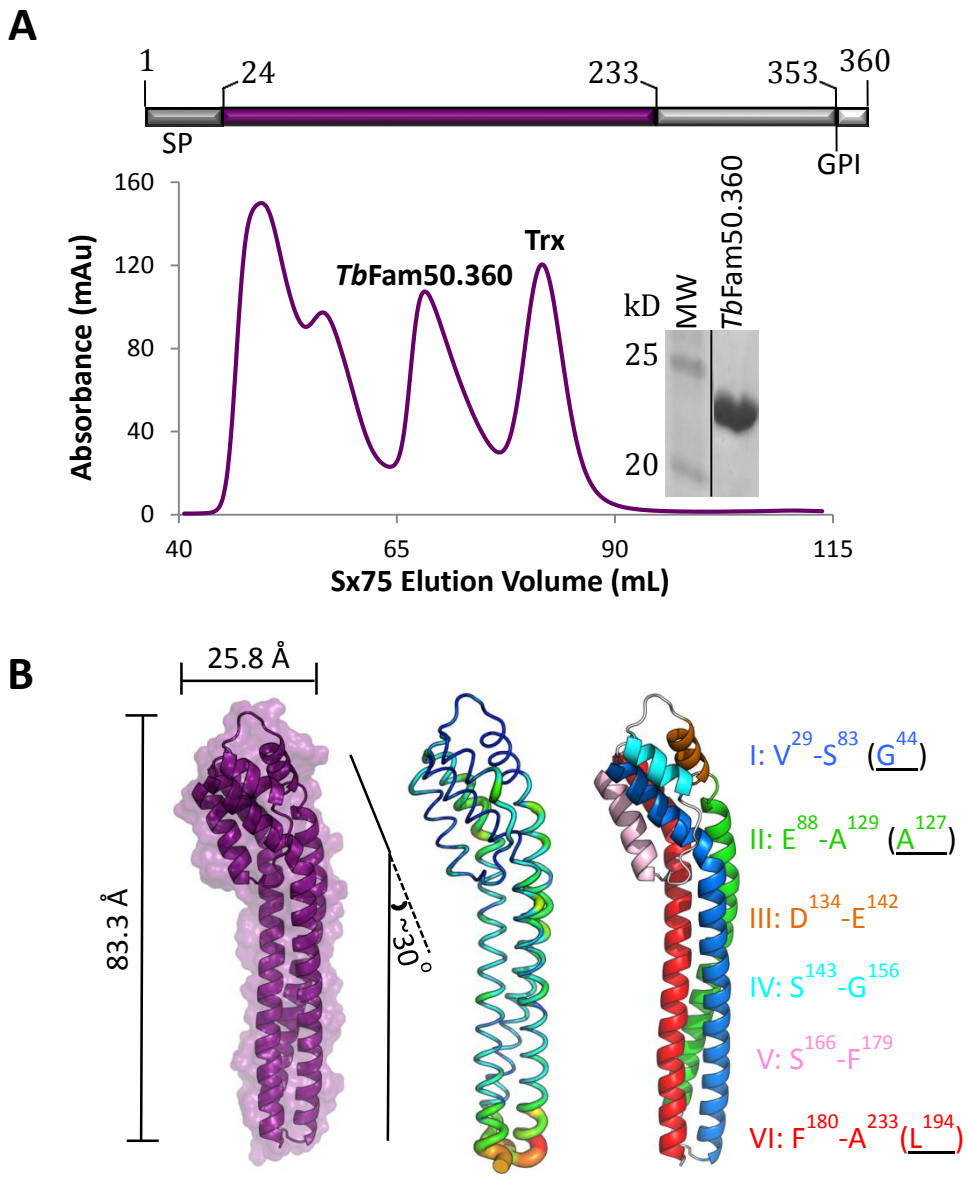

Figure 2.3: Structural and functional analysis of *Tb*Fam50.360. (A) Top: Construct encoding the *Tb*Fam50.360, which was recombinantly produced in *E.coli* is indicated in *deep purple* from Asn^{24} to Glu^{233}. Superdex 75 column size exclusion chromatogram of *Tb*Fam50.360 (*deep purple* curve). *Tb*Fam50 eluted at ~67.5 ml, consistent with the molecular weight of 22kDa *Inset*, SDS-PAGE analysis of the column fractions, with *Tb*Fam50.360 migrating at ~22kDa. **(B)** *Left*- Surface representation of *Tb*Fam50.360 shown in *deep purple*. The structure was found to be 83.3 Å tall and 25.5 Å in width. *Middle- B*-factor putty model of the *Tb*Fam50.360 ectodomain. Ordered regions are shown in *thin blue ribbons*, flexible regions- in *red* and *larger diameter tubes. Right*- Secondary structure depiction of *Tb*Fam50.360. The color schemes used were *white* (coil), *blue* (Helix I; V^{29}-S^{83}), *green* (Helix II; E^{88}-A^{129}), *orange* (Helix III; D^{134}-E^{142}), *cyan* (Helix IV; S^{143}-G^{156}), *light pink* (Helix V; S^{166}-F^{179}), *red* (Helix VI; F^{180}-A^{233}). The underlined residues enclosed in brackets indicate the bend forming residues in helix I (G^{44}), II (A^{127}) and VI (L^{194}) respectively.

2.3.2 Surface analysis reveals a pocket likely to bind a ligand in *Tb*Fam50.360

The lack of significant sequence identity between TbFam50.360 and any protein of known function led us to perform a DALI (Holm and Park, 2000, Holm and Rosenstrom, 2010) search to identify structural homologs. DALI identified TcGARP (PDB id: 2Y44) as the top hit with a Z-score of 21.4. A least squares superposition between the two structures resulted in an rmsd of 1.7 Å over 183 Cα atoms (Fig. 2.5 – left panel). The DALI also revealed structural homology to previously characterized haptoglobin-hemoglobin receptor (HpHbR) from T. congolense and T. brucei (PDB id: 4E40 and 5HU6 (not shown in fig. 2.4)) (Higgins et al., 2013, (Lane-Serff et al., 2015) (Lane-Serff et al., 2015), and variant surface glycoprotein (VSG) from with T. brucei (PDB id: 1VSG) with Z scores of 18.0, 17.7 and 6.5 respectively (Fig. 2.4).

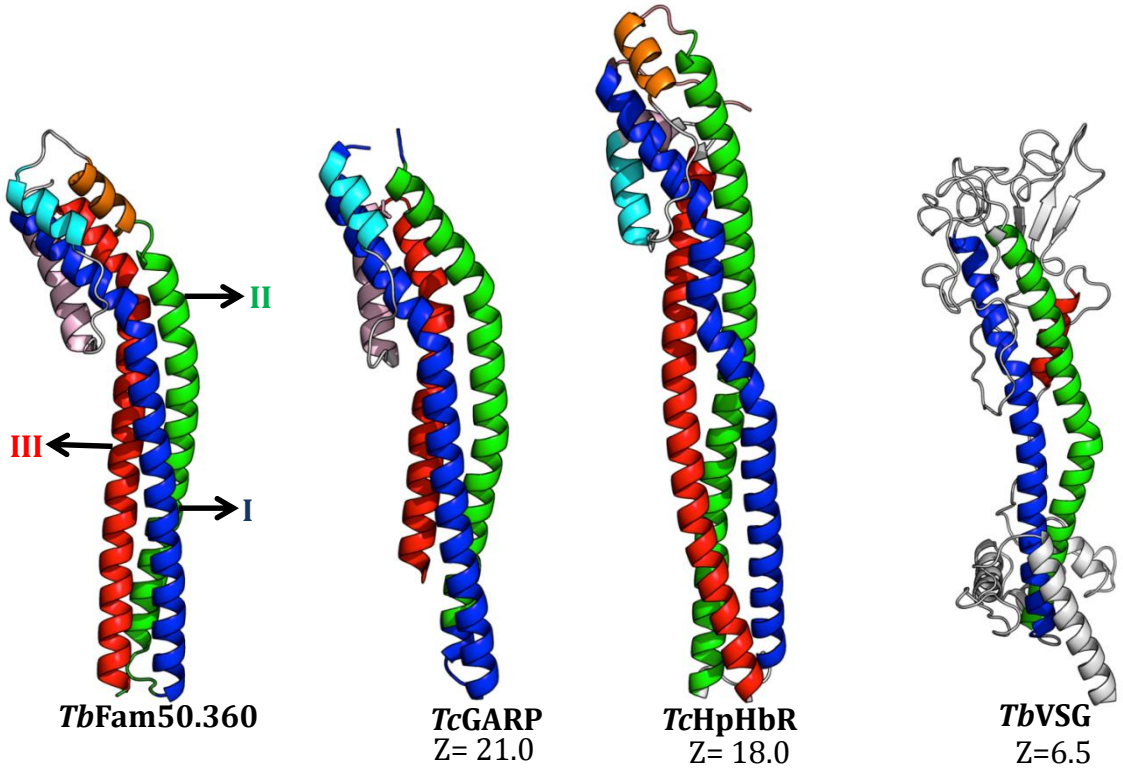

Figure 2.4: *Tb*Fam50.360 closely resembles *Tc*GARP. The figure shows the DALI comparison of *Tb*Fam50.360 with previously characterized surface proteins. *Tb*Fam50.360 shares a three-helical architecture with the N-terminal domain of *Tc*GARP (PDB id: 2Y44), *Tc*HpHbR (PDB id: 4E40) and *Tb*VSG (PDB id:1VSG). Helices I (*blue*), II (*green*), and III (*red*) form the core architecture of all three proteins. *White* color indicates loops and coils. Structural comparison shows that *Tb*Fam50.360 is structurally most homologous to *Tc*GARP with a high Z-score of 21.0. It has the least resemblance to *Tb*VSG with Z-score of 6.5.

All these proteins exhibit a complementary core of twisted three helical bundles. Structural comparison, however, indicates a closer architectural similarity with *Tc*HpHbR, and *Tb*HpHbR compared to *Tb*VSG. This is because of the breakdown of the third helical strand into loops and extensions allowing a greater structural diversity (Freymann et al., 1990). Moreover, *Tc*HpHbR, *Tb*HpHbR, and *Tb*Fam50.360 are monomeric in contrast to a dimeric VSG (Freymann et al., 1990). Despite the general architectural similarity with of *Tb*Fam50.360 with *Tc*HpHbR, the lack of key ligand binding residues (Higgins et al., 2013, Lane-Serff et al., 2015), indicates a different biological role for *Tb*Fam50.360.

A close analysis of the *Tb*Fam50.360 and *Tc*GARP structures revealed a similar distribution of acidic and basic residues along the entire length of the structure and no clear localized charge densities that would indicate a molecular recognition site. However, both structures present a surface pocket at the membrane distal end near the region where the core helices in both proteins bend (Fig. 2.5 – middle panel).

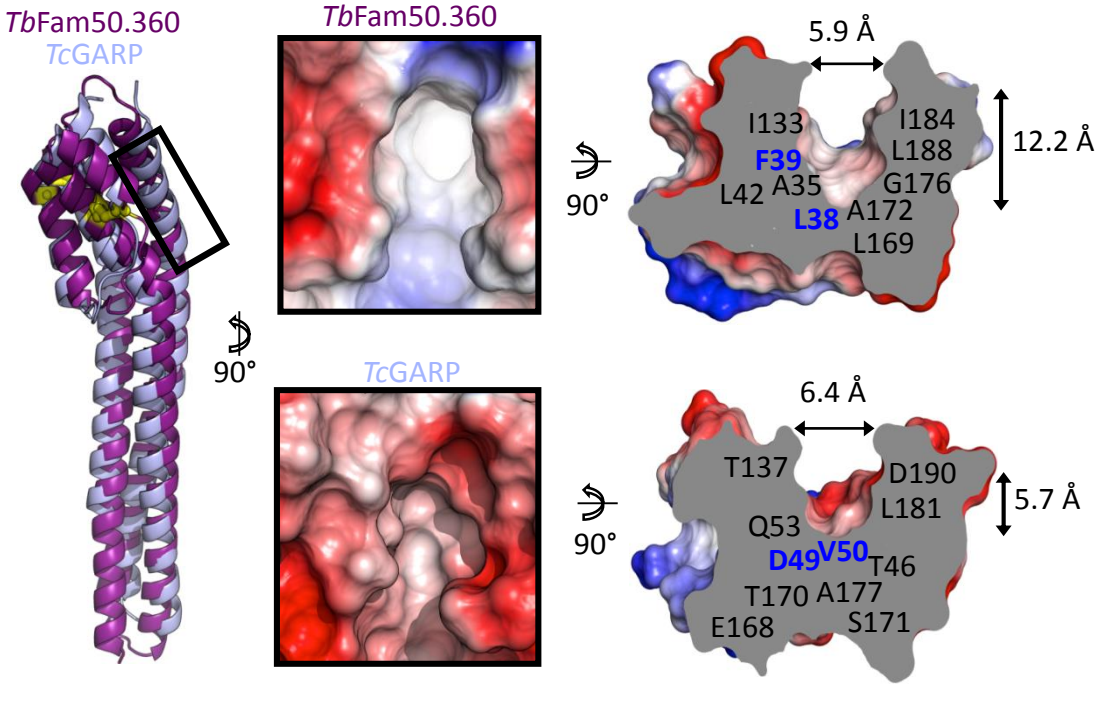

Figure 2.5: Comparison of the structures of *Tb*Fam50.360 and *Tc*GARP. *Left*: Superimposition of *Tb*Fam50.360 (*deep purple*) with *Tc*GARP *(light blue)* (PDB ID: 2Y44). The disulphides are shown in yellow ball and stick model. *Middle*: GRASP image of *Tb*Fam50.360 indicates a predominantly hydrophobic pocket and (bottom) a corresponding polar pocket in *Tc*GARP. The orientation of the images was rotated by 90° from that in *left*. *Right*. Side view of the structure of *Tb*Fam50.360 and *Tc*GARP (bottom), showing the depth and diameter of the cylindrical cavity. Residues within the central cavity are predominantly hydrophobic in *Tb*Fam50.360 and polar in *Tc*GARP. The residues involved contributing to the depth of the cavity are highlighted in *blue*. The pocket present in *Tb*Fam50.360 was deep having a depth of 12.2 Å, whereas *Tc*GARP had a shallower pocket measuring 5.7 Å in depth.

N-terminal portion of helix I, the loop connecting helix II to III, helix V and N-terminal region of helix VI form the pocket in *Tb*Fam50.360. The secondary structures contributing to the pocket formation in *Tc*GARP were similar, with the N-terminal portion of helix I, C-terminal region of helix II, the loop connecting helix III to IV, and N-terminal region of helix V forming the pocket. Intriguingly, the overall dimensions of the pockets are quite different. In *Tb*Fam50.360, the pocket, measures approximately 12.2 Å in depth and 5.9 Å in diameter (Fig. 2.5 – right panel) and is lined by ten hydrophobic residues (A35, L38, F39, L42, I133, L169 A172, G176, I184, and L188). In contrast, the pocket of *Tc*GARP is significantly smaller, measuring approximately 5.7 Å in depth and 6.4 Å in diameter and formed by predominantly polar (Fig. 2.5 – right panel) residues (T46, D49, V50, Q53, T137, E168, T170, S171, A177, L181, and D190). It is intriguing to speculate that these pockets in *Tb*Fam50.360 and *Tc*GARP may coordinate a different subset of ligands consistent with a putative role for these proteins as adhesins.

2.4 Discussion

*Tb*Fam50.360 belongs to the Fam50 family of surface anchored proteins. Currently, the functions of all members of Fam50 family remain unknown. Among the Fam50 family of proteins, *Tb*Fam50.360 appears to be important due to its predominant expression in the MF of the parasite life cycle as revealed by proteomics. Based on the expression of *Tb*Fam50.360 in the infectious stage of the life cycle, it was hypothesized to enable the parasite transition from insect vector to the mammalian host. Besides this, the function of *Tb*Fam50.360 is largely unknown. To get an insight into the function *Tb*Fam50.360 we structurally characterized the N-terminal domain of *Tb*Fam50.360. This is only the second protein, after *Tc*GARP, to be structurally characterized from the Fam50 family.

The crystal structure of *Tb*Fam50.360 N-terminal domain revealed structural homology with previously characterized trypanosome surface proteins. The mature protein, however, contains an additional, 127 residue C-terminal tail predicted to be highly disordered, which could not be crystallized. The C-term suggested that this region might allow the protein to extend longer from the parasite cell surface. Interestingly, long C-terminal regions have also been observed in apicomplexans adhesive micronemal proteins, where they are hypothesized to play a role in protein-protein interactions, cell signaling, or facilitating proteolytic processing. For example, in *Tg*AMA4, the C-terminal region (547 residue linker) functions as a tether to initially contact the parasite cell membrane and enables its homolog *Tg*AMA3 (93 residue linker) to interact with parasite cell membrane, thereby enabling a staged process (Parker et al., 2016).

*Tb*Fam50.360 could play a role similar to *Tg*AMA4. *Tb*Fam50.360 has 4 additional paralogs, (*Tb*Fam50.380, *Tb*Fam50.440, *Tb*Fam50.400, and *Tb*Fam50.420) (Fig. 2.6A and B). The N-terminal regions of these paralogs are highly identical to *Tb*Fam50.360 but the C-terminal regions have varying lengths (Fig. 2.6A and B). We do not know if they are differentially expressed

or co-expressed. Since the sequence identity of these proteins is high, it allowed us to generate high confidence models using the IntFOLD server (Roche et al., 2011). All the models look strikingly similar to *Tb*Fam50.360 with an rmsd range of 1.2-1.5 Å over 195 Cα atoms and comparable isoelectric points of approximately 5.2. Despite high sequence identity, the pocket forming residues lacked sequence conservation, thereby distorting the pocket in all the homologs. The residues F36, T131, G175, and L187, were replaced by other amino acids in the paralogs suggesting a unique function for *Tb*Fam50.360 in comparison to its paralogs (Fig. 2.6B). This region could offer the parasite a broad mechanism to engage other cellular partners.

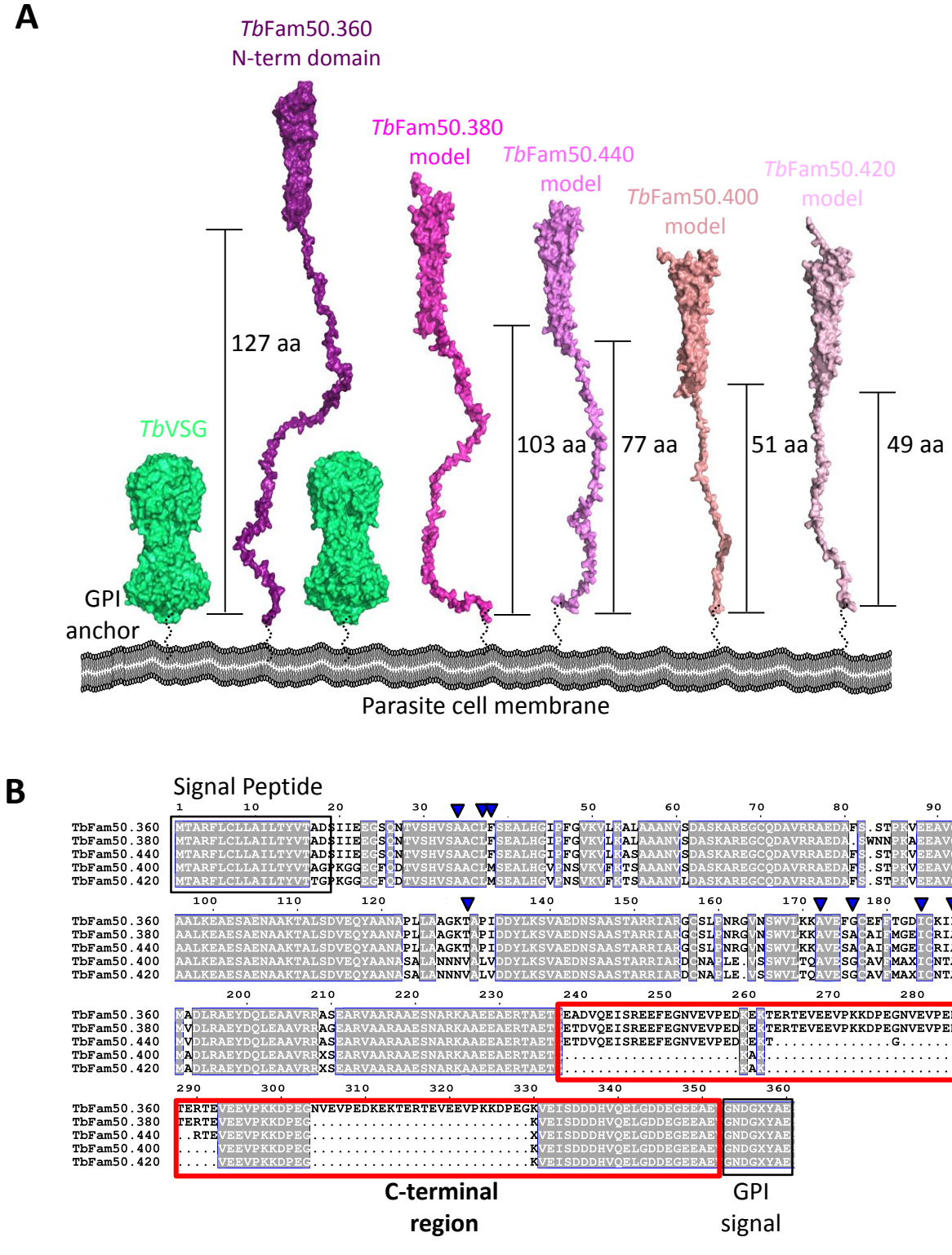

Figure 2.6: Model depicting the *Tb*Fam50.360 family of proteins in the context of the metacyclic stage of the trypanosome. (A) The C-terminal region of *Tb*Fam50.360 *(deep purple)* and full length of *Tb*Fam50.380 (*light purple*), *Tb*Fam50.440 (*pink*), *Tb*Fam50.400 *(wheat)* and

*Tb*Fam50.420 *(light pink*) were modeled using the IntFold server. *Tb*Fam50.360 N-terminal domain is shown in blue. *Tb*VSG (PDB ID: 1VSG), also expressed in the metacyclic stage, is shown in lime green. The length of the C-terminal region is also shown; *Tb*Fam50.360- 127aa, *Tb*Fam50.380-103aa, *Tb*Fam50.440-77aa, *Tb*Fam50.400-51aa, *Tb*Fam50.420-49aa. All the proteins are membrane anchored by the GPI. The figure shows that the proteins belonging to the *Tb*Fam50.360 family extend much more from the parasite cell membrane when compared to *Tb*VSG. **(B)** Sequences for *Tb*Fam50.360, *Tb*Fam50.380, *Tb*Fam50.440, *Tb*Fam50.400 and *Tb*Fam50.420 were aligned in Clustal Omega and illustrated in ESPript3.0. The red box indicates the C-terminal region. The black boxes indicate the signal peptide and the GPI signal regions. The inverted blue triangles indicate the pocket forming residues found in *Tb*Fam50.360. The residues conserved across the family are indicated in light grey. Accession numbers: *Tb*Fam50.360; *Tb*427.07.360, *Tb*Fam50.380; *Tb*427.07.380, *Tb*Fam50.440; *Tb*427.07.400, *Tb*Fam50.420; *Tb*427.07.420.

Furthermore, the presence of many lysines and arginines in the C-terminal regions suggested a susceptibility to proteolysis. Since the C-terminal regions are predicted to be disordered this scenario is highly likely. Based on this information, we decided to investigate the C- terminal region for the presence of potential proteolytic sites using the PROSPER server (Song et al., 2012). PROSPER analysis revealed the C- terminal regions of all homologs to be replete with several protease cleavage sites. This analysis corroborates in the case of *Tb*Fam50.360, where investigations into the saliva of the infected tsetse fly, found that *Tb*Fam50.360 is present in the soluble fraction (unpublished data). The reason for this cleavage remains unknown. It could be a part of a homeostatic process as many of these parasites are dying and lysing, releasing proteases and cleaving the protein. The cleavage could also be a strategic process employed by the parasite to release the adhesive function using specific proteases.

Conclusion

HAT continues to be a major health burden in Africa. Improved therapeutic strategies are desperately needed as parasites continue to develop resistance to front-line drugs and evade the development of highly effective vaccine. In this study, we report the first structural characterization of *Tb*Fam50.360 that has been hypothesized to be involved in insect-host transition based on its predominant expression in the MF of the parasite life cycle. Our structural characterization reveals the first high resolution insight into the N- terminal domain of *Tb*Fam50.360, which adopts an extended helical bundle with an overall shape that highly resembles *Tc*HpHbR, *Tb*VSG and *Tc*GARP. Structural analysis reveals a possible ligand binding function for *Tb*Fam50.360 and provides a high resolution framework for future epitope mapping studies and a foundation for targeted drug discovery. Currently binding studies are underway in an attempt to identify its molecular binding partner and establish its function. Unraveling the function of *Tb*Fam50.360 may have important implications for designing novel transmission blocking therapeutics.

Chapter 3 : Structural characterization of *Tb*PSSA2 and *Tc*CISSA

Contributions:

Sean, and Sarah Goomeshi Nobary obtained the preliminary crystals for *Tb*PSSA2 and *Tc*CISSA (Materials and Methods section 3.2.2-3.2.4). Dr. John Burke and Meredith Jenkins performed the HDX-MS (Materials and Methods section 3.2.6). Dr. Marty Boulanger solved and refined the structures (Materials and Methods section 3.2.2-3.2.4), and I completed the interpretation (Materials and methods section 3.2.1 and 3.2.5 (Refinement and validation)).

3.1 Introduction

Surface proteins are crucial in contributing to the success of trypanosomes. A majority of surface proteins identified to date are the GPI-anchored proteins. These proteins are present in abundance on the surface of the trypanosome. Due to their sheer abundance, the identification of less abundant albeit important TM proteins gets impeded. Very few proteins have been identified in life cycle stages of trypanosome so far, the majority of which have unknown functions. TM proteins play a major role in signal transduction by virtue of their structure and thus may facilitate cell differentiation and migration (Ullrich and Schlessinger, 1990, Juliano and Haskill, 1993). Investigating these proteins may provide insights into their role and potential to support the design and development of novel therapeutics and vaccines. In this chapter, I describe the study of two such TM proteins, *Tb*PSSA2 and *Tc*CISSA expressed by *T. brucei* and *T. congolense* respectively.

*Tb*PSSA2 was the first TM protein identified in the vector stages of trypanosome by screening expressed *T. brucei* PCF plasma membrane extracts with antiserum made against *T. brucei* PCF plasma membrane extracts (Jackson et al., 1993). A later study by Fragoso et al. (Fragoso et al., 2009) confirmed the expression of *Tb*PSSA2 in PCF and EMF at the protein level

using an immunoassay with genetically modified trypanosomes. Furthermore, the authors also studied the phenotype resulting from *Tb*PSSA2 knockout (KO). Though the *Tb*PSSA2 KO mutants established infections in the midgut, they were unable to progress to the salivary glands resulting in fewer and weaker salivary gland infections. Additionally, they identified a phosphorylation site at a Threonine-Proline motif (residue 305-306) that is phosphorylated by MAPK (mitogen activated protein kinase). This phosphorylation was important for facilitating the transit of *Tb*PSSA2 from the ER to the cell surface (Fragoso et al., 2009). While the function of *Tb*PSSA2 is unknown, the fact that it is an integral membrane protein with both intra- and extracellular domains suggests that this protein may act as cytoskeletal scaffold protein or a receptor/sensor.

The BLAST analysis revealed a homolog of *Tb*PSSA2 in *T. congolense* with 61% sequence identity (Fig. 3.1 B). Interestingly, a recent proteomic study conducted by the Pearson lab investigating the differential expression of surface proteins by *Trypanosoma congolense* in the different life cycle stages also revealed high expression of this protein in PF and EMF, similar to *Tb*PSSA2. They named this protein *Tc*CISSA (Congolense-Insect Stage Specific Antigen) (Eyford et al., 2011). The amino acid sequence of *Tc*CISSA revealed that a pair of glutamates replaces the Ser-Pro motif (residue 305-306) that was phosphorylated by MAPK in *Tb*PSSA2 and involved in cellular localization. The glutamates can act as a functional mimic to phosphorylation. Therefore, it is highly likely that *Tc*CISSA may have a localization pattern similar to *Tb*PSSA2.

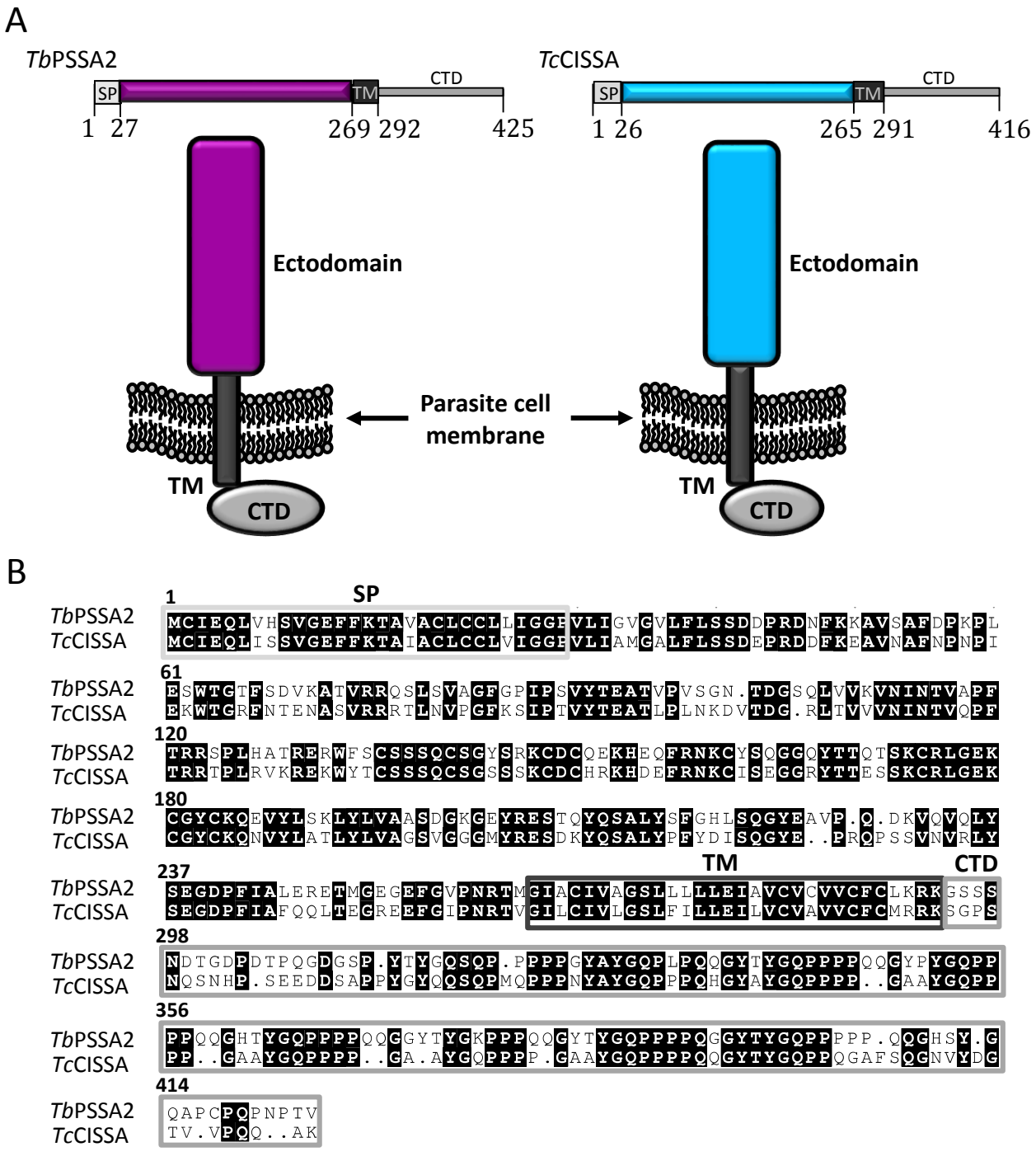

Figure 3.1: Domain organization and sequence alignment of *Tb*PSSA2 and *Tc*CISSA. (A) Top *left*: Schematic representation of *Tb*PSSA2 with predicted structural domains and signal peptide (SP). SP is shown in *white*; Transmembrane domain (TM)- *dark grey*; Cytoplasmic domain (CTD) – *light grey*. The entire ectodomain is indicated in *deep purple*. Top *right:* Schematic representation of *Tc*CISSA with the same color scheme as *Tb*PSSA2 except the ectodomain is colored in *blue*. (B) Sequences for *Tb*PSSA2 and *Tc*CISSA was aligned in Clustal Omega and illustrated in ESPript3.0. The conserved amino acids are colored black. The predicted signal peptide (SP) and transmembrane domain (TM) are boxed. (Accession numbers: *Tb*PSSA2-*Tb*427.10.11220, *Tc*CISSA-*Tc*IL3000.10.9440)

Both *Tb*PSSA2/*Tc*CISSA have relatively non-descriptive database annotation of a surface protein with no functional motifs as determined by InterProScan (Zdobnov and Apweiler, 2001)(data not shown). Therefore, functional prediction through amino acid sequence analysis provides very little insight. Both proteins are predicted to have a signal peptide for the secretory pathway, an N-terminal extracellular domain, a single transmembrane helix and a C-terminal cytoplasmic domain (Fig. 3.1 A). Besides the genetic knockout study on *Tb*PSSA2, these proteins are largely unstudied. Due to their structure, with a single membrane-spanning domain, their role in sensing and transmitting signals from tsetse to the parasite could be speculated.

In this chapter, I describe the cloning, recombinant expression and structural characterization of the extracellular domain of *Tc*CISSA and *Tb*PSSA2. Intriguingly, both structures adopt a unique bilobed architecture. Our study represents the first detailed insight into a previously understudied trypanosome surface antigen, with intriguing functional consequences.

3.2 Materials and Methods

3.2.1 Bioinformatics

Sequences of *Tb*PSSA2/*Tc*CISSA were aligned using Clustal Omega (Sievers et al., 2011) with BLOSUM62 matrix and pairwise alignment. The gap penalty at the beginning and the end was assigned a value of 1. The sequences were illustrated in ESPript 3.0 (Robert and Gouet, 2014). Accession numbers for aligned sequences were obtained from TriTrypDB (Aslett et al., 2010) and are as follows; *Tb*PSSA2-*Tb*427.10.11220, *Tc*CISSA -*Tc*IL3000.10.9440. The HINGEprot program (Emekli et al., 2008) was used to predict rigid body domains and flexible hinge regions in *Tb*PSSA2/*Tc*CISSA. The structure figures were made in PYMOL (DeLano, 2002).

3.2.2 Construct design and cloning

The sequences corresponding to *Tb*PSSA2 and *Tc*CISSA from *Trypanosoma brucei* (*Tb*427.10.11220) and *Trypanosoma congolense* (*Tc*IL3000.10.9440) were obtained from TriTrypDB (Aslett et al., 2010). *Tb*PSSA2 (Glu40 to Val262, with numbering based on the initiation methionine) and *Tc*CISSA (Ser40 to Val264) were codon optimized for *E. coli* and synthesized by GenScript. The construct for *Tb*PSSA2 begins a few residues after the predicted signal peptide, as the secondary structure prediction for the region between Gly28-Leu39 was weak with low confidence, and ended before the predicted transmembrane domain. The *Tc*CISSA construct aligns with the predicted start of the crystallization construct of *Tb*PSSA2. Signal peptides were predicted using SignalP 4.0 (Petersen et al., 2011) and the TM domain boundaries of *Tb*PSSA2/*Tc*CISSA were predicted using TMPred (Hofman, 1993). The constructs were then sub-cloned into a modified pET32a vector (Novagen) that contained a N-terminal thioredoxin (Trx) and hexa-histidine tags separated from the gene of interest by a thrombin protease site.

3.2.3 Protein expression and purification

*Tb*PSSA2 and *Tc*CISSA were produced recombinantly in *E. coli* Rosetta-gami 2 (DE3) cells (Novagen) grown in autoinduction medium (Invitrogen) from a 5% inoculum. Following four hours of growth at 37 °C and 64 hours at 16 °C, the cells were harvested and the pellet re-suspended in 20 mM HEPES pH 8.3, 1 M NaCl, 30 mM imidazole. The cell suspension was lysed using a French Press, and centrifugation was performed to remove the insoluble material. The soluble fraction was allowed to batch-bind with Ni-agarose beads at 4°C for 1 hour. Both proteins were eluted with buffer containing 250 mM imidazole, and fractions were analyzed by SDS-PAGE and pooled based on purity. The Trx-His_6 tag was removed by thrombin cleavage overnight at 18 °C, and the proteins were further purified by size exclusion chromatography (Superdex 16/600 75) in 20mM HEPES pH 6.8, 150mM NaCl. Fractions containing monomeric *Tc*CISSA and *Tb*PSSA2 were pooled and further purified by cation-exchange chromatography on a HiTrap SP FF column (GE Healthcare) in 20mM Tris-HCl pH 6.2, 10mM NaCl. *Tc*CISSA and *Tb*PSSA2 were eluted with an increasing concentration of NaCl. The protein concentrations were determined by absorbance at 280 nm with calculated extinction coefficients of 28000 M^{-1} cm^{-1} and 29380 M^{-1} cm^{-1} for *Tb*PSSA2 and *Tc*CISSA respectively (The extinction coefficients were calculated using Protparam (Gasteiger et al., 2005)).

3.2.4 Crystallization and Data Collection

*Tb*PSSA2

Crystals of purified recombinant *Tb*PSSA2 were initially identified in the Index screen (Molecular Dimensions) and MCSG-1 screen (Microlytic) using the sitting drop method at 18 °C.

MCSG-1 produced highest quality crystals. The final drops consisted of 0.3 µL of *Tb*PSSA2 at 15mg ml^{-1} with 0.3 µL of reservoir solution and were equilibrated against 60 µL of reservoir solution. Diffraction quality crystals grew in 48 hours in 0.2M $(NH_4)_2SO_4$, 0.1M Bis-tris pH 6.5, and 25% PEG 3350. A single crystal was looped, cryopreserved in 25% glycerol for 20 seconds, and flash-cooled to -173.15 °C directly in the cryostream. Diffraction data were collected on beamline 9-2 at Stanford Synchrotron Radiation Laboratory (SSRL) at a wavelength of 0.9795 Å.

***Tc*CISSA**

Initial crystallization trials for *Tc*CISSA were performed using the commercial screens Index and PEG/Ion screen (Hampton Research), Wizard I and II and MCSG-1 screens (Emerald BioSystems). The protein was set in 96-well sitting drop Intelli-Plates (Hampton Research) with different drop ratios containing of 0.2 or 0.3 µl protein solution (25$mg.ml^{-1}$) and 0.2µl reservoir solution and equilibrated against 55µl reservoir solution. A single crystal of *Tc*CISSA was observed in 20% PEG 3350 with 0.2*M* ammonium sulfate after overnight incubation at 18 °C; the crystals developed equally well using both drop ratios. Diffraction data were collected on beamline 9-2 at Stanford Synchrotron Radiation Laboratory (SSRL) at a wavelength of 0.9795 Å.

3.2.5 Data processing, structure solution and refinement

***Tb*PSSA2**

Diffraction data for native and iodide derivative crystals were processed to 1.65 and 1.85 Å resolution, respectively, using Imosflm (Battye et al., 2011, Leslie, 1992) and Scala (Evans, 2006) in the CCP4 suite of programs (Winn et al., 2011). The structure of *TbPSSA2* was phased by Iodine derived single wavelength anomalous dispersion. A total of twenty-eight iodide sites

was identified and refined using the SHELX C/D/E pipeline (Sheldrick, 2010). High-quality phases were obtained after density modification in dm (Cowtan, 1994) and enabled building and registering of 60% of the backbone using buccaneer (Cowtan, 2007). The remaining structure was built manually and used as a molecular replacement model for the higher resolution native data using Phaser (McCoy, 2007). Solvent atoms were selected using COOT (Emsley and Cowtan, 2004) and refined in Phenix Refine (Afonine et al., 2012). The overall structure of *Tb*PSSA2 was refined to an R_{free} of 20.4%. Stereo-chemical analysis performed with PROCHECK and SFCHECK in CCP4 (Collaborative Computational Project, 1994) showed excellent stereochemistry with more than 99% of the residues found in the conformationally favorable regions and no residues modeled in the unfavourable regions of the Ramachandran plot. For calculating $R_{free,}$ 5% of the reflections were set aside. Data collection and refinement statistics are presented in Table 3.1.

***Tc*CISSA**

Diffraction data for native crystals were processed to 2.45 Å resolution using the methods described for *Tb*PSSA2. Initial phases were obtained by molecular replacement using Phaser (McCoy, 2007) with *Tb*PSSA2, pruned with CHAINSAW (Schwarzenbacher et al., 2004) to better reflect the *Tc*CISSA sequence, as a search model. Refinement, structural validation, and stereochemical analysis was performed similar to *Tb*PSSA2 described above. Data collection and refinement statistics are presented in Table 3.1.

Table 3.1: Data collection and refinement statistics

	***Tb*PSSA2_I SAD**	***Tb*PSSA2**	***Tc*CISSA**
A. Data collection statistics			
Spacegroup	P1	P1	$P2_1$
a, b, c (Å)	32.13, 57.82, 60.61	31.82, 57.81, 61.24	75.81, 87.65, 98.89
α, β, γ (deg.)	74.24, 76.45, 89.76	73.55, 76.43, 89.98	90, 113.27, 90
Wavelength	1.1807	1.1807	0.9795
Resolution range (Å)	56.59 – 1.80 (1.84 – 1.80)	56.95 – 1.65 (1.67 – 1.65)	39.47 – 2.45 (2.49 – 2.45)
Measured reflections	184718 (10800)	171475 (8353)	146837 (11384)
Unique reflections	35093 (2079)	45400 (2236)	42734 (2110)
Redundancy	5.3 (5.2)	3.8 (3.7)	3.4 (3.5)
Completeness (%)	93.2 (91.9)	92.3 (90.7)	97.6 (98.3)
I/σ(I)	11.3 (2.4)	10.3 (2.0)	17.8 (1.2)
R_{merge}[a]	0.087 (0.822)	0.066 (0.542)	0.044 (0.517)
$CC_{1/2}$	0.997 (0.771)	0.995 (0.739)	0.960 (0.821)
Number of I sites	28	N/A	N/A
B. Refinement statistics			
Resolution (Å)		47.12 – 1.65	39.48 – 2.45
R_{crys}*Tb*/R_{free}[c]		0.170/0.204	0.257/0.289
No. of atoms			
Protein (A/B/C/D/U)		1728/1692	1630/1568/1468/1486/74
Sulfate/Glycerol		5	40/12
Solvent		492	68
B-values ($Å^2$)			
Protein (A/B/C/D/U)		21.2/24.8	65.8/70.0/87.1/84.8/68.4
Sulfate/Glycerol		31.0	75.3/62.4
Solvent		31.0	51.9
r.m.s. deviation from ideality			
Bond lengths (Å)		0.006	0.003
Bond angles (deg.)		0.991	0.763
Ramachandran statistics (%)			
Most favoured		98.8	96.0
Allowed		1.2	4.0
Disallowed		0.0	0.0

Values in parentheses are for the highest resolution shell

[a] $R_{merge} = \sum_{hkl}\sum_i|I_{hkl,i} - [I_{hkl}]| / \sum_{hkl}\sum_i I_{hkl,i}$, where $[I_{hkl}]$ is the average of symmetry related observations of a unique reflection

[b] $R_{cryst} = \sum|F_{obs} - F_{calc}|/\sum F_{obs}$, where F_{obs} and F_{calc} are the observed and the calculated structure factors, respectively

[c] R_{free} is R using 5% of reflections randomly chosen and omitted from refinement

3.2.6 Hydrogen- Deuterium Exchange Mass Spectroscopy (HDX-MS) measurements

*Tc*CISSA was prepared at a concentration of 10μM in a buffer composed of 20 mM Tris/HCl, 100 mM NaCl (pH 7.0). The HDX was initiated by mixing 10 μl of purified protein with 40 μl of the 2H2O buffer (10 mM HEPES pH 7.5, 50 mM NaCl, 2 mM TCEP, 92% $^{2}H_2O$); the mixtures were incubated at various time intervals 0-300s on ice. At the indicated time-points, the mixtures were quenched by adding ice-cold quench buffer (final concentration 0.6 M guanidine-HCl, 0.8% formic acid). For non-deuterated samples, 10 μl of the purified protein was mixed with 40 μl of $^{1}H_2O$ buffer (20 mM Tris-HCl, pH 7.0, and 100 mM NaCl in water) and ice-cold quench buffer (final concentration 0.6 M guanidine-HCl, 0.8% formic acid) was added. The quenched samples were injected into UPLC (ultra-pressure liquid chromatography) and digested online by passing them through an immobilized pepsin column (Applied Biosystems; porosyme, 2-3131-00) at a flow rate of 130 μL/min for 5 min with with a 5-36% B (buffer A 0.1% formic acid, buffer B 100% acetonitrile) gradient over 20 min. Peptide fragments were subsequently collected on a VanGuard precolumn trap for de-salting. The trap was subsequently eluted in line with an Acquity 1.7 μm particle, 100 × 1 mm2 C18 UPLC column (Waters), with a 5-36% B (buffer A 0.1% formic acid, buffer B 100% acetonitrile) gradient over 20 min.

Peptides were identified by running tandem MS/MS experiments using a 5–36% B gradient over 120 min. The resulting MS/MS datasets were analyzed with the Mascot search within Mascot distiller (Matrix Science). The MS tolerance in mascot was set to 3 ppm with an MS/MS tolerance at 0.1 Da. Peptides showing a Mascot score >20 were analyzed using HD-Examiner Software (Sierra Analytics). The full list of peptides was then validated manually by comparing it with a non-deuterated protein's MS scan to check for the correct m/z state, and the presence of overlapping peptides. All ambiguous peptides were excluded from further analysis. The HD-

Examiner software performed the first round of analysis and identification, but it was ensured that all peptides (both deuterated and non-deuterated) were verified manually at every state and time point to check the correct charge state, m/z range, presence of overlapping peptides, and any deviation from the expected retention time. Corrections for back exchange were generated from a fully deuterated sample. All experiments were performed in triplicates.

3.3 Results and Discussions

3.3.1 The ectodomains of *Tb*PSSA2 and *Tc*CISSA adopt previously uncharacterized architecture

Recombinantly produced N-terminal extracellular domains of *Tb*PSSA2 and *Tc*CISSA from *E. coli* were initially purified using nickel affinity chromatography, followed by thrombin cleavage to remove the TRX tag. The size exclusion profile indicated a monomeric organization for both *Tb*PSSA2 and *Tc*CISSA (Fig. 3.2A). Native *Tb*PSSA2 and *Tc*CISSA were crystallized but in the absence of sequence identity with reported structures, no suitable molecular replacement models were identified. The phases for *Tb*PSSA2 were ultimately obtained from an iodide based crystal soak using single wavelength anomalous dispersion. Iodine phasing solved the structure of *Tb*PSSA2 to a resolution of 1.65 Å. Two molecules were present in the asymmetric unit (AU) of the primitive unit cell. While *Tb*PSSA2 crystallized in space group P1 with two molecules in AU, *Tc*CISSA crystallized in space group $P2_1$ with four molecules in AU. The structure of *Tb*PSSA2 emerged as a suitable molecular replacement model for *Tc*CISSA, which crystallized as a monomer with four molecules in the AU of the primitive monoclinic unit cell. The structure of *Tc*CISSA was solved to a resolution of 2.45 Å.

Both *Tb*PSSA2 and *Tc*CISSA have similar overall structure, in agreement with their sequence homology (sequence identity ~61%). They adopt a bi-lobed architecture consisting of 2 lobes (hereafter L1 and L2) separated by a hinge region (Fig. 3.2B). Since several residues (13 in *Tb*PSSA2 and 10 in *Tc*CISSA) in the C-terminal were hard to model, predicting the precise orientation of *Tb*PSSA2 and *Tc*CISSA with respect to the membrane was difficult.

L1 is the larger lobe consisting of 2 polypeptide segments from both the amino terminus (residues 40-123 in *Tb*PSSA2 and residues 44-123 in *Tc*CISSA) and carboxyl terminus (residues 188-256 in *Tb*PSSA2 and residues 188-255 in *Tc*CISSA). Interestingly, both N and C terminus fall at the same end. The intervening segment (residues 126-186) constitutes L2. A single elongated β-strand, β8, spans the entire length of the protein, including both lobes and hinge region. This strand extends to a maximum length of 49 Å.

The individual lobes L1 and L2 of *Tb*PSSA2 and *Tc*CISSA exhibit a high correlation with each other with a root mean square deviation (RMSD) ranging from 0.234-0.321 Å over 113 Cα atoms (L1) and an RMSD range of 0.31-0.42 Å over 52 Cα in L2. L1 adopts an identical configuration in both *Tb*PSSA2/*Tc*CISSA. The individual lobes L1 exhibit a high correlation with each other with a root mean square deviation (RMSD) ranging from 0.234-0.321 Å over 113 Cα atoms. N-terminal and C-terminal α-helical periphery and a 7-stranded β-sandwich core consisting of β1, 2, 3, 4, 5, 9, 10 and 11 define the overall structure of L1. The β sandwich packs against β-8 on one face in addition to a 4-residue α helix (α-2) on the opposite face. L2 forms a β-α-β motif. The β strands, β-7, and β-6, coordinate the β-8 linker. This lobe is the most stable between the two lobes stabilized by four intramolecular disulphide linkages. Two disulphide bonds are formed in the loop connecting β-6 and α-3 while the remaining 2 are formed with the linker β-8 (Fig. 3.2C).

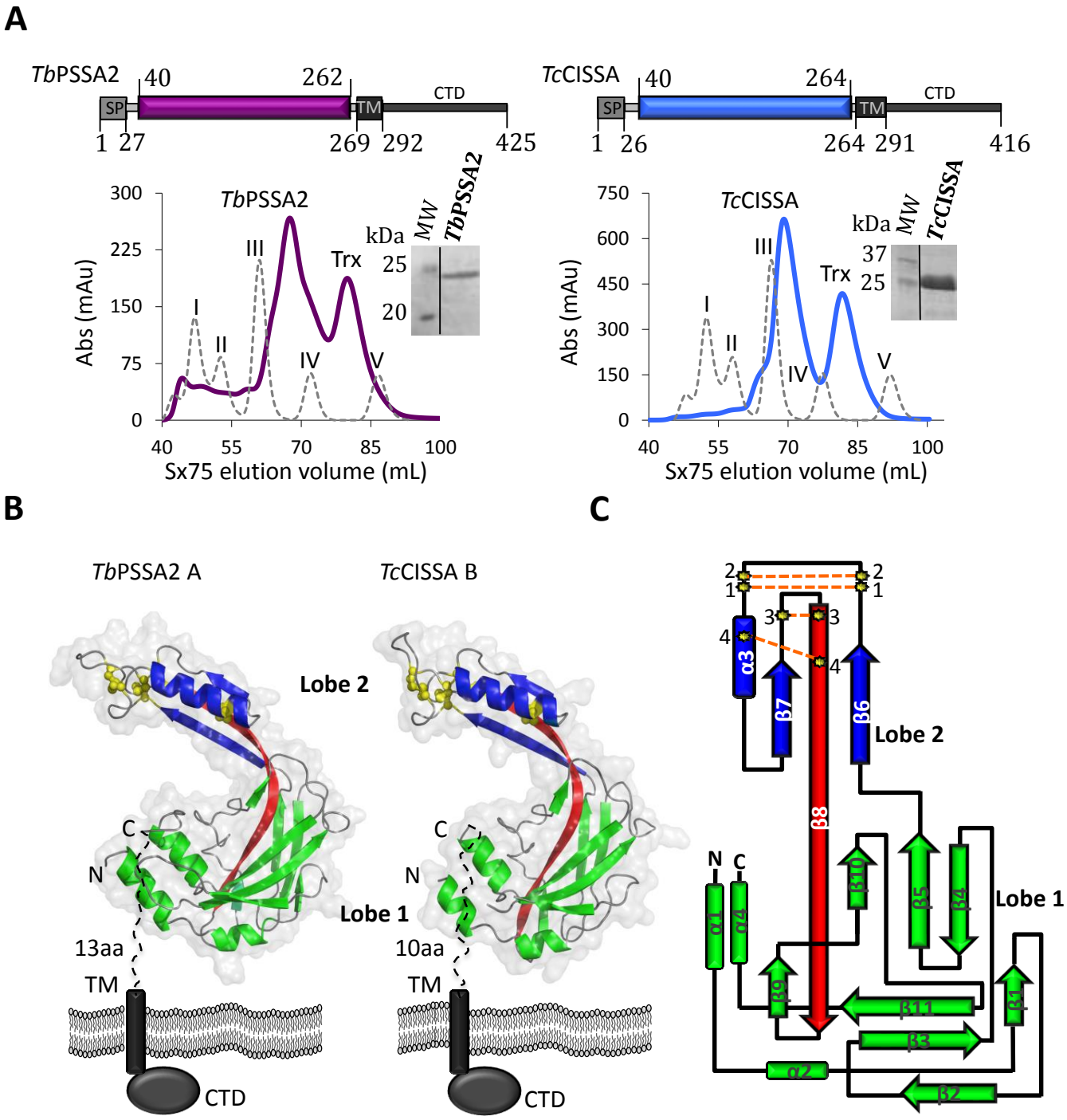

Figure 3.2: Both *Tb*PSSA2 and *Tc*CISSA behave as monomers in solution and adopt a unique bilobed architecture. (A) *Left*: Schematic representation of *Tb*PSSA2 with predicted structural domains and signal peptide. Construct for crystallization (Glu^{40} to Val^{262}) is indicated in *deep purple*. Bottom *left*: Superdex 75 size exclusion chromatogram of *Tb*PSSA2 construct (*deep purple*). The peak for *Tb*PSSA2 and thioredoxin (Trx) is indicated. The protein standards used were conalbumin (75 kDa; *peak I*), ovalbumin (43 kDa; *peak II*), carbonic anhydrase (29 kDa; *peak III*), ribonuclease (13.7kDa; *peak IV*) and aprotinin (6.500kDa; *peak V*). *Inset*, SDS-PAGE

of purified *Tb*PSSA2. Top *right*: Schematic representation of *Tc*CISSA as shown in A with construct for crystallization (Ser^{40} to Val^{264}) indicated in *blue.* Bottom *right:* Superdex 75 size exclusion chromatogram of *Tc*CISSA construct (*blue*) with SDS-PAGE of purified *Tc*CISSA indicated in *inset.* (B) (A) Secondary structure depiction of *Tb*PSSA2 *(left)* and *Tc*CISSA is shown in the predicted orientation to the parasite membrane. The chain A of *Tb*PSSA2 and chain B of *Tc*CISSA is shown as they were the most extensively modeled chains. 2 lobes (L1 and L2) connected by a hinge region divide the structure. L1 is colored in *green*, L2 in *blue*. The segment connecting both the strands is colored in *red*. The four disulphide bonds in L2 are indicated in *yellow ball and sticks*. (B) Topology diagram of *Tb*PSSA2 and *Tc*CISSA colored as in (A).

To investigate the similarity of *Tb*PSSA2 and *Tc*CISSA to other structurally characterized proteins, a DaliLite search (Holm and Rosenstrom, 2010) was performed for the entire crystallization construct. Dali returned hits with low Z scores (Z-score<4.0) and weak alignments with RMSD scores greater than 2 A. Since the entire structure returned low Dali scores, a search with the individual lobes L1 and L2 alone was carried out. However, both L1 and L2 also gave results similar to the entire structure. Therefore, with Dali not returning any strong matches, it appears that the structures adopted by *Tb*PSSA2 /*Tc*CISSA are previously uncharacterized.

3.3.2 *Tb*PSSA2 and *Tc*CISSA have conformational flexibility

There are 2 copies of *Tb*PSSA2 and 4 copies of *Tc*CISSA within the ASU. Both *Tb*PSSA2 and *Tc*CISSA exhibited flexibility among the individual monomers in the crystal structure. This was evidenced by a higher RMSD value overlay of the bilobed monomers (0.35-0.52 Å in *Tb*PSSA2 over 137 C_α and 0.3-0.57 Å in *Tc*CISSA over 132 C_α), while L1 overlay with a lower RMSD (0.21-0.25Å over 117-119 C_α in *Tb*PSSA2 and 0.27-0.31 Å over 107-119 C_α in *Tc*CISSA). The inclusion of L2 resulted in an increase in RMSD thereby suggesting an inter-lobe flexibility, visually supported by a L1 overlay and the observation of resulting relative rotation in L2 (Fig. 3.3A, B). Moreover, hinge prediction bioinformatics algorithm (Emekli et al., 2008) identified residues H126, A127, T128, R129, E130, K184, Q185, and N186 in *Tb*PSSA2; R126, V127, K128, R129, E130, K184, Q185, and N186 in *Tc*CISSA as the hinge residues with very high probability.

On comparing the degree of flexibility between different monomers of *Tb*PSSA2 and *Tc*CISSA, it was found that the individual L2 rotates to an angle of 5° (into the plane of the paper) in *Tb*PSSA2 (Fig. 3.3A), whereas it rotates to maximum angle of 30° (into the plane of the paper) in *Tc*CISSA. The existence of distinct conformers of *Tb*PSSA2 and *Tc*CISSA suggests an inherent flexibility around the lobes. The range of flexibility that we observe is generated by crystal packing forces.

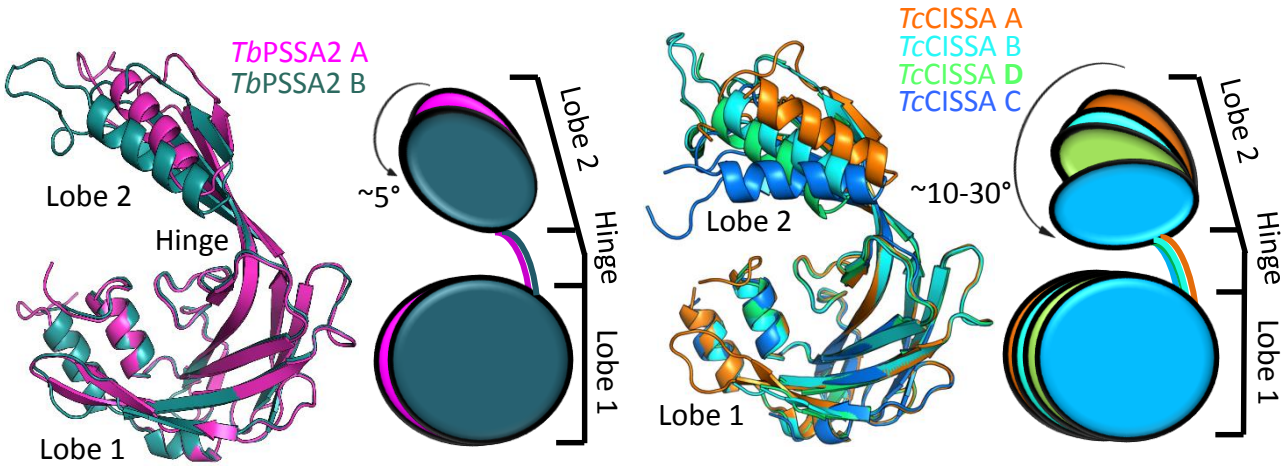

Figure 3.3: *Tb*PSSA2 and *Tc*CISSA display conformational flexibility between their lobes. (A) *Left*: Structural superposition of 2 monomers of *Tb*PSSA2 (chain A- *deep purple*; chain B- *dark teal*). The hinge and lobes are indicated. The lobes L2 are displaced to a maximum of 6 Å. *Right*: pictorial representation of chain A and B overlay depicting the degree of rotation (~5°) between the L2. (B) Structural superposition of the four monomer chains of *Tc*CISSA (chain A- *orange*; chain B- *cyan*; chain C- *marine*; chain D- *limegreen*) showing a maximum swing of ~18 Å between the lobes L2. *Right*: pictorial representation of *Tc*CISSA same as in (A) with the angle rotation ranging from ~10-30°.

To investigate whether the dynamic structural conformation observed in the crystal structure are also present in solution, we performed Hydrogen-deuterium exchange (HDX). HDX-MS measures the rate of exchange of peptide amide hydrogen with deuterium in the solvent. The rate of deuterium uptake is affected by secondary structure, flexibility and the dynamics of the

protein conformation. Flexible regions show a high exchange rate, whereas ordered regions such as regions with secondary structures show slow exchange rate. We performed HDX-MS on *Tc*CISSA since it showed the highest conformational flexibility (Fig. 3.4- left panel). Mass spectrometric identification of peptic peptides from *Tc*CISSA resulted in about ~95% of sequence coverage. HDX-MS clearly suggested a possibility of multiple conformers in solution consistent with the crystal structure. Interestingly HDX-MS did not reveal flexibility between the lobes albeit it revealed flexibility around the β6-α3 region (lobe 2) with a distinct slow and a fast exchanging population (Fig. 3.4- right panel). Though the reason for this flexibility is unknown, it is tempting to speculate that this may be in response to molecular recognition of binding partner or in facilitating ligand binding. Further experimentations should be performed to confirm this.

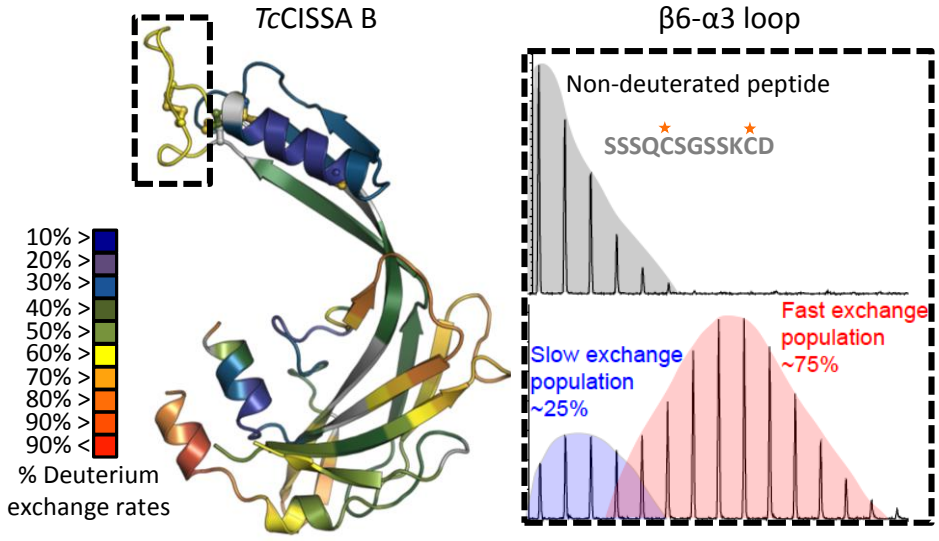

Figure 3.4: HDX-MS reveals conformational herogeneity of *Tc*CISSA in solution. *Left Panel*: The figure shows the heat map at 300s^{-} overlaid on to the X-ray crystal structure of *Tc*CISSA. The β6-α3 loop is indicated by the dotted box. The color key shows the HDX levels. *Right panel:* Shows the analysis of the β6-α3 loop region. The HDX-MS at the 300s^{-} exchange time point is shown. HDX-MS clearly indicates the existence of 2 distinct populations in the loop region (blue-fast exchanging; red-slow exchanging), consistent with the crystal structure. The non-deuterated peptide is shown in grey and disulphide connectivity is shown as orange star.

3.3.3 Hinge region in *Tc*CISSA may be capable of closing around a ligand

The hinge region in chain A of *Tc*CISSA accommodates a peptide with unmapped side chains due to difficulties in registering the sequence in electron density (Fig. 3.5A). This peptide nicely wraps around the hinge region, thereby mimicking a ligand-binding region. The contours around the strand-binding region reveal intriguing structural evidence that suggests a binding event at the hinge region (Fig. 3.5A). Interestingly, this is the only chain in *Tc*CISSA that accommodates a strand and coordinates a sulphate. A closer inspection of the hinge region revealed an R128 and K130 pair at the base of the hinge where the divergence between different monomers of *Tc*CISSA begins. In chain A, R128 is pinned to sulphate along with K130 (Fig. 3.5B-middle panel). This sulphate might be a side chain of a biological binding partner. However, the other conformers do

not accommodate this strand because R128 and K130 (to a lesser extent) in these conformers have steric hindrance with this strand (Fig. 3.5B-right panel).

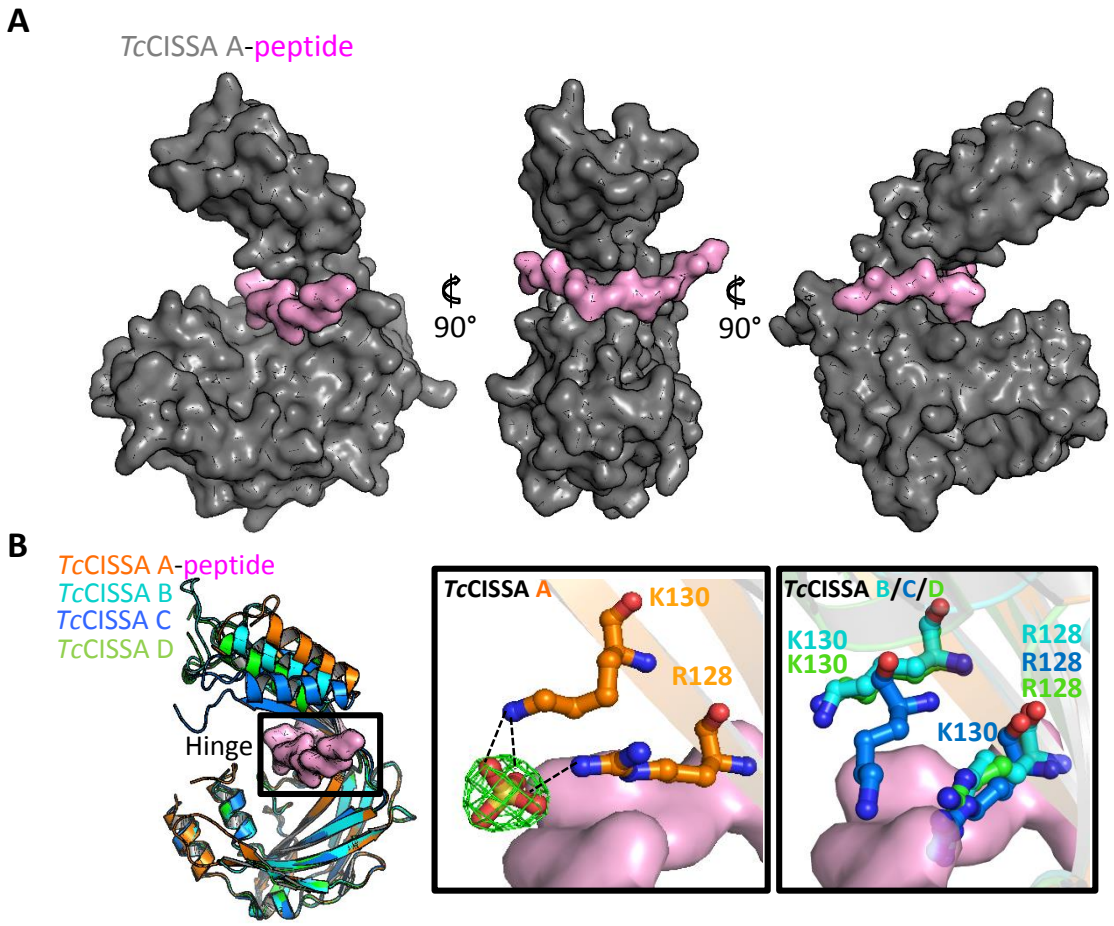

Figure 3.5: Analysis of the hinge region in *Tc*CISSA. The figure shows the analysis of strand bound region in *Tc*CISSA. (A) Surface representation of the *Tc*CISSA Chain A (*grey*) highlighting the assembly of bound strand (*lightpink).* It can be seen that the strand nicely wraps around the protein. (B) (*Left panel*), shows superimposition of different monomers of *Tc*CISSA with the same coloring scheme used in (Fig. 3.3). The hinge region is indicated by black box. *Middle panel*, Top: R127 and K130 in the hinge region of chain A coordinating with sulphur is shown. The sulphur is depicted as stick model with a 2Fo-Fc electron density map (*green* mesh) contoured at 1.5σ. *Right panel*: shows R128 and K130 of chains B, C, and D. These amino acids sterically clash with the strand and not coordinated by sulphur.

While the exact reason for orientations between chains around the hinge region is unknown, the orientation of the membrane distal loop could be having an impact in the hinge region organization. Biologically, this hinge region may be involved in interacting with a binding partner, which is consistent with it being on the surface and recognizing a partner.

Conclusions

Surface proteins play an important role in enabling a successful transmission of the parasites causing trypanosomiasis. In this study, we report the first structural characterization of the ectodomains of *Tb*PSSA2 and *Tc*CISSA, two membrane spanning surface proteins expressed differentially in the PF and EMF stages of the trypanosome life cycle. Currently very little is known about the role of these proteins. A previous study showed that the KO of *Tb*PSSA2 results in a phenotype incapable of infecting the salivary glands (Fragoso et al., 2009). A recent proteomic study showed an up-regulation of *Tc*CISSA in PF and EMF (Eyford et al., 2011). These are intriguing pieces of evidence to suggest their possible role in the molecular crosstalk between tsetse and trypanosome. Our structures reveal a previously uncharacterized bilobed architecture for these proteins with an inherent flexibility between the lobes suggesting that they may be able to accommodate a binding partner. Interestingly, we found that *Tc*CISSA accommodated a peptide between the lobes, which re-enforces our hypothesis that these proteins may bind their ligands between the lobes. In conclusion, we have taken the proteins previously postulated to be involved in sensing and shown that through a unique architecture has an inherent capacity to bind a ligand and have their flexible lobes re-orient relative to each other. Our structures give an insight of how these proteins may be involved in sensing and their capacity to act as a sensor to recognize the extracellular environment. Ultimately, identifying the binding partners of these proteins would be key to fully understanding their functions.

Chapter 4 : General discussion and future studies

The pathogens causing VBDs express an arsenal of surface proteins, enabling their successful transmission. However, the molecular mechanism underlying their transmission is unknown. Therefore, the aim of my thesis was to understand the molecular mechanism underlying vector-pathogen interaction. Using tsetse-trypanosome as the model system, I elucidated the structure of important surface proteins involved in the transmission of trypanosomes inside tsetse and attempted to infer the functions of these proteins using the structure.

Chapter 2 describes the structural characterization of *Tb*Fam50.360. It adopted a three helical architecture, similar to previously characterized trypanosome surface proteins. Structural analysis revealed the presence of a hydrophobic pocket that may play a role in ligand binding. Moreover, the presence of a C-terminal linker region that may allow the protein to extend from the VSG layer on the parasite surface and in interacting with the ligand, makes us believe that *Tb*Fam50.360 is likely to be an adhesin. However, in the absence of a binding partner, we cannot test this hypothesis. Therefore, future studies must involve in identifying binding partners to this protein. Currently, our collaborators are trying to identify possible *Tb*Fam50.360 binding proteins by pull-down experiments. Studies may also involve investigating the C- terminal region using NMR. Additionally, studies can also be carried out by truncating the C-terminal region, and observing its effect on parasite survival and differentiation. Insights into the characterization of the C - terminal region may shed more light on the function of this novel protein.

Chapter 3 describes the structural characterization of *Tb*PSSA2 and *Tc*CISSA. Structural analysis revealed that both proteins adopted a previously uncharacterized bilobed architecture with one of the lobes articulating movements in and out of the plane of the paper. Bilobed proteins

described in the past bind their partner between their lobes. Interestingly, one of the monomers of *Tc*CISSA (chain A) crystallized with a strand suggesting a capacity to bind ligand. Since the electron density of the strand was ambiguous, it was not possible to identify the ligand. In such a scenario, characterizing the binding partners is imperative. Our collaborators are working to identify possible binding partners using pull downs.

In this study, we structurally characterized important surface proteins previously hypothesized to play an important role in tsetse-trypanosome interaction. These proteins can be added to the repertoire of structurally characterized surface proteins expressed by *T. brucei* and *T. congolense*. Our structural analysis shows that they adopt interesting structural architecture and based on the structure we were able to identify putative ligand-binding regions. However, deciphering their functions was difficult without the presence of a binding partner. Once the binding partners are identified, co-crystallization studies with these partners would shed more insight into the mechanism underlying binding and also their function.

Bibliography

ACOSTA-SERRANO, A., VASSELLA, E., LINIGER, M., RENGGLI, C. K., BRUN, R., RODITI, I. & ENGLUND, P. T. 2001. The surface coat of procyclic Trypanosoma brucei: programmed expression and proteolytic cleavage of procyclin in the tsetse fly. *Proceedings of the National Academy of Sciences,* 98**,** 1513-1518.

AFONINE, P.V, GROSSE-KUNSTLEVE, R.W, ECHOLS, N, HEADD, J.J, MORIARTY N.W, MUSTYAKIMOV, V, TERWILLIGER, T.C, URZHUMTSEV, A, ZWART, P.H, AND ADAMS P.D. 2012. Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr D Biol Crystallogr* 68, 352-67.

AGNANDJI, S. T., LELL, B., SOULANOUDJINGAR, S. S., FERNANDES, J. F., ABOSSOLO, B. P., CONZELMANN, C., METHOGO, B., DOUCKA, Y., FLAMEN, A. & MORDMÜLLER, B. 2011. First results of phase 3 trial of RTS, S/AS01 malaria vaccine in African children. *The New England journal of medicine,* 365**,** 1863-75.

ALLSOPP, R. 2001. Options for vector control against trypanosomiasis in Africa. *Trends in parasitology,* 17**,** 15-19.

ALPHEY, L., BENEDICT, M., BELLINI, R., CLARK, G. G., DAME, D. A., SERVICE, M. W. & DOBSON, S. L. 2010. Sterile-insect methods for control of mosquito-borne diseases: an analysis. *Vector-Borne and Zoonotic Diseases,* 10**,** 295-311.

ASLETT, M., AURRECOECHEA, C., BERRIMAN, M., BRESTELLI, J., BRUNK, B. P., CARRINGTON, M., DEPLEDGE, D. P., FISCHER, S., GAJRIA, B. & GAO, X. 2010. TriTrypDB: a functional genomic resource for the Trypanosomatidae. *Nucleic acids research,* 38**,** D457-D462.

BARRY, J. D. & MCCULLOCH, R. 2001. Antigenic variation in trypanosomes: enhanced phenotypic variation in a eukaryotic parasite. *Advances in parasitology,* 49**,** 1-70.

BATTYE, T. G., KONTOGIANNIS, L., JOHNSON, O., POWELL, H. R. & LESLIE, A. G. 2011. iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr D Biol Crystallogr,* 67**,** 271-81.

BAYNE, R. A., KILBRIDE, E. A., LAINSON, F. A., TETLEY, L. & BARRY, J. D. 1993. A major surface antigen of procyclic stage Trypanosoma congolense. *Molecular and biochemical parasitology,* 61**,** 295-310.

BEECROFT, R. P., RODITI, I. & PEARSON, T. W. 1993. Identification and characterization of an acidic major surface glycoprotein from procyclic stage Trypanosoma congolense. *Molecular and biochemical parasitology,* 61**,** 285-294.

BOUTEILLE, B., OUKEM, O., BISSER, S. AND DUMAS, M., 2003. Treatment perspectives for human African trypanosomiasis. *Fundamental & clinical pharmacology,* 2**,** 171-181.

BRUN, R., BLUM, J., CHAPPUIS, F. & BURRI, C. 2010. Human african trypanosomiasis. *The Lancet,* 375**,** 148-159.

BUDD, L. T. 1999. DFID-funded tsetse and trypanosome research and development since 1980. *Economic analysis,* Aylesford, UK, DFID Livestock Production, Animal Health and Natural Resources Systems Research Programmes. 123 pp.

BURRI, C. 2010. Chemotherapy against human African trypanosomiasis: is there a road to success? *Parasitology,* 137**,** 1987-1994.

BÜTIKOFER, P., VASSELLA, E., BOSCHUNG, M., RENGGLI, C. K., BRUN, R., PEARSON, T. W. & RODITI, I. 2002. Glycosylphosphatidylinositol-anchored surface molecules of Trypanosoma congolense insect forms are developmentally regulated in the tsetse fly. *Molecular and biochemical parasitology,* 119**,** 7-16.

CHAPPUIS, F., LOUTAN, L., SIMARRO, P., LEJON, V. & BÜSCHER, P. 2005. Options for field diagnosis of human African trypanosomiasis. *Clinical microbiology reviews,* 18**,** 133-146.

COLLABORATIVE COMPUTATIONAL PROJECT, N. 1994. *Acta Crystallogr. Sect. D Biol. Crystallogr.,* 50**,** 760–763.

COUSTOU, V., GUEGAN, F., PLAZOLLES, N. & BALTZ, T. 2010. Complete in vitro life cycle of Trypanosoma congolense: development of genetic tools. *PLoS Negl Trop Dis,* 4**,** e618.

COWTAN, K. 1994. DM: an automated procedure for phase improvement by density modification. *Joint CCP4 and ESF-EACBM newsletter on protein crystallography,* 31**,** 34-38.

COWTAN, K. 2007. Fitting molecular fragments into electron density. *Acta Crystallographica Section D: Biological Crystallography,* 64**,** 83-89.

CROFT, S. 1997. The current status of antiparasite chemotherapy. *Parasitology,* 114**,** 3-15.

DE VRIES, P. J. & DIEN, T. K. 1996. Clinical pharmacology and therapeutic potential of artemisinin and its derivatives in the treatment of malaria. *Drugs,* 52**,** 818-836.

DE VALDEZ, M.R.W., NIMMO, D., BETZ, J., GONG, H.F., JAMES, A.A., ALPHEY, L. AND BLACK, W.C., 2011. Genetic elimination of dengue vector mosquitoes. *Proceedings of the National Academy of Sciences,* 108, pp.4772-4775.

DELANO, W. L. 2002. The PyMOL molecular graphics system. *DeLano Scientific,* San Carlos, CA, USA.

DELESPAUX, V. & DE KONING, H. P. 2007. Drugs and drug resistance in African trypanosomiasis. *Drug Resistance Updates,* 10**,** 30-50.

DODSON, E. J., WINN, M. & RALPH, A. 1997. Collaborative Computational Project, number 4: providing programs for protein crystallography. *Methods in enzymology,* 277**,** 620-634.

EMEKLI, U., SCHNEIDMAN-DUHOVNY, D., WOLFSON, H. J., NUSSINOV, R. & HALILOGLU, T. 2008. HingeProt: automated prediction of hinges in protein structures. *Proteins,* 70**,** 1219-1227.

EMSLEY, P. & COWTAN, K. 2004. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr,* 60**,** 2126-2132.

EVANS, P. 2006. Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr,* 62**,** 72-82.

EYFORD, B. A., SAKURAI, T., SMITH, D., LOVELESS, B., HERTZ-FOWLER, C., DONELSON, J. E., INOUE, N. & PEARSON, T. W. 2011. Differential protein expression throughout the life cycle of Trypanosoma congolense, a major parasite of cattle in Africa. *Molecular and biochemical parasitology,* 177**,** 116-125.

FRANZ, A.W., SANCHEZ-VARGAS, I., ADELMAN, Z.N., BLAIR, C.D., BEATY, B.J., JAMES, A.A. AND OLSON, K.E., 2006. Engineering RNA interference-based resistance to dengue virus type 2 in genetically modified Aedes aegypti. *Proceedings of the National Academy of Sciences,* 103, pp.4198-4203.

FRAGOSO, C. M., BURKARD, G. S., OBERLE, M., RENGGLI, C. K., HILZINGER, K. & RODITI, I. 2009. PSSA-2, a Membrane-Spanning Phosphoprotein of Trypanosoma brucei, Is Required for Efficient Maturation of Infection. *Plos One,* 4**,** e7074.

FREYMANN, D., DOWN, J., CARRINGTON, M., RODITI, I., TURNER, M. & WILEY, D. 1990. 2.9 Å resolution structure of the n-terminal domain of a variant surface glycoprotein from Trypanosoma brucei. *Journal of molecular biology,* 216**,** 141-160.

FU, G., LEES, R.S., NIMMO, D., AW, D., JIN, L., GRAY, P., BERENDONK, T.U., WHITE-COOPER, H., SCAIFE, S., PHUC, H.K. AND MARINOTTI, O., 2010. Female-specific flightless phenotype for mosquito control. *Proceedings of the National Academy of Sciences,* 107, pp.4550-4554.

GASTEIGER, E., HOOGLAND, C., GATTIKER, A., DUVAUD, S.E., WILKINS, M.R., APPEL, R.D. AND BAIROCH, A., 2005. *Protein identification and analysis tools on the ExPASy server,* pp. 571-607. Humana Press

GOODWIN, L. G. 1984. Chemotherapy. *Transactions of the Royal Society of Tropical Medicine and Hygiene,* 78**,** 1-8.

GRAHAM, S., MATTHEWS, K., SHIELS, P. & BARRY, J. 1990. Distinct, developmental stage-specific activation mechanisms of trypanosome VSG genes. *Parasitology,* 101**,** 361-367.

GREENBERG, B. 1973. Flics and disease. Vol. II. II. Biology and disease transmission. *Flics and disease. Vol. II. II. Biology and disease transmission.*

GU, J., LIU, M., DENG, Y., PENG, H. AND CHEN, X., 2011. Development of an efficient recombinant mosquito densovirus-mediated RNA interference system and its preliminary application in mosquito control. *Plos one*, 6, p.e21329.

HAJDUK, S. L. 1984. Antigenic variation during the developmental cycle of Trypanosoma brucei. *J Protozool,* 31**,** 41-47.

HIGGINS, M. K., TKACHENKO, O., BROWN, A., REED, J., RAPER, J. & CARRINGTON, M. 2013. Structure of the trypanosome haptoglobin–hemoglobin receptor and implications for nutrient uptake and innate immunity. *Proceedings of the National Academy of Sciences,* 110**,** 1905-1910.

HOFMAN, K. 1993. TMbase-A database of membrane spanning protein segments. *Biol. Chem. Hoppe-Seyler,* 374**,** 166-166.

HOFFMANN, A.A., MONTGOMERY, B.L., POPOVICI, J., ITURBE-ORMAETXE, I., JOHNSON, P.H., MUZZI, F., GREENFIELD, M., DURKAN, M., LEONG, Y.S., DONG, Y. AND COOK, H., 2011. Successful establishment of Wolbachia in Aedes populations to suppress dengue transmission. *Nature*, 476, pp.454-457.

HOLM, L. & PARK, J. 2000. DaliLite workbench for protein structure comparison. *Bioinformatics,* 16**,** 566-567.

HOLM, L. & ROSENSTROM, P. 2010. Dali server: conservation mapping in 3D. *Nucleic Acids Res,* 38**,** W545-W549.

ILEMOBADE, A. 2009. Tsetse and trypanosomosis in Africa: the challenges, the opportunities. *Onderstepoort J Vet Res,* 76**,** 35-40.

JACKSON, A. P., ALLISON, H. C., BARRY, J. D., FIELD, M. C., HERTZ-FOWLER, C. & BERRIMAN, M. 2013. A cell-surface phylome for African trypanosomes. *PLoS Negl Trop Dis,* 7**,** e2121.

JACKSON, D. G., SMITH, D. K., LUO, C. & ELLIOTT, J. F. 1993. Cloning of a novel surface antigen from the insect stages of Trypanosoma brucei by expression in COS cells. *J Biol Chem,* 268**,** 1894-1900.

JULIANO, R. & HASKILL, S. 1993. Signal transduction from the extracellular matrix. *Journal of Cell Biology*, 120**,** 577-577.

KARIITHI, H.M., VAN OERS, M.M., VLAK, J.M., VREYSEN, M.J., PARKER, A.G. AND ABD-ALLA, A.M., 2013. Virology, epidemiology and pathology of Glossina hytrosavirus, and its control prospects in laboratory colonies of the tsetse fly, Glossina pallidipes (Diptera; Glossinidae). *Insects*, 3**,** pp.287-319.

KENNEDY, P. G. 2008. The continuing problem of human African trypanosomiasis (sleeping sickness). *Annals of neurology,* 64**,** 116-126.

KRISSINEL, E. & HENRICK, K. 2005. Detection of protein assemblies in crystals. *Springer Berlin, Heidelberg*, 3695, 163-174

KRISSINEL, E. & HENRICK, K. 2007. Inference of macromolecular assemblies from crystalline state. *J Mol Biol,* 372**,** 774-797.

LACROIX, R., MCKEMEY, A.R., RADUAN, N., WEE, L.K., MING, W.H., NEY, T.G., AA, S.R., SALMAN, S., SUBRAMANIAM, S., NORDIN, O. AND ANGAMUTHU, C., 2012. Open field release of genetically engineered sterile male Aedes aegypti in Malaysia. *PloS one*, 7, p.e42771

LESLIE, A. 1992. Recent changes to the MOSFLM package for processing film and image plate data. *Joint CCP4+ESF-EAMCB Newslett Prot Crystallogr,* 26

LIU, L. X. & WELLER, P. 1996. Antiparasitic drugs. *New England Journal of Medicine,* 334**,** 1178-1184.

LOFGREN, C., DAME, D., BREELAND, S., WEIDHAAS, D. E., JEFFERY, G., KAISER, R., FORD, H., BOSTON, M. & BALDWIN, K. 1974. Release of chemosterilized males for the control of Anopheles albimanus in El Salvador III. Field methods and population control. *The American journal of tropical medicine and hygiene,* 23**,** 288-297.

LONGNECKER, M. P., ROGAN, W. J. & LUCIER, G. 1997. The human health effects of DDT (dichlorodiphenyltrichloroethane) and PCBs (polychlorinated biphenyls) and an overview of organochlorines in public health. *Annual review of public health,* 18**,** 211-244.

LOVELESS, B. C., MASON, J. W., SAKURAI, T., INOUE, N., RAZAVI, M., PEARSON, T. W. & 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. *Journal of Biological Chemistry,* 286**,** 20658-20665.

MATTHEWS, K.R., 2011. Controlling and coordinating development in vector-transmitted parasites. *Science*, 331, pp.1149-1153.

MCGRAW, E.A. AND O'NEILL, S.L., 2013. Beyond insecticides: new thinking on an ancient problem. *Nature Reviews Microbiology*, 11, pp.181-193.

MCCOY, A. J. 2007. Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallogr D Biol Crystallogr,* 63**,** 32-41.

MCCOY, A. J., GROSSE-KUNSTLEVE, R. W., ADAMS, P. D., WINN, M. D., STORONI, L. C. & READ, R. J. 2007. Phaser crystallographic software. *J Appl Crystallogr,* 40**,** 658-674.

MOREIRA, L., SAIG, E., TURLEY, A., RIBEIRO, J., O'NEILL, S. AND MCGRAW, E., 2009. Human probing behavior of aedes aegypti when infected with a life-shortening strain of wolbachia: Lay knowledge about Aedes aegypti in northern Australia and its implications for policy and practice. *Public Library of science: Neglected Tropical Diseases* [P], 3, pp.1-6.

MURSHUDOV, G. N., VAGIN, A. A. & DODSON, E. J. 1997. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallographica Section D: Biological Crystallography,* 53**,** 240-255.

MUTABINGWA, T. K. 2005. Artemisinin-based combination therapies (ACTs): best hope for malaria treatment but inaccessible to the needy! *Acta tropica,* 95**,** 305-315.

NOLAN, D. P., JACKSON, D. G., BIGGS, M. J., BRABAZON, E. D., PAYS, A., VAN LAETHEM, F., PATURIAUX-HANOCQ, F., ELLIOT, J. F., VOORHEIS, H. P. & PAYS, E. 2000. Characterization of a novel alanine-rich protein located in surface microdomains in Trypanosoma brucei. *Journal of Biological Chemistry,* 275**,** 4072-4080.

NOSTEN, F. & WHITE, N. J. 2007. Artemisinin-based combination treatment of falciparum malaria. *The American journal of tropical medicine and hygiene,* 77**,** 181-192.

O'CONNOR, L., PLICHART, C., SANG, A.C., BRELSFOARD, C.L., BOSSIN, H.C. AND DOBSON, S.L., 2012. Open release of male mosquitoes infected with a Wolbachia biopesticide: field performance and infection containment. *PLoS Negl Trop Dis*, 6, p.e1797.

PARKER, M. L., PENARETE-VARGAS, D. M., HAMILTON, P. T., GUÉRIN, A., DUBEY, J. P., PERLMAN, S. J., SPANO, F., LEBRUN, M. AND BOULANGER, M. J., 2016. Dissecting the interface between apicomplexan parasite and host cell: Insights from a divergent AMA–RON2 pair. *Proceedings of the National Academy of Sciences*, 113**,** 398-403.

PAYS, E. & NOLAN, D. P. 1998. Expression and function of surface proteins in Trypanosoma brucei. *Molecular and biochemical parasitology,* 91**,** 3-36.

PETERSEN, T. N., BRUNAK, S., VON HEIJNE, G. & NIELSEN, H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature methods,* 8**,** 785-786.

PHUC, H.K., ANDREASEN, M.H., BURTON, R.S., VASS, C., EPTON, M.J., PAPE, G., FU, G., CONDON, K.C., SCAIFE, S., DONNELLY, C.A. AND COLEMAN, P.G., 2007. Late-acting dominant lethal genetic systems and mosquito control. *BMC biology*, 5, 11-15.

PIERLEONI, A., MARTELLI, P. & CASADIO, R. 2008. PredGPI: a GPI-anchor predictor. *BMC bioinformatics,* 9**,** 392.

POLITZAR, H. & CUISANCE, D. 1984. An integrated campaign against riverine tsetse, Glossina palpalis gambiensis and Glossina tachinoides by trapping, and the release of sterile males. *International Journal of Tropical Insect Science,* 5**,** 439-442.

RICHARDSON, J. P., BEECROFT, R. P., TOLSON, D. L., LIU, M. K. & PEARSON, T. W. 1988. Procyclin: an unusual immunodominant glycoprotein surface antigen from the procyclic stage of African trypanosomes. *Molecular and biochemical parasitology,* 31**,** 203-216.

ROBERT, X. & GOUET, P. 2014. Deciphering key features in protein structures with the new ENDscript server. *Nucleic acids research,* 42**,** W320-W324.

ROCHE, D. B., BUENAVISTA, M. T., TETCHNER, S. J. & MCGUFFIN, L. J. 2011. The IntFOLD server: an integrated web resource for protein fold recognition, 3D model quality assessment, intrinsic disorder prediction, domain prediction and ligand binding site prediction. *Nucleic acids research,* 39**,** W171-W176.

RODITI, I., FURGER, A., RUEPP, S., SCHÜRCH, N. & BÜTIKOFER, P. 1998. Unravelling the procyclin coat of Trypanosoma brucei. *Molecular and biochemical parasitology,* 91**,** 117-130.

RODITI, I. & LEHANE, M. J. 2008. Interactions between trypanosomes and tsetse flies. *Current Opinion in Microbiology,* 11**,** 345-351.

RODITI, I. AND LINIGER, M., 2002. Dressed for success: the surface coats of insect-borne protozoan parasites. *Trends in microbiology*, 3**,** 128-134.

ROZENDAAL, J.A., 1997. Vector control: methods for use by individuals and communities. *World Health Organization*.

RUEPP, S., FURGER, A., KURATH, U., RENGGLI, C. K., HEMPHILL, A., BRUN, R. & RODITI, I. 1997. Survival of Trypanosoma brucei in the tsetse fly is enhanced by the expression of specific forms of procyclin. *The Journal of cell biology,* 137**,** 1369-1379.

SABCHAREON, A., WALLACE, D., SIRIVICHAYAKUL, C., LIMKITTIKUL, K., CHANTHAVANICH, P., SUVANNADABBA, S., JIWARIYAVEJ, V., DULYACHAI, W., PENGSAA, K. & WARTEL, T. A. 2012. Protective efficacy of the recombinant, live-attenuated, CYD tetravalent dengue vaccine in Thai schoolchildren: a randomised, controlled phase 2b trial. *The Lancet,* 380**,** 1559-1567.

SAKURAI, T., SUGIMOTO, C. & INOUE, N. 2008. Identification and molecular characterization of a novel stage-specific surface protein of Trypanosoma congolense epimastigotes. *Molecular and biochemical parasitology,* 161**,** 1-11.

SAVAGE, A. F., CERQUEIRA, G. C., REGMI, S., WU, Y., EL SAYED, N. M. & AKSOY, S. 2012. Transcript expression analysis of putative Trypanosoma brucei GPI-anchored surface proteins during development in the tsetse and mammalian hosts. *PLoS Negl Trop Dis,* 6**,** e1708.

SCHELLENBERG, J. R. A., ABDULLA, S., NATHAN, R., MUKASA, O., MARCHANT, T. J., KIKUMBIH, N., MUSHI, A. K., MPONDA, H., MINJA, H. & MSHINDA, H. 2001. Effect of large-scale social marketing of insecticide-treated nets on child survival in rural Tanzania. *The Lancet,* 357**,** 1241-1247.

SCHWARZENBACHER, R., GODZIK, A., GRZECHNIK, S. K. & JAROSZEWSKI, L. 2004. The importance of alignment accuracy for molecular replacement. *Acta Crystallographica Section D: Biological Crystallography,* 60**,** 1229-1236.

SHELDRICK, G. M. 2010. Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallogr D Biol Crystallogr,* 66**,** 479-485.

SIEVERS, F., WILM, A., DINEEN, D., GIBSON, T. J., KARPLUS, K., LI, W., LOPEZ, R., MCWILLIAM, H., REMMERT, M. & SÖDING, J. 2011. Fast, scalable generation of high quality protein multiple sequence alignments using Clustal Omega. *Molecular systems biology,* 7, 539.

SIMARRO, P. P., CECCHI, G., FRANCO, J. R., PAONE, M., DIARRA, A., RUIZ-POSTIGO, J. A., FÈVRE, E. M., MATTIOLI, R. C. & JANNIN, J. G. 2012. Estimating and mapping the population at risk of sleeping sickness. *PLoS Negl Trop Dis,* 6**,** e1859.

SIMARRO, P. P., CECCHI, G., FRANCO, J. R., PAONE, M., FÈVRE, E. M., DIARRA, A., POSTIGO, J. A. R., MATTIOLI, R. C. & JANNIN, J. G. 2011. Risk for human African trypanosomiasis, Central Africa, 2000–2009. *Emerg Infect Dis,* 17, 2322–2324.

SIMARRO, P. P., JANNIN, J. & CATTAND, P. 2008. Eliminating human African trypanosomiasis: where do we stand and what comes next? *PLoS Med,* 5**,** e55.

SONG, J., TAN, H., PERRY, A. J., AKUTSU, T., WEBB, G. I., WHISSTOCK, J. C. & PIKE, R. N. 2012. PROSPER: an integrated feature-based tool for predicting protease substrate cleavage sites. *PloS one,* 7**,** e50300.

THOMÉ, R. C., YANG, H. M. & ESTEVA, L. 2010. Optimal control of Aedes aegypti mosquitoes by the sterile insect technique and insecticide. *Mathematical Biosciences,* 223**,** 12-23.

ULLRICH, A. & SCHLESSINGER, J. 1990. Signal transduction by receptors with tyrosine kinase activity. *Cell,* 61**,** 203-212.

UNITED NATIONS, E. P. 1991. Joint FAO/UNEP Programme for the operation of prior informed consent. 1991 ed. Rome - Geneva: Food and Agriculture Organization of the United Nations

URWYLER, S., STUDER, E., RENGGLI, C. K. & RODITI, I. 2007. A family of stage specific alanine rich proteins on the surface of epimastigote forms of Trypanosoma brucei. *Molecular microbiology,* 63**,** 218-228.

UTZ, S., RODITI, I., RENGGLI, C. K., ALMEIDA, I. C., ACOSTA-SERRANO, A. & BÜTIKOFER, P. 2006. Trypanosoma congolense procyclins: unmasking cryptic major surface glycoproteins in procyclic forms. *Eukaryotic cell,* 5**,** 1430-1440.

VAN DEN ABBEELE, J., CLAES, Y., VAN BOCKSTAELE, D., LE RAY, D. & COOSEMANS, M. 1999. Trypanosoma brucei spp. development in the tsetse fly: characterization of the post-mesocyclic stages in the foregut and proboscis. *Parasitology,* 118**,** 469-478.

VASSELLA, E., BÜTIKOFER, P., ENGSTLER, M., JELK, J. & RODITI, I. 2003. Procyclin Null Mutants of Trypanosoma bruceiExpress Free Glycosylphosphatidylinositols on Their Surface. *Molecular biology of the cell,* 14**,** 1308-1318.

VICKERMAN, K. 1985. Developmental cycles and biology of pathogenic trypanosomes. *British medical bulletin,* 41**,** 105-114.

VICKERMAN, K., TETLEY, L., HENDRY, K. A. & TURNER, C. M. R. 1988. Biology of African trypanosomes in the tsetse fly. *Biology of the Cell,* 64**,** 109-119.

VINCENDEAU, P. & BOUTEILLE, B. 2006. Immunology and immunopathology of African trypanosomiasis. *Anais da Academia Brasileira de Ciências,* 78**,** 645-665.

VOLKMAN, S. K., HARTL, D. L., WIRTH, D. F., NIELSEN, K. M., CHOI, M., BATALOV, S., ZHOU, Y., PLOUFFE, D., LE ROCH, K. G. & ABAGYAN, R. 2002. Excess polymorphisms in genes for membrane proteins in Plasmodium falciparum. *Science,* 298**,** 216-218.

VREYSEN, M. J., SALEH, K. M., ALI, M. Y., ABDULLA, A. M., ZHU, Z.-R., JUMA, K. G., DYCK, V. A., MSANGI, A. R., MKONYI, P. A. & FELDMANN, H. U. 2000. Glossina austeni (Diptera: Glossinidae) eradicated on the island of Unguja, Zanzibar, using the sterile insect technique. *Journal of economic entomology,* 93**,** 123-135.

WANG, L., BERNE, B. & FRIESNER, R. 2011. Ligand binding to protein-binding pockets with wet and dry regions. *Proceedings of the National Academy of Sciences,* 108**,** 1326-1330.

WELBURN, S. & MAUDLIN, I. 1999. Tsetse–trypanosome interactions: rites of passage. *Parasitology Today,* 15**,** 399-403.

WHITE, N. J. 1985. Clinical pharmacokinetics of antimalarial drugs. *Clinical pharmacokinetics,* 10**,** 187-215.

WINN, M. D., BALLARD, C. C., COWTAN, K. D., DODSON, E. J., EMSLEY, P., EVANS, P. R., KEEGAN, R. M., KRISSINEL, E. B., LESLIE, A. G., MCCOY, A., MCNICHOLAS, S. J., MURSHUDOV, G. N., PANNU, N. S., POTTERTON, E. A., POWELL, H. R., READ, R. J., VAGIN, A. & WILSON, K. S. 2011. Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr,* 67**,** 235-42.

WORLD HEALTH ORGANIZATION 1992. WHO Expert Committee on Vector Biology and Control. 15th Report. *WHO Tech. Rep. Ser,* 818**,** 67.

WORLD HEALTH ORGANIZATION 2013. *Sustaining the drive to overcome the global impact of neglected tropical diseases. S*econd WHO report on neglected tropical diseases. Geneva, Switzerland: World Health Organization.

WORLD HEALTH ORGANIZATION 2014a. A global brief on vector-borne diseases. WHO Press, Geneva, Switzerland.

WORLD HEALTH ORGANIZATION 2014b. World malaria report. WHO Press, Geneva, Switzerland.

YAKOB, L., ALPHEY, L. & BONSALL, M. B. 2008. Aedes aegypti control: the concomitant role of competition, space and transgenic technologies. *Journal of applied Ecology,* 45**,** 1258-1265.

YIP, K. 2000. Malaria eradication: the Taiwan experience. *Parassitologia,* 42**,** 117-126.

ZDOBNOV, E. M. & APWEILER, R. 2001. InterProScan–an integration platform for the signature-recognition methods in InterPro. *Bioinformatics,* 17**,** 847-848.