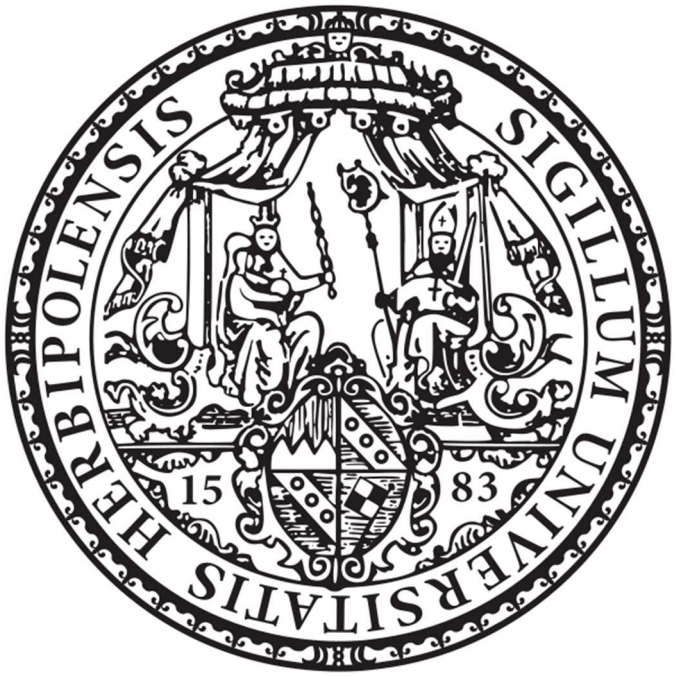

Trans-regulation of *Trypanosoma brucei* variant surface glycoprotein (VSG) mRNA and structural analysis of a *Trypanosoma vivax* VSG using X-ray crystallography

Dissertation zur Erlangung des
naturwissenschaftlichen Doktorgrades
der Julius-Maximilians-Universität Würzburg

vorgelegt von

Erick Onyango Aroko

geboren in Rachuonyo, Kenia

Würzburg, 2021

Eingereicht am:

Mitglieder der Promotionskommission:

Vorsitzender:

1. Gutachter: Prof. Dr. Markus Engstler

2. Gutachter: Dr. Daniel K. Masiga

Tag des Promotionskolloquiums:

Doktorurkunde ausgehändigt am:

Table of Contents

List of Figures

List of Tables

Summary

African trypanosomes are unicellular parasites that cause nagana and sleeping sickness in livestock and man, respectively. The major pathogens for the animal disease include *Trypanosoma vivax, T. congolense,* and *T. brucei brucei,* whereas *T. b. gambiense* and *T. b. rhodesiense* are responsible for human infections. Given that the bloodstream form (BSF) of African trypanosomes is exclusively extracellular, its cell surface forms a critical boundary with the host environment. The cell surface of the BSF African trypanosomes is covered by a dense coat of immunogenic variant surface glycoproteins (VSGs). This surface protein acts as an impenetrable shield that protects the cells from host immune factors and is also involved in antibody clearance and antigenic variation, which collectively ensure that the parasite stays ahead of the host immune system. Gene expression in *T. brucei* is markedly different from other eukaryotes: most genes are transcribed as long polycistronic units, processed by trans-splicing a 39-nucleotide mini exon at the 5′ and polyadenylation at the 3′ ends of individual genes to generate the mature mRNA.

Therefore, gene expression in *T. brucei* is regulated post-transcriptionally, mainly by the action of RNA binding proteins (RBPs) and conserved elements in the 3′ untranslated regions (UTR) of transcripts. The expression of VSGs is highly regulated, and only a single *VSG* gene is expressed at a time from one of the ~15 subtelomeric domains termed bloodstream expression sites (BES). When cells are engineered to simultaneously express two VSGs, the total *VSG* mRNA do not exceed the wild type amounts. This suggests that a robust *VSG* mRNA balancing mechanism exists in *T. brucei*. The present study uses inducible and constitutive expression of ectopic *VSG* genes to show that the endogenous *VSG* mRNA is regulated only if the second *VSG* is properly targeted to the ER. Additionally, the endogenous *VSG* mRNA response is triggered when high amounts of the *GFP* reporter with a VSG 3′UTR is targeted to the ER. Further evidence that non-VSG ER import signals can efficiently target VSGs to the ER is presented. This study suggests that a robust *trans*-regulation of the *VSG* mRNA is elicited at the ER through a feedback loop to keep the *VSG* transcripts in check and avoid overshooting the secretory pathway capacity.

Further, it was shown that induction of expression of the *T. vivax* VSG ILDat1.2 in *T. brucei* causes a dual cell cycle arrest, with concomitant upregulation of the protein associated with differentiation (PAD1) expression. It could be shown that *T. vivax* VSG ILDat1.2 can only be sufficiently expressed in *T. brucei* after replacing its native GPI signal peptide with that of a *T. brucei* VSG. Taken together, these data indicate that inefficient VSG GPI anchoring and expression of low levels of the VSG protein can trigger differentiation from slender BSF to stumpy forms. However, a second *T. vivax* VSG, ILDat2.1, is not expressed in *T. brucei* even after similar modifications to its GPI signals. An X-ray crystallography approach was utilized to solve the N-terminal domain (NTD) structure of VSG ILDat1.2. This is first structure of a non-*T. brucei* VSG, and the first of a surface protein of *T. vivax* to be solved. VSG ILDat1.2 NTD maintains the three-helical bundle scaffold conserved in *T. brucei* surface proteins. However, it is likely that there are variations in the architecture of the membrane proximal region of the ILDat1.2 NTD and its CTD from *T. brucei* VSGs. The tractable *T. brucei* system is presented as a model that can be used to study surface proteins of related trypanosome species, thus creating avenues for further characterization of trypanosome surface coats.

Zusammenfassung

Afrikanische Trypanosomen sind einzellige Parasiten, die Nagana in Nutzvieh und die Schlafkrankheit im Menschen verursachen. Zu den Hauptverursachern der Tierkrankheit gehören *Trypanosoma vivax*, *T. congolense* und *T. brucei brucei*, während *T. b. gambiense* und *T. b. rhodesiense* für Infektionen im Menschen verantwortlich sind. Da die Blutstromform (BSF) der afrikanischen Trypanosomen rein extrazellulär vorkommt, bildet die Zelloberfläche eine kritische Grenzregion mit der Wirtsumgebung. Die Zelloberoberfläche der BSF afrikanischer Trypanosomen ist mit einem dichten Mantel an immunogenen variablen Oberflächenglykoproteinen (*variant surface glycoprotein*, VSG) umgeben. Dieses Oberflächenprotein dient als Barriere zum Schutz gegen Faktoren des Wirtsimmunsystems und spielt ebenfalls eine Rolle in Antikörper-Clearance und antigener Variation, welche gemeinsam dafür sorgen, dass der Parasit dem Wirtsimmunsystem stets einen Schritt voraus bleibt. Die Genexpression von *T. brucei* weist dezidierte Unterschiede im Vergleich zu anderen Eukaryoten auf: Die meisten Gene werden als lange polyzystronische Einheiten transkribiert, die durch trans-Splicing eines Miniexons aus 39 Nukleotiden am 5′ und Polyadenylierung am 3′ Ende der individuellen Gene prozessiert wird.

Daher wird die Genexpression in *T. brucei* posttranskriptionell reguliert, zumeist durch RNA Bindeproteine (RBPs) und konservierte Elemente in der 3′ untranslatierten Region (UTR). Die Expression der VSGs ist stark reguliert, so wird zu einer gegebenen Zeit stets nur ein *VSG* Gen aus einer von ~15 Subtelomerregionen, die Blutstrom Expressionsorte (*bloodstream expression sites*, BES) genannt werden, exprimiert. Zellen, die gentechnisch manipuliert wurden um zwei VSGs zu exprimieren, produzieren die gleiche Menge an VSG mRNA wie Wildtyp Zellen. Dies deutet auf die Existenz eines robusten Mechanismus zur Regulierung der Gesamt-VSG mRNA Menge in *T. brucei* hin. Diese Arbeit verwendet induzierbare sowie konstitutive Expression eines ektopischen *VSG* Gens um zu zeigen, dass die endogene *VSG* mRNA nur reguliert wird, wenn das zweite *VSG* zum ER gelangt. Außerdem wird die endogene VSG mRNA Antwort auch ausgelöst, wenn hohe Mengen eines GFP Reporters, der eine VSG 3′UTR enthält, zum ER geleitet wird. Weiterhin, wird gezeigt, dass ER Importsignale anderer Proteine VSGs effizient zum ER dirigieren können. Das Ergebnis dieser Studie deutet darauf hin, dass eine Rückkopplungsschleife am ER eine robuste *trans*-Regulation der *VSG* mRNA auslöst, die die VSG Transkripte limitiert und somit eine Überlastung des sekretorischen Wegs verhindert.

Weiterhin konnte gezeigt werden, dass es nach Induktion der Expression des *T. vivax* VSGs ILDat1.2 in *T. brucei* zu einem doppelten Zellzyklusarrest mit gleichzeitiger Hochregulation der Expression des *protein associated with differentation* (PAD1) kam und dass dieses *T. vivax* VSG nur nach Austausch des GPI Signalpeptids durch das eines *T. brucei* VSGs effizient exprimiert werden konnte. Zusammengenommen suggerieren diese Daten, dass eine ineffiziente GPI-Verankerung und wenig abundante Expression des VSGs die Differenzierung der sogenannten *slender* BSF zur sogenannten *stumpy* Form einleiten kann. Ein zweites *T. vivax* VSG, ILDat2.1, konnte hingegen auch nach Austausch des GPI Signals nicht in *T. brucei* exprimiert werden. Mit Hilfe der Röntgenstrukturanalyse wurde die

Struktur der N-terminalen Domäne (NTD) des ILDat1.2 VSGs gelöst. Es handelt sich hierbei um die erste Proteinstruktur eines VSGs, welches nicht aus *T. brucei* stammt und die erste Struktur eines Oberflächenproteins von *T. vivax*. Das in *T. brucei* Oberflächenproteinen konservierte drei-Helix Grundgerüst ist auch in der NTD des ILDat1.2 VSGs enthalten. Die Architektur der Membranproximalen Gegend der IlDat1.2 NTD und CTD unterscheiden sich aber vermutlich von der der *T. brucei* VSGs. Das leicht handhabbare *T. brucei* System bietet somit ein geeignetes Modell um die Oberflächenproteine anderer afrikanischer Trypanosomen Spezies zu untersuchen und eröffnet neue Wege zur Charakterisierung ihrer Oberflächenmäntel.

1.0 Introduction

1.1 Classification and distribution of African trypanosomes

Trypanosomes are obligate uniflagellate protozoan parasites belonging to the order *Kinetoplastida*, family *Trypanosomatidae,* and genus *Trypanosoma*. A distinctive feature of kinetoplastids is the presence of a dense concatenated extranucleolar DNA material known as the kinetoplast (Vickerman, 1976). Kinetoplastids consist of many free-living organisms as well as plant and animal parasites. The genus *Trypanosoma* encompasses different species of African and South American parasites that infect a wide range of vertebrates (Stevens & Gibson 1999). African trypanosomes are extracellular parasites causing Human African Trypanosomiasis (HAT) also known as sleeping sickness and Animal African Trypanosomiasis (AAT) which is also known as nagana in humans and livestock, respectively. HAT is caused by *Trypanosoma brucei* subspecies while AAT is mainly caused by *T. vivax* and *T. congolense.* These parasites belong to the Salivaria group trypanosomes as they are mainly transmitted via the infected saliva of *Glossina* spp. during bloodmeal acquisition. HAT and AAT cases are confined to sub-Saharan Africa, where tsetse flies are prevalent (Figure 1).

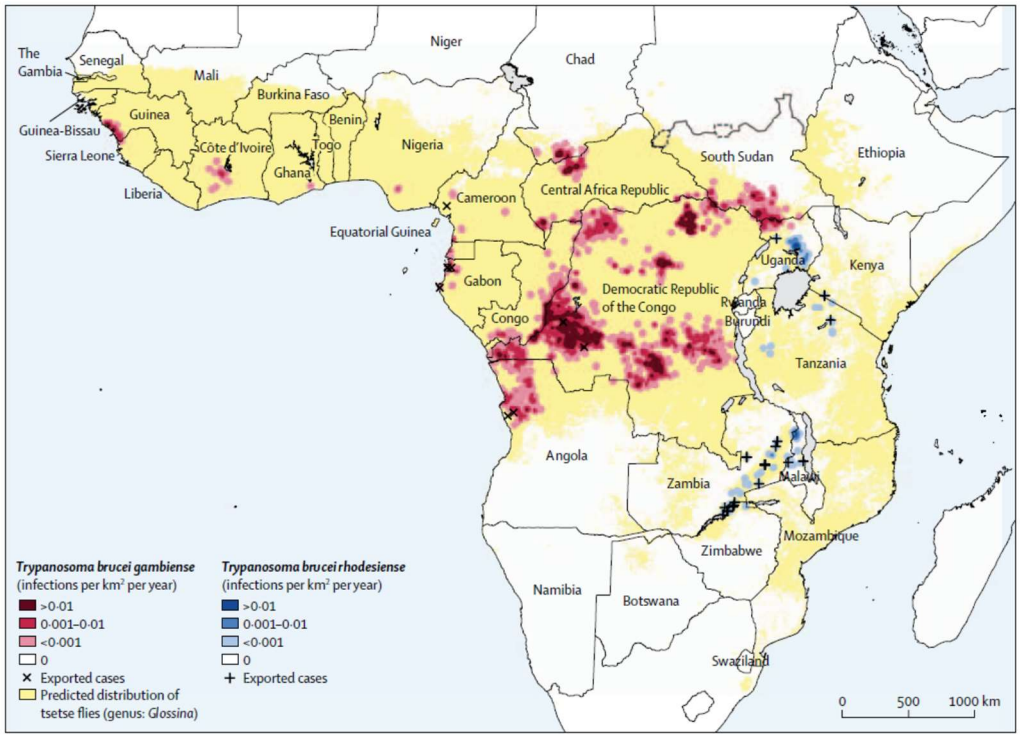

Figure 1: Geographical distribution of tsetse flies and HAT infections in sub-Saharan Africa.
The predicted tsetse fly distribution is indicated with yellow color and was provided by the Program Against African Trypanosomiasis (PAAT, 2014). Reported western and central Africa *T. b. gambiense* HAT cases are represented in shades of red while the eastern and southern Africa *T. b. rhodesiense* HAT cases are represented in shades of blue (reporting period is 2010 – 14). The image is adapted from (Büscher et al., 2017).

The Scottish pathologist David Bruce observed trypanosomes for the first time in blood samples of infected horses in present-day South Africa (Bruce, 1895). Bruce later observed *T. brucei* in cattle blood and in 1903, together with his colleagues, provided the first evidence that tsetse flies transmitted trypanosomes (Bruce et al., 1903).

1.2 Human African trypanosomiasis (HAT)

HAT is transmitted to humans by the bite of a tsetse fly infected with either of the two subspecies of *T. brucei*: *T. b. gambiense* or *T. b. rhodesiense*. Different tsetse species and subspecies have varying vectorial capacities for the transmission of human infective *T. brucei* subspecies. Savannah tsetse species such as *G. pallidipes* and *G. morsitans* are the main vectors of *T. b. rhodesiense* whereas the riverine *palpalis* group including *G. palpalis* and *G. fuscipes* are mostly responsible for *T. b. gambiense* transmission (Lehane et al., 2016; Solano et al., 2013). Infection by *T. b. gambiense* results in a chronic form of the disease (responsible for about 95% of HAT cases) and is confined to central and western Africa, whereas *T. b. rhodesiense* infection is acute and is responsible for the disease in eastern and southern Africa (Simarro et al., 2010). On the other hand, a third subspecies, *T. b. brucei*, is not infective to humans but is to livestock.

The natural resistance of humans to *T. b. brucei* is attributed to the presence of trypanosome lytic factors (TLF1 and 2) in the human serum (Hajduk et al., 1989; Raper et al., 1999; Rifkin, 1978). TLF is composed of two proteins: apolipoprotein L1 (ApoL1) and haptoglobin-related protein (Hpr). TLF1 uptake by *T. brucei* is facilitated by the haptoglobin hemoglobin receptor (TbHpHbR) and clathrin-mediated endocytosis, while TLF2 uptake is independent of the TbHpHbR (Vanhollebeke et al., 2008; Vanhollebeke & Pays, 2010). When TLF is taken up, the ApoL1 component interacts with the lysosome membrane forming pores that interfere with the integrity of the lysosomes. These pores lead to the discharge of lysosomal contents into the cytoplasm causing cell lysis (Pérez-Morga et al., 2005; Vanhamme et al., 2003).

Human infective *T. b. rhodesiense* and *T. b. gambiense* parasites are resistant to ApoL1-mediated lysis. *T. b. rhodesiense* expresses a VSG-like protein called serum resistance-associated protein (SRA) from one of the variable surface glycoprotein (VSG) expression sites (ES). SRA confers resistance to TLF lysis by binding to ApoL1, blocking its incorporation into the lysosomal membrane and with that its pore-forming ability (Uzureau et al., 2013). *T. b. gambiense,* on the other hand, appears to evade TLF activity by different mechanisms. Polymorphism in the *T. b. gambiense* HpHbR lowers its ability to bind TLF, decreasing the amounts taken up by endocytosis (Kieft et al., 2010; Vanhollebeke et al.,

2008). The parasites also express a *T. b. gambiense*-specific glycoprotein (TbgGP), which neutralizes ApoL1 toxicity and induces membrane stiffening upon interaction with lipids. *T. b. gambiense* also has a reduced sensitivity to ApoL1 through altered cysteine protease activity (Capewell et al., 2015; Uzureau et al., 2013).

HAT disease progression is divided into two stages. In the first or early stage of infection, the parasites are restricted to the peripheral circulatory system, and it is characterized by joint pains, headaches, and intermittent fever. The second or late stage of the disease is marked by the parasites crossing the blood-brain barrier and colonizing the central nervous system. This stage of infection is characterized by motor-sensory and sleep disorders, accompanied by visual impairment and if left untreated, the parasites eventually kill the host (Jamonneau et al., 2012). Suramin and melarsoprol are used for the treatment of stage 1 and 2 *T. b. rhodesiense* HAT, respectively, while stage 1 and 2 *T. b. gambiense* HAT is treated by pentamidine and nifurtimox-eflornithine combination therapy (NECT), respectively, (Büscher et al., 2017; Yun et al., 2010). Recently, fexinidazole was approved as the first oral drug for the treatment of stage 1 and stage 2 HAT (Lindner et al., 2020). Another compound, acoziborole, has shown remarkable efficacy and safety in clinical trials (Dickie et al., 2020). Heightened control measures in various endemic countries are bearing fruit as less than 1,000 cases of HAT were reported in 2018 (Franco et al., 2020).

1.3 Animal African trypanosomiasis (AAT)

AAT is one of the common livestock diseases of economic importance in sub-Saharan Africa. The main causative agents of AAT are *T. congolense* and *T. vivax*, while *T. b. brucei, T. evansi*, *T. equiperdium, T. suis, T. godfreyi* and *T. simiae* also infect livestock. *T. vivax* has a wide host range infecting mainly cattle and various other species of livestock as well as wildlife (Fetene et al., 2021). Infection with *T. vivax* causes a wasting disease characterized by weight loss, abortions, and reduced milk production and is the leading cause of AAT in the region of western Africa and is mainly transmitted by tsetse flies (Gardiner & Wilson, 1987; Osório et al., 2008). However, mechanical transmission of *T. vivax* by biting flies, tabanids and stable flies (*Stomoxys)* has also been reported (Desquesnes & Dia, 2004; Wells, 1972). Indeed, *T. vivax* has also been isolated in cattle and horses in southern America where the primary tsetse vector is absent, indicating that mechanical transmission is the only mode of transmission in this region (Giordani et al., 2016; Jones & Dávila, 2001). *T. congolense* is highly pathogenic and is responsible for AAT in eastern and southern Africa. The tsetse fly is the only known *T. congolense* vector.

T. evansi, a derivative of *T. brucei*, parasitize a wide range of livestock and wild animals in tropical Africa, Asia and South America. In these regions, the highest prevalence of *T. evansi* was observed in dromedary camels where it causes sura (Aregawi et al., 2019). *T. evansi* can be transmitted by the bite of several hematophagous insects, bats and biting flies. Cases of human infections with *T. evansi* have also been reported (Joshi et al., 2005; van Vinh Chau et al., 2016) indicating a zoonotic potential.

Chemotherapy to treat infection and vector control strategies applied to manage AAT have not been effective. The drugs used for livestock treatment mainly include diminazene aceturate, isometamidium chloride and suramin, which were developed more than 50 years ago. The drugs are highly toxic, and cases of drug resistance have also been reported (Dagnachew et al., 2015; Giordani et al., 2016; Küpper & Wolters, 1983; Mbwambo et al., 1988). AAT leads to annual losses exceeding 1 billion USD due to direct effects on cattle production while the combined losses in terms of agricultural gross domestic product (GDP) amass to over 4 billion USD (FAO, 2017).

1.4 The life cycle of *Trypanosoma brucei*

T. brucei parasites must shuttle between the mammalian host and the insect vector to complete their life cycle (Figure 2). This digenetic life cycle presents enormous challenges to the parasites; for instance, they must adapt to differences in pH and temperatures between the mammalian bloodstream and insect midguts, and the difference in micronutrients available in these environments. The slender bloodstream form (BSF) parasites utilize glucose, present in the blood of their mammalian hosts, as the carbon source for the glycolytic pathway (Hannaert et al., 2003); the TCA cycle and the electron transport chain are reduced substantially in these cells (Zíková et al., 2017). Recently, glycerol was shown to be the carbon source for slender BSF parasites residing in the mammalian skin and adipose tissues (Kovářová et al., 2018; Nakane et al., 2020). When the parasitaemia reaches a certain critical threshold, parasites perceive and respond to extracellular oligopeptides via a signaling cascade (Rojas et al., 2019) to induce differentiation of the slender BSF to the G0/G1 growth-arrested short stumpy-forms (Reuner et al., 1997). The stumpy parasites are tsetse fly infective and die off after 2 – 3 days if not taken up by the fly (Turner et al., 1995). It has been reported that differentiation to stumpy stage is required to maintain the cell density at a level that does not kill the host and pre-adapts the parasites to transition or progress to the next life cycle stage (Rico et al., 2013). However, Schuster et al., 2019 have shown that slender BSF parasites can also infect flies.

Differentiation to the stumpy forms is accompanied by morphological and molecular changes: the parasites become transcriptionally inactive, save for the up-regulation of some mitochondrial proteins, expression site associated gene 9 (ESAG9) protein family (Barnwell et al., 2010) and the carboxylate-transporters protein associated with differentiation 1 and 2 (PAD1 and 2) (Dean et al., 2009). ESAG9 genes, as the name suggests, were initially thought to be expressed exclusively from the BES (Florent et al., 1991). However, ESAG9 genes are mostly found in other non-BES locations and rarely in the VSG BES (Hertz-Fowler et al., 2008). Apart from the density-dependent differentiation of slender BSF to stumpy forms, treatment of slender BSF with membrane-permeable cAMP analogues has been shown to induce differentiation to stumpy forms *in vitro* (Laxman et al., 2006). Additionally, attenuation of the active BES, caused by high level expression of an ectopic VSG, also triggers differentiation to stumpy forms in both monomorphic and pleomorphic *T. brucei* in a cell density-independent manner (Batram et al., 2014; Zimmermann et al., 2017).

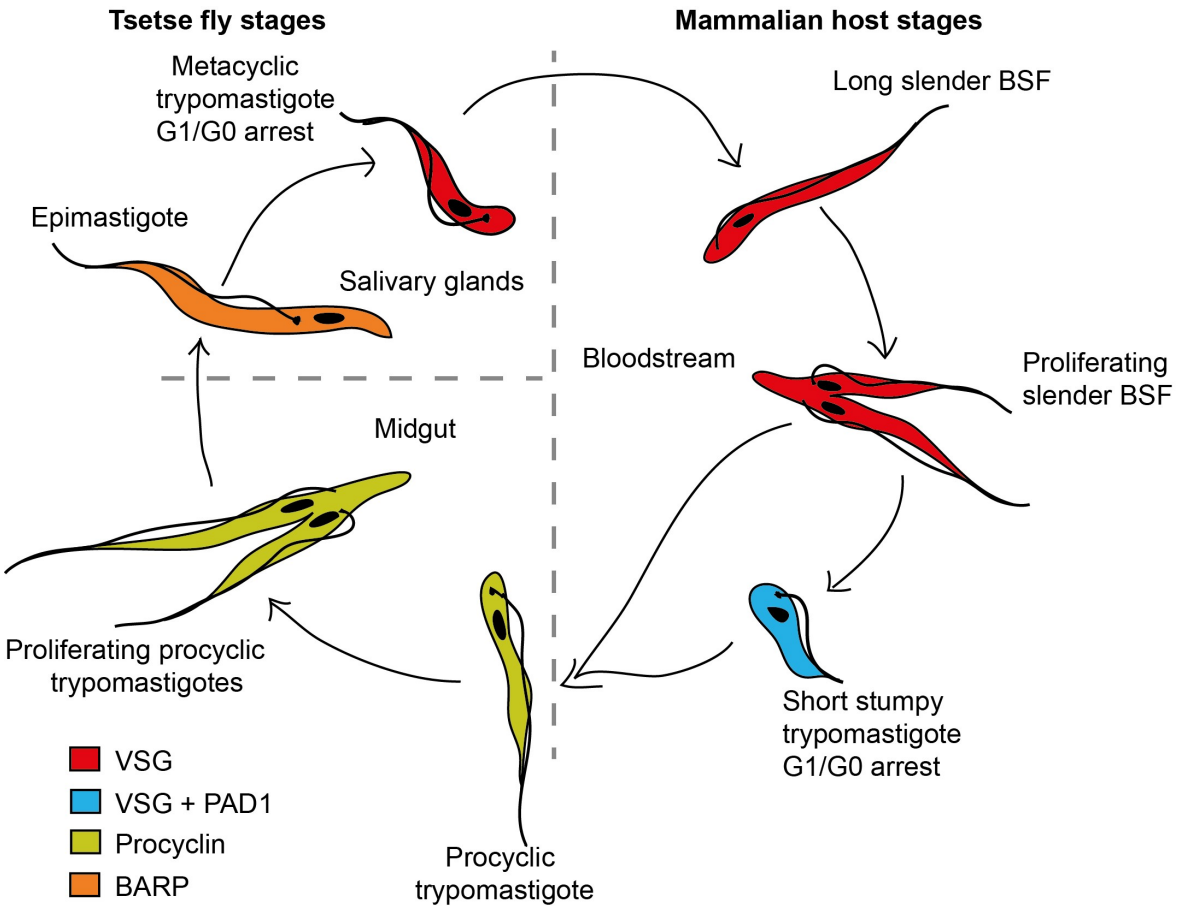

Figure 2: The life cycle of *T. brucei*.
The cartoon depicts the trypanosome journey through the tsetse fly vector and the mammalian host. The color coding represents the different surface proteins expressed as the parasite differentiates from one form to the other. Adapted with modifications from (Clayton, 2019).

Upon the uptake of stumpy parasites by the tsetse flies, the parasites differentiate to procyclic forms; a transition marked by a cascade of molecular, metabolic and morphological changes (Clayton, 2002; Matthews et al., 2004; Smith et al., 2017; van Grinsven et al., 2009). The early hallmarks of differentiation from stumpy to procyclic stage include the timely replacement of the surface VSGs with procyclin protein (Matthews et al., 2004). Different procyclin isoforms exist, the main ones being the GPEET and EP procyclins (Acosta-Serrano et al., 2001; Roditi et al., 1987; Roditi et al., 1989). After 4 – 7 days in the midguts, the GPEET procyclin is lost while the EP procyclin is retained as the primary surface protein (Vassella et al., 2000). Though the function of procyclins is not well known, procyclin null mutants are less efficient in establishing infections in flies in comparison to wild type parasites (Acosta-Serrano et al., 2001; Ruepp et al., 1997). In the absence of glucose, proline is the primary carbon source of procyclic trypanosomes (Lamour et al., 2005).

For establishment of infection, the parasites must cross the peritrophic membrane (PM) into the ectoperitrophic space where they develop into long procyclics and mesocyclic forms (Schuster et al., 2017). The parasites must then colonize the proventriculus, a junction between the mid- and foregut, and migrate to the foregut as long epimastigotes. They then colonize the salivary glands where they mature as short epimastigotes (van den Abbeele et al., 1999). The parasites then differentiate to metacyclic forms expressing metacyclic VSGs (mVSG) on their surface, which re-establishes infectivity (Vickerman, 1985).

1.5 The surface coat of BSF African trypanosomes

The cell surface of BSF African trypanosomes is covered by a dense coat of GPI anchored immunogenic VSGs. Among the different species of mammalian trypanosomes, the *T. brucei* surface coat is the best characterized. In this species, the VSG constitute approximately 10% of total protein and 90% of the cell surface proteins (Grünfelder et al., 2002). There are approximately 5.7 x 10^6 VSG dimers on the cell surface (Grünfelder et al., 2002; Jackson et al., 2013). Additionally, the number of VSG dimers on the cell surface can vary depending on the degree of packing (Bartossek et al., 2017). The *T. brucei* genome encodes for a vast repertoire of approximately 2500 VSG genes and pseudogenes (Cross et al., 2014); the expression of which is precisely regulated to ensure that only a single VSG protein is expressed at a time. In contrast to many other organisms, genomic architecture of *T. brucei* is unique in the sense that there are considerable differences in sizes even between individual chromosomes of a diploid pair (Berriman et al., 2005). *VSG* genes are found in the 15 – 20 polycistronic transcription units known as bloodstream expression sites (BES), which are mainly located in the large megabase chromosomes, whereas a small subset can

be found within the intermediate chromosomes (Berriman et al., 2005; Hertz-Fowler et al., 2008). BES harbor expression site associated genes (ESAGs) of known and unknown functions, and subtelomeric *VSG* genes that are organized in arrays. Additionally, *VSG* genes and pseudogenes are contained in the small minichromosomes, further contributing to the rich *VSG* archive (Cross et al., 2014). The active VSG is transcribed from one of the BES that occupies an extranucleolar polymerase I (RNAP I) transcribed unit known as the expression site body (ESB) (Budzak et al., 2019; Navarro & Gull, 2001).

Other studied *T. brucei* surface proteins include the transferrin receptor (TbTfR) that localizes at the flagellar pocket surface, the haptoglobin-hemoglobin receptor (TbHpHbR), the PAD proteins and the invariable surface glycoproteins (ISGs). The TbTfR receptor genes (ESAG6 and 7) are located in the expression site forming part of the polycistron known as the expression site-associated genes (ESAGs) (Schell et al., 1991; Steverding et al., 1994). TbTfR facilitates receptor-mediated endocytosis of host transferrin. The receptor is a heterodimer localized in the flagellar pocket and attached by a single GPI anchor to the cell membrane (Mehlert et al., 2012; Salmon et al., 1994). Approximately 3 x 10^3 TbTfR heterodimers are expressed under basal conditions (Salmon et al., 1994), however, in conditions of iron deficiency, TbTfR protein expression is upregulated (Fast et al., 1999). Regulation of TbTfR expression is driven by its 3′UTR (Benz et al., 2018). Homologs of *T. brucei* ESAG6 and 7 gene families are present in the genome of *T. congolense* while they appear to have been lost in *T. vivax* (Jackson et al., 2012), suggesting a different route for iron acquisition in *T. vivax*.

The TbHpHbR receptor mediates cellular uptake of sequestered hemoglobin (Hb) in the form of haptoglobin-hemoglobin (HpHb) and TLF acquisition from the host bloodstream (Lane-Serff et al., 2016; Vanhollebeke et al., 2008). However, HpHbR is expressed by the insect stage epimastigotes but not the BSF *T. vivax* or *T. congolense* parasites (Jackson et al., 2015; Lane-Serff et al., 2016). Another essential surface protein in *T. brucei* is the PAD family, which codes for a surface carboxylate transporter essential for perception of citrate/*cis*-aconitate and is required for differentiation when parasites are taken up by the tsetse flies (Dean et al., 2009). This gene family is absent in *T. vivax* while orthologs in *T. congolense* are expressed at low levels and are not upregulated at high parasitaemia as is the case in *T. brucei* (Jackson et al., 2013; Silvester et al., 2018).

The *T. congolense* cell surface architecture is similar to *T. brucei. T. vivax*, however, shows an overall expanded cell surface protein repertoire compared to *T. brucei* and *T. congolense* (Jackson et al., 2013). Additionally, it has been shown recently that immunization of mice

with a *T. vivax* protein that localizes to the flagellum confers immunity against infection with *T. vivax* (Autheman et al., 2021). Based on these studies, it is suggested that the *T. vivax* VSG surface coat is probably not as dense as the *T. brucei* coat.

1.6 Modulation of host immune responses

Infectious microorganisms face constant attack by the host's immune responses. Therefore, these microbes have devised different mechanisms to thrive in the hostile environments encountered within their hosts. One of the strategies utilized by *T. brucei* to overcome the immune system of the host is antibody clearance, a process by which hydrodynamic drag forces generated by the parasites' motility causes the removal of VSG-antibody complexes from the cell surface (Engstler et al., 2007). The complexes are moved to the flagellar pocket, located at the posterior end, where they are internalized by clathrin-mediated endocytosis (Figure 3 A, B). The VSG-antibody complexes together with defective VSGs are transported to the lysosomes for degradation, while functional VSGs are recycled back to the cell surface (Pal et al., 2003; Wang et al., 2003). It has been shown that the entire cell surface VSG pool is internalized and recycled to the cell surface within 12 min (Engstler et al., 2004). Glycosylphosphatidylinositol- (GPI) anchoring of the membrane form VSG (mfVSG) contributes to the fluid nature of the VSG coat and VSG lateral diffusion (Bülow et al., 1988; Engstler et al., 2004). The internalized proteins and newly synthesized VSGs are reported to be exported in TbRAB11-positive exocytic carriers and released to the cell surface via the flagellar pocket. However, different studies reported conflicting results on the role of TbRAB11 in exocytosis of newly synthesized VSGs (Hall et al., 2005; Jeffries et al., 2001; Umaer et al., 2018). The actual mechanism of VSG exocytosis remains to be characterized in detail.

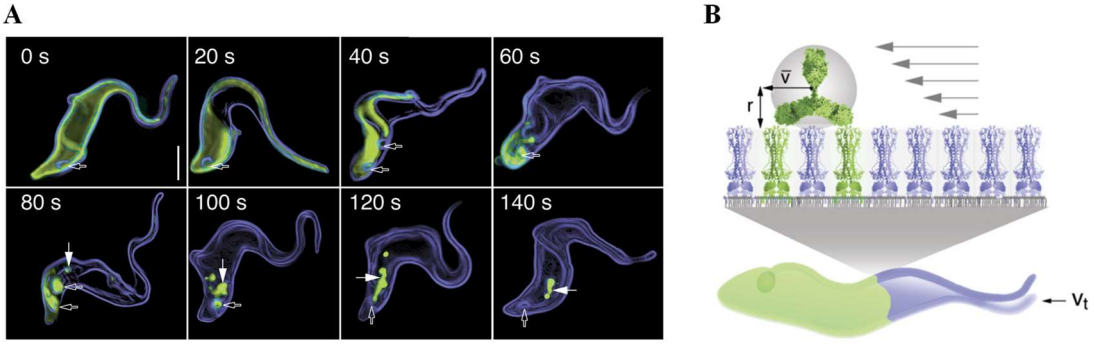

Figure 3: Removal of host-derived antibodies from the cell surface of *T. brucei*.
(A) Antibody removal from the cell surface is visualized. Anti-VSG antibodies (conjugated with Alexa Fluor 488, green) are localized to the posterior end by 80 s and are completely endocytosed in 120 s. Open arrows indicate the position of the flagellar pocket and filled arrows point to the lysosome. **(B)** Schematic model indicating the movement of VSG-IgG-complexes due to hydrodynamic flow forces. Images adapted from (Engstler et al., 2007).

Another immune evasion strategy is referred to as antigenic variation, which involves switching on and off of surface antigens or a complete replacement of the initially expressed surface proteins with a new antigenically distinct protein (Deitsch et al., 2009). Antigenic variation was first discovered in trypanosomes and has since been reported in many intracellular and extracellular parasites. *Plasmodium falciparum* erythrocyte membrane protein 1 (PfEMP1), encoded by approximately 60 *var* genes, is expressed on the surface of plasmodium-infected erythrocyte (Baruch et al., 1995; Su et al., 1995). Each parasite expresses only one isoform of this surface protein at a time while maintaining the rest in a transcriptionally silent state. Switching the expression of *var* gene to a new isoform allows the parasite to evade host immunity (Kraemer & Smith, 2006; Scherf et al., 2008).

T. brucei VSG molecules are highly immunogenic, and over time, the mammalian immune system is able to clear the parasites. Therefore, the parasites have developed sophisticated mechanisms to continuously change their surface coats. Antigenic variation occurs either by transcriptional mechanisms or by DNA rearrangement/recombination events (McCulloch & Barry, 1999; Pays, 1991; Taylor & Rudenko, 2006).

Recombination events include telomere *VSG* exchange, array *VSG* conversion and segmental *VSG* conversion (Stockdale et al., 2008). In telomere *VSG* exchange, a telomere resident *VSG* is replaced by a telomeric *VSG* from a silent expression site (ES) or minichromosome. Array *VSG* conversion involves the replacement of the ES resident *VSG* with one from the subtelomeric array while segmental *VSG* conversion involves the replacement of the ES resident *VSG* by several *VSG* pseudogenes that recombine to generate the so-called mosaic VSG (Figure 4). Recombination events leading to VSG switching are driven by DNA double-strand breaks (DSBs), occurring adjacent to the 70-bp repeats, upstream of the *VSG* gene (Boothroyd et al., 2009; Glover et al., 2013; Li, 2015).

Transcriptional control of antigenic variation occurs by an *in situ* switch, which involves transcription activation of a new ES with a concomitant shut down of the previously active ES. It is thought that this transcriptional control is regulated by epigenetic mechanisms (Figueiredo et al., 2008; Horn & Cross, 1997). Batram et al. showed that attenuation and eventual silencing of the formerly active ES could be achieved by overexpression of a second *VSG* gene from a non-telomeric locus (Batram et al., 2014).

In cases of prolonged infection, a small subset of the parasites that switch their surface antigens evade the host immune responses and maintain infection, while the rest are eliminated. This cycle goes on for a long time and is responsible for the waves of parasitaemia observed in infected individuals and livestock.

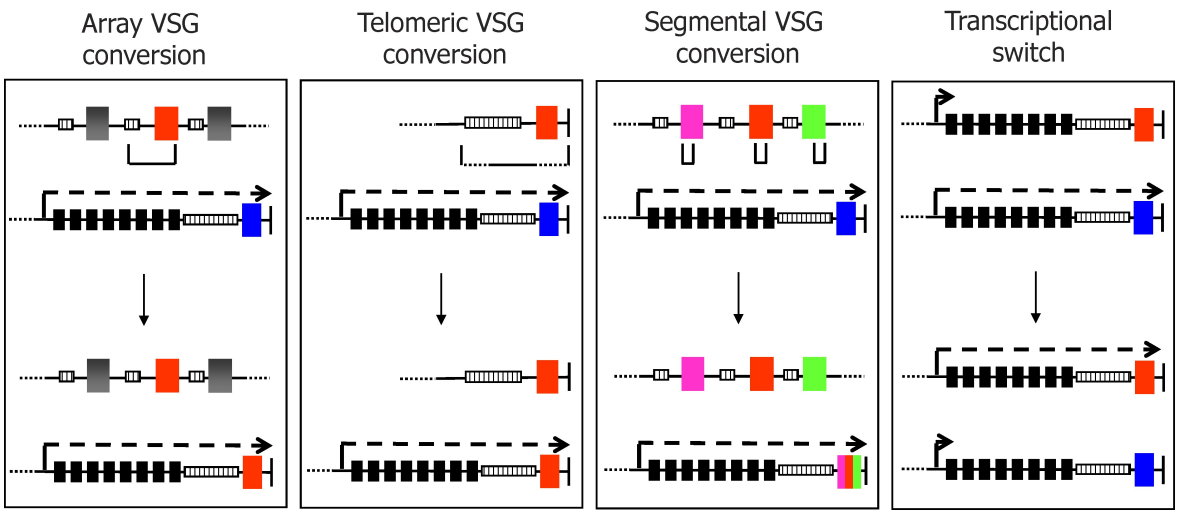

Figure 4: Schematic illustration of VSG switching mechanisms.
Recombination events lead to VSG switching by either array, telomeric or segmental VSG gene conversion. The transcriptional switch involves an *in situ* inactivation of the active ES and simultaneous activation of a previously repressed ES. Image adapted from (Stockdale et al., 2008).

1.7 Evolution of *VSG* genes

The order of events leading to the evolution of VSGs in African trypanosomes remains unclear. The African-trypanosome ancestors likely expressed a small repertoire of surface proteins, similar to the *Leishmania* gp63, which is conserved in kinetoplastids including *T. brucei* (El-Sayed & Donelson, 1997), and the insect-infective *Crithidia* and *Herpetomonas* (Etges, 1992). Selection pressure occasioned by the different challenging host environments must have driven the paralogous expansion of surface genes, with African trypanosomes and *T. cruzi* evolving to express VSGs and mucins, respectively.

A characteristic feature of kinetoplastids surface proteins is GPI-anchoring (Schlagenhauf et al., 1998). However, there is no evidence of monoallelic gene expression (a mechanism driving *VSG* gene expression in African trypanosomes) in *Leishmania*, *T. cruzi* or *Crithidia*. This suggests that the evolution of monoallelic expression of *VSG* genes was preceded by a paralogous expansion of surface proteins in kinetoplastids (Manna et al., 2014). *VSG* genes are categorized into 'functional' (encoding all the hallmarks of a functional VSG), 'pseudogenes' (containing in-frame stop codons and frameshifts) and 'incomplete' (containing partial sequences) (Berriman et al., 2005).

The availability of the *T. brucei* reference genome and the draft genomes of *T. congolense* and *T. vivax* has enabled an analysis of the evolution of the surface proteins of these parasites to some extent. A comparison of the genomic architecture of these three species revealed that there is high conservation in genome organization. The major differences were observed in the genes encoding cell-surface proteins (Jackson et al., 2012).

The African trypanosomes have a rich archive of *VSG* genes. As the *T. brucei* VSGs are the best characterized, *T. vivax* and *T. congolense VSG* genes are classified in the same manner as *T. brucei* VSGs. These VSGs are classified based on the conservation patterns of the cysteine residues in their N-terminal domains (NTD) and C-terminal domains (CTD). The NTDs were initially classified into three groups (type A, B and C), based on the distribution of conserved cysteine residues (Carrington et al., 1991). However, group C was later reclassified as a subtype of A-type VSGs, leaving only two groups: A and B (Marcello & Barry, 2007). Similarly, the CTDs were initially classified into type 1 – 4 (Carrington et al., 1991). Later, two additional types (5 and 6) were added to the classification based on genomic data (Berriman et al., 2005). Types 2, 4 and 5 consist of one conserved four-cysteine subdomain while types 1, 3 and 6 are di-domains containing two such subdomains (Figure 5) (Marcello & Barry, 2007). It has been reported that functional VSGs can be formed by similar NTDs and different CTD types (Hutchinson et al., 2003).

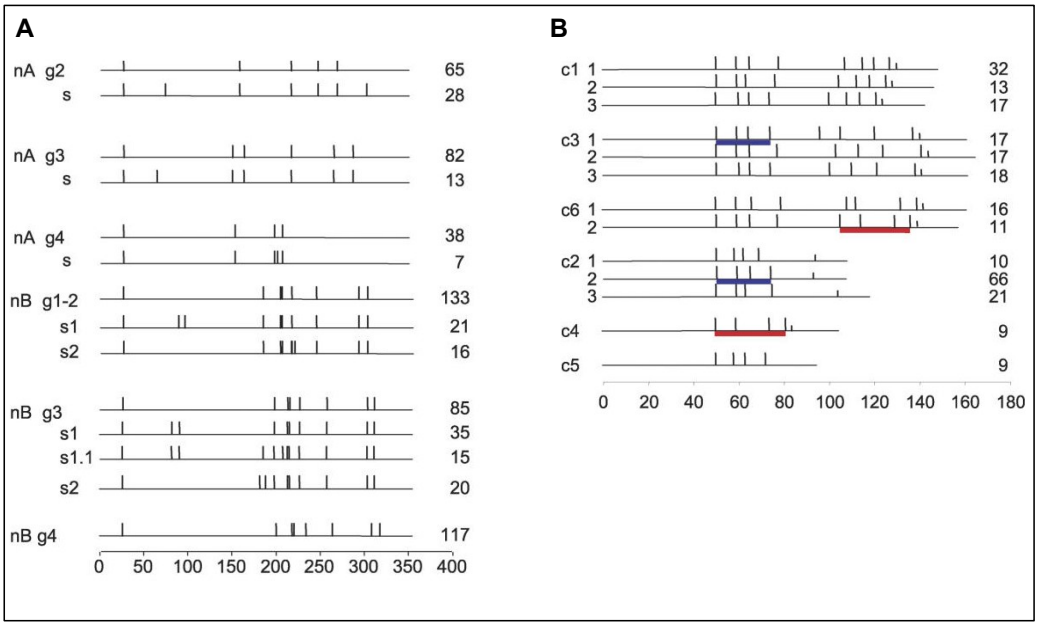

Figure 5: Scheme of conserved cysteine residue patterns in VSG domain types.
The vertical bars represent cysteine residues while the half-vertical bars represent the putative start of the GPI signal peptide. The number of sequences in each category is shown on the right, and the scale bar represents the number of amino acid residues. **(A)** N-terminal domain classification with g, group and s, subgroup. **(B)** C-terminal domain types. Same color underlines shared cysteine conservation between domain types. The image is adapted from (Marcello and Barry, 2007), with minor modifications.

The entire repertoire of *T. congolense* VSGs belongs to the B-type VSG group and is further divided into two subgroups known as Fam13 and Fam16. Despite the absence of the A-type VSG group in *T. congolense*, homologs of A-type VSG-like TFR genes of *T. brucei* are present in this species. *T. vivax* VSGs are classified as either A or B-type, though only A-type

VSGs have been experimentally confirmed. However, this species has additional unique subgroups of *VSG*-like genes designated Fam25 and Fam26. The unique subgroups may have been acquired by *T. vivax* later in evolution or were not inherited by *T. brucei* and *T. congolense* from the common ancestor (Jackson et al., 2012). The model in Figure 6 summarizes the evolution of VSGs in African trypanosomes. It must be noted, however, that homology of *T. congolense* and *T. vivax* VSGs to *T. brucei* VSGs alone does not qualify them as functional VSGs. More of these VSGs need to be further characterized to ascertain whether they are indeed functional VSGs. To date, only two *T. vivax* VSGs, ILDat1.2 and ILDat2.1, have been characterized to some extent (Chamond et al., 2010; Gardiner et al., 1996; Gardiner & Wilson, 1987). A third VSG, *TvY486_0027060*, is not well studied (Jackson et al., 2013).

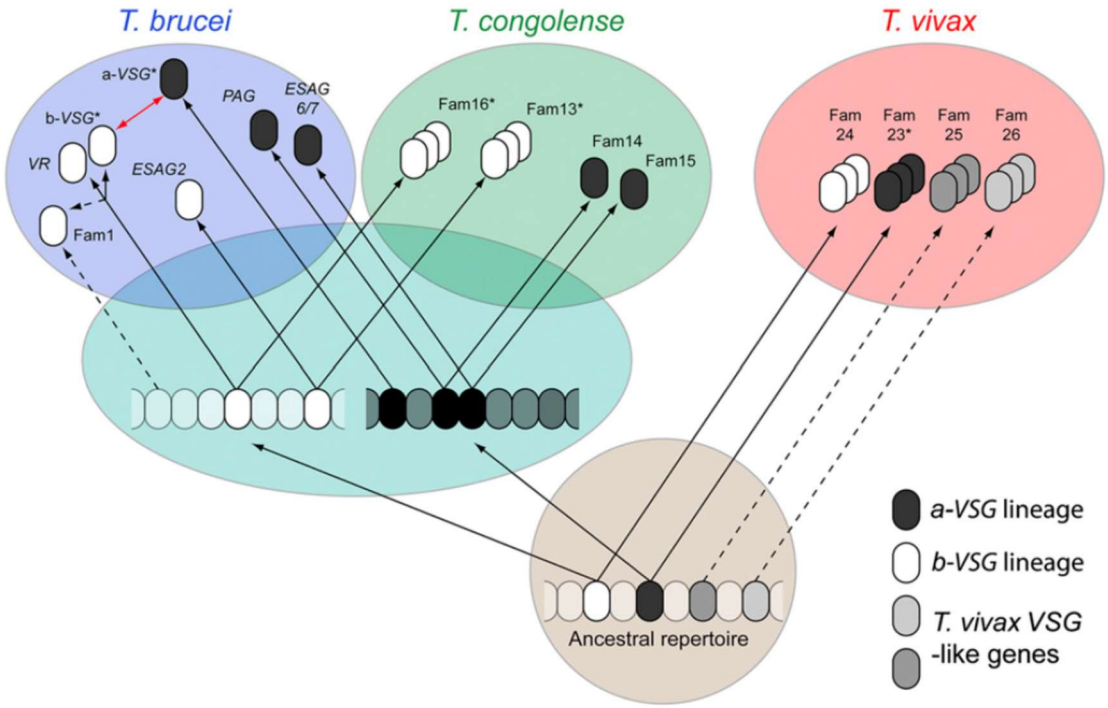

Figure 6: A model for the evolution of *VSG* genes in African trypanosome.
The schematic depicts the distribution of VSG groups and subgroups in ancestral and contemporary genomes. *A*-type and b-type *VSG* lineages are marked in black and white respectively, while *T. vivax*-specific *VSG*-like gene families are shown in different shades of grey. An asterisk indicates groups with published *VSGs*. The red arrow shows shared CTD between the a-type and b-type *T. brucei* subfamilies. The illustration is adapted from (Jackson et al., 2012).

1.8 VSG architecture

Of all the African trypanosome species, *T. brucei* VSGs are the best characterized. *T. brucei* VSGs are homodimeric proteins (Figure 7A, B) of between 50 – 60 kDa per monomer and extend about 15 nm from the cell surface (Bartossek et al., 2017; Vickerman, 1969). The

membrane distal NTD is made up of between 350 – 400 amino acids and is highly variable from one VSG to the other. This high amino acid sequence variability forms the basis of antigenic variation. The CTD, which is more secluded, is much smaller and harbors flexible linkers and structured regions of between 20 – 40 amino acids (Chattopadhyay et al., 2005; Jones et al., 2008). Bartossek et al., 2017 showed that the CTD confers flexibility to the VSG that allows it to exist either in a relaxed or compact conformation. Despite the low sequence similarity of VSG genes, their tertiary structure is highly conserved (Bartossek et al., 2017; Blum et al., 1993; Freymann et al., 1990). However, it is not clear whether the B-type VSGs exist as monomers or form dimers. Additionally, recently solved structures of two *T. brucei* VSGs, VSGsur and VSG13, show additional extensions on the 'upper lobe' of the NTD (Zeelen et al., 2021) indicating the presence of divergent subgroups of VSGs.

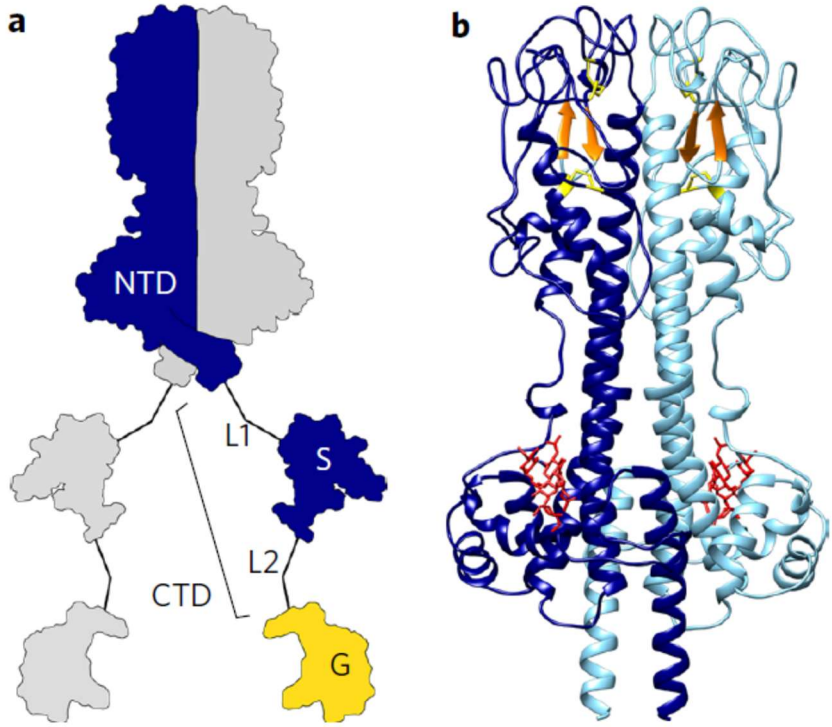

Figure 7: Structure of the NTD of *T. brucei* MITat1.1 VSG.
(A) Illustration of the domain organization of MITat1.1 VSG dimer showing the NTD, and the CTD. Linker L1 connects the NTD to the CTD structured domain (S), and Linker L2 connects the S domain to the GPI anchor (G). **(B)** Crystal structure of MITat1.1 NTD homodimer. The two monomers are colored in dark blue and light blue. The cysteine residues are shown in yellow, the antiparallel β-sheets are highlighted in orange, and the N-glycan is colored red. Image adapted from (Bartossek et al., 2017).

The majority of *T. brucei* VSGs are *N*-glycosylated (Schwede et al., 2015). This post-translational modification appears to be required for the expression of some but not all VSGs (Nicola Jones, unpublished). Additionally, studies on the lateral diffusion of VSGs have shown that *N*-glycosylation modulates the molecular crowding threshold in order to maintain

fluidity of the VSG coat (Hartel et al., 2016). Recently, Pinger et al. reported *O*-glycosylation of a *T. brucei* VSG molecule for the first time. VSG3 (MITat1.3) has an *O*-linked α-glucose moiety on serine 317, which is exposed at the topmost accessible surface of the VSG. This *O*-linked glycosylation of VSG3 increases virulence: parasites expressing VSG3 with a point mutation where serine 317 was replaced with alanine, thus abolishing *O*-glycosylation, showed reduced evasion of mice IgM immune response. Mice infected with parasites expressing this mutant VSG3 cleared the parasites and survived what would have been the first peak of parasitaemia, while 90% of mice infected with the wild type VSG3 succumbed to infection during the first peak of parasitaemia (Pinger et al., 2018).

VSGs of *T. vivax* and *T. congolense* are not well characterized and their molecular structures have not been experimentally determined. The known *T. vivax* VSGs are smaller in comparison to the *T. brucei* VSGs with a molecular mass of the monomers ranging from 40 – 46 kDa (Gardiner et al., 1996; Gardiner & Wilson, 1987). It has been proposed that both *T. vivax* and *T. congolense* VSGs have only minimal polypeptide extensions after the NTD (Schwede et al., 2015).

1.9 Regulation of *VSG* mRNA levels in *T. brucei*

Spatiotemporal regulation of gene expression is essential in driving a myriad of cellular functions in prokaryotes and eukaryotes. In eukaryotic cells, DNA is localized in the nucleus packaged as chromatin (Ridgway & Almouzni, 2001). A big chunk of eukaryotic genomes comprises introns, which are located between gene coding sequences. Most eukaryotic genes are transcribed individually under the control of specific promoters. However, some level of gene clustering is present in some eukaryotes, including *Caenorhabditis elegans*, tunicates and kinetoplastids (Ganot et al., 2004; Spieth et al., 1993).

Most genes in *T. brucei* are arranged in polycistronic transcription units (PTUs) transcribed mainly by RNAP II (Clayton, 2019; Palenchar & Bellofatto, 2006; Vanhamme & Pays, 1995). Strikingly, virulence genes coding for the major surface proteins in both *T. brucei* BSF and PCF parasites are arranged in PTUs that are transcribed by the highly processive RNAP I (Günzl et al., 2003). Most *T. brucei* genes lack introns and the pre-mRNA is co-transcriptionally processed by *trans*-splicing a capped 39-nucleotide exon at the 5′ end of the mRNA, and polyadenylation of the 3′ end of the upstream mRNA into monocistronic mRNAs (Bangs et al., 1992; Matthews et al., 1994; Ullu et al., 1993).

mRNA stability is an important aspect of gene expression regulation. Most *T. brucei* mRNAs have half-lives of between 5 – 20 minutes whereas mRNAs for abundant proteins such as tubulin and VSG have longer half-lives exceeding 2 hours (Manful et al., 2011). The high

abundance and longer *VSG* mRNA half-lives relative to the *ESAGs* in the VSG PTU points to the existence of a unique mechanism to regulate the expression of *VSG* gene. *T. brucei VSG* 3′UTRs harbor conserved 8-mer and 16-mer motifs. The conserved 16-mer motif is essential for stabilizing the *VSG* transcript (Berberof et al., 1995; Ridewood et al., 2017). How exactly the 16-mer motif stabilizes the *VSG* transcripts remained elusive until the recent finding that the 16-mer associates with an F-box mRNA binding protein CFB2, which recruits a stabilizing complex and a cap-binding translation complex that act to stabilize the *VSG* mRNA (do Nascimento et al., 2020). Additionally, it is suggested that 16-mer-dependent enrichment of N6-methylation of adenosine nucleotides in the *VSG* mRNA poly(A) tails could also be essential for stabilizing *VSG* transcripts (Viegas et al., 2020).

VSG mRNA is the most abundant in BSF *T. brucei,* constituting approximately 10 – 20% of total transcripts (Kraus et al., 2019; Nilsson et al., 2010). When a second *VSG* is integrated upstream of the BES-resident *VSG*, the mRNA and protein of the endogenous and ectopic VSGs are expressed on a 50:50 basis (Muñoz-Jordán et al., 1996). A similar expression pattern was observed when the ectopic VSG was integrated downstream of the BES promoter (Ridewood et al., 2017; Smith et al., 2009). When the transcript of one of the *VSGs* is depleted by RNAi, the second *VSG* is expressed at wild type levels (Smith et al., 2009). Inducible overexpression of an ectopic *VSG* from the rDNA spacer supports these observations: the total transcripts of both the endogenous and the ectopic *VSG* do not exceed the wild type levels following induction of expression (Batram et al., 2014) . Therefore, there appears to be a robust regulation mechanism that keeps the *VSG* mRNA at ideal levels. Maudlin et al. propose that the regulation of *VSG* mRNA is dependent on the production of functional GPI-anchored VSG proteins (Maudlin et al., 2021).

Codon usage bias has been shown to influence gene expression. Different combinations of base-triplets (codons) encode different amino acids. A single amino acid can be encoded for by up to six codons. The choice or preference of a particular codon for an amino acid, called codon-usage bias, can influence eukaryotic gene expression (Mahlab & Linial, 2014; Pechmann et al., 2014; Quax et al., 2015). In *T. brucei*, there is a strong bias towards G/C in the third position in many genes, synonymous to high mRNA and protein abundance (Horn, 2008; Jeacock et al., 2018). Whereas this is the case for most *T. brucei* genes, VSG genes show a rather unusual codon-usage bias. In these genes, C or A appears to be favored in the third position (Cross et al., 2014). Interestingly, this unusual VSG codon-usage bias was also observed in *T. vivax* and *T. congolense* VSG gene families. Whether this unusual VSG-gene codon-usage bias is critical for the regulation of VSG expression remains to be determined.

1.10 VSG biosynthesis

VSGs are essential for the parasite's survival both *in vivo* and *in vitro*. As the VSG constitutes ~10% of total cellular protein and ~90% of the cargo for the secretory pathway, its synthesis is likely finetuned. However, studies have reached contrasting conclusions on whether an excess or just sufficient VSG protein is synthesized (Field et al., 2010; Tiengwe et al., 2016). The VSG biosynthetic pathway can be divided into five stages: VSG translocation into the endoplasmic reticulum (ER), folding and glycosylation, the addition of GPI anchors, ER exit, and processing in the Golgi, and delivery to the cell surface. Here the focus is drawn to translocation of VSGs into the ER and GPI anchoring.

1.10.1 Translocation of VSGs into the ER

The translocation of proteins in eukaryotes is complex compared to prokaryotes. This is due to subcellular compartmentalization and the need for post-translational protein modifications. Mammals and yeast utilize two pathways for trafficking of polypeptides into the ER lumen: the signal recognition particle (SRP)-dependent and -independent pathways (Ast et al., 2013; Rapoport, 2007). These pathways are conserved in kinetoplastids including the African trypanosomes, leishmania and South American *T. cruzi* (Cordero et al., 2019; Goldshmidt et al., 2008; Silverman et al., 2008). Nascent polypeptides of secretory proteins are synthesized with cleavable signals targeting the proteins to the ER.

Nascent VSG polypeptides contain cleavable signal peptides at the N- and C-terminal ends. The N-terminal signals of the BSF VSGs are approximately 15 – 30 residues long and conform to the general physicochemical characteristics of ER import signal peptides (Boothroyd et al., 1981; Cross, 1984). Signal peptides have three conserved physicochemical regions: a positively charged N-region, a central hydrophobic core (h-region) and a more polar C-terminal cleavage region (Heijne, 1988; Heijne, 1998; Wickner & Schekman, 2005). As the case is in other organisms, a clear conservation pattern of the VSG signal peptide residues appears to be lacking (Al-Qahtani et al., 1998; Boothroyd et al., 1981; Kunze & Berger, 2015). However, it has been shown that the h-regions of *T. brucei* and *Saccharomyces cerevisiae* ER import signals contain conserved residue motifs which appear to contribute to the overall activity of the signal peptides (Duffy et al., 2010). Other aspects of the h-region including, the length and the overall hydrophobicity, can also determine the efficiency of signal peptides (Huber et al., 2005; Nilsson et al., 2015). Whether the VSG is imported into the ER via the SRP-dependent or independent one is not known. However, procyclic *T. brucei* preferentially utilizes the SRP-dependent pathway for the translocation of transmembrane proteins and the SRP-independent one for GPI anchored surface protein

(Goldshmidt et al., 2008). Whether this applies to translocation of VSGs remains to be determined. Concomitant with translocation is the cleavage of the signal peptide, presumably by ER-resident signal peptidase enzyme, after which maturation of the VSG protein begins, and glycosylation takes place.

1.10.2 GPI anchoring

GPI-anchors and GPI-anchored proteins (GPI-APs) are ubiquitous in eukaryotes, having been identified from simple unicellular protozoan to complex mammalian systems. African trypanosome research has contributed immensely to the knowledge of GPI-anchoring of surface proteins. The first GPI anchor structure to be solved was from *T. brucei* (Ferguson et al., 1988).

In the ER, a preassembled GPI precursor is transferred *en bloc* to newly synthesized proteins as a post-translational modification (Walter & Johnson, 1994). This modification occurs by a transamidation reaction involving the GPI precursor and the carboxyl-terminal GPI signal peptide of a partially folded protein. The GPI signal peptide is cleaved in the process, and the carboxyl-terminal is joined to the amino group of the ethanolamine residue in the GPI precursors (Ikezawa, 2002; Nagamune et al., 2003; Walter & Johnson, 1994). From the ER, proteins are transported to the Golgi apparatus where the glycans and the GPI anchors are further modified. Subsequently, vesicles containing the proteins bud off, delivering the proteins to the plasma membrane (Kelly, 1985).

Eukaryotic GPI signal peptides show low sequence conservation though it appears there is a selection mechanism for amino acids to occupy specific positions. The signal peptide consists of three domains: three small amino acids at the ω, ω +1 and ω +2 positions (where ω is the amino acid at which the preformed GPI anchor is added), a polar core of between 5 – 10 amino acids and a hydrophobic core of between 15 – 20 amino acids (Ferguson et al., 2015). The *T. brucei* VSG GPI signal peptides show a higher degree of sequence conservation (Böhme & Cross, 2002; Cross, 1984). These VSG GPI signal peptides are classified into two groups. In the first group, the ω position is almost always a serine, and the signal peptide sequence is 17 amino acids in length. The second group has either asparagine or aspartate in the ω position, with a signal peptide length of 23 amino acids. Other conserved features include the ω+2 and ω+7 positions, which are always a serine and a lysine, respectively. A majority of the GPI signal peptides terminate in three hydrophobic amino acids Leu-Leu-Phe (Böhme & Cross, 2002; Cross, 1984).

T. vivax and *T. congolense* VSG signal peptides have not been well characterized. However, from the few confirmed VSG sequences of these species, it appears that their GPI signal peptides do not share the characteristic features of the *T. brucei* VSG GPI signal peptides.

1.11 Protein structure determination

Proteins carry out their regulatory or structural functions independently or by interacting with other proteins. To understand how they do this at the molecular level, it is important to experimentally determine their exact three-dimensional (3D) structure. Over the years different methods for solving the 3D atomic structure of proteins have been developed. The experimental approaches used for protein structure determination can be divided into those that require protein crystals and those that do not. The widely used approaches that do not require protein crystals include nuclear magnetic resonance (NMR) spectroscopy (Wüthrich, 2001), cryo-electron microscopy (Cryo-EM) (Yu et al., 2008) and small-angle X-ray and neutron scattering (SAXS and SANS) (Ibel & Stuhrmann, 1975; Petoukhov & Svergun, 2006; Svergun & Koch, 2003). The advantage of these methods is that they can be used to solve structures of proteins that do not crystallize or form poor quality crystals. Of these methods, NMR has the best resolution (atomic level), but it is only applicable for small proteins. Structure determination by X-ray crystallography requires crystals and can yield atomic resolution.

1.11.1 X-ray crystallography

Since the discovery of X-rays in 1895 by Wilhelm Röntgen and the determination of the myoglobin protein structure (Kendrew et al., 1958) , X-ray crystallography has become the cornerstone of protein structure determination. At present, only X-ray and neutron diffraction, when applied to crystals, together with NMR can resolve the structures of biological macromolecules to atomic levels. The success of X-ray crystallography and the quality of diffraction data is mainly dependent on the ability to grow sizeable crystals of appropriate physicochemical properties. This requirement makes crystallization experiments the rate-limiting step in macromolecular structure determination by this method (McPherson, 1989). A sample of very high purity and concentration ranging from ~2 – 50 mg/ml is required to set up a crystallization experiment. Since it is difficult to purify such high protein amounts from their natural environments, molecular cloning techniques and heterologous protein expression systems are frequently used to generate sufficient material for crystallization.

Different proteins and macromolecules in solution respond differently to changes in pH, temperature and salt concentration. How these factors affect the ability of a particular protein to crystallize is not known. As a result, macromolecular crystal growth is mostly by trial (McPherson & Gavira, 2014). Easy-to-use commercially available kits are increasingly being used for initial screening trials of appropriate crystal growth conditions. These kits have different buffer, pH, salt, and precipitant formulations that offer a broad spectrum of crystal growth conditions. Once some conditions for crystal growth have been identified in the trials, they are optimized before scaling up are to produce large crystals. The widely used methods to promote macromolecular crystal growth include liquid-liquid diffusion, vapor and batch-diffusion (Dessau & Modis, 2011; Ducruix & Giegé, 1992). In vapor diffusion, the protein sample is first mixed with an equal volume of the mother liquor (reservoir). When this sample mixture is placed as a drop in a sealed compartment containing the reservoir solution, the concentration difference drives the diffusion of water molecules from the protein solution drop to the reservoir. This diffusion, in turn, decreases the droplet size while increasing the protein concentration and when a state of supersaturation is achieved, crystals start to grow (Figure 8 A). Equilibration is achieved when the precipitant concentration in the drop is approximately the same as in the reservoir.

There are two methods for promoting crystal growth by vapor diffusion; sitting drop and hanging drop vapor diffusion (McPherson, 1998; Weber). Both methods use the same principle with the only difference being the positioning of the drops. In sitting drop vapor diffusion, the drop is placed on a small well adjacent to a well containing the reservoir while in hanging drop vapor diffusion the drop is suspended on top of the reservoir on an inverted coverslip (Figure 8 B). Crystals are normally grown at temperatures ranging between 4 and 37 °C The sitting drop crystallization method has the advantage of allowing automation.

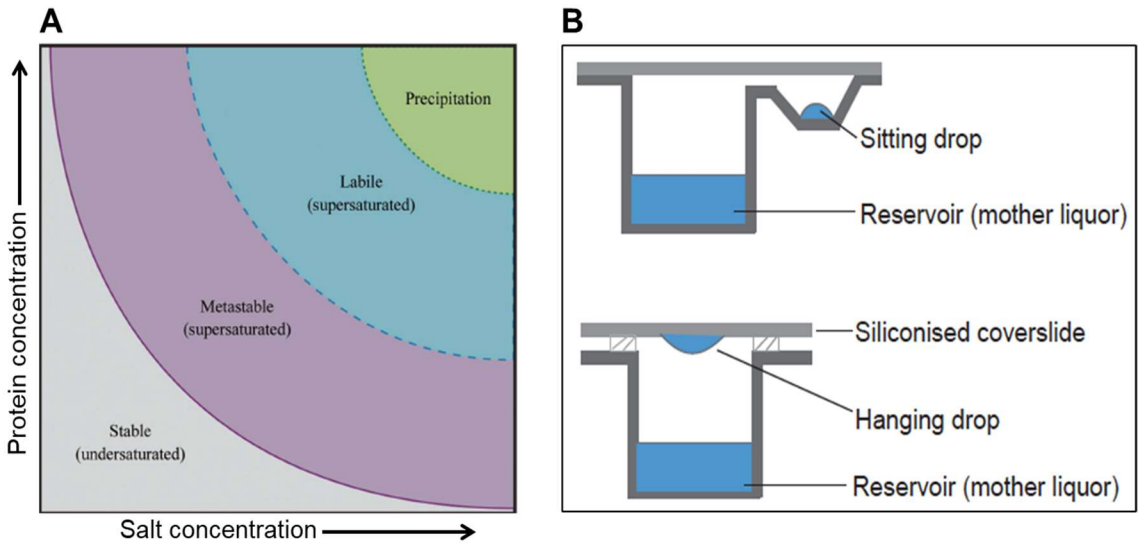

Figure 8: The principle of protein crystallization.
(A) Phase diagram of macromolecule crystallization. In vapor diffusion experiments, crystals grow when a state of supersaturation is achieved. In the stable region, no crystals grow as the protein is undersaturated and as the vapor diffuses from the sample, a supersaturated metastable state is achieved where nuclei and crystals may be formed. Nucleation occurs in the labile region, and crystals grow, while only precipitates can occur in the precipitation region. Image adapted from (McPherson and Gavira, 2014). **(B)** Illustration of sitting drop (top) and hanging drop (bottom) vapor diffusion set up.

1.11.2 X-ray diffraction

Once crystals of appropriate size and homogeneity have been obtained, the next critical step is diffraction data acquisition. To obtain useful diffraction data for protein structure determination, one may be required to screen several crystals, depending on the complexity of the macromolecules. Generally, multi-domain protein complexes and membrane proteins require a high number of crystals (Abola et al., 2000). A crystal is made up of periodic arrangements of a unit cell to form a lattice. A unit cell is the smallest 3D unit of a molecule that may contain a single atom or atoms in a fixed angular arrangement. The unit cell dimensions are described by the Miller indices *h, k* and *l* (Miller, 1839). When crystals are exposed to an X-ray beam, the incident radiation is diffracted and recorded by photodetectors. This is governed by Bragg's law, which states that when a crystal is bombarded with X-rays of a fixed wavelength at certain angles, intense reflected X-rays are produced that constructively interfere (Jauncey, 1924). This is represented by the equation: $n\lambda = 2dsin\theta$ where n is the order of reflection, λ is the wavelength of the incident X-ray radiation, d is the interplanar spacing of the crystal and θ is the angle of incident radiation. For each unit cell in the path of the incident radiation, the diffracted wave angle and intensity is recorded by a photodetector. This diffraction pattern and intensity is interpreted by the so-called *Ewald sphere,* which gives the reciprocal position effects of the atom(s) within a unit

cell. Once, the unit cell and space group have been defined, an electron density map is generated, from which the protein structure is then determined. However, the phase information is lost during this process. Without this phase information, reconstruction of the macromolecule structure from the unit cell diffractions is not tenable. This phenomenon has been christened the “phase problem”, and over the years, crystallographers have devised different methods to solve this problem. The most used methods for solving the phase problem include molecular replacement and isomorphous replacement. Molecular replacement relies on the availability of a structurally similar model with sequence similarity of more than 25% (Evans & McCoy, 2008; Rossmann, 1990, 2001). The rotation and translation function of the known molecule are computationally calculated and placed in the unit cell. From this model, the density map of the new protein can be interpreted.

On the other hand, the isomorphous replacement method of structure determination relies on the availability of heavy-atom structures and works well for larger molecules like proteins. A reactive heavy metal ion group is added to the protein by either soaking the crystals in a solution containing low concentrations of heavy metal salts or by co-crystallization of the protein with heavy metal salts (Dickerson et al., 1961; Ke, 1997) . Incorporation of the heavy metals should be in a way that does not interfere with the unit cell of the protein. Since the heavy metal atoms diffract strongly relative to the unit cell of the protein of interest, their position within the protein can be easily determined. The first protein structures to be resolved by this method were those of hemoglobin (Green et al., 1954) and myoglobin (Kendrew et al., 1958). Other phase determination methods that have come with the advancement of instrumentation and computational methods include multiwavelength anomalous scattering or dispersion and density modification (Hendrickson & Ogata, 1997; Taylor, 2010).

1.12 Aims of the study

African trypanosomes are exclusively extracellular, and therefore, their surface coats form an essential interface between the parasite and the host. The parasites use their surface for evasion of the host immune responses and acquisition of nutrients and host-derived macromolecules. A dense monolayer of VSGs covers the cell surface of BSF parasites. VSG expression is maintained at roughly constant level — including during VSG switching — to always ensure a dense VSG coat. of the *VSG* mRNA is highly regulated to ensure wild type levels of the protein is always expressed on the cell surface — including during VSG switching. Due to polycistronic expression of genes in trypanosomes, the regulation of gene expression is mostly post-transcriptional. Different studies have reported that the total *VSG*

mRNA amount in *T. brucei* does not exceed the wild type levels even when a second *VSG* is expressed (Berberof et al., 1995; Smith et al., 2009). Thus, it has been suggested that a robust regulation mechanism operates to keep the total *VSG* mRNA amounts in the cell in check, and that such a mechanism operates in *trans*. The conserved 16-mer motif in the VSG 3′UTR and the CFB2 RBP are proposed to be involved in this regulation of *VSG* mRNA levels. However, the mechanistic aspects of this *VSG* mRNA regulation mechanism are not known. Therefore, the first aim of this thesis was to determine how *VSG* mRNA levels are regulated in *T. brucei.*

Despite the many years of research on African trypanosomes, studies of the major pathogens responsible for AAT, *T. vivax* and *T. congolense*, has been lagging. This can be attributed to the lack of both forward and reverse genetic manipulation tools. Only recently has the genetic manipulation of *T. congolense* become possible. Whereas BSF *T. congolense* and insect stages can be cultured in media, only the insect stage *T. vivax* epimastigotes have been successfully cultivated *in vitro* to date. Thus, the cell surface of BSF *T. vivax* is not well characterized. The few studied *T. vivax* VSGs are smaller (40 – 46 kDa) in size compared to the *T. brucei* VSGs (50 – 60 kDa), and they appear to lack structured regions present in the CTD of *T. brucei* VSGs. The possible effect of these structural differences on the cell surface architecture is not known. Therefore, I sought to utilize the tractable *T. brucei* system as a platform for forward-surface engineering to study *T. vivax* VSGs. Specifically, the second and third sections of this thesis aimed to: (i) determine whether *T. vivax* VSGs can form a surface coat on BSF *T. brucei* and (ii) solve the three-dimensional structure of a *T. vivax* VSG using X-ray crystallography.

2.0 Materials and Methods

2.1 Materials

2.1.1 Antibodies for western and dot blots

Table 1: Antibodies used for Western and dot blots

Antibody	Organism	Dilution	Source
Primary antibodies			
Anti-PFR L13D6	Mouse	1:20	P. Bastin, Paris (FR)
Anti-GFP	Mouse	1:1000	Roche, Rotkreuz (CH)
Anti-VSG 221	Rabbit	1:5000	N. Jones, Würzburg (DE)
Anti-VSG 121	Rabbit	1:2000	M. Carrington, Cambridge (UK)
Anti-ILDat1.2	Rabbit	1:10000	Pineda Antibody services, Berlin (DE)
Secondary antibodies			
Anti-mouse IgG *(IRDye680LT)*	Goat	1:10000	LI-COR Biosciences, Lincoln (USA)
Anti-rabbit IgG *(IRDye800CW)*	Goat	1:10000	LI-COR Biosciences, Lincoln (USA)

2.1.2 Buffers and Solutions

DNA analyses

1x TAE buffer

40 mM Tris, 40 mM glacial acetic acid, 1 mM EDTA (pH 8.0)

6x DNA loading dye

10 mM Tris-HCl (pH 7.6), 0.03% bromophenol blue, 0.03% xylene cyanol FF, 60% glycerol 60 mM EDTA

P1 resuspension buffer

50 mM Tris-HCl (pH 8.0), 10 mM EDTA, 100 µg/ml RNase A

P2 lysis buffer

200 mM NaOH, 1% SDS (w/v)

P3 neutralization buffer

3 M potassium acetate (pH 5.5)

RNA analyses

Denhardt's solution (DEN, 50x)

1% (w/v) Ficoll 400, 1% (w/v) polyvinylpyrrolidone, 1% (w/v) bovine serum albumin (BSA)

Glyoxal solution

75.4% (v/v) DMSO, 8.5% (v/v) Glyoxal (deionized), 15 mM NaH_2PO_4 (pH 6.9)

Northern blot hybridization solution

5x SSC, 5x DEN, 0.1% (w/v) SDS, 0.1% (w/v) tetrasodium pyrophosphate, 0.01% (w/v) heparin

Northern blot washing solution

0.1% (w/v) SDS, 0.1x SSC

RNA buffer (50x)

0.5 M NaH_2PO_4 (pH 6.9)

20x SSC

3 M NaCl, 0.3 M trisodium citrate (pH 7.0)

***T. brucei* cultivation**

HMI-9 medium

For 1 l of media, 17.66 g Iscove's modified Dulbecco's MEM (IMDM) powder, complemented with 3.024 g sodium bicarbonate, 136 mg hypoxanthine, 28.2 mg bathocuproine sulfonate, 14 µl β-mercaptoethanol, 39 mg thymidine, 100,000U penicillin, 100 mg streptomycin, 182 mg cysteine. The medium was prepared using filtered and deionized water (ddH_2O) and the pH adjusted to 7.5. Finally, the medium was sterilized by filtration (pore size 0.2 µm), and 10% (v/v) heat-inactivated (1 h, 56 °C) fetal calf serum (FCS) added.

2x trypanosome freezing mix

20% (v/v) glycerol in HMI-9 medium, filter sterilized

Trypanosome dilution buffer (TDB)

5 mM KCl, 80 mM NaCl, 1 mM $MgSO_4$, 20 mM Na_2HPO_4, 2 mM NaH_2PO_4 and 20 mM glucose in ddH_2O (pH 7.6)

Protein analyses

Coomassie solution

10% (v/v) 2-propanol, 5% (v/v) glacial acetic acid, 0.05% (w/v) Coomassie R-250 in ddH_2O

Destaining solution

10% (v/v) 2-propanol, 5% (v/v) glacial acetic acid in ddH_2O

2x Laemmli sample buffer

120 mM Tris-HCl, pH 6.8, 4% (w/v) SDS, 20% (v/v) glycerol, 0.004% (w/v) bromophenol blue, 2% (v/v) β-mercaptoethanol

Laemmli running buffer

25 mM Tris-HCl, 192 mM glycerol, 0.1% (w/v) SDS

Stacking gel buffer

0.5 M Tris-HCl (pH 6.8), 0.4% (w/v) SDS

Stacking gel

0.39 ml 30% acrylamide/bis-acrylamide 37.5:1, 0.75 ml stacking gel buffer, 1.83 ml ddH_2O, 15 µl 10% (w/v) ammonium persulfate (APS), 3 µl TEMED

Resolving gel buffer

1.5 M Tris-HCl (pH 6.8), 0.4% (w/v) SDS

Resolving gel 12.5%

2 ml 30% acrylamide/bis-acrylamide 37.5:1, 1.5 ml resolving gel buffer, 2.5 mL ddH_2O, 30 µl 10% APS (w/v), 10 µl TEMED

Phosphate buffered saline (PBS)

10 mM Na_2HPO_4, 1.7 mM KH_2PO_4, 137 mM NaCl, 2.7 mM KCl, pH 7.4

Enzymes

Enzymes used in this study were either obtained from Thermo Fisher Scientific or New England BioLabs.

Commercial crystallization screens

Table 2: Overview of commercial crystallization screens

Screen	Manufacturer
Index HT	Hampton Research (USA)
JCSG+	Molecular Dimensions (USA)
OptiMix PEG	Fluidigm (USA)
Protein complex	Qiagen (NL)
Wizard Classic I/II	Rigaku (USA)

2.1.3 Equipment and devices

Table 3: Overview of equipment used

Equipment	Manufacturer
Airstream Class II BSC	Esco global, Changi (SG)
AMAXA Nucleofector II	Lonza, Basel (CH)
Äktaprime plus	GE Healthcare, Little Chalfont (UK)
Binder CO_2 incubator	Binder, Tuttlingen (DE)
EM AFS2 freeze substitution system	Leica Microsystems, Wetzlar (DE)
EM HPM100 HPF-system	Leica Microsystems, Wetzlar (DE)
FACScalibur Flow Cytometer	BD Biosciences, Franklin Lakes (USA)
HPLC Ultimate 3000	Dionex softron, Germering (DE)
iBright CL1000 image scanner	Invitrogen, Waltham (USA)
iMIC	Till Photonics, Munich (DE)
JEM-2100 TEM	JEOL, Tokyo (JP)
Li-COR Clx Imager	Li-COR Biosciences, Lincoln (USA)
Lyophilisation device	CHRIST, Osterode (DE)
ManiFold dot blotter	GE Healthcare, Little Chalfont (UK)
Neubauer chamber	Marienfeld, Lauda-Königshofen (DE)
NT8® Drop setter crystallization robot	FORMULATRIX® Bedford, MA (USA)
Rock Imager	FORMULATRIX® Bedford, MA (USA)
T-100 Thermo cycler	Bio-Rad Laboratories, Munich (DE)
Tecan Infinite M200 Plate reader	Tecan, Männedorf (CH)
TemCam F416 4k x 4k	Tietz video & Imaging, Gauting (DE)
Trans-Blot semi-dry transfer cell	Bio-Rad Laboratories, Munich (DE)

2.1.4 Fluorescent probes and dyes

Table 4: List of fluorescent RNA probes

Probe	Fluorophore	Sequence (5′ – 3′)
Tubulin	DY-782	ATCAAAGTACACATTGATGCGCTCCAGCTGCAG
GFP	DY-682	GTCGCCGTTCTTCTGCTTGTCGGCCATGATATA
VSG221	DY-682	CAGCGTAAACAACGCACCCTTCGGTTGGTCGTC
GPEET 5′UTR (pLew82v4)	DY-682	TAGGATATCAAGCTTGTGAATTTTACTTTTTGG

2.1.5 Kits

Table 5: Kits used

Kit	Source
Amaxa Basic Parasite Nucleofector Kit	Lonza, Basel (CH)
Clone JET PCR Cloning Kit	Thermo Fisher Scientific, Waltham (USA)
NucleoSpin® Gel and PCR Clean-up	Macherey-Nagel, Düren (DE)
NucleoSpin® Plasmid miniprep Kit	Macherey-Nagel, Düren (DE)
Nucleobond® PC 100 midiprep Kit	Macherey-Nagel, Düren (DE)
Phusion Human Specimen Direct PCR Kit	Thermo Fisher Scientific, Waltham (USA)
RevertAid First Strand Synthesis Kit	Thermo Fisher Scientific, Waltham (USA)
RNeasy Mini Kit	Qiagen, Venlo (NE)

2.1.6 Oligonucleotides

Table 6: List of primers for PCR.

The list indicates the primer name, sequence, manufacturer, and the name of the designer.

Name	5′ – 3′ Sequence	Source	Designed by
M13 For	GTAAAACGACGGCCAGT	Promega	Commercial
M13 Rev	AACAGCTATGACCATG	Promega	Commercial
T3	AATTAACCCTCACTAAAGG	MWG	Commercial
T7	TAATACGACTCACTATAGG	MWG	Commercial
D63	GCTGCACGCGCCTTCGAG	MWG	N. Jones
D64	TGCTGTGCCATCAGATTA	MWG	N. Jones
D80	GAATTCCCCTAACTAACCTAACTA	Sigma	N. Jones
D81	GAATTCTAAGCGATGGCCGTGCAC	Sigma	N. Jones
D128	GAATTCGAACAGTTTCTGTACTATATTG	Sigma	N. Jones
D129	GAATTCGTGTTAAAATATATC	Sigma	N. Jones
D130	GAATTCTCAAGAATGACTAGTAGC	Sigma	E. Aroko
D131	GTTTAAAAAGAGGGGGAATTAAAAAAGT AAGGCCGC	Sigma	E. Aroko
D132	GAATTCCCCTAACTAACCTAAC	Sigma	E. Aroko
D133	GAATTCATGGCACCTCGTTCCCTT	Sigma	N. Jones
D134	GTTTAAAAAGAGGGGGAATTAGAATGCG GCAACGAG	Sigma	N. Jones
D135	ACTCTGCATTGCAGTCTC	Sigma	N. Jones
D136	ATAAGCACTAGCCGTGTC	Sigma	N. Jones
D138	GGGAAATTTGAAGTTTTAGAAACAGAAG GCAGCTGC	Sigma	N. Jones
D139	GGGAAATTTGAAGTTTTAGAAACAGAAG GCAGCTGC	Sigma	N. Jones
D146	GAATTCATGTGGCAGCGAATTTTG	Sigma	E. Aroko
D147	CTTACTAGAAAACTGGAATCGCGCGCTT TGCGGTCTTG	Sigma	E. Aroko
D148	GATTCCAGTTTTCTAGTAAG	Sigma	E. Aroko
D149	GAAATAGCGCAATCGCAAAG	Sigma	E. Aroko

D150	TTGCGATTGCGCTATTTCTTCTTTTAGTTC TTTAGTTAAGG	Sigma	E. Aroko
D151	TGGGATCGGGTAGAAGAC	Sigma	E. Aroko
D152	GTCTTCTACCCGATCCCACATTCG	Sigma	E. Aroko
D159	CTTGTGCTTCTGGGTCATTCGTTGCCGTT CATTTTC	Sigma	E. Aroko
D160	ACCCAGAAGCACAAGCC	Sigma	E. Aroko
EA4	GCGGGCTTGTGCTTCTGGGTGACGAGCC TTTCTGTGTT	Sigma	E. Aroko
EA5	GCGGAAGGCGGCAACGCAG	Sigma	E. Aroko
EA6	CTGCGTTGCCGCCTTCCGCTGCTCCCGAT TTTCTGG	Sigma	E. Aroko
EA7	ACAGACAAAGGCGCGATCAAG	Sigma	E. Aroko
EA8	GATCGCGCCTTTGTCTGTCTCCGCGTGCC TCGCGAAG	Sigma	E. Aroko
EA9	ATGGCACCTCGTTCCCTTTATCTGCTCGC TGTTCTTCTGTTCAGCGCGAACCTCTTCG CTGGCGTGGGATTTGCCGCAGCCATGCT GAGCAAGGGCGAGGAGCTG	Sigma	E. Aroko
EA10	GGCTGCGGCAAATCCCACGCCAGCGAAG AGGTTCGCGCTGAACAGAAGAACAGCGA GCAGATAAAGGGAACGAGGTGCCATGCC GCGTTCGTGTCGCGTAGGAATAAC	Sigma	E. Aroko
EA11	GTAAAAGATCTAGTATATAGGAGCAACG C	Sigma	E. Aroko
EA12	CTCCTCGCCCTTGCTCAGCATGGTGGCGA CCGGCCGGTGGATC	Sigma	E. Aroko
EA13	GATCCACCGGCCGGTCGCCACCATGCTG AGCAAGGGCGAGGAG	Sigma	E. Aroko
EA14	TGCTCCCGATTTTCTGGGTAGTAAAACGT AAAGACTAATCGCGTACGCCGCTAGGGC TCTGTGCACGGCCATGCCGCGTTCGTGTC GCGTAGGAATAAC	Sigma	E. Aroko
EA15	ATGGCCGTGCACAGAGCCCTAGCGGCGT ACGCGATTAGTCTTTACGTTTTACTACCC AGAAAATCGGGAGCAATGcTGAGCAAGG GCGAGGAGCTG	Sigma Sigma	E. Aroko E. Aroko
EA16	ATGGCACCTCGTTCCCTTTATCTGCTCGC TGTTCTTCTGTTCAGCGCGAACCTCTTCG CTGGCGTGGGATTTGCCGCAGCCACAGA CAAAGGCGCGATCAAGTTTG	Sigma	E. Aroko
EA17	ATGATGATTTCCCGCGCTTTGTTTGTGCT CGGCCTCAGTCTTCACTTCATATTCGCAC CGACCGCCTACGCGACAGACAAAGGCGC GATCAAGTTTG	Sigma	E. Aroko
EA18	CGCGTAGGCGGTCGGTGCGAATATGAAG TGAAGACTGAGGCCGAGCACAAACAAA	Sigma	E. Aroko

	GCGCGGGAAATCATCATGCCGCGTTCGT GTCGCGTAGGAATAAC		
EA19	ATGAGCATCACTTTTCATAGTTTATGGCT ACTTTTGACAGTGTTATGCACTGCAGGTA TTCGTGCTACAGACAAAGGCGCGATCAA GTTTG	Sigma	E. Aroko
EA20	AGCACGAATACCTGCAGTGCATAACACT GTCAAAAGTAGCCATAAACTATGAAAAG TGATGCTCATGCCGCGTTCGTGTCGCGTA GGAATAAC	Sigma	E. Aroko

2.1.7 Software

Table 7: List of software

Program	Source
4Peaks	Macentosj
Abode Illustrator Creative Suite 6	Adobe Systems Inc., San Jose (USA)
Adobe Photoshop Creative Suite 6	Adobe Systems Inc., San Jose (USA)
CellQuest™ pro	BD Biosciences, Franklin Lakes (USA)
CLC Main Workbench 7	CLC bio, Venlo (NE)
iControl	Tecan, Männedorf (CH)
INTAS	INTAS Science Imaging, Göttingen (DE)
ImageJ64	National Institute of Health (USA)
Image Studio Lite	LI-COR Biosciences, Lincoln (USA)
Live Acquisition	TILL-Photonics, Gräfelfing (DE)
Prism 5.0	GraphPad Software, La Jolla (USA)
Rock Maker®	FORMULATRIX® Bedford, MA (USA)

2.1.8 Plasmids

Table 8: List of plasmids used in this study.

The plasmid list indicates the name, origin and the selection markers used. The integration locus is indicated for plasmids used for *T. brucei* transfection.

Name	Locus	Expression	Resistance	Origin
221KO	221 ES	Constitutive	Puromycin	J. Bührdel
p3845	221 ES^{prom}	Constitutive	Blasticidin	A. Schwede
p4231	β-Tubulin	Constitutive	Blasticidin	M. Carrington
pKD4	221 ES	Constitutive	Neomycin	(Muñoz-Jordán et al 1996)
pbRN6	221 ES^{tel}	Constitutive	Blasticidin	(Horn and Cross 1997)
pLew82v4	rDNA spacer	Inducible	Phleomycin	G. Cross
pJET1.2		Cloning	Ampicillin	ThermoFischer
pBluescript II SK (+)		Cloning	Ampicillin	Snapgene

ES, Expression site; ES^{prom}, Expression site promoter-proximal; ES^{tel}, Expression site telomere-proximal; rDNA Spacer, Ribosomal DNA spacer.

2.1.9 Organisms

Bacterial strains

Escherichia coli *(*E. coli*) Top 10*

Genotype: F-mcrA Δ(mrr-hsdRMS-mcrBC) Φ80lacZΔM15 ΔlacX74 recA1 araD139 Δ(araleu)7697galU galK rpsL(StrR) endA1 λ-

Trypanosomes

Wild type strains

Molteno Institute Trypanozoon antigen type 1.3 BSF, expressing VSG 224 (also known as MITat1.3 or VSG3).

Molteno Institute Trypanozoon antigen type 1.6 BSF, expressing VSG 121 (also known as MITat1.6 or VSG6)

Transgenic cell lines

MITat1.2 13-90

Molteno Institute Trypanozoon antigenic type 1.2 BSF, expressing VSG 221 (also known as MITat1.2 or VSG2). This cell line expresses the tetracycline repressor and T7 polymerase (Wirtz et al., 1999). All transgenic cell lines generated in this study are based on this parental cell line.

2.1.10 Generated *T. brucei* cell lines

Table 9: Transgenic *T. brucei* cell lines generated during the study

Name	Parental cell line	Plasmid	Selection
221^{ES}.EP^{SP}:GFP:UTR	M1.2 13-90 (221^{ES})	pbRN6.EP^{SP}:GFP:UTR	5 BLAS, 5 HYG, 2.5 NEO
221^{ES}.EP^{SP}:GFP:ΔUTR	M1.2 13-90 (221^{ES})	pbRN6.EP^{SP}:GFP: ΔUTR	5 BLAS, 5 HYG, 2.5 NEO
221^{ES}.121^{SP}:GFP	M1.2 13-90 (221^{ES})	pbRN6.121^{SP}:GFP:UTR	5 BLAS, 5 HYG, 2.5 NEO
221^{ES}.EP^{SP}:VSG121	M1.2 13-90 (221^{ES})	pbRN6.EP^{SP}:VSG121	5 BLAS, 5 HYG, 2.5 NEO
221^{ES}.ILDat1.2^{tet}	M1.2 13-90 (221^{ES})	pLew.ILDat1.2	1 BLE, 5 HYG, 2.5 NEO
221^{ES}.ILDat2.1^{tet}	M1.2 13-90 (221^{ES})	pLew.ILDat2.1	1 BLE, 5 HYG, 2.5 NEO
221^{ES}.121^{SP}:ILDat2.1^{tet}	M1.2 13-90 (221^{ES})	pLew.121^{SP}:ILDat2.1	1 BLE, 5 HYG, 2.5 NEO
221^{ES}.ILDat2.1^{SP}:121^{tet}	M1.2 13-90 (221^{ES})	pLew.ILDat2.1^{SP}:VSG 121	1 BLE, 5 HYG, 2.5 NEO
221^{ES}.EP:GFP^{tet}	M1.2 13-90 (221^{ES})	pLew.EP:GFP	1 BLE, 5 HYG, 2.5 NEO
GFP:$PAD1_{UTR}$. ILDat1.2^{tet}	221^{ES}.ILDat1.2^{tet}	p4231	1 BLE, 5 HYG, 2.5 NEO, 5 BLA
GFP:$PAD1_{UTR}$.221^{RNAi}	M1.2 13-90 (221^{RNAi})	p4231	2 BLE, 5 HYG, 2.5 NEO, 5 BLAS
GFP:$PAD1_{UTR}$.221^{ES}	M1.2 13-90 (221^{ES})	p4231	5 BLAS, 5 HYG, 2.5 NEO
GFP^{PRO}.221^{ES}. ILDat1.2^{tet}	221^{ES}.ILDat1.2^{tet}	p3845	1 BLE, 5 HYG, 2.5 NEO, 5 BLA
GFP^{PRO}.221^{ES}.121^{tet}	221^{ES}.121^{tet}	p3845	1 BLE, 5 HYG, 2.5 NEO, 5 BLA
221^{ES}.ILDat1.2: M1.11^{tet}	M1.2 13-90 (221^{ES})	pLew.ILDat1.2:M1.11^{SP}	1 BLE, 5 HYG, 2.5 NEO
221^{ES}.ILDat2.1: M1.11^{tet}	M1.2 13-90 (221^{ES})	pLew.ILDat2.1:M1.11^{SP}	1 BLE, 5 HYG, 2.5 NEO
221^{ES}.121^{SP}:ILDat2.1: M1.11^{tet}	M1.2 13-90 (221^{ES})	pLew.VSG121^{SP}: ILDat2.1:M1.11^{SP}	1 BLE, 5 HYG, 2.5 NEO
221^{ES}.ILDat1.2	M1.2 13-90 (221^{ES})	pKD4.ILDat1.2:M1.11^{SP}	5 HYG, 30 NEO
221^{KO}.ILDat1.2	221^{ES}.ILDat1.2	VSG221 KO	5 HYG, 30 NEO, 0.3 PURO
221^{ES}.ILDat1.2: 221ML^{tet}	M1.2 13-90 (221^{ES})	pLew:ILDat1.2:221ML	1 BLE, 5 HYG, 2.5 NEO
221^{ES}.ILD:ML	M1.2 13-90 (221^{ES})	pKD4.ILDat1.2:221ML	5 HYG, 30 NEO
221^{ES}.ILD:CTD	M1.2 13-90 (221^{ES})	pKD4.ILDat1.2: 221CTD	5 HYG, 30 NEO
221^{KO}.ILD:ML	221^{ES}.ILD:ML	VSG221 KO	5 HYG, 30 NEO, 0.3 PURO

221^{KO}.ILD:CTD	221^{ES}.ILD:CTD	VSG221 KO	5 HYG, 30 NEO, 0.3 PURO

Abbreviations: ES, expression site; EP, EP1 procyclin; SP, signal peptide for ER import; GFP, green fluorescent protein; UTR, untranslated region; tet, inducible expression upon addition of tetracycline; KO, knock out; PRO, promoter-proximal; ILD, ILDat1.2; ML, minimal linker; CTD, C-terminal domain; BLA, blasticidin; BLE, phleomycin; HYG, hygromycin; NEO, neomycin (G-418); PURO, puromycin; VSG, variant surface glycoprotein; **Δ**, nucleotides 4 – 51 of the *VSG 121* 3'UTR were deleted. The antibiotics concentration is given in µg/ml.

2.2 Methods

2.2.1 Working with *E. coli*

Cultivation of *E. coli*

E. coli bacterial cells were grown in Luria Bertani (LB) medium or on LB agar plates and incubated at 37 °C. The bacterial cells cultivated in LB medium were agitated at 200 rpm.

Transformation of *E. coli*

Chemically competent *E. coli* Top10 cells were generated (Chung et al., 1989), and transformed by heat shocking. In the case of a ligation, 10 µl of the ligation reaction was added to 100 µl of chemically competent *E. coli* cells and incubated for 20 min on ice. The cells were heat-shocked at 42 °C for 50 sec, chilled on ice for 2 min and plated on LB agar plates supplemented with 100 µg/ml of ampicillin.

Isolation of plasmid DNA from transformed *E. coli*

Two strategies were used for the extraction of plasmids from transformed bacterial cells. The first approach was crude mini preps (using homemade buffers), which was primarily done to obtain a small amount of plasmid DNA material sufficient to be used in restriction digests for purposes of confirming the presence and or correctness of inserts. In this case, a 1 ml overnight bacterial culture was set up in a 2 ml Eppendorf tube and cells harvested by centrifugation (20,000x *g*, 5 min, 4 °C). The supernatant was discarded, and the pellet was resuspended in 200 µl of resuspension solution (P1). Subsequently, 200 µl of lysis solution (P2) was added, and the tube was carefully inverted 6 times to mix, followed by incubation (5 min, room temperature (RT). To the sample, 200 µl of neutralization solution (P3) was added and inverted 6 – 8 times to mix thoroughly. The sample was incubated on ice for 15 min followed by centrifugation (20,000x *g*, 15 min, 4 °C). Carefully, 500 µl of the supernatant was transferred to a new 1.5 ml Eppendorf tube and an equal volume of isopropanol added to precipitate the DNA. The sample was left to stand (RT, 5 min) and centrifuged (20,000x *g*, 20 min, 4 °C) followed by two wash steps with 1 ml of 70% ethanol (20,000x *g*, 5 min, 4 °C). All the supernatant discarded, and the pellet dried for approximately 10 min on a clean bench. The pellet was resuspended in 20 µl of ddH_2O.

The second approach was applied when high-quality plasmids were needed for sequencing or transfections. Overnight 5 ml or 30 ml bacterial cultures were set up for plasmid DNA extraction using NucleoSpin® plasmid miniprep or Nucleobond® PC100 midiprep kits

(Macherey-Nagel, DE), respectively. This was essentially done as per the manufacturer's protocol. Concentration and purity of the extracted plasmid DNA were determined by measuring the absorbance at 260 nm and 280 nm using the Tecan Infinite M200 reader (Tecan Group, CH). The plasmids were immediately used for downstream applications or stored at -20 °C.

2.2.2 Handling and manipulation of *T. brucei*

Cultivation of monomorphic *T. brucei*

All cell lines generated in this study were based on *T. brucei* Lister 427 MITat1.2 13-90 bloodstream forms expressing the tetracycline (tet) repressor and the T7 polymerase (Wirtz et al*.,* 1999). The cells were cultivated in HMI-9 medium supplemented with 10% heat inactivated fetal calf serum (Sigma-Aldrich, USA) at 37 °C and 5% CO_2, and kept below 8 x 10^5 cells ml^{-1} to maintain exponential growth. T7 polymerase and the tet repressor were maintained by selection with 5 µg/ml hygromycin and 2.5 µg/ml G418, respectively.

Cultivation of *T. brucei* to high density

For soluble VSG (sVSG) and membrane form VSG (mfVSG) harvesting, *T. brucei* were grown to a density of approximately 3.5 x 10^6 cells ml^{-1} in 400 ml HMI-9 medium (supplemented with 10% heat inactivated fetal calf serum (Gibco, USA)) using 2 l conical flasks. Cells were grown at 37 °C and 5% CO_2 with agitation on an orbital shaker at 55 rpm.

Freezing and thawing of trypanosomes

To store trypanosome cell lines for future use, 4 x 10^6 cells per stabilate were harvested by centrifugation at 1400x *g* for 10 min at 4 °C. The cells were resuspended in 500 µl of ice-cold HMI-9 media, and 500 µl of ice-cold 2x trypanosome freezing buffer added. The final concentration of glycerol in the HMI-9 medium is 10%. The cells were transferred into cryotubes and stored at -80 °C or -150 °C freezers for short and long periods, respectively.

To thaw frozen stabilates, cryotubes containing the cells were incubated at 37 °C for until just thawed and transferred into 10 ml of ice-cold HMI-9 medium in a 15 ml Falcon tube. The cells were centrifuged (1400x *g*, 10 min, 4 °C), the supernatant discarded, and the cell pellet resuspended in pre-warmed HMI-9 media. Subsequently, the cells were transferred into culture flasks and incubated for one hour, followed by density determination, dilution and addition of selection antibiotics.

Stable transformation of trypanosomes

3 x 10^7 mid-log phase (5 – 6 x 10^5 cells/ml) BSF cells were harvested (1400x *g*, 10 min, RT) and resuspended in 100 µl of Basic Parasite Nucleofector Solution 1 of the Amaxa Basic Parasite Nucleofector kit 1. 10 µg of linearized plasmid DNA was added to the cells, transferred into a cuvette and electroporated using the AMAXA Nucleofector II device (Lonza) X-001 program. The transfected cells were transferred to 30 ml of pre-warmed HMI-9 medium supplemented with parental cell line antibiotics. Serial dilutions were made (1:10, 1:50 and 1:100) in medium containing the parental cell line antibiotics and 1 ml aliquots dispensed into 24-well tissue culture plates. The cells were left to recover for 6 hours in the incubator after which 1 ml of pre-warmed medium supplemented with parental cell line antibiotics and double the amount of selection antibiotic was added to each well. The plates were transferred back to the incubator and screened under the microscope 5 or 6 days after transfection. Drug-resistant clones were transferred to tissue culture flasks.

Surface labelling of trypanosomes with Atto 488-NHS

For BSF surface labelling, 2 x 10^7 cells were harvested (1400x *g*, 10 min, 4 °C) and resuspended in 1 ml of TDB. The cells were centrifuged in a picofuge for 90 sec; the supernatant aspirated leaving 200 µl in which the pellet was resuspended. The cells were labelled by adding 2 µl of 10 mM Atto 488-NHS followed by incubation for 15 min on ice and in the dark. The cells were washed twice in 1 ml ice-cold TDB and resuspended in 20 µl of the supernatant. 750 µl of ice-cold TDB was added to the cells followed by 250 µl of 8% formaldehyde and the tubes carefully inverted 3 times. The cells were fixed for 1 hour at RT or overnight at 4 °C. The cells were centrifuged (picofuge, 90 sec), the supernatant removed and resuspended in 1 ml of TDB.

Staining of trypanosome DNA with DAPI

Fixed cells (4 µl) from the labelling and fixation procedure above were transferred to a clean 1.5 ml Eppendorf tube and 0.5 µl of a 5 µg ml^{-1} DAPI stock solution added. The cells were incubated for 2 min before viewing under the microscope.

Trypanosome viability assay

Propidium iodide DNA stain was used to determine dead or live cells in a population. Propidium iodide is membrane impermeant and will only stain the DNA material of dead cells due to a compromised cell membrane. A total of 2 x 10^6 cells was harvested by centrifugation (1400x *g*, 10 min, 4 °C), washed twice in ice-cold TDB (1400x *g*, 10 min,

4 °C) and resuspended in 400 µl of TDB. 5 µg ml^{-1} propidium iodide was added to the resuspended cells followed by incubation for 10 min on ice. Intracellular propidium iodide fluorescence was analyzed using a FACSCalibur™ flow cytometer (BD Biosciences, USA) with an excitation wavelength of 535 nm and fluorescence emission maximum of 617 nm (FL2 filter).

2.2.3 Molecular biology methods

2.2.3.1 DNA analyses

Isolation of crude genomic DNA from trypanosomes

The Phusion Human Specimen Direct PCR Kit (Thermo Fisher Scientific, USA) was used to isolate DNA from trypanosomes for use as the template for integration PCR. The manufacturer's protocol was followed with minor modifications. In brief, 1 x 10^6 cells were harvested (1400x *g*, 10 min, RT) and the supernatant discarded completely. The cell pellet was resuspended in 20 µl of dilution buffer, 0.5 µl of DNARelease™ additive added and mixed well followed by incubation (5 min, RT). Subsequently, the reaction was incubated (2 min, 98 °C) then centrifuged (2000x *g*, 5 min, RT) and 0.5 µl of the supernatant used as the template in the polymerase chain reaction (PCR).

Polymerase chain reaction

DNA fragments were amplified either in 25 or 50 µl reaction volumes. For a standard 50 µl reaction, a PCR master mix containing 10 ng plasmid DNA or 100 ng genomic DNA, 1x Phusion HF or GC buffer, 0.5 µM forward and reverse primers (Sigma-Aldrich, USA), 200 µM dNTPs, 1.6% DMSO, 1U Phusion polymerase and ddH_2O was set up. Unless otherwise stated, all components of the Phusion kit were sourced from Thermo Fisher Scientific. The cycling parameters for most reactions were as follows: initial denaturation at 98 °C for 30 sec, followed by 30 cycles of 98 °C for 10 sec, annealing at 50 °C for 30 sec, and elongation at 72 °C for 30 sec. A final elongation step at 72 °C was performed for 8 minutes.

For joining two fragments, fusion PCR was carried out based on the method by Heckman and Pease (2007). Briefly, four sets of primers are designed and used to amplify the two fragments separately. The reverse and the forward primers used to amplify the first and second fragments respectively are designed in such a way that they generate overlapping sequence overhangs to each other. The second step of PCR is conducted with equimolar amounts of the PCR products of the first PCR reactions as a template. The forward and

reverse flanking primers used in this PCR are the ones used in the first PCR reactions to amplify the first and second fragments, respectively. This second PCR generates a hybrid or fusion gene product of the two fragments.

Restriction digests and DNA modifications for cloning

Restriction enzymes were used to confirm the correct integration of DNA fragments (inserts) into plasmid vectors, to linearize plasmids for transfection and to generate sticky or blunt ends for ligation reactions. Restriction digests were carried out as per the manufacturer's protocol. Klenow fragment (Thermo Fisher Scientific, USA) was used to blunt the fragment overhangs to generate blunt ends as per the manufacturer's instructions. FastAP thermosensitive alkaline phosphatase was used to dephosphorylate vector ends to prevent recircularization. The dephosphorylation reactions were done as per the manufacturer's instructions (Thermo Fisher Scientific, USA). For both sticky-end and blunt-end ligation of inserts into linear vector DNA, T4 DNA ligase (Thermo Fisher Scientific, USA) was used with a vector to insert molar ratio of 1:3 and incubation at 22 °C for 1 hour. Cloning into pJET1.2 vector was done as per the manufacturer's protocol.

Agarose gel electrophoresis

Samples were resolved on 0.8% agarose in 1x TAE buffer (Sigma-Aldrich, USA) to visualize and separate DNA fragments following a PCR or a restriction enzyme digest. The samples were pre-mixed with 6x DNA loading dye before loading on the agarose gel, and GeneRuler DNA ladder mix (Thermo Fisher Scientific, USA) was used as a size marker. In general, DNA fragments were separated at 120 V for 35 min. The gel was then stained in a 3 µg/ml ethidium bromide bath for 15 min and visualized under UV light with the INTAS image scanner. Alternatively, 0.8% agarose in 1x TAE was directly stained with SYBR™ Safe DNA gel stain (Invitrogen, USA) prior to DNA sample separation. In this case, 40 ml of the agarose solution was stained with 0.8 µl of the 10,000x SYBR™ Safe concentrate. After the separation, the gel was imaged under green LED transilluminator light with the iBright™ CL1000 imaging system (Thermo Fisher Scientific, USA). In case DNA fragments were to be purified from the agarose gels, the bands were excised from the gels using clean scalpels and the DNA extracted using the NucleoSpin® Gel and PCR clean-up kit (Macherey-Nagel, DE) following the manufacturer's instructions.

Isopropanol precipitation of DNA for transfection

This was carried out to purify and concentrate DNA from restriction digestion reactions for transfection. An equivalent of 10 µg of plasmid DNA was linearized in a 100 µl-reaction

volume according to the above procedure. To precipitate the DNA, 1/10 volume 3 M sodium acetate (pH 5.2) and 1 volume isopropanol were added to the linearized DNA. The sample was mixed by inverting the tube at least 3 times and centrifuged (20,000x *g*, 30 min, 4 °C). The supernatant was carefully tipped off, 1 ml of 70% ethanol added and centrifuged (20,000x *g*, 10 min, 4 °C). The washing step was repeated once more. The sample was then transferred to a sterile workbench, the supernatant tipped off, and the DNA dried for 10 – 15 min followed by resuspension in sterile water. The DNA was used directly for transfection or stored at -20 °C for future use.

2.2.3.2 RNA analyses

Isolation of total RNA and cDNA synthesis

Total RNA was isolated from 1 x 10^8 BSF *T. brucei* parasites. Cells were harvested by centrifugation (1400x *g*, 10 min, 4 °C) followed by resuspension of the cell pellet in 1 ml of ice-cold serum-free HMI-9. The resuspended cells were then transferred into RNase-free 2 ml tubes and centrifuged (1400x *g*, 10 min, 4 °C). The supernatant was aspirated, leaving approximately 50 µl in which the cells were resuspended and subsequently snap-frozen in liquid nitrogen. The frozen cells were stored at -80 °C until use. The RNeasy Mini Kit (Qiagen, NL) was used for total RNA extraction. The cell pellet was removed from the -80 °C freezer and resuspended in 600 µl of RTL buffer supplemented with 6 µl of β-mercaptoethanol by pipetting up and down vigorously at 37 °C. This step is very critical and affects the amount of RNA obtained in the end. Subsequent steps were carried out as per the manufacturer's protocol and the RNA concentrations and purity determined photometrically using the Tecan reader (Tecan, CH).

RNA dot blots

Exogenous and endogenous VSG mRNA were analyzed by dot blotting. First, the RNA was denatured by adding 7.2 µl of glyoxal mix to 3 µg of RNA in 4 µl of RNase free water followed by incubation (50 °C, 40 min). 40 µl of 10x SSC buffer was then added to the sample before loading onto a nitrocellulose membrane. The sample concentration and reagent volumes indicated above apply when probing for a single VSG. If probing for two VSGs, the amounts were doubled. 20 min before completion of the sample denaturation, Amersham Hybond-N membrane (GE Healthcare, UK) of the correct size was cut and placed in the dot blotter. The membrane was washed with 200 µl of 10x SSC buffer by applying vacuum until the entire buffer had passed through the membrane. 45 µl of the sample was applied per well, vacuum applied, and the membrane washed with 200 µl of 10x SSC buffer

as above. The membrane was removed from the dot blotter and placed on a Whatman paper. The RNA was immobilized on the membrane by UV crosslinking (Stratalinker 1800, Stratagene, USA) using the autocrosslink function (1200x 100 µJ/cm^2) followed by incubation for 60 min at 80 °C. The membrane was prehybridized in 10 ml of hybridizing mix (60 min, 42 °C) with vertical rotation. Subsequently, 1 µl (10 nM) of fluorescently labelled probes for VSGs and tubulin loading control were added to the prehybridized membrane and incubated (overnight, 42 °C) in the dark. The membrane was washed twice with 20 ml of Northern wash solution (RT, 20 min) and dried between two Whatman papers. The membrane was scanned on a LI-COR Odyssey CLx scanner and analyzed with the corresponding Image Studio^{TM} Lite Quantification software (LI-COR Biosciences, USA).

2.2.4 Protein analyses

Preparation of protein samples for SDS-PAGE and Western blots

Parasites were harvested by centrifugation (1400x *g*, 10 min, 4 °C) and washed in 1 ml of ice-cold TDB (picofuge, 90 sec). The supernatant was carefully discarded, and the cell pellet resuspended in 1x protein sample buffer to a final concentration of 2 x 10^5 cells µl^{-1} and denatured at 100 °C for 5 min. The cell lysates were used directly or stored at -20 °C for a short duration until use.

Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)

Discontinuous SDS-PAGE was used to separate the proteins in a cell lysate. 5% and 12.5% acrylamide/bisacrylamide (37.5:1) stacking and resolving gels, respectively were set up. For VSG analysis, 1 x 10^6 cells were loaded into each well. A protein marker (Prestained protein marker 10 – 170 kDa, Thermo Fisher Scientific, USA) was also included in a separate well to estimate the molecular size of the separated proteins. Electrophoresis was done at 120 V (BIO-RAD, USA) in 1x Laemmli running buffer till the dye had just run out of the gel.

Staining and destaining of SDS protein gels

After separation of proteins by SDS-PAGE, gels were stained in Coomassie brilliant blue (R-250) buffer for 30 min on an orbital shaker. The buffer was then poured off, the gel rinsed with water and subsequently incubated in Coomassie destaining solution on an orbital shaker until the protein bands were clearly visible.

Western blotting

Specific antibodies were used to probe for specific proteins resolved on 12.5% SDS-PAGE gels by western blotting. The proteins were first transferred onto a nitrocellulose membrane (Amersham Protran 0.45 µm NE; GE Healthcare, UK) cut to the right size. The gel was placed on top of the membrane and six pieces of 6.5 x 8 cm Whatman papers pre-soaked in the transfer buffer stacked on the top and bottom. This stack was subsequently placed on a Trans-Blot® SD semi-dry transfer cell (BIO-RAD, USA) with the membranes facing the anode and transfer done at 60 mA (~1.1 mA/cm^2) for 1 hour. The membranes were retrieved from the transfer cell and blocked with 5% (w/v) skim milk powder in PBS for 1 hour at room temperature or overnight at 4 °C. Subsequently, the membranes were incubated with appropriate concentrations of primary antibodies diluted in PBS containing 1% (w/v) skim milk powder and 0.1% (v/v) Tween 20 for 1 hour at room temperature followed by four 5-minute washes in PBS containing 0.2% (v/v) Tween 20. The membranes were then probed with secondary antibodies diluted in in PBS containing 1% (w/v) skim milk powder and 0.1% (v/v) Tween 20 for 1 hour and unbound antibodies were removed by four 5-minute washes in PBS containing 0.2% (v/v) Tween 20. Membranes were finally washed in PBS, dried between two clean Whatman papers and analyzed with a LI-COR Odyssey Imager CLx and the corresponding ImageStudio™ Lite Quantification software.

Protein dot blot

Protein dot blots were used to quantify changes in VSG protein expression. This method is timesaving as it eliminates the need to separate the proteins before probing the protein of interest with specific antibodies, and it also cuts the transfer step which could introduce errors. The nitrocellulose membrane was cut, soaked in 1x PBS, dried on Whatman paper and placed on a dot blotter. 3 µl of the protein sample, equivalent to 6 x 10^5 cells, was loaded on each well and allowed to dry. Subsequent detection of the protein with antibodies was carried out as per the western blot protocol above.

2.2.5 High-pressure freezing and electron microscopy

Fifty-milliliter BSF culture at 6 x 10^5 cells ml^{-1} was centrifuged (750x *g*, 3 min, RT). The supernatant was discarded, leaving 1 ml in which the pellet was resuspended, and 1 ml of heat-inactivated FCS was added as a cryoprotectant. The cells were then centrifuged (750x *g*, 3 min, RT), and the supernatant was discarded, leaving approximately 200 µl in which the pellet was resuspended. The sample was transferred into PCR tubes and further compacted (10 sec, picofuge). Next, the supernatant was discarded, and the pellet was transferred into

the specimen carrier (Leica Microsystems, DE). High-pressure freezing was done at a freezing speed of >20,000 Ks^{-1} and a pressure > 2100 bar in an EM HPM100 (Leica Microsystems, DE). The samples were transferred into liquid nitrogen in an EM AFS2 freeze substitution system (Leica Microsystems, DE) and stored until freeze-substitution.

Samples were embedded in epon resin to enable thin sectioning of the specimen. First, the samples were incubated in 0.5% (v/v) glutaraldehyde and 0.1% (w/v) tannic acid in anhydrous acetone (96 hours, –90 °C). The solution was changed once after the first 24 hours of incubation and subsequently washed four times for 1 hour with anhydrous acetone (-90 °C). This was followed by fixation in 2% osmium tetroxide in anhydrous acetone (28 hours, -90 °C). Sample temperature was gradually raised to -20 °C in 14 hours and incubated at this temperature for 16 hours. Finally, the temperature was gradually raised to 4 °C within 4 hours. The samples were washed four times in anhydrous acetone within 2 – 3 hours, and the temperature raised to 20 °C within 1 hour. Samples were dislodged from the carriers and transferred into increasing epon concentrations (starting with 50% epon in acetone for 5 hours at 20 °C, 90% epon in acetone overnight at 4 °C, and finally in 100% epon for 2 – 3 hours three times at 20 °C) for embedding. The samples were incubated at 60 °C for 24 hours to polymerize, cut into 60 nm serial sections with an 'Ultra Jumbo Diamond Knife' (Diatom, CH), and transferred into pioloform coated slotted copper grids. Images were acquired with a 200 kV JEM-2100 (JEOL, JP) transmission electron microscope fitted with a TemCam F416 4k x 4k camera (Tietz Video and Imaging Processing Systems, DE).

2.2.6 Extraction and purification of soluble VSG (sVSG)

Crude sVSG was extracted from approximately 3 – 5 x 10^9 parasites by hypotonic lysis followed by purification via anion exchange chromatography. For this, the parasites were harvested in 450 ml centrifugation buckets by centrifugation (800x *g*, 30 min, 4 °C). The supernatant was discarded, the pellet resuspended in ice-cold TDB then transferred into two 50 ml Falcon tubes and centrifuged (1400x *g*, 10 min, 4 °C). This wash step was repeated once. Subsequently, the cells were resuspended in 4 ml of 10 mM sodium phosphate buffer, pH 8.0 supplemented with 0.1 mM TLCK and 1x cOmplete™ protease inhibitor cocktail supplemented with EDTA (Roche, CH). The cells were either processed directly or flash-frozen in liquid nitrogen and stored at –80 °C until use. If the cells were retrieved from the –80 °C freezer, the pellet was thawed, and the sVSG extracted in the same manner as the directly processed cells. Sample aliquots of 1.33 ml were transferred into three 1.5 ml Eppendorf tubes followed by incubation in a water bath at 37 °C for 5 min. The samples

were centrifuged (14,000x *g*, 10 min, 4 °C) and the supernatants containing sVSG collected into a fresh Falcon tube and kept on ice. The pellets were each resuspended in 1.33 ml of the phosphate buffer supplemented with protease inhibitors as above, incubated and centrifuged as in the first extraction step. The extraction step was repeated a third time. The supernatants were combined into one Falcon tube and centrifuged (15,000x *g*, 10 min, 4 °C) to remove any remaining cell debris. The supernatant was concentrated to approximately 1 ml in an Amicon Ultra-15 10K centricon by centrifugation (5,000x *g*, approximately 30 min, 4 °C). The flow-through was always stored to avoid loss of sample in case the concentrator was damaged. The buffer was exchanged by adding 11 ml of 20 mM Tris-HCl, pH 8.0 to the concentrate and centrifugation (5,000x *g*, approximately 30 min, 4 °C) until about 1 ml of sample was left. The buffer exchange step was repeated, and the 1 ml concentrate transferred to a 15 ml Falcon tube. The concentrator was rinsed with 20 mM Tris-HCl, pH 8.0 to ensure most of the protein was transferred, and the total sample volume brought to 4 ml.

Anion exchange was done using a HiTrap™ Q HP anion exchange chromatography column on the ÄKTAprime chromatographic system (GE Healthcare, UK) using the pre-set anion exchange HiTrap Q run program template. Sample application volume was set to 4 ml, the flow rate set to 1.0 ml/min and 1 ml of sample injected per run. Fractions were collected in a linear gradient of 1 M NaCl (1 – 100%). The fractions were kept on ice and 50 µl aliquots collected for SDS-PAGE analysis. Clean fractions from each run containing sVSG (fraction 7, 8 and 9) were combined and concentrated to about 1 ml (5,000x *g*, 90 min, 4 °C) in an Amicon Ultra-15 centricon with a molecular weight cut-off of 10 kDa (Millipore, USA). Buffer exchange was done to bring the protein into water. 12 ml of ultrapure water was added to the 1 ml sVSG concentrate and centrifuged until about 1 ml was left in the concentrator (5,000x *g*, 30 min, 4 °C). The buffer exchange procedure was repeated twice leaving about 300 µl of sample volume on the last spin. Protein concentration was quantified photometrically at 280 nm using the Tecan reader. An aliquot of the protein was resolved on a 12.5% SDS-PAGE gel to determine its purity.

2.2.7 X-ray crystallography

Protein crystallization

Commercial screens including Protein complex (Qiagen, NL), Optimix PEG (Fluidigm, USA), Index HT (Hampton Research, USA), JCSG+ (Molecular Dimensions, USA) and Wizard classic 1 & 2 (Rigaku Reagents, USA) were used to set up initial crystallization trials. Crystals were grown at both 4, and 20 °C by the sitting drop vapor-diffusion method in MRC-2 96-well crystallization plates (Hampton Research, USA). The crystallization

conditions of the commercial screens are provided online by the manufacturers. The NT8® drop setter for protein crystallization robot (Formulatrix, USA) was used to dispense 50 nl of VSG solution and 50 nl of the mother liquor into the first drop position in each of the 96 wells. A 1:1 dilution of the VSG protein solution was made and dispensed as above in the second drop position in each of the 96 wells. The drops were equilibrated against 50 µl of reservoir solution. The plates were sealed with crystal clear sealing tape and transferred into the Rock Imager incubator (Formulatrix, USA). The Rock Imager incubates the samples in a controlled environment and captures high-quality images on a user-defined schedule, for a maximum of 21 days. The wells where crystals grew were identified, and crystallization conditions in each of the wells were tabulated. The conditions were optimized to improve crystal growth and quality. The protein and mother liquor volume were increased to 0.3 µl, dispensed and incubated as explained above.

Crystallization in large drops was performed at 20 °C by hanging-drop vapor-diffusion in the presence or absence of cesium chloride salt. Incorporating cesium salt in the protein crystals, and subsequent location of the cesium atoms can help with solving the phase, which is required for protein structure determination. First, 1 ml aliquots of the mother liquor were manually pipetted into crystalgen SuperClear™ 24-well plates (Jena Bioscience, DE) followed by greasing the rims of the wells with KORASILON silicon grease (Jena Bioscience, DE). Siliconized 22 mm diameter circular cover slides (Jena Biosciences, DE) were then cleaned by Servisol air duster105/2 (Farnell, DE), and 1 µl of the protein sample applied as a drop to the cover slide, followed by addition of 1 µl of the mother liquor to the sample drop. The cover slides were then inverted and placed on top of the wells. Pressure was applied to fix the slides tightly to create a sealed chamber, and the plates were subsequently incubated at 20 °C. Single protein crystals were harvested from the wells by microLoops™ and transferred into the same mother liquor solution supplemented with 25% (v/v) glycerol for cryoprotection. The crystals were then flash-frozen and stored in liquid nitrogen until X-ray diffraction measurements were done.

Diffraction data collection and model building

X-ray diffraction data was collected on beamline 14.1 at the HZB synchrotron radiation source (Berlin, Germany) and the diffraction data was processed using XSD and scaled with the Aimless suite (Evans, 2011). Phasing was done by isomorphous replacement using the magic triangle I3C (5-amino-2,4,6.triiodoisophthalic acid). Location of heavy metal substructure in the crystals was carried out with Crank2 suite within CCP4 (Winn et al., 2011). Substructure phasing and refinement was done using REFMAC5 (Murshudov et al.,

2011) followed by density modification by Parrot. Automated model building was done in buccaneer and additional rounds of manual model building in Coot (Emsley et al., 2010) and final automated refinement was done using REFMAC5. The structure representations were generated in UCSF Chimera (Pettersen et al., 2004).

3.0 Results

3.1 Post-transcriptional regulation of *T. brucei VSG* expression

The VSG surface coat is an essential interface between the parasites and the host, mediating cellular processes, particularly those involved in evasion of the host's immune responses. The expression of *T. brucei* VSGs is a highly regulated process and depletion of the VSG protein is lethal, both *in vivo* and *in vitro* (Ooi et al., 2018; Sheader et al., 2005). VSG expression studies, in which a second *VSG* gene was integrated and expressed from the active BES, generated double expressor cell lines expressing equal amounts of the VSG proteins (Muñoz-Jordán et al., 1996). Later, it was shown that the double expressor cell lines expressed the endogenous and ectopic VSGs on a 50:50 basis, thereby restricting the total *VSG* mRNAs to the wild type levels (Smith et al., 2009). Additionally, the inducible overexpression of an ectopic *VSG* from the transcriptionally silent rDNA spacer caused an increase in the ectopic *VSG* mRNA levels with concomitant reduction of the endogenous *VSG* mRNA levels. The total amount of *VSG* mRNAs initially exceeded the wild type amounts but levelled at approximately 100% of the wild type quantities within 8 hours of overexpression (Batram et al., 2014). Therefore, it was proposed that a mechanism to balance *VSG* mRNA was operational in *T. brucei,* and that possibly the conserved 16-mer motif in the VSG 3'UTR was involved (Ridewood et al., 2017). However, a recent study showed a massive increase in total *VSG* mRNA levels when premature termination codons (PTCs) were introduced in different locations within the open reading frame (ORF) of the ectopic *VSG* (Maudlin et al., 2021), thus questioning the presence of a VSG balancing mechanism. In this study I sought to dissect the mechanism of VSG balancing in *T. brucei,* and further investigated if the balancing of VSG is a special feature of *T. brucei* VSGs by expressing VSGs of related African trypanosome species and non-VSG reporter genes.

3.1.1 Expression of *T. vivax VSG* reporters fused to *T. brucei VSG* 3′UTR gives varying growth phenotypes

The surface coats of *T. brucei* and the related *T. congolense* and *T. vivax* species are covered by VSGs. Though not well characterized, *T. congolense* and *T. vivax* VSGs are much smaller in comparison to *T. brucei* VSGs and lack the conserved 16-mer motif present in the 3′UTR of all *T. brucei* VSGs. This motif is required for high expression and stability of *VSG* mRNA (Berberof et al., 1995). It is further argued that the *T. vivax* VSG coat may not be as dense as the coat found in *T. brucei* (Autheman et al., 2021; Gardiner & Wilson, 1987). Therefore,

it is possible that African trypanosomes have evolved different mechanisms to regulate the expression of VSGs. I selected two well documented *T. vivax VSG* genes ***IL****RAD* ***D****uttonella* ***a****ntigen* ***t****ype 1.2* and *2.1* (*ILDat1.2: TvY486_0008160* and *ILDat2.1: Z48228.1)* for expression in *T. brucei* (Chamond et al., 2010; Gardiner et al., 1996; Gardiner & Wilson, 1987; Vos & Gardiner, 1990) with an aim to determine if the regulation of *VSG* mRNA amounts required unique features present only in *T. brucei VSGs*. For this purpose, I opted to utilize the pLew82v4 (24 009; Addgene plasmid) inducible overexpression system (Figure 9 A). The pLew82v4 vector integrates the ectopic *VSG* and the *phleomycin* (ble) resistance cassette into the transcriptionally silent ribosomal spacer region. The expression of the ectopic *VSG* is driven by an ectopic T7-promoter under the control of a tetracycline operator, and high levels of expression of the *VSG* are achieved in the presence of tetracycline. With this expression system early events after induction of ectopic *VSG* expression can be captured.

First, the *ILDat1.2* and *ILDat2.1 VSG* ORFs fused to *MITat1.1 VSG* 3′UTR were synthesized with EcoRI restriction sites at the 5′ and 3′ ends and cloned into pBSK II (+) vector (Stratagene). A start codon was added to the ILDat2.1 ORF as the reported sequence lacks the start codon (Gardiner et al., 1996). The 3′UTR of a *T. brucei* VSG was used because *T. vivax* VSG UTRs lack the conserved 16-mer motif. The fragments were mobilized with SmaI and HindIII and ligated into the pLew82v4 overexpression vector sequentially digested with XhoI (followed with blunting by filling in the overhangs using Klenow fragment) and HindIII to generate pLew.ILDat1.2 and pLew.ILDat2.1 constructs. The constructs were NotI-linearized and transfected into the parental 13-90 cell line, generating the cell lines 221^{ES}.ILDat1.2^{tet} and 221^{ES}.ILDat2.1^{tet}.

The 221^{ES}.ILDat1.2^{tet} cell line exhibited slowed growth in the first 12 hours after induction of expression, followed by stalling in growth for two days after which the cell numbers began to decline. Inducing the expression of ILDat2.1 VSG had only a mild effect on the growth of the 221^{ES}.ILDat2.1^{tet} cell line. The cells had a PDT of 7.2 hours compared to 7.0 and 6.4 hours of the uninduced and the parental 13-90 cells, respectively (Figure 9 B, C). These variation in growth phenotypes indicated that there could be differences in the regulation of *VSG* expression in the two cell lines. Next, the *VSG* transcripts were quantified.

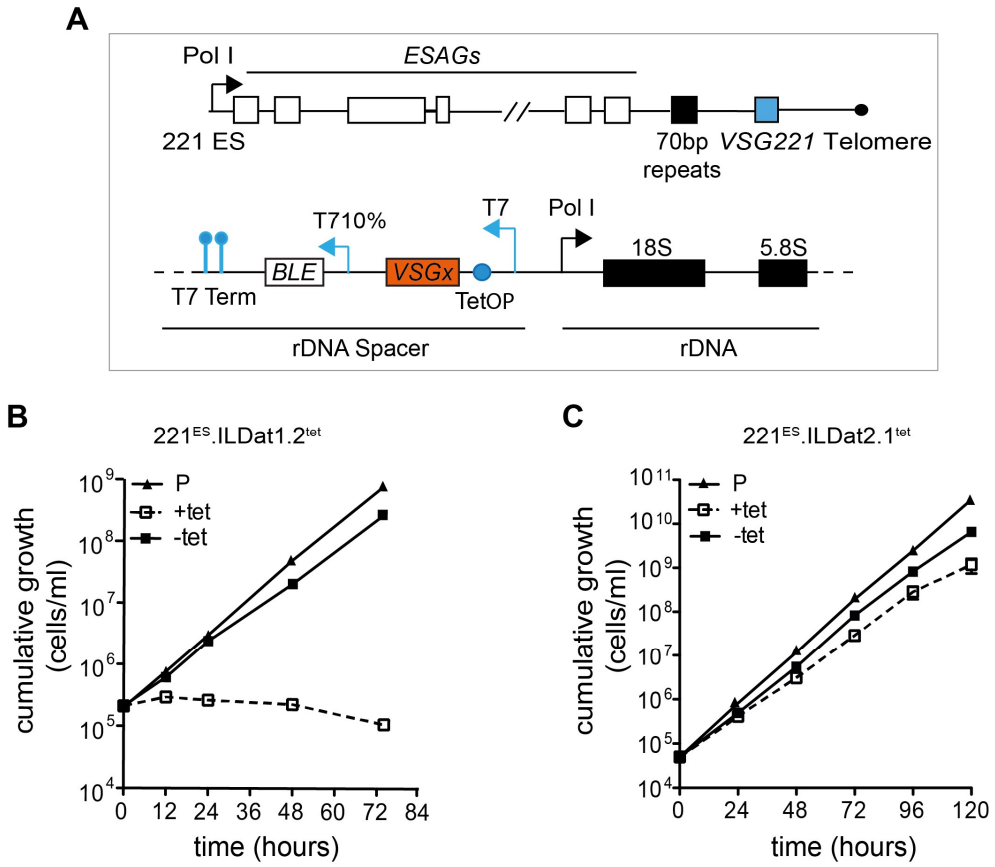

Figure 9: ILDat1.2 and ILDat2.1 expressing cells display varying growth phenotypes.
(A) Schematic of the employed ectopic VSG overexpression strategy in *T. brucei.* The upper panel depicts the active MITat1.2 VSG ES. The lower panel shows the rDNA spacer region where the ectopic VSG is integrated. The ectopic VSG expression is regulated by a T7-promoter (T7), which is controlled by a tetracycline operator (TetOP). The *phleomycin* (ble) resistance cassette used for selection is under the regulation of a reduced T7 ($T7^{10\%}$) promoter, and both the ectopic VSG and the bleomycin resistance gene transcription is terminated by the T7 terminator (T7 Term). **(B)** Growth curves of 221^{ES}.ILDat1.2^{tet} and **(C)** 221^{ES}.ILDat2.1^{tet} cell lines in the absence (–tet) and presence of 1 µg/ml tetracycline (+tet). In both cases, the 13-90 parental cell line (P) was used as a control, and the curves represent the cumulative mean cell numbers ± SD of three independent clones.

3.1.2 *ILDat1.2* and *ILDat2.1 VSG* reporter mRNAs are sufficiently transcribed

In a study by Batram et al., induction of ectopic *VSG121* expression from the transcriptionally silent rDNA spacer locus led to a rapid increase in the ectopic *VSG* mRNA transcripts, reaching approximately 80% of the wild type levels within 2 hours of inducing expression. This increase in *VSG121* mRNA was accompanied by a simultaneous decline of the endogenous *VSG221* mRNA level to ~50 and 20% of the wild type levels after 4 and 8 hours, respectively, of inducing expression (Batram et al., 2014).

Quantitative RNA dot blots showed a 10-fold increase of ILDat1.2 *VSG* mRNA within 8 hours of induction of expression, with a simultaneous decrease in the endogenous *VSG221* mRNA to ~20% of the wild type levels in the 221^{ES}.ILDat1.2^{tet} cell line (Figure 10 A, B). In the 221^{ES}.ILDat2.1^{tet} cell line, there was a 20-fold upregulation of *ILDat2.1* mRNA within

8 hours of inducing expression (Bachelor thesis, I. Jung). However, the endogenous *VSG221* mRNA transcript remained at wild type levels (Figure 10 C, D).

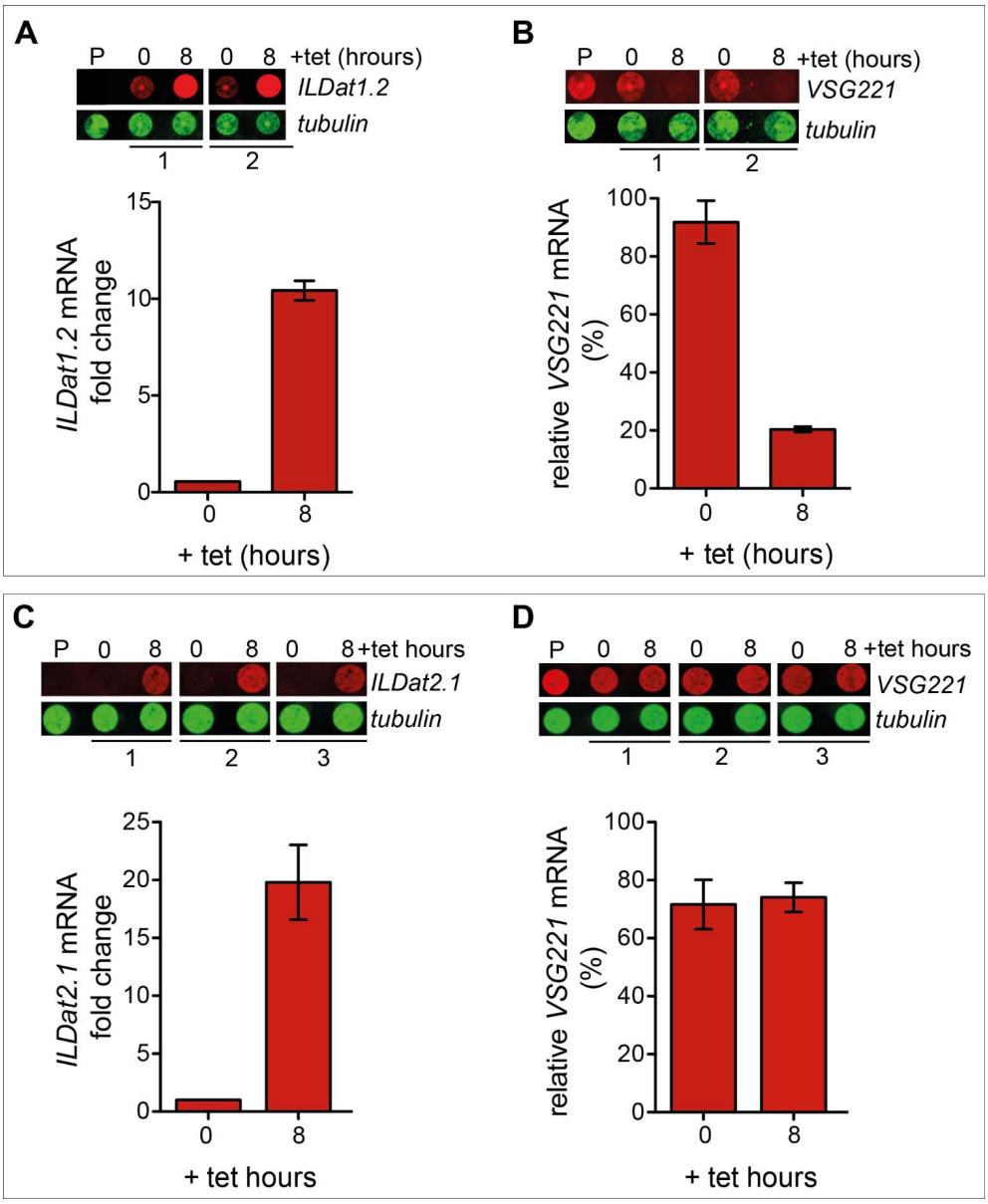

Figure 10: Ectopic *ILDat1.2* and *ILDat2.1 VSG* mRNAs are sufficiently transcribed.
(A) mRNA dot blot showing ILDat1.2 and **(B)** endogenous *VSG221* transcript levels 8 hours after induction of ectopic *ILDat1.2 VSG* expression in two clones of the 221^{ES}.ILDat1.2^{tet} cell line. The *VSG* mRNA is detected in the red channel, while *β-tubulin*, which served as the loading control, is detected in the green channel. Below the dot blot is a graphical representation of the *VSG* quantification. *ILDat1.2 VSG* expression is given in fold-change since a wild type cell line expressing *ILDat1.2 VSG* was unavailable, whereas the *VSG221* mRNA level is given relative to the *VSG221* expression of 13-90 parental cells **(C)** Quantification of ectopic *ILDat2.1* and **(D)** endogenous *VSG221* mRNA transcripts 8 hours post-induction of *ILDat2.1* overexpression in the 221^{ES}.ILDat2.1^{tet} cell line. The numbers under the blots represent the individual clones analyzed. The *VSG221* expression level is presented as a percentage relative to *VSG221* levels of the 13-90 parental cell line. The error bars represent means ± SD of two and three independent clones of the 221^{ES}.ILDat1.2^{tet} and 221^{ES}.ILDat2.1^{tet} cell lines, respectively. The blots were imaged and quantified using the Li-COR Odyssey imaging system and Image Studio Lite.

Next, expression of the *T. vivax* VSG proteins after inducing expression was analyzed by SDS-PAGE. Although it would have been better to assess ectopic VSG protein expression by western blotting, at the time these experiments were conducted ILDat1.2 and ILDat2.1 VSG antibodies were unavailable. The Coomassie gels showed that the ectopic ILDat1.2 VSG protein (~46 kDa) was expressed in low amounts only (Figure 11 A), while ILDat2.1 protein (~42 kDa) appeared not to be expressed at all (Figure 11 B). It was not clear why ILDat1.2 was poorly expressed and ILDat2.1 VSG protein was not expressed in these cell lines.

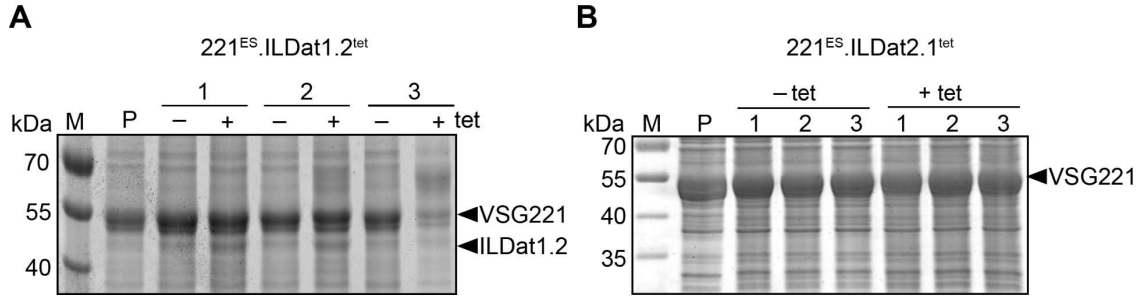

Figure 11: ILDat1.2 is poorly expressed while ILDat2.1 appears not to be expressed in *T. brucei.* **(A)** SDS-PAGE analysis of protein expression in the 221^{ES}.ILDat1.2^{tet} cell line before (-tet) and 24 hours post-induction (+tet) showed that ILDat1.2 was not sufficiently expressed in *T. brucei*. A faint band of approximately 46 kDa, which corresponds to the molecular weight of ILDat1.2, is visible in induced samples of clones 1 and 2. Clone 3 appeared to have switched to a new VSG as a prominent band of ~65 kDa was visible. **(B)** Analysis of protein expression in the 221^{ES}.ILDat2.1^{tet} cell line using SDS-PAGE. Bands corresponding to ILDat2.1 molecular mass were not visible in the induced clones. The 13-90 parental cells (P) served the control.

Taken together with the mRNA expression data, it appears that despite the transcription of sufficient ectopic *ILDat1.2 VSG* mRNA, the corresponding protein was expressed in low amounts. As the endogenous *VSG221* mRNA levels were downregulated, the total VSG protein produced was not sufficient to support growth and hence the stalling and eventual cell death observed in the 221^{ES}.ILDat1.2^{tet} cell line (Figure 9 B). On the other hand, ILDat2.1 protein seemed not to be expressed despite the transcription of sufficient corresponding mRNA in the 221^{ES}.ILDat2.1^{tet} cells. Given that the endogenous *VSG221* mRNA level was not downregulated in these cells, sufficient VSG was produced, which supported normal growth (Figure 9 C). These data showed that a VSG balancing mechanism that downregulated the endogenous *VSG221* transcript levels was operational in the 221^{ES}.ILDat1.2^{tet} cell line but not in the 221^{ES}.ILDat2.1^{tet} cell line. As the ILDat2.1 protein was not made, analysis revealed that the reported ILDat2.1 ER signal (Gardiner et al., 1996) had a low probability score for the presence of an ER import signal. Therefore, it is possible

that the ILDat2.1 ER signal is not a bona fide signal or that it simply does not function in *T. brucei.* Thus, the *ILDat2.1* mRNA was not targeted to the ER for post-transcriptional processing in the 221^{ES}.ILDat2.1^{tet} cell line explaining why no protein appeared to be expressed. These results can further indicate that for the endogenous *VSG* mRNA to be regulated, the ectopic *VSG* mRNA must be correctly targeted to the ER. This hypothesis was tested by replacing the reported ILDat2.1 ER import signal (Gardiner et al., 1996) by one of a *T. brucei* VSG.

3.1.3 *Trans*-regulation of the *VSG* mRNA is elicited at the ER

To investigate whether the *ILDat2.1 VSG* mRNA was not targeted to the ER (where the *trans*-regulation of *VSG* mRNA is possibly initiated) due to a dysfunctional ER import signal in the 221^{ES}.ILDat2.1^{tet} cell line, the ILDat2.1 VSG ER import sequence was replaced with a bona fide *T. brucei* VSG ER import signal sequence. If the hypothesis is correct, the endogenous *VSG221* mRNA should be downregulated upon induction of high-level expression of the reporter *ILDat2.1 VSG* containing a functional ER import signal.

The chimeric reporter *VSG* gene consisting of the *VSG121* ER import signal sequence and *ILDat2.1* coding sequence was generated by PCR-driven overlap extension (Heckman & Pease, 2007). For this purpose, the *VSG121* ER signal sequence was amplified from plasmid pLew.M1.6 using forward primer D81 (flanked with an EcoRI restriction site to facilitate cloning) and reverse primer EA6 (containing a few N-terminal nucleotide overhangs of the mature ILDat2.1 ORF). The ILDat2.1 coding sequence lacking the ER import sequence was amplified in a second PCR reaction from plasmid pBSK.ILDat2.1:M1.11 using forward primer EA5 and reverse primer D132 (flanked with an EcoRI restriction site). Subsequently the PCR products were used as the template in a third PCR reaction with primers D81 and D132. The reporter *VSG* gene product was inserted into the pJet1.2 vector. Next, the insert was excised with EcoRI and subcloned into pBSK II (+) and finally into the pLew82v4 overexpression vector (section 3.1.1) to generate the pLew.121^{SP}:ILDat2.1 construct (Figure 12 A). The linearized construct was transfected into *T. brucei* 13-90 cells to generate cell line 221^{ES}.121^{SP}:ILDat2.1^{tet}.

Induction of expression of the VSG reporter expression led to a slowed growth culminating in a cessation of growth after 24 hours (Figure 12 B). Quantification of the *VSG* mRNAs 8 hours after induction of expression showed an 8-fold increase in the ectopic *ILDat2.1 VSG* mRNA with simultaneous reduction of the endogenous VSG221 mRNA to 60% of the wild type levels, confirming activation of the *VSG* mRNA *trans*-regulation

response (Figure 12 C, D). Analysis of protein expression showed that the VSG reporter was not expressed 24 hours post-induction (Figure 12 E).

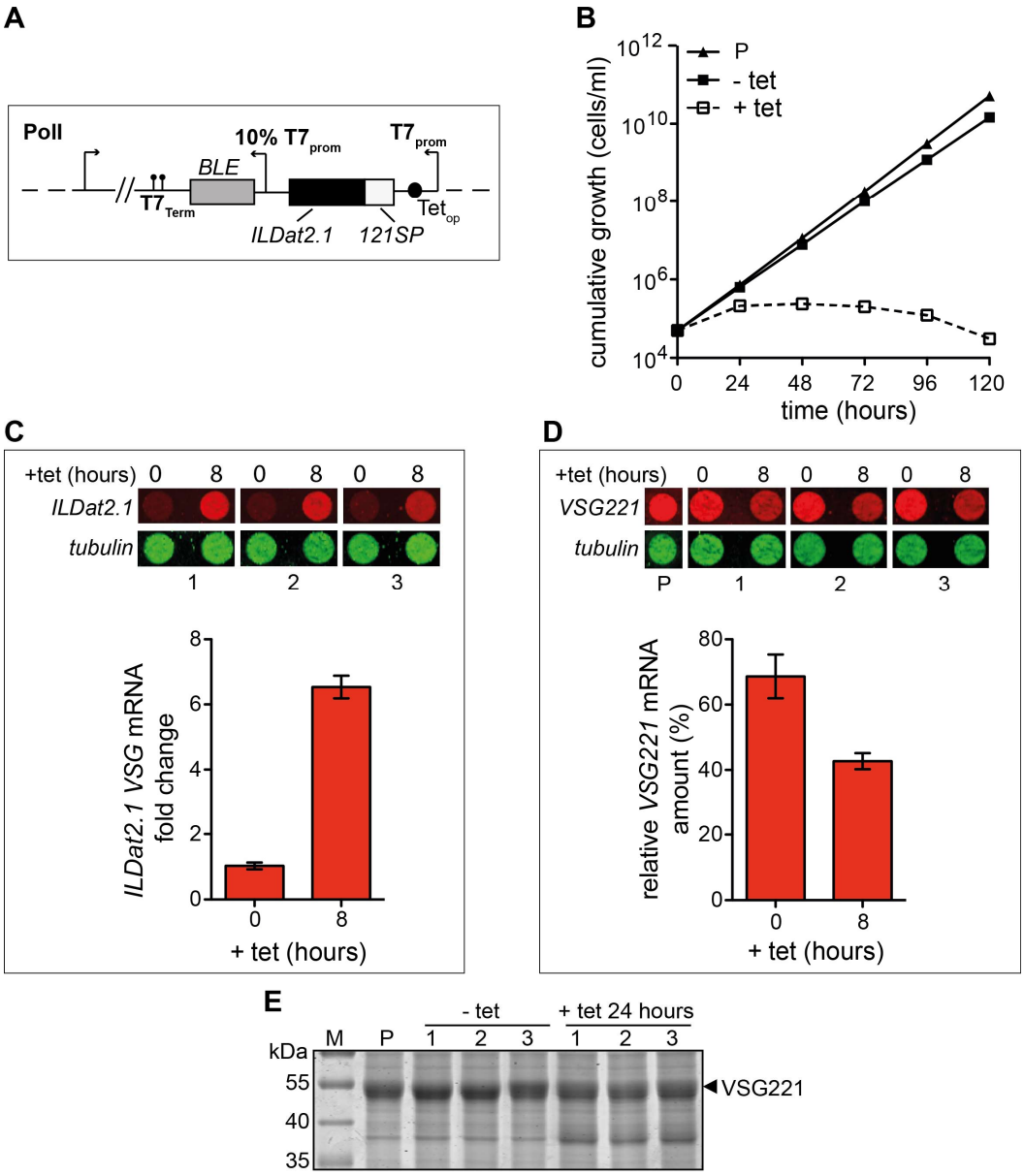

Figure 12: Replacement of the native ILDat2.1 ER import signal with that of VSG121 triggers *trans*-regulation of the *VSG* mRNA.
(A) Schematic of the hybrid *VSG* overexpression construct used to generate the $221^{ES}.121^{SP}$:ILDat2.1^{tet}. **(B)** Growth curve of the induced (+tet) and non-induced (–tet) $221^{ES}.121^{SP}$:ILDat2.1^{tet} cells. The parental 13-90 cells (P) served as a growth control and means ± SD of three clones is shown in each graph. **(C)** Quantification of *VSG221* and **(D)** *ILDat2.1 VSG* mRNA after tetracycline-induced 121^{SP}:ILDat2.1 reporter expression. The *VSG* transcripts were normalized to *β-tubulin* and the *VSG221* transcript levels are given relative to the parental cell line levels. The error bars represent mean ± SD of three independent clones. **(E)** Analysis of VSG protein expression by SDS-PAGE. The 13-90 parental cell line was used as a control, three clones (1, 2 and 3) of the $221^{ES}.121^{SP}$:ILDat2.1^{tet} cell line were analyzed.

Together, these results confirmed that the ILDat2.1 ER import signal sequence reported by (Gardiner et al., 1996) is either not a bona fide signal peptide or that it does not work in

T. brucei and thus is not appropriate for targeting ILDat2.1 VSG to the ER in *T. brucei*. The data further points to the ER as the organelle where the *trans*-regulation of *VSG* mRNA is elicited. However, the ectopic VSG protein was not produced indicating that, possibly, regulation of the *VSG* transcripts was independent of translation. To further ascertain the requirement of correct ER targeting for the regulation of *VSG* mRNA levels, the ER import signal of *T. brucei* VSG121 was replaced with the dysfunctional ILDat2.1 signal. This VSG reporter was then overexpressed in *T. brucei*, and the VSG mRNA and protein levels were analyzed.

3.1.4 A VSG121 reporter containing the ILDat2.1 VSG ER import signal is not expressed in *T. brucei*

Given that VSG121 is a functional *T. brucei* VSG, perturbation of its expression on replacement of its ER import signal sequence with the ILDat2.1 signal could be attributed to the ILDat2.1 ER import signal sequence. For this purpose, the VSG121 ER import signal sequence was replaced with the ILDat2.1 ER import signal (Gardiner et al., 1996) using PCR-driven overlap extension (section 3.1.3). The *ILDat2.1* ER signal sequence was amplified from plasmid pBSK.ILDat2.1 using forward primer IJ2 and reverse primer EA8. The *VSG121* coding sequence lacking its native ER import signal was amplified from plasmid pJET.M1.6 using forward primer EA7 and reverse primer D82. The products of the two reactions were fused in a third PCR using primer set IJ2 and D82. The generated hybrid ILDat2.1^{SP}:VSG121 gene product was subcloned into pJet1.2 (section 3.1.3) and finally into the pLew82v4 vector (section 3.1.1) to generate construct pLew.ILDat2.1^{SP}:VSG121 (Figure 13 A). The construct was transfected into *T. brucei* 13-90 cells to generate the 221ES.ILDat2.1^{SP}:121^{tet} cell line.

Induction of protein expression caused a slight decrease in the growth of the cells; the induced cells (+tet) doubled every 7.7 hours compared to the 6.4 hours doubling time of the uninduced cells (–tet) (Figure 13 B). Quantification of *VSG* mRNAs by RNA dot blotting showed that there was a 3-fold increase in the ectopic hybrid *VSG* mRNA 8 hours after inducing expression whereas the endogenous *VSG221* mRNA remained at wild type levels, (Figure 13 C, D). However, one clone of these cell lines was completely leaky, as seen on the blot, and the other two had relatively higher leakiness of expression in comparison to the other overexpression cell lines generated in this thesis. This accounted for the low fold-increase of the ectopic *VSG* mRNA.

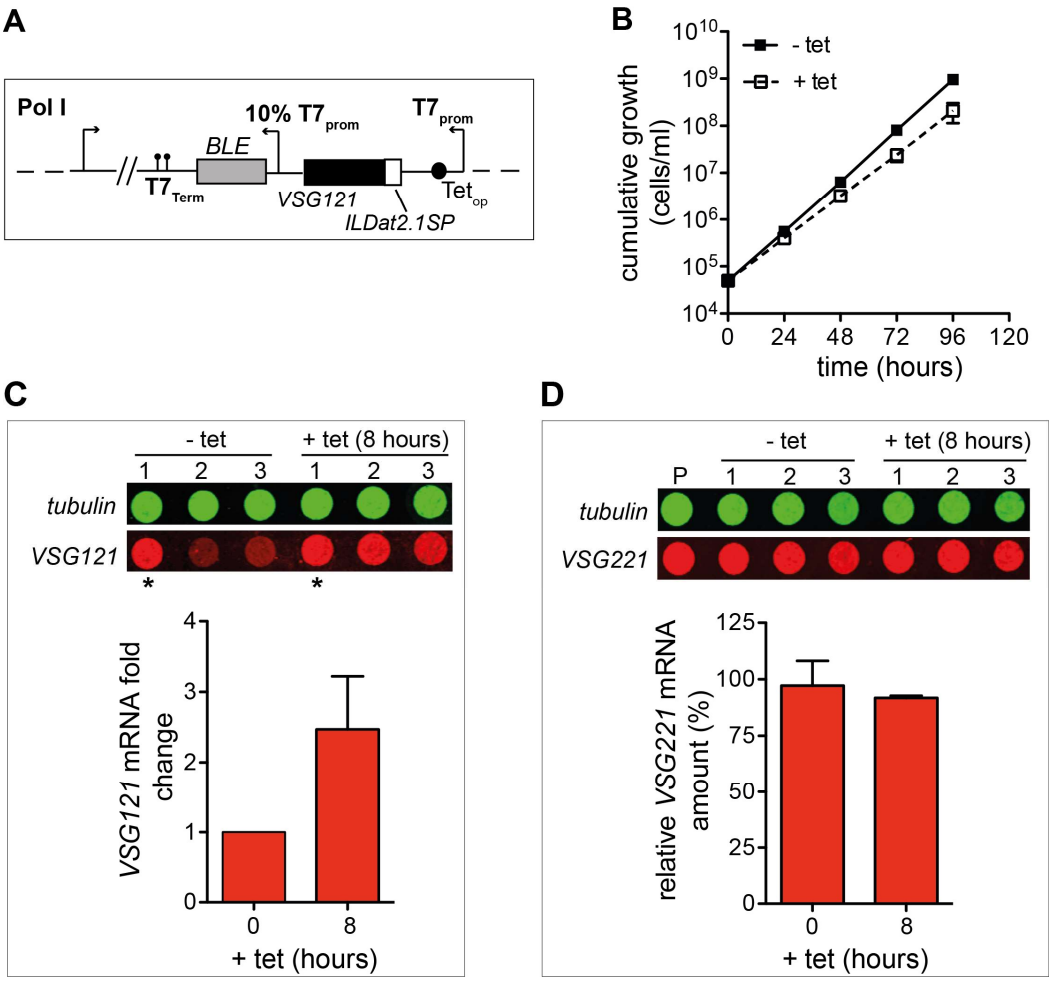

Figure 13: Endogenous *VSG* mRNA regulation response is not elicited upon the expression of VSG121 with an ILDat2.1 ER import signal.
(A) Schematic of the *VSG* hybrid overexpression construct. **(B)** Growth curve showing the induced (+tet) and non-induced (–tet) 221^{ES}.ILDat2.1^{SP}:121^{tet} cell line. **(C)** Quantification of endogenous *VSG221* and **(D)** ectopic *VSG121* mRNA transcripts after 8 hours of inducing expression using dot blotting. The *VSG221* and *VSG121* transcripts were normalized to *β-tubulin*. *VSG221* expression is represented as a percentage relative to the *VSG221* amounts of the parental 13-90 cells (P). *VSG121* amounts are represented as fold-change relative to the *VSG* levels of non-induced cells. The clone marked with an asterisk was eliminated from the analysis as it showed an abnormally high expression of the ectopic *VSG* even before induction of overexpression. The values are represented as means and the error bars represent ± SD of three independent clones for B and D, and two clones for C.

Analysis of protein expression using western blots confirmed that the ectopic VSG121 reporter containing the ILDat2.1 ER import signal was not expressed 24 hours after inducing expression (Figure 14 A, upper panel). On the other hand, the endogenous VSG221 protein remained constant upon the induction of overexpression (Figure 14 A lower panel, B).

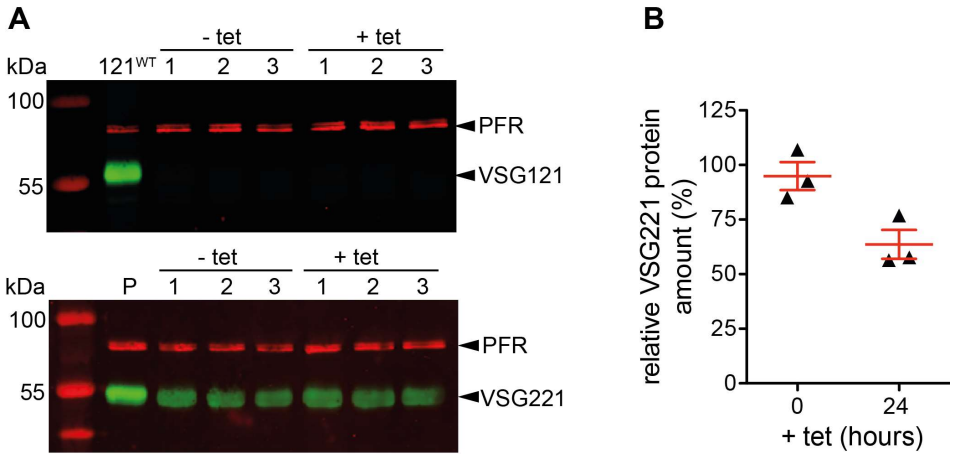

Figure 14: VSG121 protein is not expressed upon the replacement of its ER import signal with that of ILDat2.1 VSG.

(A) Western blot analysis of the ectopic VSG121 (upper panel) and endogenous VSG221 protein (lower panel) after induction of expression for 24 hours. The PFR protein served as a loading control whereas 13-90 parental cells (P) and 121 wild type cells (121^{WT}) served as controls for VSG221 and VSG121 expression, respectively. **(B)** Quantification of the endogenous VSG221 protein amounts, in A, before (0 h) and 24 hours after induction of ectopic hybrid VSG121 expression. The VSG221 protein was normalized to PFR protein and the VSG221 amounts are expressed relative to VSG221 expression in the 13-90 parental cells. Three independent clones were analyzed.

These results are consistent with the previous finding (Section 3.1.4) that the reported ILDat2.1 VSG ER import sequence does not work in *T. brucei* and may not be a bona fide ER import signal at all. Additionally, this strongly supports the hypothesis that the *trans*-regulation of *VSG* mRNA is elicited at the ER as there was no reduction of the endogenous *VSG221* mRNA levels upon overexpression of the reporter VSG121 with impaired ER targeting. The fact that expression of the ER-targeted *T. vivax ILDat1.2* and *ILDat2.1* mRNA elicited the endogenous *VSG* mRNA response despite the low level, or lack of protein production suggested that *VSG* balancing did not require translation and was independent of the *T. brucei VSG* open reading frame (ORF).

To test the possibility that *VSG* mRNA balancing was independent of the *VSG* mRNA and only required the expression of abundant ER-targeted mRNA, GFP reporters with and without the ER-targeting signal were overexpressed in *T. brucei* (Batram, 2013). The reporter for cytosolic expression of GFP contained the *GFP* ORF and a *VSG121* 3′UTR, for high levels of expression (Berberof et al., 1995), and was inducibly expressed from the transcriptionally silent rDNA spacer by T7 polymerase. Quantification of the mRNAs showed that there was a 9-fold increase in the expression of *GFP* mRNA within 4 hours of inducing expression, followed by a ~50% decline after 24 hours. The endogenous *VSG221* mRNA remained unchanged within 24 hours after inducing expression. However, the high levels of GFP expression appeared to be toxic, leading to cell death (Batram, 2013). For the

expression of ER-targeted GFP reporter, the ER import signal of EP1 procyclin (GPI-anchored major surface coat protein of the insect stage procyclics) was added at the 5′ end of the *GFP* reporter ORF that was coupled to the *VSG121* 3′UTR and expressed as above. Expression of high amounts of the reporter caused a rapid decrease of the *VSG* mRNA to 20% of the wild type level within 2 hours and the cells started dying off within 4 hours of GFP expression. Though these experiments showed that the regulation of *VSG* mRNA did not require the ectopic VSG ORF, the early onset of cell death was of concern as it remained unclear whether a global transcriptional down-regulation, possibly occasioned by the toxicity of high levels of expressed GFP or targeting GFP to the ER could have led to the reduction of the endogenous *VSG* mRNA amounts (Batram, 2013). Therefore, it was necessary to use an alternative strategy for the expression of high but non-toxic levels of the GFP reporter transgene.

Pol I-driven transcription of the BES ensures high levels of VSG expression. It has been shown that when an ectopic VSG is integrated upstream of the endogenous VSG in the active BES, the endogenous and ectopic *VSG* mRNAs are expressed on a 50:50 basis (Muñoz-Jordán et al., 1996; Smith et al., 2009). The total amount of *VSG* mRNA in these double expressor cell lines does not exceed wild type levels. It has further been shown that integration of the *VSG121* gene downstream of the ES resident *VSG221* resulted in a reduction in *VSG221* mRNA to ~70%, while the ectopic *VSG121* was expressed at ~30% of the wild type levels (Majeed Bakari Soale, unpublished). Therefore, whether the integration of the *GFP* reporter downstream of *VSG221* would result in the expression of slightly lower non-toxic GFP levels compared to overexpression from the rDNA spacer was examined. Two GFP reporter cell lines were generated for this purpose; in one of the cell lines, the reporter consisted of the *GFP* ORF and the 3′UTR of *VSG121* with an intact 16-mer motif, and in the second cell line, the reporter consisted of the *GFP* ORF and a mutant *VSG121* 3′UTR in which the first seven nucleotides of the 16-mer motif were deleted. The reporter with a mutated 16-mer was generated to ensure lower levels of GFP expression compared to the reporter with wild type 16-mer. As such, these cell line could be used if the reporter with a wild type 16-mer produced toxic amounts of the GFP reporter. The GFP reporter with an intact 16-mer was expressed at higher levels compared to the one with the mutant 16-mer sequence, and the expressed GFP reporter amounts were tolerable in both cell lines. Additionally, the endogenous *VSG* mRNA amounts were unaffected. (Majeed Bakari Soale, unpublished data). Having demonstrated that high levels of expression of the GFP reporter

from the BES was not lethal, I sought to generate cell lines where the ES-integrated GFP reporter was targeted to the ER.

3.1.5 A GFP reporter flanked with the 5′ EP1 ER import and VSG 3′UTR sequence is not expressed from the BES

Encouraged by the possibility to express non-toxic levels of cytosolic GFP reporter from the BES, the next step was to add ER signal sequences to the reporters and investigate whether the endogenous VSG regulation response was elicited in the reporter cell lines. Therefore, the *EP1* procyclin ER import signal was fused at the 5′ end of (i) a *GFP* reporter coupled to the native *VSG121* 3′UTR (with an intact 16-mer motif), and (ii) a *GFP* reporter coupled to *VSG121* 3′UTR deletion mutant in which the first seven nucleotides of the 16-mer were deleted (Figure 15 A, B). The reporter construct containing the *EP1* ER import signal sequence and the *GFP* ORF coupled to *VSG121* 3′UTR was generated by PCR-driven overlap extension (Section 3.1.3). For this purpose, the *VSG221* intergenic region — upstream of the *GFP* reporter in the modified pbRN6 plasmid (Janzen et al., 2004), which was modified to include VSG processing sequences by M. Bakari Soale — was amplified using the forward primer EA11 (containing an Eco105I restriction site) and reverse primer EA10 (containing an *EP1* ER signal sequence overhang). The pbRN6 vector integrates the gene of interest downstream of the active VSG in the VSG221 expression site. A *GFP* ORF coupled to the *VSG121* 3′UTR sequence was amplified in a second PCR reaction from the modified pbRN6 plasmid, using forward primer EA9 (Containing an *EP1* ER signal sequence overhang) and reverse primer HZ32 (flanked with a HindIII restriction site). Subsequently the PCR products were used as the template in a third PCR reaction with primer EA11 and HZ32. The reporter sequence was ligated into the pJet1.2 cloning vector to generate pJet.EP^{SP}:GFP:UTR plasmid. The reporter fragment was excised from pJet.EP^{SP}:GFP:UTR plasmid using HindIII and Eco105I, followed with ligation into HindIII and Eco105I-linearized modified pbRN6 plasmid to yield the pbRN6.EP^{SP}:GFP:UTR construct. The same steps as above were followed using plasmid pbRN6.GFP:ΔUTR to generate the pbRN6.EP^{SP}:GFP:ΔUTR construct.

For stable transformations, pbRN6.EP^{SP}:GFP:UTR and pbRN6.EP^{SP}:GFP:ΔUTR constructs were linearized with a SacI and SalI double digest, and separately transfected into BSF 13-90 parasites to generate 221ES.EP^{SP}:GFP:UTR and 221ES.EP^{SP}:GFP:ΔUTR cell lines, respectively. These cell lines constitutively expressed the GFP reporter as a polycistronic transcription unit together with the VSG. Both cell lines doubled every 6.2 hours, which was slightly slower than the 5.9 hours doubling time of the parental cell line (Figure 15 C).

Analysis of the transcripts using dot blots showed that *VSG221* mRNA was expressed at wild type levels while the *GFP* reporter mRNA could not be detected in the two cell lines (Figure 15 D, E). Fluorescence microscopy showed that neither of the two cell lines expressed detectable GFP protein amounts, and this was additionally confirmed using western blotting (Figure 15 F).

Confirmatory integration PCR revealed that both constructs were correctly integrated into the *VSG221* BES (chapter 6.1 Figure 34). It was intriguing that *GFP* reporter mRNA was not transcribed, yet the antibiotic resistance cassette immediately upstream of the *GFP* reporter was transcribed as the parasites were resistant to the selection antibiotic. A possible explanation of this observation is that the *EP1* leader sequence possibly harbors elements that downregulate the transcription of downstream genes when integrated in the BES. Therefore, an alternative ER import signal — that of VSG121, was used to target the GFP reporter to the ER.

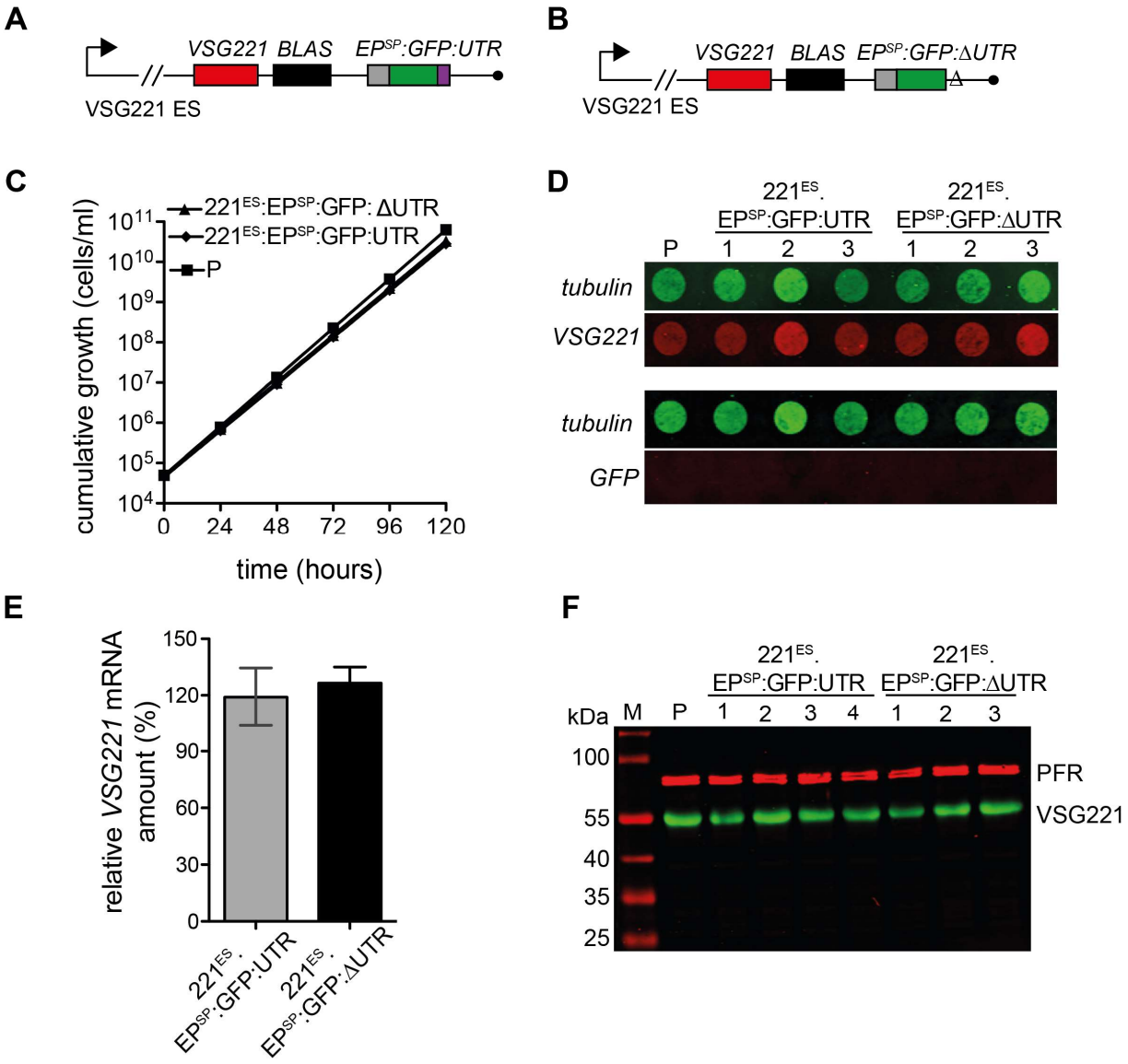

Figure 15: A *GFP* reporter flanked with the 5′ *EP1* ER import signal and *VSG* 3′UTR is not expressed from the active BES locus.

(A, B) Schematic of the modified active expression site of 221ES.EP^{SP}:GFP:UTR and 221ES.EP^{SP}:GFP:ΔUTR cell lines, respectively. **(C)** Growth profiles of the generated cells lines and the parental 13-90 cells (PC) which served as a growth control. **(D)** RNA dot blot showing endogenous *VSG221* (upper panel, red) and *GFP* (lower panel, red) mRNA levels in the two cell lines and parental 13-90 cells using dot blotting. *β-tubulin* (green) gene served as the loading control. **(E)** Quantification of *VSG221* mRNA expression from the dot blot in E. *VSG221* mRNA was normalized to *β-tubulin*, and the expression is given relative to the parental 13-90 cell line (P) *VSG221* expression level. The values are given as means and error bars represent ± SD of three independent clones shown. **(F)** Western blot of showing GFP, VSG and PFR protein expression in the two cell lines. The GFP signal could not be detected, while the endogenous VSG221 and PFR loading control were detected in green and red, respectively.

3.1.6 A GFP reporter flanked with VSG121 ER import signal and VSG121 3′UTR is expressed from the BES

To test whether the EP1 ER import sequence could be responsible for the downregulation of *GFP* gene expression when integrated in the BES, a *GFP* reporter construct in which the

EP1 ER import signal was replaced with a VSG121 ER import signal was generated. PCR-driven overlap extension (section 3.1.3) was used to generate the reporter construct. First, the *VSG221* intergenic region upstream of the *GFP* gene in the modified pbRN6 plasmid was amplified using forward primer EA11 and reverse primer EA14 (containing a *VSG121* ER signal sequence overhang). The *GFP* gene together with the *VSG121 3'UTR* was amplified from the same plasmid using forward primer EA15 (with *VSG121* ER import sequence overhang) and reverse primer HZ32. The third PCR amplification, to fuse the fragments of the first two PCRs, was done using primer EA11 and HZ32, and the fragment was subcloned into pJet1.2 then pbRN6 to generate the pbRN6.121^{SP}:GFP:UTR construct (section 3.1.5). Transformation of 13-90 cells with the linearized construct yielded the 221^{ES}.121^{SP}:GFP reporter cell line (Figure 16 A). Fluorescence microscopy analysis showed that the GFP protein was expressed by the cells (Figure 16 B).

Clones of the 221^{ES}.121^{SP}:GFP cell line had a slightly reduced growth, with a PDT of 6.5 hours compared to the parental cell line's PDT of 5.8 hours (Figure 16 C). Quantification of the transcripts using dot blots showed that the *GFP* mRNA was expressed and that the endogenous *VSG* mRNA levels decreased to 70% of the wild type amounts (Figure 16 D, E). However, the GFP protein appeared to be partially localized in the cytosol. The expression of GFP was further confirmed with Western blot analysis, and subsequent quantification of the endogenous VSG221 protein revealed a 40% reduction in the VSG protein levels (Figure 16 F, G).

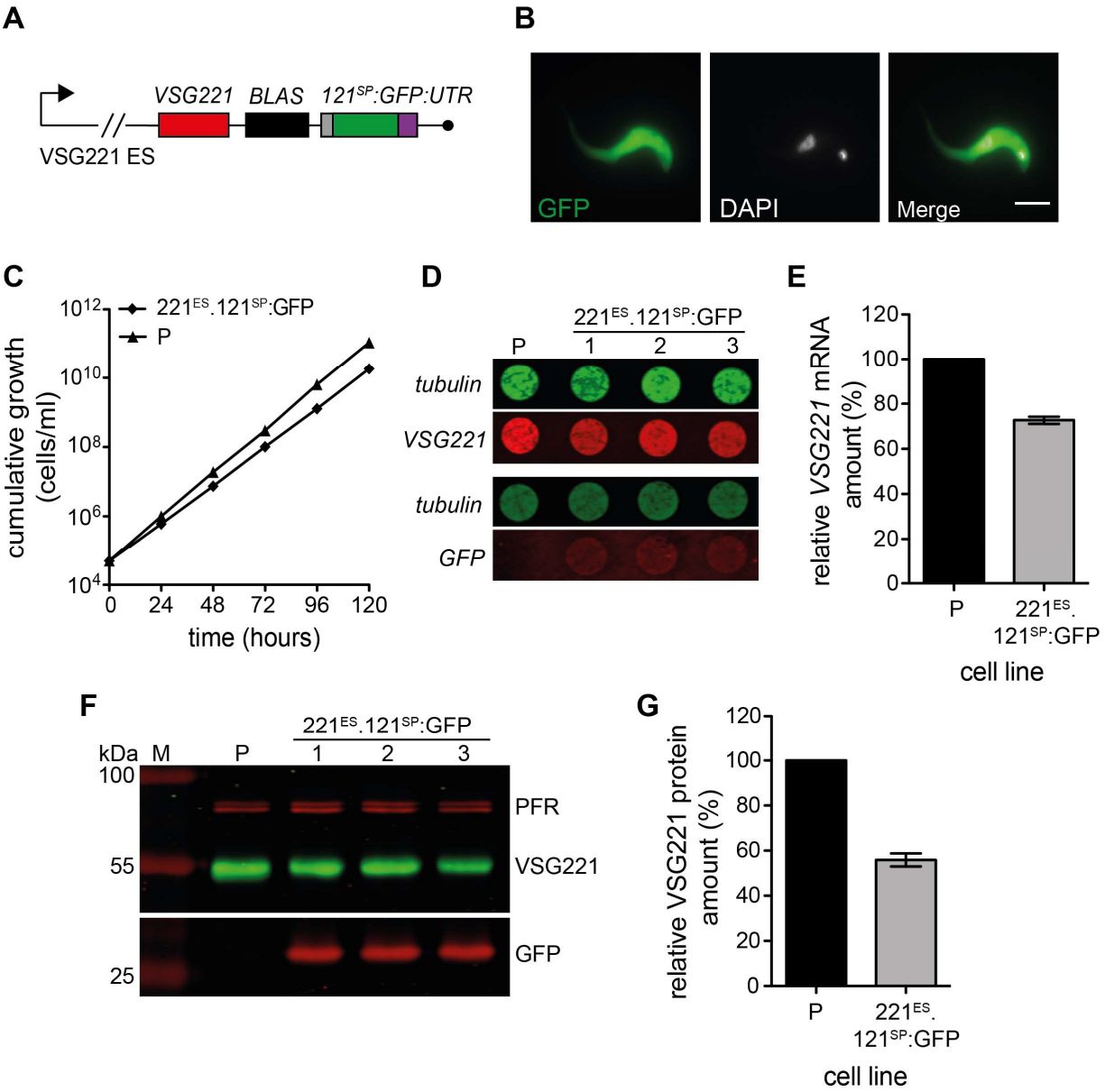

Figure 16: A GFP reporter coupled to VSG121 ER import signal is expressed from the active 221 ES.

(A) Illustration of the modified expression site of the 221ES.121^{SP}:GFP cell line. **(B)** Fluorescence microscopy images of the cell line illustrated in (A). The GFP fluorescence of formaldehyde-fixed cells is shown in green while the DAPI-stained nucleus and kinetoplast is shown in white. The scale bar represents 5 µm. **(C)** Growth curves of the three clones of 221ES.121^{SP}:GFP cell line and the parental 13-90 cells. The error bars are not visible due to the low standard deviation of the means. **(D)** RNA dot blots showing the expression of endogenous *VSG221* (upper panel, red) and *GFP* (lower panel, red) mRNA in the 221ES.121^{SP}:GFP cell line. *β-tubulin* (green) served as a loading control. **(E)** Quantification of the *VSG* expression from the dot blot in F. The *VSG221* mRNA was normalized to *β-tubulin* and expression is represented relative to the 13-90 parental cell line *VSG221* expression level. **(F)** Western blot showing VSG221 (green) and GFP (red) protein expression. Parental 13-90 cells (P) served as the negative control for GFP expression, whereas the PFR protein served as the loading control. **(G)** Quantification of VSG expression from the blot in (F). VSG protein was normalized against the PFR protein and the expression is presented relative to parental 13-90 VSG221 expression. The values are given as means and the error bars represent ± SD of three independent clones.

3.1.7 Overexpression of EP:GFP chimera in BSF *T. brucei* causes cell cycle arrest and aberrant ER morphology

It has been demonstrated that high but non-toxic amounts of the GFP reporter that is fused to a VSG 3′UTR can be expressed from the ES. However, inducible expression of high levels of a GFP reporter containing the EP1 procyclin ER import signal from the rDNA region was toxic to the parasites (Batram, 2013). Here, I tested whether an EP:GFP chimeric protein (consisting of the *GFP* ORF that was inserted in the EP1 procyclin head region, the EP repeats and a VSG 3′UTR (Figure 17 A)) could be expressed in *T. brucei*. Further, the effect of expression of the chimeric protein on the production of VSG was analysed. The EP:GFP chimera was coupled to MITat1.1 3′UTR by PCR-driven overlap extension (section 3.1.1). For this purpose, the EP:GFP chimera was amplified from the pLew82.PARP:EGFP plasmid (Markus Engstler) using forward primer D133 and reverse primer D134. Forward primer D21 and reverse primer D132 was used to amplify the MITat1.1 3′UTR from plasmid pJet.BeNat1.1:M1.1. The two fragments were fused in a second PCR using forward primer D133 and reverse primer D132 and cloned into the pLew82v4 overexpression vector (section 3.1.4) to generate the pLew.EP:GFP construct. Transfection of the construct into 13-90 cells generated the 221^{ES}.EP:GFP^{tet} cell line.

Growth curves showed that cells began to die after induction of EP:GFP expression. However, after 2 – 3 days, a subset of the population re-initiated growth in two out of the three clones. (Figure 17 B). Analysis of protein expression within 24 hours of inducing expression by Western blotting using mouse anti-GFP antibodies showed that the EP:GFP protein was expressed (Figure 17 C). However, bands of lower molecular weights compared to the EP:GFP protein were also observed in the blots, indicating that possibly the protein was being degraded. Seven days post-induction of EP:GFP overexpression, the protein, together with its degradation products, were not detectable on Western blots in the two clones that grew after initially stalling. The VSG protein appeared to be expressed at low levels, 40% of the wild type amounts, before and after inducing the expression of the ectopic protein. While this massive reduction of the VSG protein would explain the cell death observed in one clone, it is not clear why two clones re-initiated growth (Figure 17 D). Cell cycle analysis after 0, 6, 12 and 24 hours of EP:GFP expression showed that EP:GFP overexpression triggered a dual cell cycle arrest at the G1-phase and precytokinesis stages of the cell cycle (Figure 17 E).

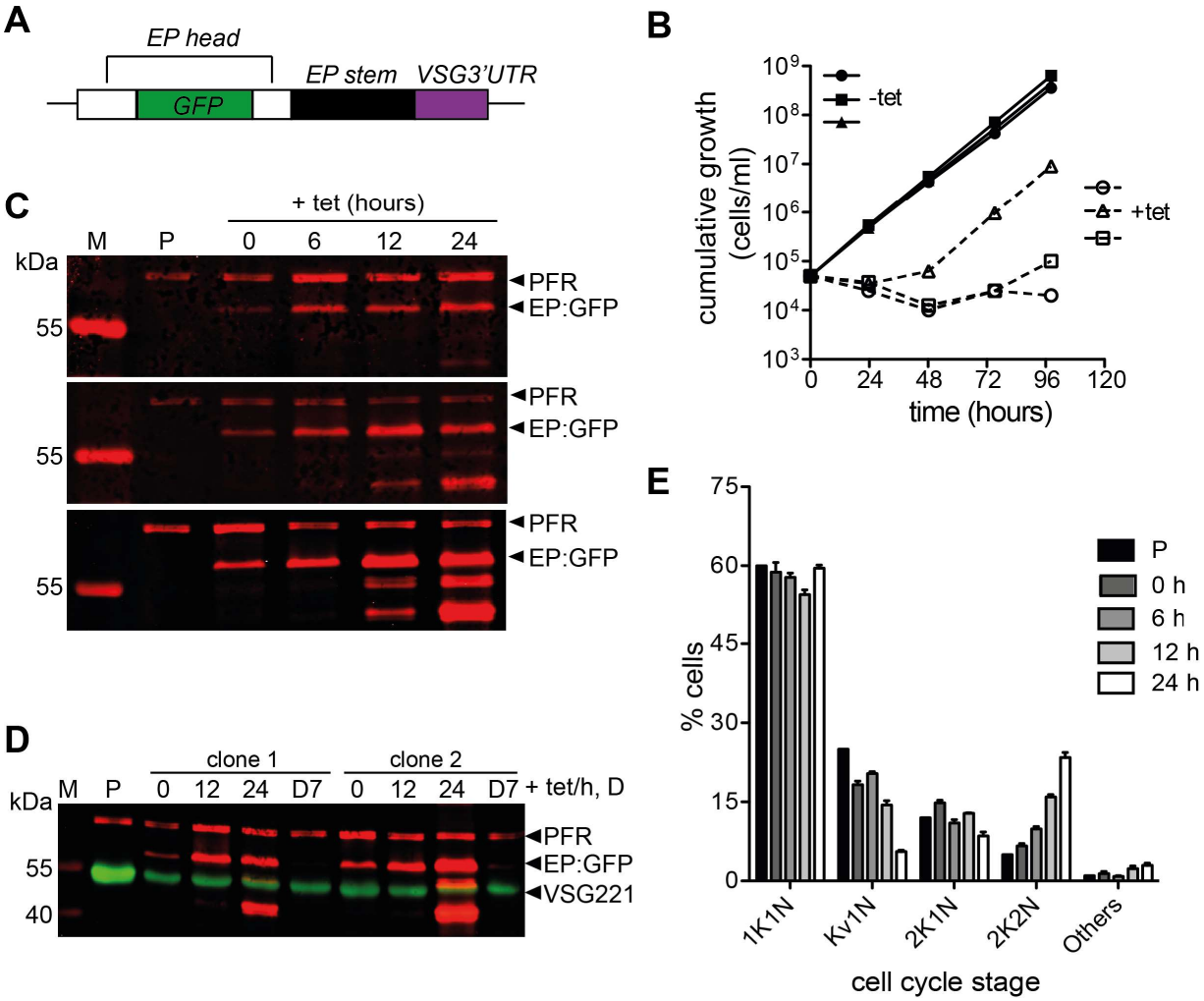

Figure 17: Overexpression of an *EP:GFP* chimera in BSF *T. brucei.*
(A) Schematic of the EP:GFP chimera. The *GFP* reporter was inserted into the EP procyclin head region. **(B)** Cumulative proliferation curve of 221^{ES}.EP:GFP^{tet} cell line cultivated in the presence (+tet) or absence (–tet) of 1 µg/ml tetracycline for 4 days. **(C)** Western blot showing the expression EP:GFP protein using GFP specific mouse antibody at intervals between 0 and 24 hours after inducing expression. The three panels represent different clones. The parental 13-90 cell line (P) served as the negative control while PFR served as a loading control. Both anti-PFR and anti-GFP are mouse antibodies. The EP:GFP degradation products are below the 55 kDa mark. **(D)** Gradual degradation and loss of EP:GFP was shown by Western blotting up to 7 days (D7) of cultivation in the presence (+tet) or absence (–tet) of tetracycline. The parental 13-90 cells served as a negative control while PFR protein served as a loading control. **(E)** The cell cycle was analyzed by counting the number of nuclei (N) and kinetoplasts (K) at different timepoints after inducing expression. DNA was visualized by DAPI staining followed by examination by fluorescence microscopy. The parental 13-90 cell line served as a control. The values are given as means and the error bars show ± SD three independent clones. More than 300 cells of each clone were analyzed.

Transmission electron microscopy (TEM) images revealed an aberrant ER morphology of the induced cells (Figure 18).

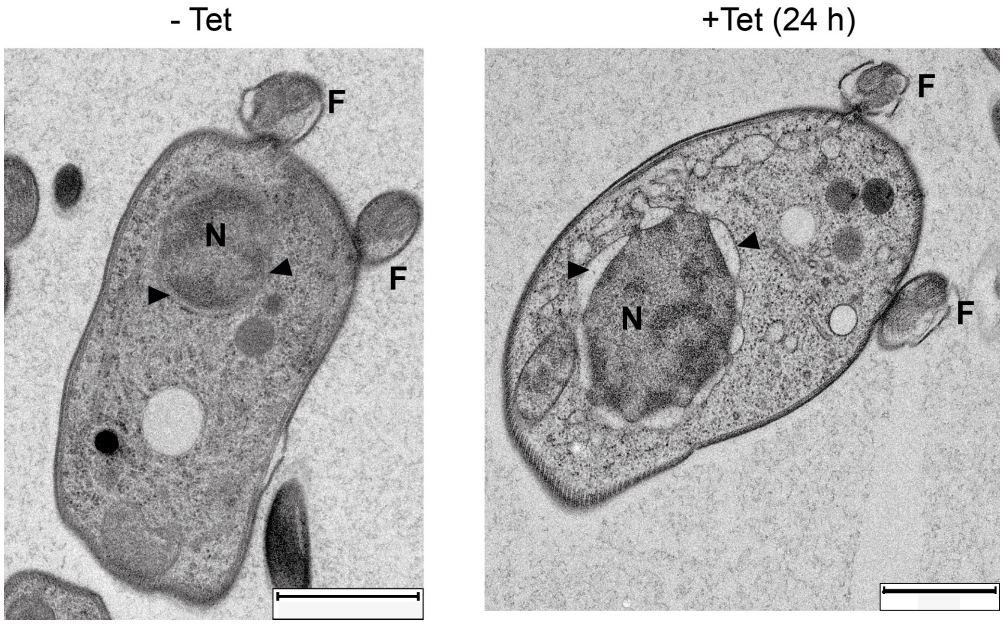

Figure 18: Transmission electron micrographs of 221ES.EP:GFP^{tet} cell line.
The left and right panels represent non-induced and induced (24 hours) cells, respectively. The arrows show the ER while N and F indicate the nucleus and the flagella, respectively. Scale bars: 1µm.

The cells that arrested at the pre-cytokinesis stage did not initiate further division of the two nuclei and kinetoplasts. This arrest is similar to the precise precytokinesis cell cycle stage arrest observed upon ablation of *VSG* mRNA by RNAi and blocking translation of *VSG* transcripts using morpholino oligonucleotides (Sheader et al., 2005; Ooi et al., 2018). In future, it will be necessary to analyze the *VSG* mRNA levels at early time points (such as 2 – 8 hours) after inducing expression, and to ascertain whether the EP:GFP protein was anchored on the cell surface.

3.2 Molecular and phenotypic characterization of the ILDat1.2 VSG overexpression cell line

ILDat1.2 VSG protein was expressed in low amounts despite the transcription of high levels of the mRNA post induction of expression in the 221^{ES}.ILDat1.2^{tet} cell line (Section 3.1.4). The endogenous *VSG221* mRNA levels reduced to 20% of the wild type levels leading to production of low levels of the VSG221 protein. This production of low levels of the VSG protein caused a stagnation in growth for 24 – 48 hours followed by cell death. It was not known why the ectopic ILDat1.2 VSG was only expressed in low amounts. It is possible this low expression of ILDat1.2 VSG protein was due to inefficient protein folding or addition of the GPI-anchor. This section covers an in-depth characterization of the 221^{ES}.ILDat1.2^{tet} cell line. Further, the site-directed mutagenesis approach is applied to examine whether *T. vivax* VSG signal sequences may hinder the efficient expression of *T. vivax* VSGs in *T. brucei*.

3.2.1 Induction of ILDat1.2 VSG expression causes a dual cell cycle arrest

Since the ILDat1.2 overexpression cell line (221^{ES}.ILDat1.2^{tet}) stalled in growth 24 – 48 hours post-induction of *VSG* expression, I investigated whether the cells were still viable during this period. For this purpose, the cells were stained with propidium iodide (PI). PI is a fluorescent agent with an excitation and emission maximum of 535 nm and 617 nm, respectively, when bound to nucleic acids. As such, PI can be used for quantifying viable or dead cells in a population as it is cell membrane impermeant and can only stain the DNA of cells with compromised cell membranes. In this study, cells were collected at different time points after induction of ectopic ILDat1.2 VSG expression, stained with 2.5 µg/ml PI and detected by flow cytometry. The results showed that the cells were viable for at least 48 hours (Figure 19 A), suggesting that the cells were arrested at some stage of the cell cycle. Analysis of the cell cycle at different time points within the first 24 hours following induction of ectopic ILDat1.2 VSG expression was carried out. For this, the cell surface was stained with Atto 488-NHS dye and cells fixed in 2% formaldehyde. The DNA material was stained with DAPI and the cells were viewed with a widefield iMIC microscope. Scoring of the kinetoplast and nucleus configuration showed that the cells were arrested at both the G1-phase (1K1N) and the precytokinesis (2K2N) cell cycle stages (Figure 19 B). Approximately 50% of the cells remained in the G1-phase of the cell cycle. In contrast, cells with dividing kinetoplasts decreased from approximately 30% to 10% of the population. The precytokinesis cells constituted approximately 30% of the population, up from below 5% before induction of ectopic VSG expression.

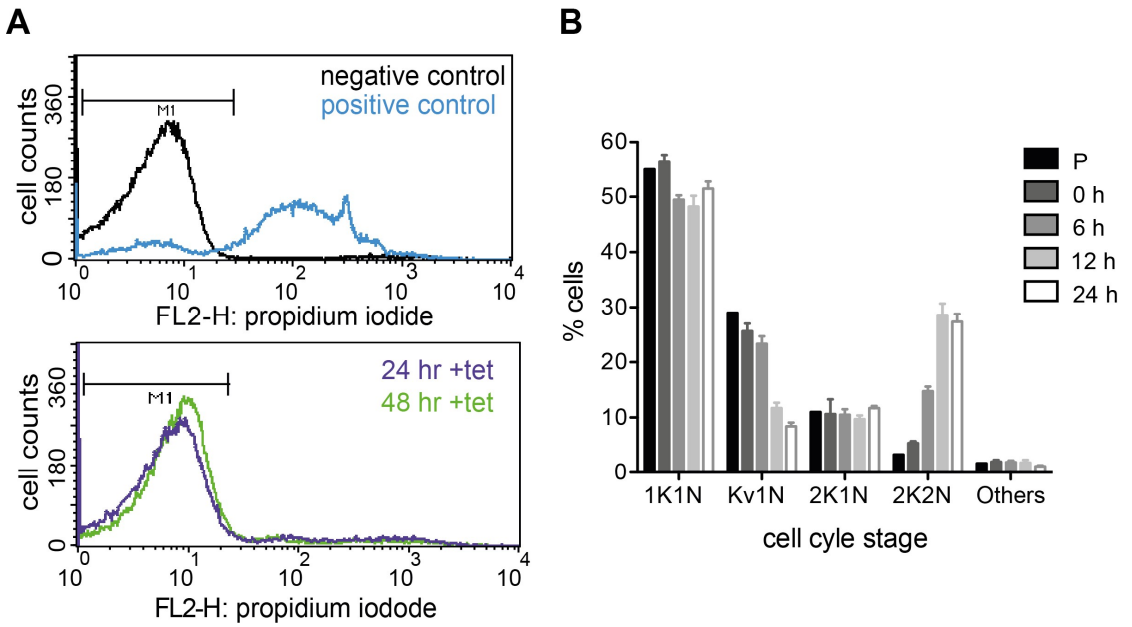

Figure 19: ILDat1.2 *VSG* overexpression causes a dual cell cycle arrest.
(A) Cell viability assay after inducing ILDat1.2 VSG overexpression in the 221^{ES}.ILDat1.2^{tet} cell line. The upper panel represents controls. Parental 13-90 cells at the exponential growth phase served as the negative control. In contrast, parental 13-90 cells treated by addition of 0.3 mg/ml of puromycin, followed by incubation for 48 hours served as the positive control. In the lower panel, the cells were induced to express ILDat1.2 *VSG* for 24 and 48 hours by addition of 1 µg/ml tetracycline. The cells were stained with propidium iodide and detected by flow cytometry. 25,000 events were counted. **(B)** Cell cycle analysis after induction of ILDat1.2 *VSG* expression. Cells were fixed at 0, 6, 12, and 24 hours after induction of ectopic ILDat1.2 *VSG* expression, stained with 5 µg/ml of DAPI and the N-K configuration was viewed on an iMIC microscope and tabulated. The parental 13-90 cell line (P) served as the control. The values are given as means and error bars represent ± SD of two independent clones. More than 300 cells were counted at each time point

It has been shown that inducing the expression of *T. vivax* ILDat1.2 VSG in *T. brucei* expressing endogenous VSG221 causes stagnation of growth, followed by cell death after 48 hours (Section 3.1.1). Additionally, the parasites undergo a dual cell cycle arrest at the G1-phase and precytokinesis cell cycle stages. It has previously been reported that overexpression of VSG118 and VSG121 in endogenous VSG221 background causes stagnation of growth at early timepoints after inducing expression, before resuming normal proliferation (Batram et al., 2014). This stagnation of growth is characterized by an accumulation of G1-phase cells, which upregulate the expression of the stumpy-specific PAD1 (Batram et al., 2014). Therefore, I sought to investigate whether the ILDat1.2 expression-induced G1-phase arrested cells upregulated the expression of PAD1.

3.2.2 Stumpy-specific PAD1 is expressed in growth-arrested cells

PAD1 is a carboxylate-transporter family protein that is expressed on the surface of transmission-competent stumpy-form parasites (Dean et al., 2009) . This surface transporter is essential for the perception of citrate/cis-aconitate (CCA) signal required to initiate life cycle development once the parasites are taken up by the tsetse-fly vector. Elements within the *PAD1* 3′UTR regulate expression of the protein. These regulatory elements repress PAD1 protein expression in pleomorphic slender BSF and derepress its expression on the developmental transition to the stumpy stage in response to SIF (Macgregor & Matthews, 2012) . Stumpy cells are growth-arrested in G0/G1-phase of the cell cycle (Fenn & Matthews, 2007; Matthews et al., 2004; Shapiro et al., 1984). Studies by Batram et al. and Zimmermann et al. showed that overexpression of VSG121 in monomorphic and pleomorphic *T. brucei* parasites resulted in a prolongation of the G1-phase of the cell cycle with a concomitant expression of the PAD1 stumpy marker (Batram et al., 2014; Zimmermann et al., 2017).

A GFP reporter cell line was generated to test whether the growth-arrested cells observed in 221^{ES}.ILDat1.2^{tet} cell line on the induction of ILDat1.2 overexpression expressed the PAD1 stumpy marker. For this purpose, plasmid p4231 (M. Carrington, Cambridge, UK) consisting of a *GFP* open reading frame sandwiched between a nuclear localization signal and the *PAD1* 3′UTR at the 5′ and 3′ end, respectively (Figure 20 A), was linearized with PaeI and NotI. The construct was transfected into the 221^{ES}:ILDat1.2^{tet} cell line generating the GFP:$PAD1_{UTR}$.ILDat1.2^{tet} reporter cell line. The construct integrates into the *tubulin* intergenic region, and green fluorescence of the nucleus is an indication of stumpy-specific PAD1 expression in this cell line.

A *VSG221* RNAi cell line, which shows a precise pre-cytokinesis cell cycle arrest phenotype on the induction of *VSG221* RNAi (Sheader et al., 2005) and the parental 13-90 cell line were also transfected with linearized plasmid p4231 to generate GFP:$PAD1_{UTR}$.221^{RNAi} and GFP:$PAD1_{UTR}$.221^{ES} cell lines, respectively. It is known that slender BSF can be induced to differentiate to the quiescent stumpy forms by cAMP analogues, mediated by their intracellular hydrolysis products (Laxman et al., 2006; Vassella et al., 1997). Therefore, the monomorphic GFP:$PAD1_{UTR}$.221^{ES} cell line was treated with 100 µM of the cAMP analogue 8-(4-chlorophenylthio)-cAMP (pCPT-cAMP) for 48 hours. There was an accumulation of the 1K1N cells with over 95% of the cells expressing the PAD1 stumpy marker at this timepoint as indicated by the green fluorescence of the nucleus. This cell line served as a positive control for flow cytometry analyses (Figure 18 B, left panel). Induction of

ILDat1.2 VSG overexpression in the GFP:PAD1$_{UTR}$.ILDat1.2^{tet} led to the expression of PAD1, marked by the appearance of a green fluorescent nucleus, in the growth-arrested parasites. Approximately 40% of the cells were GFP-positive within 24 hours of inducing the expression of ILDat1.2. After 32 hours, the population of GFP-positive cells increased to 60% (Figure 20 B, right panel). Analysis of this cell line with fluorescence microscopy corroborated the flow cytometry data (Figure 20 C).

To determine whether PAD1 expression was cell cycle stage-specific, the GFP:PAD1$_{UTR}$.ILDat1.2^{tet} cell line was treated with DAPI to stain the DNA material for cell cycle staging. The cells were viewed under the Nikon fluorescent microscope for DAPI and GFP fluorescence. More than 60% of the 1K1N cells were GFP-positive 24 hours after induction of ILDat1.2 overexpression. At the same timepoint, approximately 60% of the pre-cytokinesis 2K2N cells were GFP-positive, increasing to 80% after 32 hours (Figure 20 D). It is likely that the GFP-positive 2K2N cells emanate from a subpopulation of IKIN cells that escaped arrest but were committed to express the PAD1 protein. Inducing *VSG221* RNAi in the GFP:PAD1$_{UTR}$:221^{RNAi} cell line, led to a 2K2N pre-cytokinesis cell cycle arrest; however, PAD1 was not expressed in these cells as they were GFP negative (Figure 20 E). Together, these data show that ILDat1.2 VSG overexpression-induced cell cycle arrest is accompanied by the expression of stumpy specific PAD1 stumpy marker in monomorphic slender BSF. These findings agree with the study conducted by Batram et al., 2014.

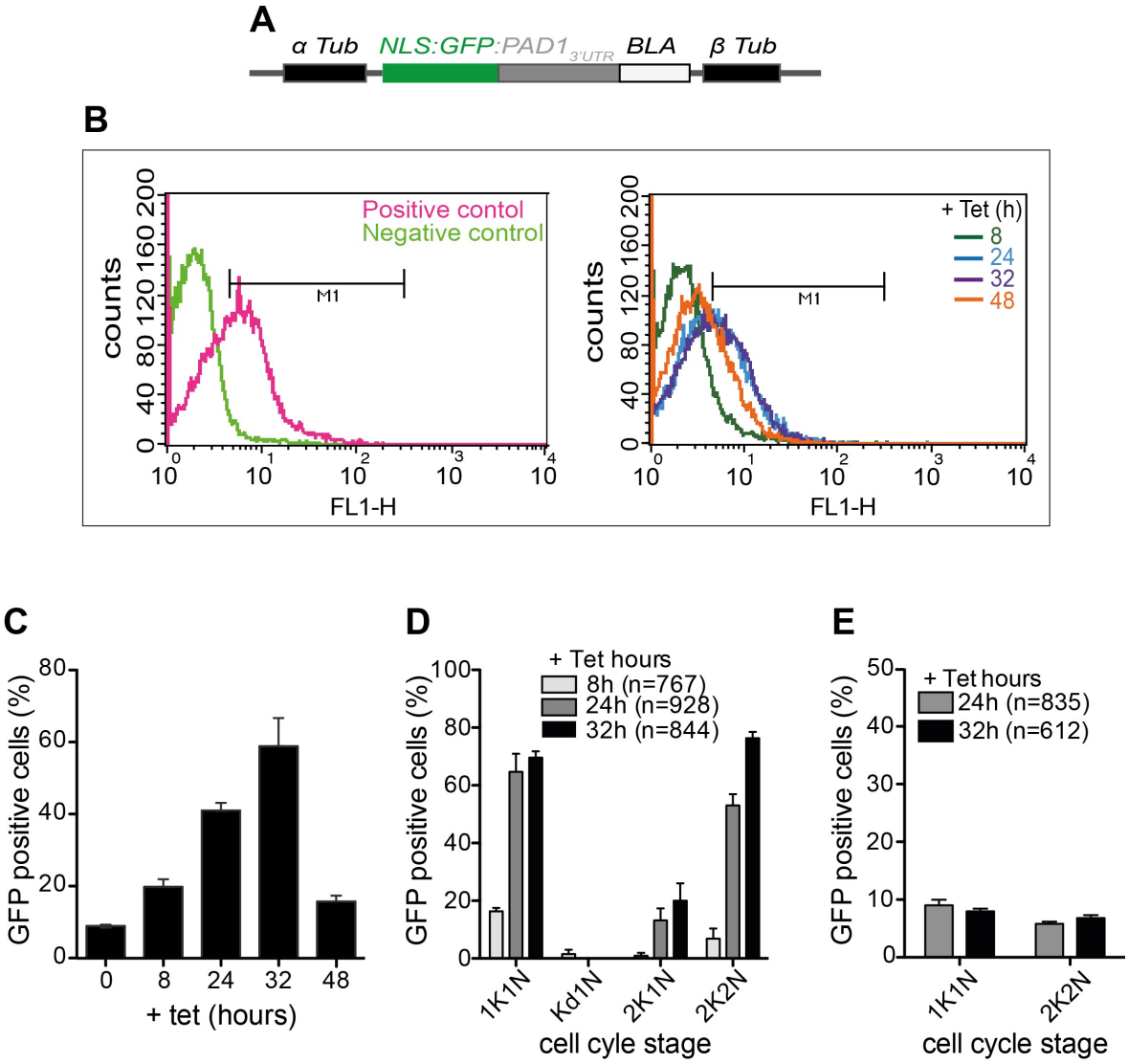

Figure 20: Stumpy-specific *PAD1* is expressed upon the induction of ectopic ILDat1.2 overexpression.

(A) Illustration of the GFP:PAD1$_{UTR}$ construct. **(B)** Quantification of GFP fluorescence of the nucleus by flow cytometry Parental 13-90 cells served as the negative control whereas GFP:PAD1$_{UTR}$.221ES cells, induced to differentiate by pCPTcAMP treatment for 48 hours, served as the positive control (left panel). The GFP:PAD1$_{UTR}$.ILDat1.2^{tet} cell line was induced to express ILDat1.2 VSG for 8, 24, 32 and 48 hours and the fluorescence of the nucleus was analyzed (left panel). A total of 25,000 events were counted at each time point by flow cytometry. **(C)** Quantification of the total number of cells with a fluorescent nucleus and **(D),** Quantification of cells with green fluorescent nuclei at different cell cycle stages after induction of ILDat1.2 VSG expression for 8, 24 and 32 hours by fluorescence microscopy. **(E)** Quantification of GFP fluorescence in the nucleus of the GFP:PAD1$_{UTR}$.221^{RNAi} cell line after inducing *VSG221* RNAi for 24 and 32 hours. Values are given as mean percentages and the error bars represent ±SD of three independent clones.

3.2.3 The active BES is not attenuated after induction of ILDat1.2 expression

Only one of the ~15 telomeric BES, harboring a single *VSG* gene, in *T. brucei* is active at a time. This ensures that only one *VSG* gene is expressed at any time (Glover et al., 2013). Two main mechanisms are used by the parasites to change the expressed *VSG* protein: the

parasites can use recombination events to replace the BES resident VSG with a new VSG, or they can silence the active BES with concomitant activation of a second BES (Hertz-Fowler et al., 2008; Taylor & Rudenko, 2006).

It has been shown that overexpression of VSG121 from the rDNA spacer led to a directional attenuation of the active BES and a prolongation of the G1-phase in monomorphic *T. brucei* cells (Batram et al., 2014). Given that overexpression of ILDat1.2 caused a G1-phase cell cycle arrest as well, I examined whether inducing ILDat1.2 VSG overexpression from the rDNA spacer caused the attenuation of the active VSG221 BES. For this purpose, a GFP reporter cell line based on the 221^{ES}.ILDat1.2^{tet} cell line was generated. The 221^{ES}.ILDat1.2^{tet} cell line was transformed with BamHI/KpnI-linearized plasmid p3845 (M. Carrington, Cambridge, UK), which integrates the *GFP* reporter together with a *blasticidin* resistance cassette downstream of the active BES promoter, yielding the GFP^{PRO}.221^{ES}.ILDat1.2^{tet} cell line (Figure 21 A, B). This cell line enabled evaluation of the BES activity by quantification of the GFP signal intensity using flow cytometry. A positive control cell line GFP^{PRO}.221^{ES}.121^{tet} was also generated in a similar way.

The generated cell lines had comparable growth phenotypes to their respective parental cell lines (Figure 21 C, D). Quantification of the GFP signal intensities after induction of ectopic VSG121 expression in the GFP^{PRO}.221^{ES}.121^{tet} cells showed a gradual attenuation of the active BES. The highest reduction of the GFP signal was recorded after 48 hours of VSG121 overexpression. After 72 hours, the GFP intensity reverted to the 24-hour levels (Figure 21 E). On the other hand, the GFP signal remained constant for 72 hours after inducing the expression of ILDat1.2 VSG in the GFP^{PRO}.221^{ES}.ILDat1.2^{tet} cell line (Figure 21 F).

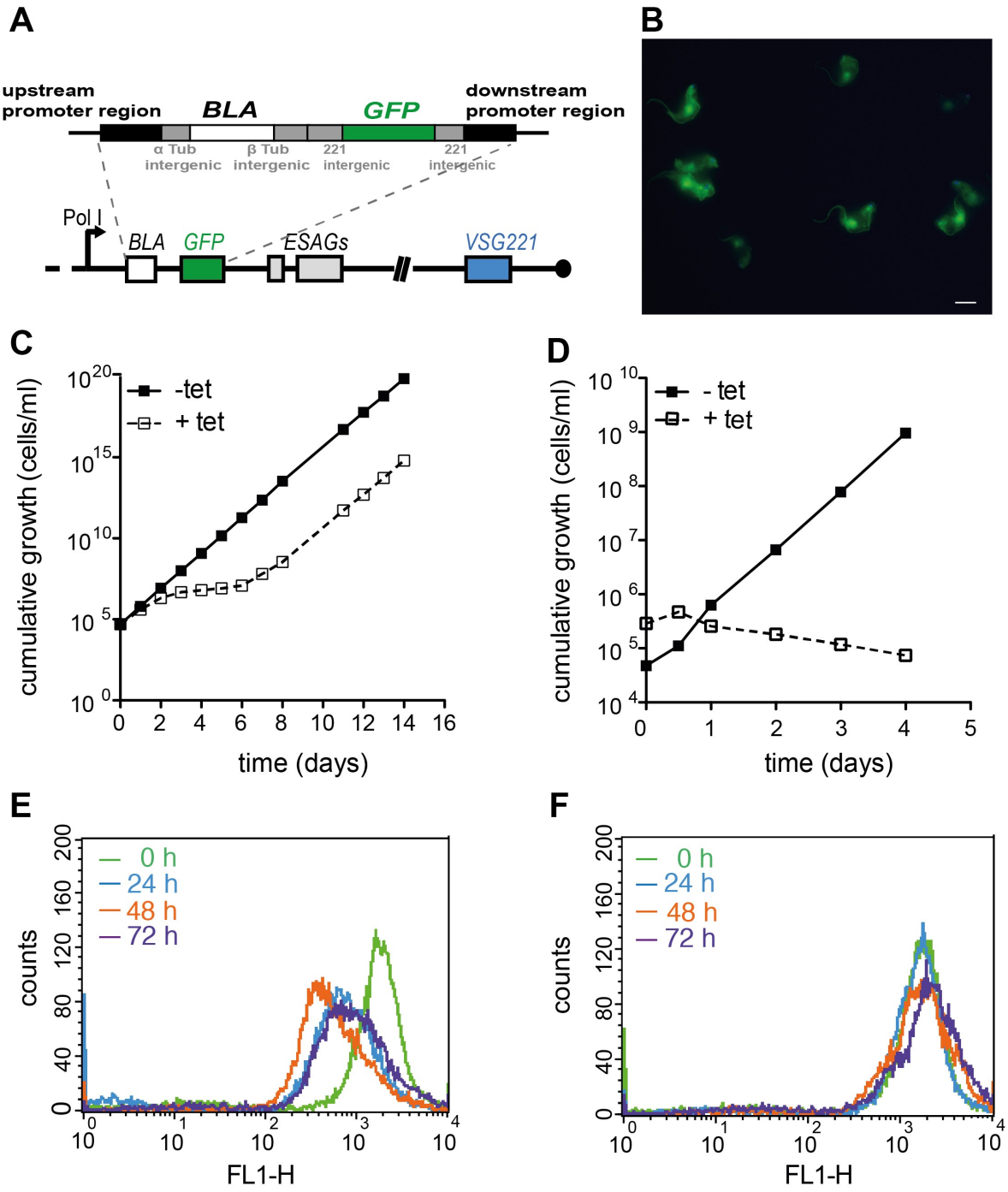

Figure 21: The active BES is not attenuated upon the induction of ILDat1.2 VSG overexpression. (A) Schematic representation of the construct used for integrating a *GFP* transgene downstream of the active VSG221 BES promoter. (B) Fluorescence microscopy images of the $GFP^{PRO}.221^{ES}.121^{tet}$ cell line constitutively expressing GFP. Scale bar: 5 µm. (C) Growth curves of the $GFP^{PRO}.221^{ES}.121^{tet}$ and (D) $GFP^{PRO}.221^{ES}$:ILDat1.2^{tet} cell lines in the presence (+tet) or absence (-tet) of tetracycline. The values represent the mean values of three independent clones. Error bars are not visible due to the small standard deviation. (E) Quantification of GFP signal by flow cytometry 0, 8, 24, 48 and 72 hours after inducing the expression of VSG121 and (F) ILDat1.2 VSG. A total of 25,000 events were counted at each time point. The data provided is for one clone of each cell line; data for two more clones is provided in Chapter 6.1 Figure 35.

3.2.4 ILDat1.2 VSG is sufficiently expressed in *T. brucei* on replacement of its GPI signal peptide with that of MITat1.11 VSG

In most eukaryotes, proteins destined to the endoplasmic reticulum (ER) for posttranslational modifications are translated with a cleavable N-terminal signal peptide. Additionally, proteins that are destined to be GPI-anchored contain a C-terminal signal peptide. This signal sequence is cleaved, and a preassembled GPI anchor added *en bloc* by a transamidase to the C-terminal of the nascent protein (Walter et al., 1984; Walter & Johnson, 1994). *T. brucei* VSG proteins also have cleavable N-terminal ER signal peptides and a rather conserved C-terminal GPI signal peptide (Section 1.10.1).

To determine whether the *T. vivax* VSG GPI signal peptides share similar sequence elements to *T. brucei* VSG GPI signal peptides, the *T. vivax* VSG GPI signal peptides were predicted using the SignalP online server (Nielsen, 2017). The predicted ILDat1.2 VSG signal was 26 residues in length, longer than the *T. brucei* VSG signals, whereas the ILDat2.1 GPI signal was 23 residues long. Alignment of the predicted signal sequences with a few characterized *T. brucei* VSG GPI signal sequences revealed the absence of the conserved elements present in *T. brucei* VSG GPI signal peptides indicated by asterisks (Figure 22). Additionally, the C-terminal region of the ILDat1.2 and ILDat2.1 VSG GPI signals terminate in a mixture of hydrophobic and hydrophilic residues instead of the pure hydrophobic core as observed in VSG GPI signal peptides.

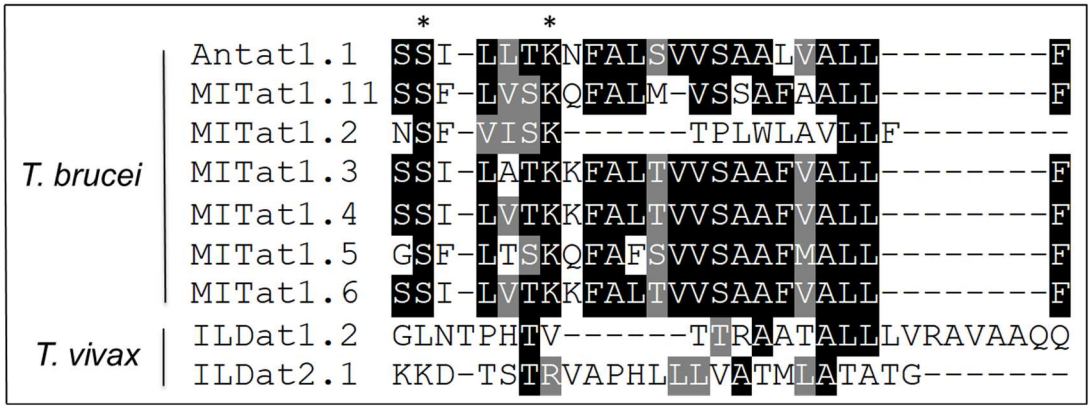

Figure 22: Multiple sequence alignment of *T. brucei* and *T. vivax* VSG GPI signal peptides.
The predicted *T. vivax* VSG GPI signal peptides were compared to a few characterized *T. brucei VSG* GPI signal peptides. Sequence alignment was done using T-Coffee (Di Tommaso et al., 2011) with default settings and rendered with BOXSHADE (ExPASy). The single letter codes represent the different amino acids; identical amino acids are shaded black while similar residues are shaded grey. The conserved serine and lysine residues at position ω+2 and ω+7, respectively, of *T. brucei VSG* GPI signal peptides, are highlighted with asterisks.

These differences between the *T. brucei* and *T. vivax* VSG GPI signal peptides informed the decision to replace the predicted *T. vivax* ILDat1.2 and ILDat2.1 VSG GPI signal peptides with a *T. brucei* MITat1.11 VSG GPI signal peptide. The hybrid sequences were generated by PCR-driven overlap extension (Section 3.1.3). First, *ILDat1.2* ORF lacking the predicted native GPI signal peptide was amplified from plasmid pBSK.ILDat1.2:M1.1 using forward primer D146 and reverse primer D147. The MITat1.11 GPI signal peptide coupled to MITat1.1 3′UTR was amplified using forward primer D148 and reverse primer D132 from plasmid pKD4.M1.11:M1.1UTR. The PCR products of the first two reactions were fused in a second PCR using primer set D146 and D132. A similar approach was used to replace the predicted ILDat2.1 GPI signal peptide with that of MITat1.11 VSG. Forward primer IJ2 and reverse primer IJ3 were used to amplify the *ILDat2.1* ORF lacking the GPI signal peptide from plasmid pBSK.ILDat2.1:M1.1. MITat1.11 GPI signal sequence together with *MITat1.1* 3′UTR was amplified from plasmid pKD4.M1.11:M1.1UTR using forward primer D146 and reverse primer D132. Products of the first two amplifications were joined in a second amplification step using primer set IJ2 and D132. The generated hybrid gene fragments were cloned into the pLew82v4 overexpression vector (Section 3.1.1) to generate the pLew.ILDat1.2:M1.11^{SP} and pLew.ILDat2.1:M1.11^{SP} constructs. Transfection of NotI-linearized pLew.ILDat1.2:M1.11 and pLew.ILDat2.1:M1.11 constructs into parental *T. brucei* 13-90 cells yielded the 221ES.ILDat1.2:M.11^{tet} and 221ES.ILDat2.1:M1.11^{tet} cell lines, respectively (Figure 23 A, B).

The growth profiles showed that the ILDat1.2 overexpressing cell line had a slight decrease in growth with a population doubling time (PDT) of 7.5 hours compared to 6.4 hours in the parental cell line. The induction of ILDat2.1 overexpression led to a similar growth reduction from 6.4 hours (parental cell line) to 7.4 hours PDT (Figure 23 C, D). Protein expression analysis by SDS-PAGE and Western blotting revealed that ILDat1.2 VSG was expressed in high amounts after induction of expression (Figure 23 E). In contrast, ILDat2.1 VSG appeared not to be expressed (Figure 23 F). These results suggest that possibly *T. vivax* GPI signal peptides are not compatible with the *T. brucei* GPI anchoring machinery, which adds the pre-formed GPI anchors to the mature VSG polypeptide.

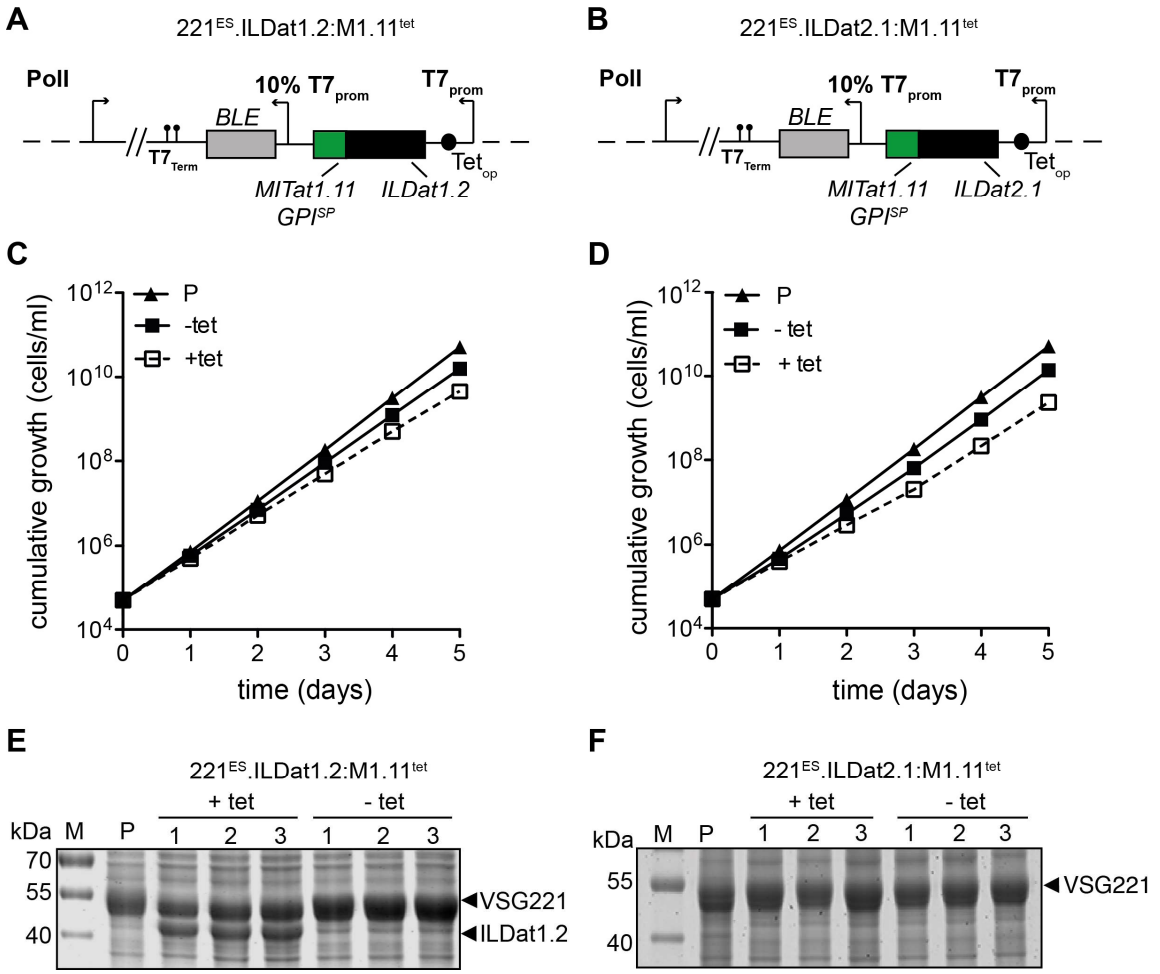

Figure 23: ILDat1.2 is highly expressed on replacement of its GPI signal peptide with that of *T. brucei* MITat1.11.

(A) Schematic of the modified rDNA spacer of the 221ES.ILDat1.2:M.11^{tet} and **(B)** 221ES.ILDat2.1:M1.11^{tet} cell lines. Predicted ILDat1.2 and ILDat2.1 VSG GPI signal peptides were replaced with the GPI signal peptide of *T. brucei* MITat1.11 VSG (green). **(C)** Cumulative growth of 221ES.ILDat1.2:M.11^{tet} and **(D)** 221ES.ILDat2.1:M1.11^{tet} cell lines upon the induction of protein expression with tetracycline for 5 days. Induced and uninduced cells are denoted by (+tet) and (–tet) respectively, while the parental 13-90 cell line (P) served as a growth control (n=3). **(E** and **F)** SDS-PAGE analysis of ectopic ILDat1.2 and ILDat2.1 VSG protein expression before (–tet) and 24 hours after induction of expression (+tet). Parental cells served as the loading control.

It was intriguing that ILDat1.2 VSG could be highly expressed in *T. brucei* by replacement of its predicted native GPI signal peptide with a *T. brucei* VSG signal peptide while the same modification could not result in the expression of ILDat2.1 VSG. To test whether the ectopic *ILDat2.1 VSG* mRNA was sufficiently expressed for VSG protein expression, the mRNA was quantified 8 hours after induction of *ILDat2.1 VSG* expression by RNA dot blotting. The ectopic *ILDat2.1* mRNA amount was upregulated while the endogenous *VSG221* mRNA level remained unchanged while the (Figure 24 A, B).

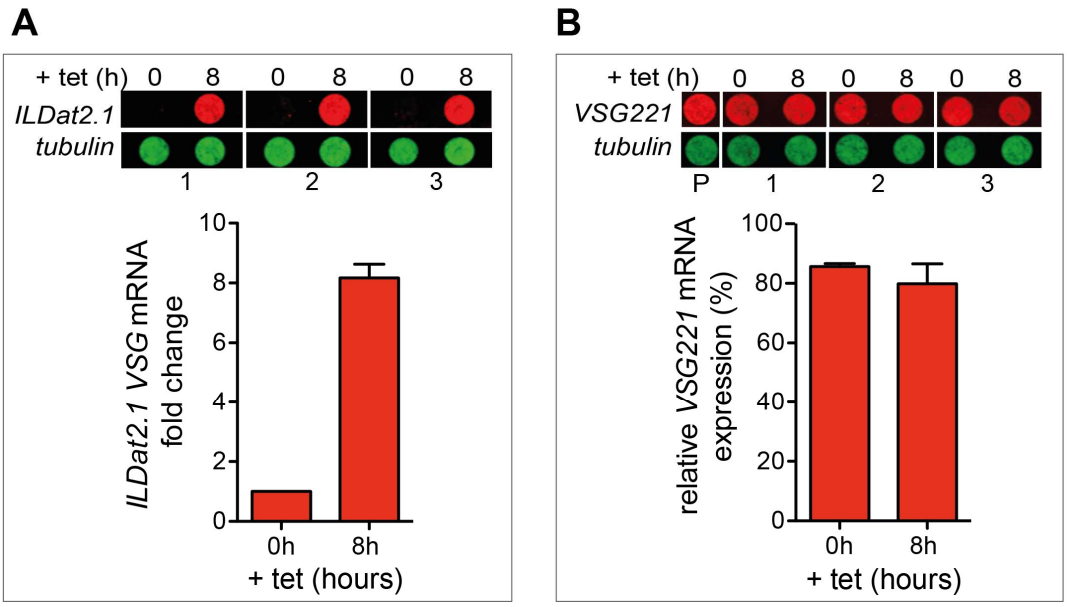

Figure 24: Expression of the hybrid *ILDat2.1* mRNA has no effect on the endogenous *VSG221* mRNA levels.
Quantification of **(A)** ectopic *ILDat2.1 VSG* and **(B)** endogenous *VSG221* mRNA by RNA-dot blots before and after induction of *ILDat2.1* overexpression. *β-tubulin* served as the loading control. *VSG* mRNA was normalized to *β-tubulin* and the *VSG221* mRNA level is expressed relative to the *VSG* level of the parental 13-90 cells (P). The *ILDat2.1* mRNA level is given as fold change relative to the uninduced cells (0 h). The values are given as means and the error bars represent ±SD of three independent clones.

3.2.5 ILDat2.1 VSG with bona fide *T. brucei* ER import and GPI signals is not adequately expressed

Previous data revealed that the predicted ILDat2.1 VSG ER import signal may not be a bona fide ER import signal and thus does not target the VSG to the ER (section 3.1.2). It was therefore not surprising that regulation of the endogenous *VSG221* mRNA was not elicited in the 221^{ES}.ILDat2.1:M1.11^{tet} cell line after induction of expression (Figure 23 F). To ensure that the hybrid gene containing the ILDat2.1 ORF and the MITat1.11 GPI-anchor signal was correctly targeted to the ER, the native ILDat2.1 ER import signal was replaced with the VSG121 ER import signal using sequential overlap extension PCR amplification (section 3.1.3). First, the *VSG121* ER import sequence was amplified from plasmid pLew.M1.6wt using forward primer D81 and reverse primer EA6. The ILDat2.1:M1.11 hybrid coding sequence lacking the 'dysfunctional' ER import sequence was amplified from plasmid pLew.ILDat2.1:M1.11 using forward primer EA6 and reverse primer D132. The products of the first PCR reactions were fused in a second PCR amplification step using primer set D81 and D132. The hybrid fragment was ligated into the pLew82v4 expression vector (section 3.1.4) to generate the construct pLew.VSG121^{SP}:ILDat2.1:M1.11^{SP}.

Transfection of the NotI-linearized construct into the *T. brucei* 13-90 cell line yielded cell line $221^{ES}.121^{SP}$:ILDat2.1:M1.11^{tet} (Figure 25 A).

Analysis of the growth showed that the cells doubled at least once within the first 24 hours after inducing the expression of the ectopic VSG, after which the cell numbers began to decline. (Figure 25 B). Analysis of protein expression by SDS-PAGE 24 hours post-induction of expression showed that the ectopic hybrid VSG was not expressed (Figure 25 C). Quantitative RNA dot blots showed an 8-fold increase in the ectopic *VSG* mRNA levels with a simultaneous reduction in the endogenous VSG221 mRNA levels to 30% of the wild type level, 8 hours after inducing the expression (Figure 25 D, E).

Taken together, these results further support our finding that the ILDat2.1 ER import signal sequence is not a bona fide ER targeting signal. Inducing the expression of *ILDat2.1 VSG* consisting of the *VSG121* ER import signal sequence and a *MITat1.11 VSG* GPI anchor signal sequence elicited the regulation of the endogenous *VSG* mRNA. However, the ectopic VSG protein appeared not to be expressed. It is plausible that the *T. vivax ILDat2.1 VSG* sequence (Gardiner et al., 1996) lacks some sequence elements that are essential for VSG processing that would qualify it as a VSG in *T. brucei*. No further attempts were made to express the *T. vivax* VSG ILDat2.1 in *T. brucei* parasites.

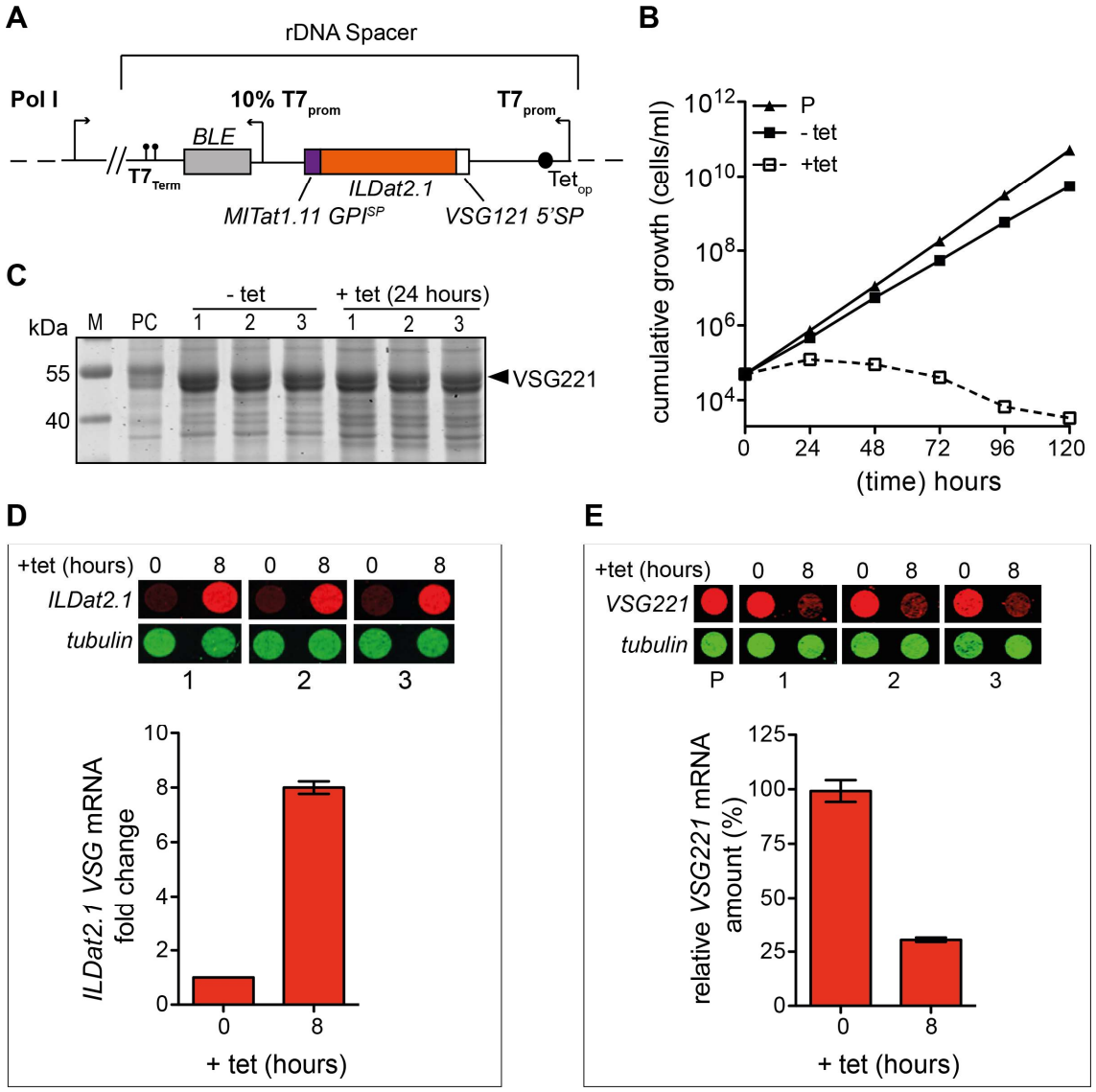

Figure 25: Expression of ILDat2.1 VSG is not rescued upon replacement of both its ER import and GPI anchor signal sequences.

(A) Schematic representation of the hybrid *ILDat2.1* gene with *T. brucei VSG* signal sequences and the overexpression construct. **(B)** Cumulative growth curves of tetracycline-induced (+tet) and non-induced (-tet) parasites. The parental 13-90 cell line (P) served as a growth control. The growth curve represents cumulative average cell numbers (±SD) of 3 independent clones. **(C)** Protein samples were resolved on Coomassie gels **(D)** RNA dot blot and quantification of the ectopic *VSG* and **(E)** endogenous *VSG221* mRNA 8 hours after induction of the ectopic *VSG* overexpression. The *VSG* mRNA was normalized against *β-tubulin* mRNA. The expression of VSG221 is presented as a percentage relative to the *VSG221* mRNA amount of parental 13-90 cells, whereas ectopic *VSG* is presented as fold change. Values are given as means and the error bars represent ±SD of three independent clones.

3.2.6 Generation of a *T. brucei* cell line constitutively expressing *T. vivax* ILDat1.2 VSG

Encouraged by the successful inducible overexpression of ILDat1.2 in *T. brucei* after replacement of its wild type GPI signal peptide, I set out to constitutively express the ectopic ILDat1.2 VSG using the approach of Muñoz-Jordán et al., 1995. The chimeric *ILDat1.2 VSG*

ORF containing the *MITat1.11* GPI signal peptide and *MITat1.1* 3′UTR was excised from plasmid pBSK.ILDat1.2:M1.11:M1.1 using EcoRI. The DNA fragment was cloned into EcoRI-digested pKD4 plasmid, to generate pKD4.ILDat1.2:M1.11^{SP} construct which was BglII-linearized and transfected into parental 13-90 cells. A double expressor (DEX) 221ES.ILDat1.2 cell line was thus generated, which constitutively expressed the ectopic *ILDat1.2 VSG* upstream of the endogenous BES-resident *VSG221* (Figure 26 A).

Next, XhoI-linearized *VSG221* knockout plasmid (Bührdel, 2010) was transfected into the DEX 221ES.ILDat1.2 cell line. Integration of this plasmid into the BES resulted in deletion of the endogenous *VSG221* by replacing it with a *puromycin* antibiotic resistance cassette, to generate the VSG221 knockout (KO) 221^{KO}.ILDat1.2 cell line (Figure 26 A). Evaluation of the 221^{KO}.ILDat1.2 knockout cell line growth showed that the cells grew slower; doubling at 9.2 hours compared to both the DEX and parental cell lines, which had a PDT of 8.2 and 6.4 hours respectively (Figure 26 B). Western blot confirmed that the endogenous VSG221 protein was replaced and that ILDat1.2 VSG was the only expressed VSG (Figure 26 C, D). In conclusion, the *T. vivax* ILDat1.2 VSG can be constitutively expressed from the BES and form a coat on *T. brucei* despite its relatively short primary sequence compared to characterized *T. brucei* VSGs.

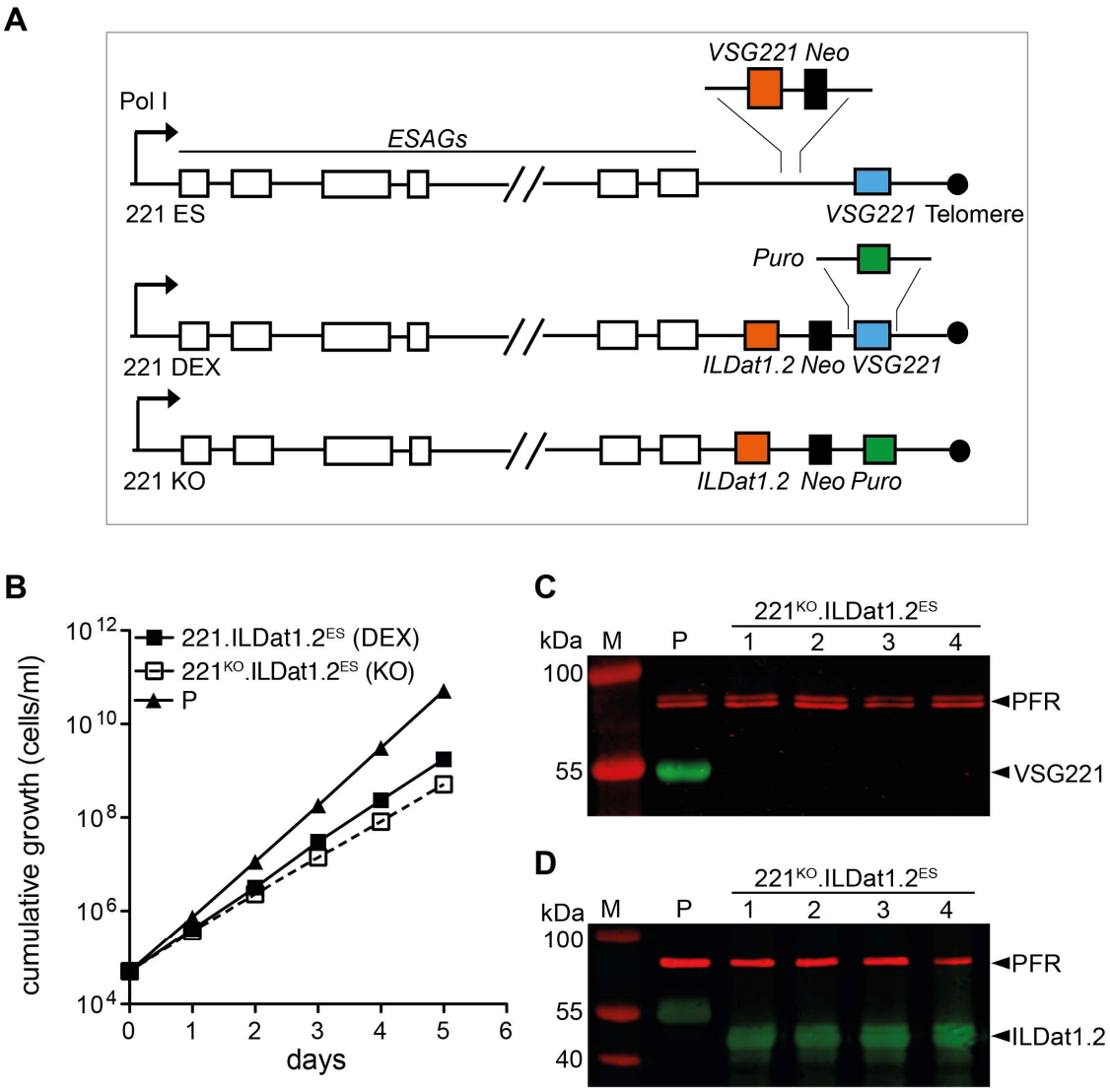

Figure 26: Generation of a *T. brucei* cell line constitutively expressing ILDat1.2 VSG.
(A) Schematic showing the generation of the *T. brucei* 221ES.ILDat1.2^{ES} (DEX) cell line and deletion of the endogenous *VSG221* to generate 221^{KO}.ILDat1.2^{ES} cell line. The ES promoter is indicated with an arrow, ESAGs with open boxes and telomere with a black circle. **(B)** Growth curves of the 221ES.ILDat1.2^{ES} (DEX) and 221^{KO}.ILDat1.2^{ES} cell line. The parental 13-90 cell (PC) served as the control. The graph represents the cumulative mean values of 3 clones for each cell line. The error bars are not visible due to the small standard deviation. **(C)** Western blot showing endogenous VSG221 expression in the 221^{KO}.ILDat1.2^{ES} cell line using *VSG221*-specific rabbit antibody. The endogenous VSG was successfully deleted in 4 clones as indicated by the absence of the ~55 kDa band. **(D)** Western blot showing the expression of ectopic ILDat1.2 VSG in 221^{KO}.ILDat1.2^{ES} cell line using ILDat1.2 rabbit antibody. There was cross-reactivity of the ILDat1.2 antibody with the endogenous VSG221 protein. However, the size difference between VSG221 and ILDat1.2 VSG (~46 kDa) makes it possible to distinguish the two proteins easily. A parental 13-90 cell line expressing VSG221 served as the positive control, and the PFR protein served as a loading control.

3.2.7 Expression of VSG chimeras in *T. brucei*

Despite the high sequence variability between *VSG* genes, their tertiary protein structure is highly conserved. The N-terminal domain (NTD) that is extended on the cell surface is vital for antigenic variation and virulence through glycosylation of some residues (Pinger et al.,

2018). The C-terminal domain (CTD) plays a structural role (Bartossek et al., 2017). It is possible to have different combinations of NTD and CTD types (Hutchinson et al., 2003), and these appear to have been interchangeably mixed during VSG repertoire evolution (Cross, 1984). During segmental *VSG* gene conversion, the NTD of a previously expressed *VSG* can be replaced by the NTD encoding region of a different *VSG* to form a novel *VSG* (Hall et al., 2013). We investigated whether VSG chimeras, generated by fusing different domains of presumably smaller VSGs of *T. vivax* and those of *T. brucei*, would be expressed and form a coat on *T. brucei* parasites.

The full length and truncated *T. brucei* CTDs were used to generate chimeric VSGs of different sizes (Figure 27 A). First, the pLew82v4 inducible expression system was used to express a VSG chimera that consisted of ILDat1.2 VSG NTD and a truncated VSG221 CTD lacking the structured region. It has been shown that VSG221 lacking the structured CTD (referred to as a minimal linker) region can be expressed in *T. brucei* (Nicola Jones, in preparation). The ILDat1.2 secondary structure was predicted using Phyre2 web portal for protein modeling, prediction and analysis (Kelley et al., 2015). The region I assigned to be the NTD was amplified from plasmid EA36 using the forward primer D146 and reverse primer D159. Next, the *VSG221* minimal linker with the intact 3′UTR was amplified from plasmid pJet.M1.2ΔSnoGS using forward primer D160 and reverse primer D107. The products of the first PCR amplifications were hybridized in a second PCR using primer set D146 and D107 (section 3.1.3), followed by ligation into pLew82v4 vector (section 3.1.4) to generate construct pLew:ILDat1.2:221ML. The construct was NotI-linearized and transfected into *T. brucei* 13-90 cells yielding the 221^{ES}.ILDat1.2:221ML^{tet} cell line.

Analysis of cell growth after induction of the ectopic VSG expression for five days showed that the cells doubled at 7.2 hours (Figure 27 B). Protein expression analysis by SDS-PAGE after 24 hours of inducing expression showed that the chimeric VSG was expressed (Figure 27 C). Next, the *ILDat1.2:221ML* chimera was excised from plasmid pBSK.ILDat1.2:221ML with EcoRI and ligated into EcoRI-linearized pKD4 plasmid to generate the pKD4.ILDat1.2:221ML construct for constitutive expression.

A second VSG chimera, consisting of the predicted ILDat1.2 NTD and a complete VSG221 CTD, was also generated by PCR-driven overlap extension (section 3.1.3). The predicted *ILDat1.2* NTD was PCR amplified as above, and the *VSG221* CTD was amplified from the plasmid pBSK.VSG221wt using forward primer D160 and reverse primer D107. Primer pair D146 and D107 was used to join the two fragments in a subsequent PCR, followed by ligation into EcoRI-linearized pKD4 to generate pKD4.ILDat1.2:221CTD construct. The

pKD4.ILDat1.2:221ML and pKD4.ILDat1.2:221CTD constructs were linearized using BglII and transfected into the parental *T. brucei* cells to generate DEX 221^{ES}.ILD:ML and 221^{ES}.ILD:CTD cell lines, respectively.

Analysis of growth profiles for five days showed that 221^{ES}.ILD:ML and 221^{ES}.ILD:CTD cell lines had a PTD of 6.7 and 7.0 hours, respectively (chapter 6.1 Figure 36 A). Analyses of protein expression by SDS-PAGE showed that both chimeric VSGs were constitutively expressed (chapter 6.1 Figure 36 B). Transfecting the VSG221 knockout construct into 221^{ES}:ILD:ML and 221^{ES}:ILD:CTD cell lines generated the 221^{KO}.ILD:ML and 221^{KO}.ILD:CTD cell lines, respectively. Western blot analysis with a VSG221 antibody confirmed that the endogenous VSG221 was not expressed while the chimeric VSGs were constitutively expressed (Figure 27 D, E). Growth analysis showed that the 221^{KO}.ILD:ML cell line had a slightly slower PDT of 10.6 hours compared to 9.6 hours of the 221^{KO}.ILD:CTD cell line (Figure 27 F, G, H).

Together, these data show that VSG chimeras consisting of ILDat1.2 and VSG221 domains, can form a coat on *T. brucei*. The 221^{KO}.ILD:ML cell line, expressing the smallest VSG chimera, had the slowest growth compared to the 221^{KO}.ILD:CTD and the 221^{KO}.ILDat1.2 cell lines, respectively.

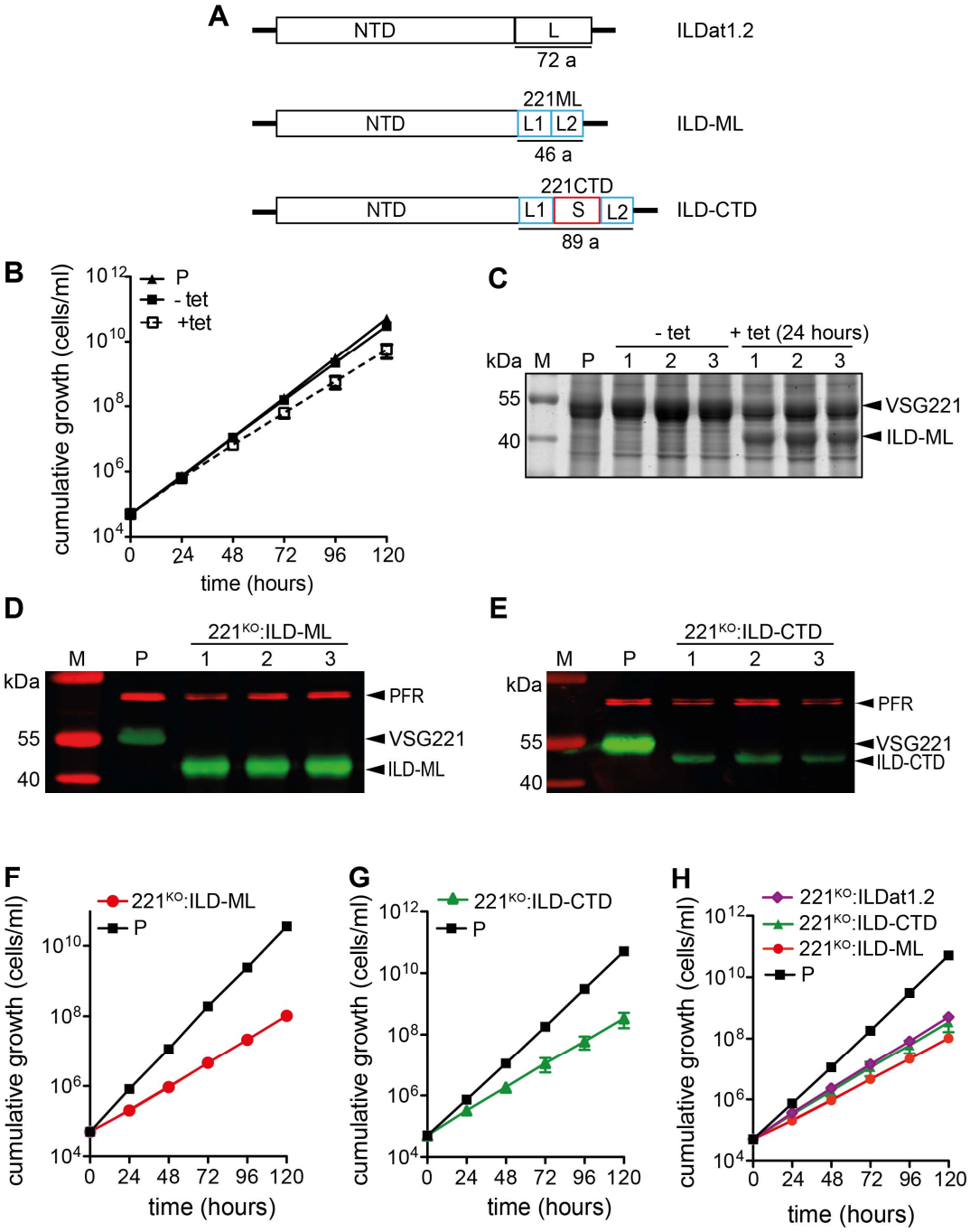

Figure 27: *T. vivax* ILDat1.2 and *T. brucei VSG221* chimeras form a coat on *T. brucei*.
(A) Schematic of ILDat1.2 VSG and ILDat1.2-VSG221 chimeras. The NTD is ILDat1.2 in all cases, and the CTD is either a truncated minimal linker (ML) or the complete VSG221 C-terminal domain (CTD). **(B)** Cumulative growth after induction of ectopic VSG overexpression for 5 days in the 221^{ES}.ILDat1.2:221ML^{tet} cell line. The parental 13-90 cell line (P) served as a control. **(C)** SDS-PAGE gel showing ectopic VSG expression of the 221^{ES}.ILDat1.2:221ML^{tet} cell line after induction of overexpression for 24 hours (+tet). **(D, E)** Western blot showing endogenous VSG221 knockout and chimeric VSG expression of the 221^{KO}.ILD:ML and the 221^{KO}:ILD-CTD cell line, respectively, by western blotting. **(F, G)** Cumulative growth of the 221^{KO}.ILD:ML and 221^{KO}:ILD-CTD cell lines, respectively. **(H)** Combined growth curves of 221^{KO}.ILD:ML and 221^{KO}.ILD:CTD cell lines. The parental 13-90 cell line served as a growth control, and three independent clones of each cell line were analyzed. Error bars represent standard deviation of the means.

3.3 Structural analysis of ILDat1.2 VSG using X-ray crystallography

BSF African trypanosomes have a dense VSG coat on their cell surface. The VSG protein, comprising approximately 95% of the BSF cell surface, is essential for protecting the parasites against the host immune system. Only *T. brucei* VSG proteins have been studied in detail, and even so, the atomic structure of just a handful of *T. brucei* VSGs has been solved. Here, I sought to solve the 3D structure of the *T. vivax* ILDat1.2 VSG protein. This VSG was selected because it is the most cited and was initially purified and characterized by Gardiner et al., 1987. Additionally, I was able to generate a transgenic *T. brucei* cell line expressing ILDat1.2 VSG (section 3.2.6).

3.3.1 Purification of soluble ILDat1.2 VSG

Numerous protein expression and purification strategies have been devised to enable the characterization of proteins. The choice of an expression and purification strategy is dependent on the characteristics of the protein and its intended application. In most cases, the protein cannot be obtained in sufficient quantities directly from the host organism necessitating the use of heterologous protein expression systems. Studies involving protein structure determination using X-ray crystallography require homogenous protein sample of very high purity (Dale et al., 2003). Approximately 5 million copies of VSG homodimers (constituting about 10% of total proteins) are expressed on the surface of BSF *T. brucei* parasites (Cross, 1975). This high level of VSG protein expression by individual cells and the possibility to cultivate cells to high densities *in vitro* enables direct purification of sufficient material for crystallization experiments (Bartossek et al., 2017). We generated a *T. brucei* based cell line expressing the *T. vivax* ILDat1.2 VSG (Section 3.2.6), thereby eliminating the need to use heterologous expression systems.

The parasites were grown to a density of ~3.5 x 10^6 cells/ml and the VSG was harvested from a total of ~4.8 x 10^9 cells. VSG was released from the membrane by osmotic lysis and activation of the glycosylphosphatidylinositol-phospholipase C (Hereld et al., 1986). The supernatant containing the VSG was purified by binding to an anion exchange HiTrap Q HP Sepharose column (GE Healthcare, UK) and subsequent elution by a linear gradient of 1M NaCl (30 – 40%) (Figure 28 A**)**. With a flow rate of 1 ml/min, the VSG was eluted after 20 – 22 min (peak 'a') corresponding to fractions 7 – 9 as confirmed by SDS-PAGE (Figure 28 B). Fractions containing the pure protein from multiple purifications were identified, pooled and buffer exchanged to water. This was necessary as high salt concentrations was shown reduce the stability of VSGs in solution (Bartossek, 2017). The protein was then concentrated to 15.3 mg/ml as determined using UV absorbance at 280 nm.

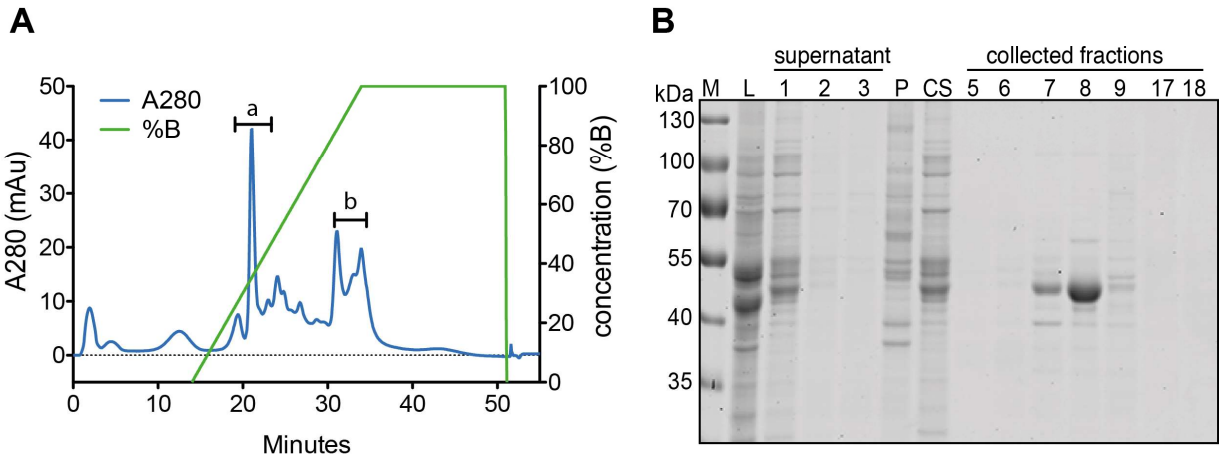

Figure 28: Purification of soluble ILDat1.2 VSG.
(A) Chromatogram of protein purification showing VSG protein elution from the anion-exchange column. The absorbance values (blue graph) are reflected on the left Y-axis while the NaCl elution gradient (green graph) is indicated on the right Y-axis. The run time indicated on the X-axis is at a constant flow rate of 1 ml/min. The peaks marked 'a' and 'b' correspond to collected fractions 7 – 9 (VSG) and 17 – 18, respectively. **(B)** Analysis of VSG extraction and purification by SDS-PAGE. M: molecular weight marker, L: cell lysate, 1 – 3: extracted supernatants, P: pellet, CS: combined supernatants, 5 – 18: eluted fractions. Most of the VSG was extracted from the parasites in the first extraction step.

3.3.2 Optimization of ILDat1.2 VSG protein crystallization using the sitting drop method

Five commercial screens (Protein complex, Index HT, Optimix PEG, Wizard I/II and JCSG+) were used for the initial crystallization trial screening using the sitting drop method at both 4 and 20 °C. For each screen, a 15.3 mg/ml sample and a 1:2 dilution (7.65 mg/ml) was used to set up the crystallization experiment. Diffraction-quality crystals grew after 8 days in the Protein Complex (well D1) and Wizard 1&2 (well C4) screens at 20 °C. The conditions consisted of 0.2 M sodium chloride and 0.1 M HEPES pH 7.5 buffer for the two wells, while the precipitant was 25% (w/v) polyethylene glycol 4000 (PEG 4K) and 20% (w/v) PEG 3K for protein complex and Wizard 1&2 screens, respectively. Fine screens were formulated (chapter 6.2 Table 13) to further improve crystal quality by titrating the conditions in which crystals, needles and quasi-crystals grew. Fresh crystallization experiment was then set up using the fine screen formulations, and crystals grew after 5 – 13 days in different conditions (Table 10).

Table 10: Successful fine screen conditions for the crystallization of ILDat1.2

Well position	Drop	Precipitant	Salt	Buffer
D1	1 and 3	25% w/v PEG 4000	0.2 M KCl	0.1 M MES pH 6.5
D2	1 and 3	23% w/v PEG 4000	0.2 M KCl	0.1 M MES pH 6.5
D3	1 and 3	20% w/v PEG 4000	0.2 M KCl	0.1 M MES pH 6.5
D4	1	18% w/v PEG 4000	0.2 M KCl	0.1 M MES pH 6.5
D6	1 and 3	25% w/v PEG 4000	0.1 M KCl	0.1 M MES pH 6.5
D8	1 and 3	20% w/v PEG 4000	0.1 M KCl	0.1 M MES pH 6.5
D11	1	23% w/v PEG 4000	0.2 M KCl	0.1 M MES pH 6.0
D12	1	20% w/v PEG 4000	0.2 M KCl	0.1 M MES pH 6.0
E4	1 and 3	23% w/v PEG 4000	0.1 M KCl	0.1 M MES pH 6.5
E5	1 and 3	20% w/v PEG 4000	0.1 M KCl	0.1 M MES pH 6.5
E6	1 and 3	18% w/v PEG 4000	0.1 M KCl	0.1 M MES pH 6.5

Drop 1 and 3 represent the concentrated (15.3 mg/ml) and diluted protein sample (7.65 mg/ml), respectively.

Crystals appeared after an equilibration time of one week on average, and the crystals of highest protein order and size ~(100 – 250 µm) grew in wells D2, D3, D11, D12 and E6. The drops with the highest protein concentrations (drop 1) gave larger crystals compared to the less concentrated sample (drop 3). In most cases the crystals had pyramidal, trigonal and tetragonal prism morphology.

3.3.3 Crystallization of ILDat1.2 VSG using the hanging drop method

The hanging drop vapor diffusion method was used to grow large diffraction-quality crystals. A homemade fine screen formulation based on the trial commercial screens was used in this crystallization experiment (Chapter 6.2 Table 14). Hanging drop vapor diffusion crystallization was set up at 20 °C using 1 µl of ILDat1.2 protein sample (15.3 mg/ml) mixed with 1 µl of the mother liquor against 1000 µl of reservoir (mother liquor). Molecular replacement does not appear to work well with VSGs. Therefore, in a separate experiment, cesium chloride was incorporated in the fine screen formulations to aid in solving the phase. This is a prerequisite for structure determination. Crystals were only obtained in the experimental set up without cesium chloride salt. The conditions that aided crystal growth are given in Table 11.

Table 11: Successful conditions for ILDat1.2 crystal growth by the hanging drop method

Well position	Precipitant	Salt	Buffer
A1	25% w/v PEG 4000	0.2 M KCl	0.1 M MES pH 6.5
A2	23% w/v PEG 4000	0.2 M KCl	0.1 M MES pH 6.5
A3	20% w/v PEG 4000	0.2 M KCl	0.1 M MES pH 6.5
A4	18% w/v PEG 4000	0.2 M KCl	0.1 M MES pH 6.5
B2	23% w/v PEG 4000	0.1 M KCl	0.1 M MES pH 6.5
B3	25% w/v PEG 3350	0.1 M KCl	0.1 M MES pH 6.5
B6	23% w/v PEG 4000	0.2 M NaCl	0.1 M HEPES pH 7.5
B9	23% w/v PEG 3350	0.2 M NaCl	0.1 M HEPES pH 7.5
B10	25% w/v PEG 4000	0.2 M NaCl	0.1 M HEPES pH 7.5

Crystal growth was observed after 7 – 13 days of incubation. All the conditions above produced diffraction quality crystals. The crystals were carefully extracted from the crystallization drops using CryoLoops and soaked for 5 – 60 seconds in the mother liquor solution supplemented with 100 mM 5-amino-2,4,6-triiodoisophthalic acid (I3C). The amino, carboxylate and iodide functional groups of I3C interact with the protein via hydrogen bonds. The three I atoms in I3C give strong anomalous signal that can be used for phasing (Beck et al., 2008). After soaking, the crystals were cryoprotected by transfer into reservoir solution augmented with 30% (v/v) glycerol and transferred to liquid nitrogen, for storage until data acquisition.

3.3.4 X-ray diffraction data acquisition and model building

The diffraction data was collected at the HZB synchrotron radiation source (Berlin, Germany) on beamline 14.1 and crystals diffracting to 2.2 Å resolution were used for phasing and structure determination. The data was processed using XDS and scaled with the Aimless suite. Location of the heavy atom substructure in the crystals and phase improvement were carried out with the Crank2 suite belonging to the CCP4 suite (Winn et al., 2011). Within Crank2, substructure solution was performed using ShelxC, and D (Sheldrick, 2010). Substructure phasing and refinement was done using REFMAC5 (Murshudov et al., 2011) followed by density modification with Parrot. Automated model building was done using buccaneer. The final model was generated by additional rounds of manual model building in Coot (Emsley et al., 2010) and automated refinement using REFMAC5. The final model showed good stereochemistry as indicated by the

Ramachandran statistics and refined to final R-values of 22.8 and 26.6 for R and Rfree, respectively.

Table 12: X-ray data collection and refinement statistics

Data collection	
Wavelength [Å]	1.7
Space group	I 222
Unit cell parameters a, b, c [Å] α, β, γ [°]	 46.6, 174.2, 204.2 90, 90, 90
Resolution range [Å]	66.3 – 2.2 (2.27 – 2.2)
Unique reflections	42883 (3649)
R_{merge} [%]	14.6 (341.1)
R_{pim} [%]	4.2 (95.6)
I/σI	14.6 (1.0)
CC (1/2)	0.999 (0.746)
Multiplicity	25.2 (26)
Completeness [%]	100 (100)
Refinement	
Resolution range [Å]	66.3 – 2.2
Unique reflections	40798
Number of atoms	5782
R_{work} /R_{free} [%]	22.8/26.6
Mean B-factor [$Å^2$]	38.7
RMS Deviations Bond lengths [Å] Bond angles [°]	 0.011 1.82
Ramachandran statistics [%]	Favored 98.1%; Allowed 1.8%; Outliers 0.1%

Values in parentheses correspond to highest resolution shell.

3.3.5 Crystal structure of the NTD of *T. vivax* ILDat1.2 VSG

The full length ILDat1.2 sVSG was used for crystallization. However, the crystal structure contained only 261 residues of the 356-residue full length peptide. This shows that ILDat1.2 VSG was cleaved during crystallization experiments, consistent with findings in previous studies. This truncation is presumably mediated by endogenous peptidases (Freymann et al., 1984) that cleave the flexible linker that connects the NTD to the CTD. Therefore, presumably, only the NTD of ILDat1.2 VSG was crystallized.

The NTD of the ILDat1.2 VSG molecule is a 9.35 nm long homodimer (Figure 29 A). The central structure of the monomer consists of a coiled-coil made up of two antiparallel 50 residue containing helices of approximately 7.8 nm in length (residues 6 – 56; 60 – 110). As the C-terminal of the ILDat1.2 molecule terminated midway through the central core, it is likely that part of the NTD of the ILDat1.2 molecule was also cleaved. The membrane distal head region comprises an elongated surface loop (residues 111 – 242) that extends midway through the length of the VSG molecule before returning to the head region. The variable surface loop is broken by three α-helices and four short antiparallel β-strands, and form a compact head structure that is further stabilized by two disulfide bridges formed between the conserved cysteine residues C13 and C144 and residues C122 and C191 (Figure 29 B). The loop region is followed by a long helix (residues 243 – 260) that travels downwards, almost half the length of the central coiled-coil. A single oligosaccharide is linked to residue N59 at the membrane proximal end of the coiled-coil. All protein structure representations in this thesis section were generated in UCSF Chimera (Pettersen et al., 2004).

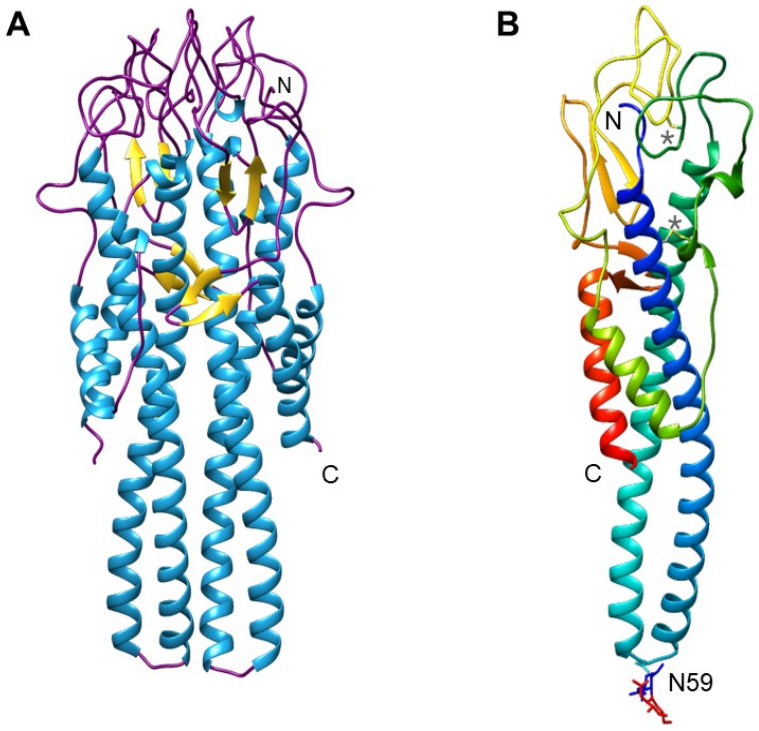

Figure 29: Structure of the ILDat1.2 NTD.
(A) A ribbon diagram of the structure of ILDat1.2 NTD dimer. The coiled-coil forming the central core of the VSG and the other α-helices are shown in light blue, the β-sheets in yellow, and the connecting loops are in purple. The N-glycan is not shown. **(B)** Structure of the ILDat1.2 monomer in rainbow color scheme from the N- (blue) to C-terminal ends (red). The disulfide bonds formed between the two cysteine pairs are marked with asterisks. The single N-glycan is attached to residue N59 at the base of the coiled-coil.

3.3.6 Comparison of the NTD of ILDat1.2 with those of *T. brucei* VSGs

The canonical *T. brucei* VSG structure was considered a homodimer that consists of an elongated NTD and a more secluded CTD attached to the cell membrane by a GPI anchor. However, recent studies have uncovered VSG structures that deviate from the norm: B-type VSG3 (MITat1.3) exists in solution and the crystal asymmetric unit as a monomer rather than a dimer, and a single *O*-linked α-glucose moiety is present on the top surface (Pinger et al., 2018). On the other hand, the recently reported A-type VSGsur and MITat1.13 homodimers have an additional "upper lobe" extension on top of the head region (Zeelen et al., 2021), making them slightly taller. A comparison of the ILDat1.2 VSG structure with the A-type MITat1.2 and ILTat1.24 (Blum et al., 1993; Freymann et al., 1990), and B-type MITat1.3 *T. brucei* VSGs showed that the ILDat1.2 architecture was more similar to the A-type VSGs, with minor structural differences (Figure 30).

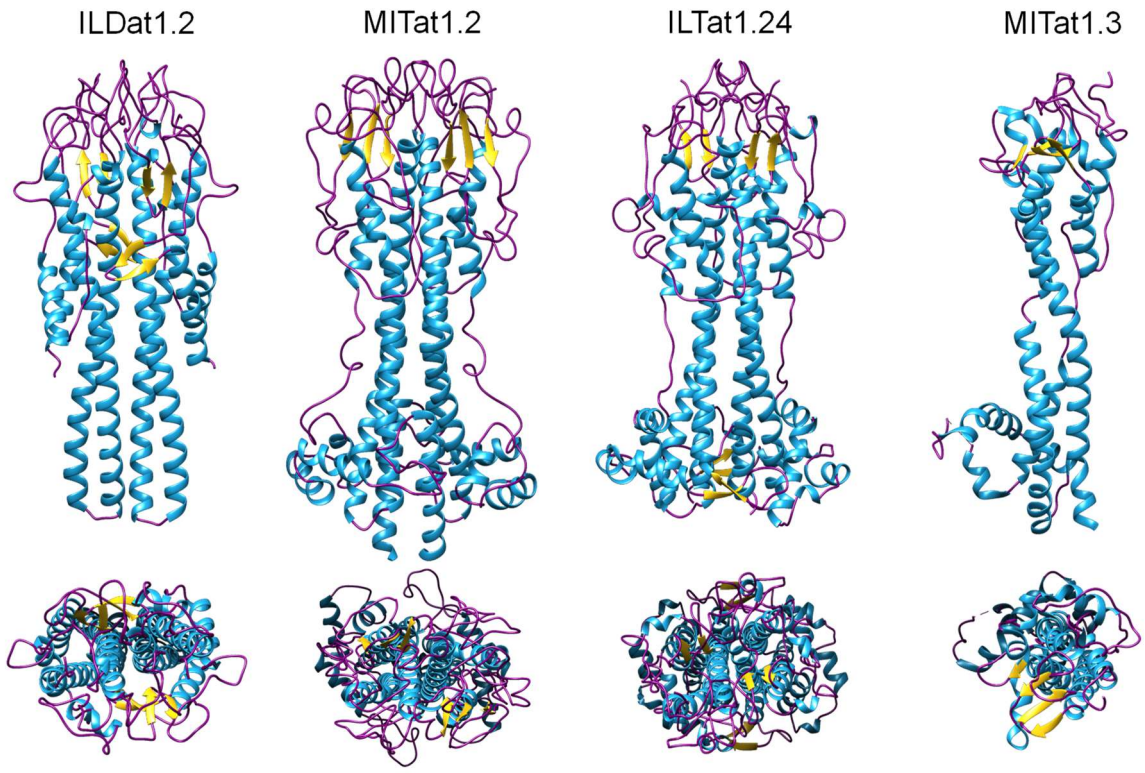

Figure 30: The structure of the ILDat1.2 NTD is similar to those of A-type *T. brucei* VSGs.
The structure of ILDat1.2 VSG N-terminal domain in comparison to those of A-type (MITat1.2; PDB: 1VSG and ILTat1.24; PDB: 2VSG) and B-type (MITat1.3; PDB: 6ELC) *T. brucei* VSGs. The α-helices are marked light blue, β-sheets yellow, and the loops in purple. A-type MITat1.1 VSG was omitted in the comparison as it is highly similar to MITat1.2 (Bartossek et al., 2017).

Noteworthy is the short α-helix that extends just below the coiled-coil kink in each monomer. These helices extend farther downwards to create a wider ILDat1.2 NTD trunk (between the head and the hip regions) compared to the other VSGs. Additionally, the antiparallel β-sheets appear to be buried deep in the head region of ILDat1.2 VSG. The surface hydrophobicity of ILDat1.2 NTD was compared with the other VSGs to determine whether there were any striking variations. The hydrophobicities were based on the Kyte-Doolittle scale (Kyte & Doolittle, 1982) and visualized with UCSF Chimera (Pettersen et al., 2004). The monomers of one A-type (MITat1.2) and the B-type MITat1.3 VSG were selected for analysis. Hydrophilic residues make the bulk of the VSG surface (Figure 31 upper panel) whereas the dimerization interface of the monomers consists of mostly hydrophobic residues (Figure 31 lower panel).

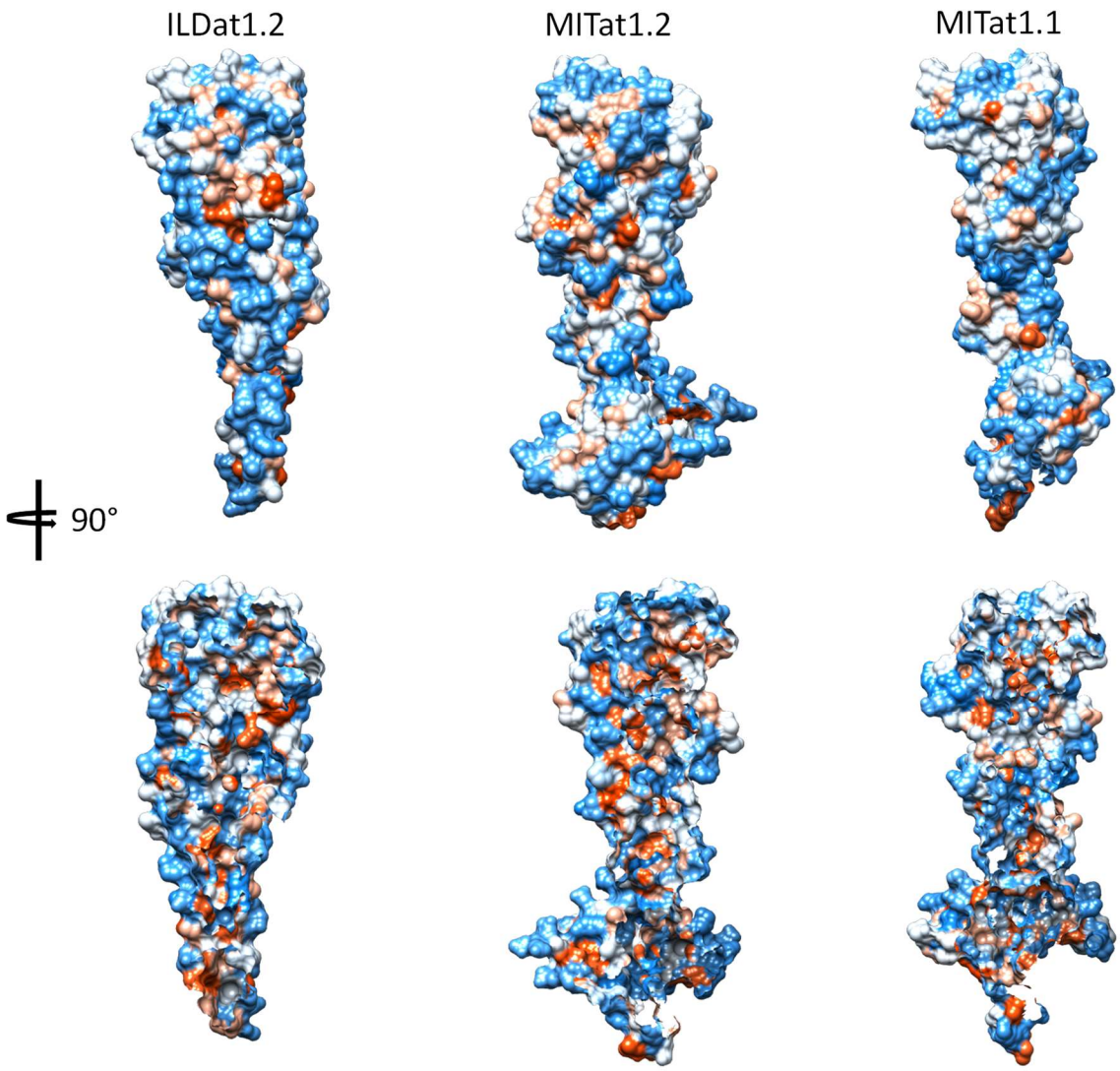

Figure 31: Comparison of surface hydrophobicity of VSG NTD monomers.
Depiction of the VSG monomers in two orientations related by a rotation of 90° along the vertical axis. The frontal view (lower panel) shows the dimerization interface. The amino acid residues are shown by color coding; hydrophilic residues are highlighted in blue while hydrophobic residues are in red. The stronger the hydrophobicity/hydrophilicity, the stronger the color. The surfaces of the VSGs have a high distribution of hydrophilic residues (upper panel) while the dimerization interface shows a strong distribution of hydrophobic residues to the rest of the molecule (lower panel).

The architecture of the solved ILDat1.2 VSG was more similar to the A-type *T. brucei* VSGs (Figure 30). The ILDat1.2 NTD consists of only 261 residues, however it is likely that some portion of the NTD was cleaved during the crystallization process. The NTD of MITat1.2 VSG consists of 362 residues, 101 more than the ILDat1.2 NTD. Therefore, it is unlikely that the cleaved portion of the ILDat1.2 VSG would be sufficient to form structural elements at the base of the VSG molecule, in addition to the linker required for GPI anchoring as observed in the A-type *T. brucei* VSGs. Given that most characterized trypanosome surface proteins share the conserved 3-helical bundle structure, and no structure of a *T. vivax* surface protein has been reported, ILDat1.2 was compared to the VSG-like *T. brucei* protein TFR

(PDB: 6SOY) and HpHbR (4E40) from *T. congolense*. The ILDat1.2 NTD was structurally more similar to the *T. brucei* TFR protein. However, the TcHpHbR markedly differs from the ILDat1.2 NTD: it is longer than ILDat1.2 NTD and the TbTFR and consists of short loops connecting the three helices in the head region. Additionally, TcHpHbR is free of β-sheets (Figure 32).

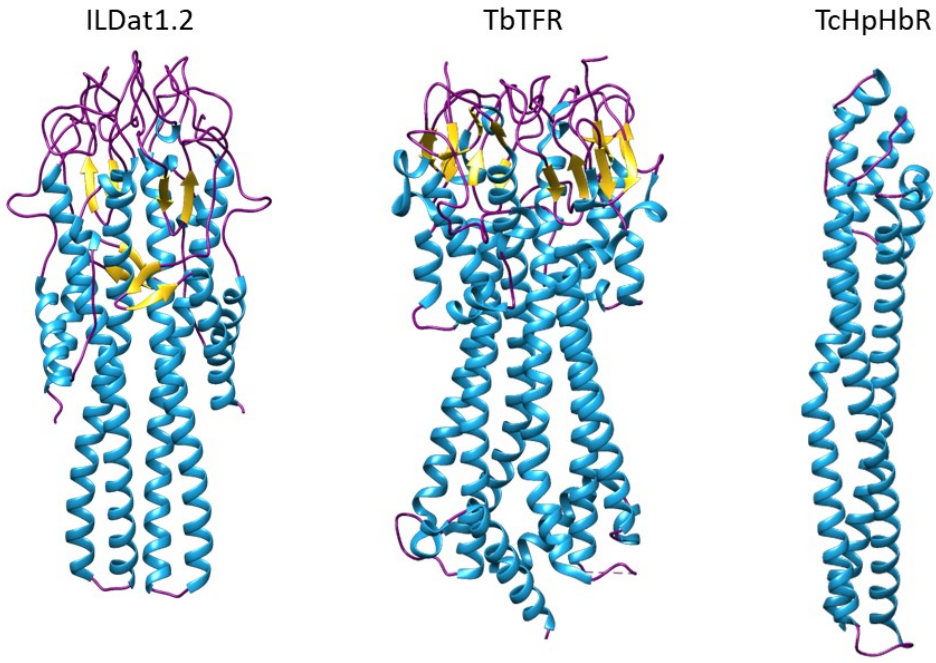

Figure 32: Structural comparison of ILDat1.2 with known non-VSG trypanosome surface proteins. The helices are shown in light blue, β-sheets in yellow, and the loops regions in purple All the surface proteins have the conserved three-helical bundle structure. However, whereas the third helix in the TbTFR and TcHpHbR proteins are complete and almost equal in length to the two helices making the coiled-coil, the third helix in ILDat1.2 is shorter. The NTDs of ILDat1.2 and TFR are approximately equal in length (9.35 nm), whereas the TcHpHbR is ~1.8 nm longer.

Distance-based similarities were measured in UCSF Chimera by the superimposition of ILDat1.2 NTD to MITat1.2, MITat1.3 and TFR structures over 42, 19 and 27 pruned pairs, respectively. The RMSD for ILDat1.2 NTD structure to MITat1.2, MITat1.3 and the TFR was 1.192, 0.885 and 1.204 Å, respectively. These data indicate that the structures are highly dissimilar since low RMSD values were only obtained when a few residue pairs were considered. The sequence identity of ILDat1.2 to MITat1.2, MITat1.3 and TFR was calculated to be 20, 18 and 20%, respectively, using Clustal Omega (Sievers et al., 2011) in UCSF Chimera. Further structural comparison was done by the superimposition of ILDat1.2 atomic coordinates with the MITat1.2, MITat1.3, and TFR coordinates in UCSF Chimera using the Needleman-Wunsch algorithm (Needleman & Wunsch, 1970) and BLOSUM-62

Matrix (Henikoff & Henikoff, 1992) for alignment. The pruning distance was set at 2 Å (Figure 33).

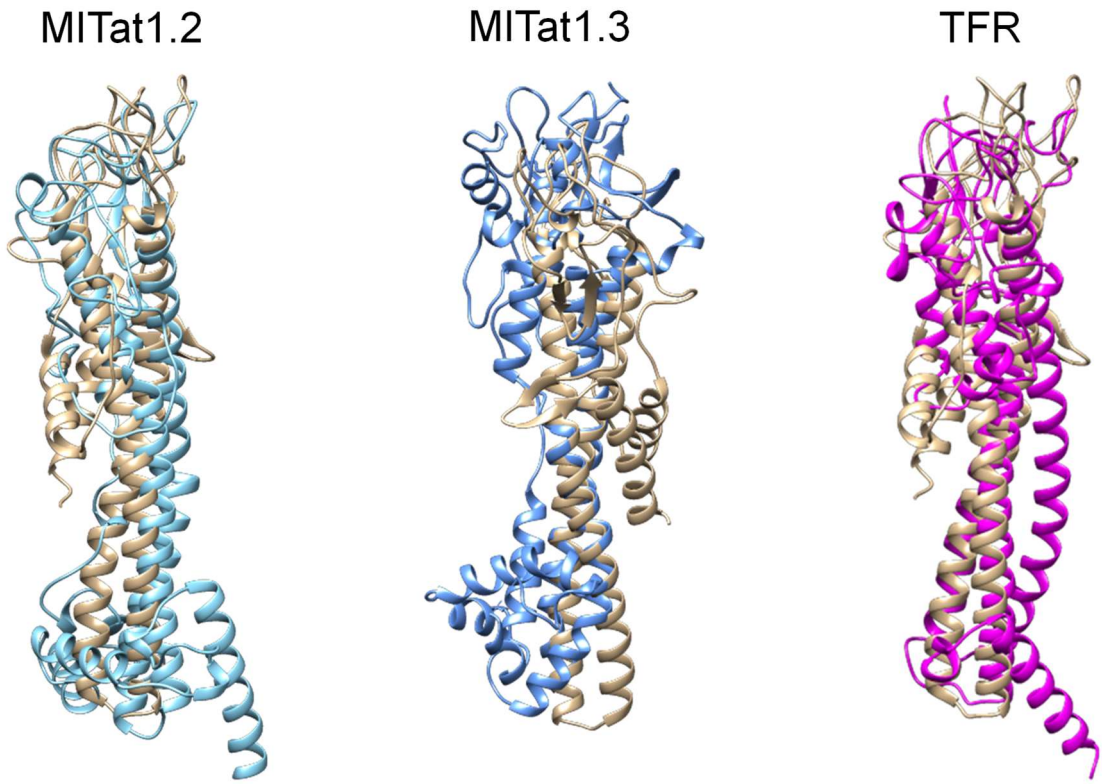

Figure 33: Superposition of ILDat1.2 VSG and selected *T. brucei* surface proteins.
MITat1.2 (cyan), MITat1.3 (blue) and TFR (magenta) are superimposed on the experimental ILDat1.2 structure (brown). Whereas the 3-helix bundle scaffolding is conserved in all the surface proteins, there are variations in other structured and non-structured regions.

4.0 Discussion

4.1 Post-transcriptional *trans*-regulation of *VSG* mRNA expression

The trypanosome VSG protein is essential for the evasion of host immune factors. Monoallelic transcription of the *VSG* gene from a single polymerase I transcribed BES ensures that only a single VSG antigen is expressed at a time from a vast repertoire of silent genes (Borst, 2002). Under natural conditions, trypanosomes have to handle two VSG antigens on the cell surface during switching, which occurs by recombination events or transcriptional switching of the ES (Horn, 2014). It has been shown that trypanosomes can be engineered to constitutively express two VSGs from the same BES simultaneously (Muñoz-Jordán et al., 1996). In these parasites, the ectopic VSG was integrated upstream of the BES-resident VSG, and the mRNA and protein of the two VSGs were expressed at approximately equal amounts (Muñoz-Jordán et al., 1996). A similar balancing of the VSG transcript and protein was observed when an ectopic *VSG* gene was inserted downstream of the BES promoter, approximately 60kb from the telomere (Smith et al., 2009). We also know that the inducible overexpression of a second VSG from the rDNA spacer leads to a downregulation of the endogenous *VSG* mRNA, with the total *VSG* mRNA population not exceeding wild type levels (Batram et al., 2014; Zimmermann et al., 2017). These studies imply that a robust mechanism that keeps the *VSG* mRNA at optimum levels is operational in *T. brucei.* Batram proposed that the 16-mer sequence in the VSG 3′UTR is perceived and "counted" to regulate the *VSG* transcript levels (Batram, 2013). It was further suggested that an RBP that binds to the 16-mer could be the limiting factor regulating the *VSG* transcript level (Ridewood et al., 2017). Recently, an F-box RBP, CFB2, was shown to interact with the VSG 16-mer motif to stabilize the *VSG* transcript (do Nascimento et al., 2020). Thus, CFB2 is proposed to be the limiting factor that regulates the *VSG* mRNA levels. Recently, a study that involved the introduction of premature termination codons in the ectopic VSG genes, reported that regulation of the *VSG* mRNA levels is dependent on the production of functional VSG proteins.

This study explored whether the balancing of *VSG* mRNA requires *T. brucei* VSGs and hence if it is species-specific. To this end, two *T. vivax* VSGs were overexpressed in a background of *T. brucei* cells expressing VSG221 from the BES. *T. vivax* and the related *T. congolense* parasites also undergo antigenic variation of the VSG coats as is the case in *T. brucei*. However, it is suggested that their VSG proteins may be divergent. Inducing the expression of *T. vivax ILDat1.2* and *ILDat2.1 VSG* reporters resulted in distinct growth

phenotypes; Whereas the *ILDat1.2* overexpression cell line stalled in growth after inducing expression, the *ILDat2.1* overexpressors had no growth defect. Consistent with previous data (Batram et al., 2014), inducing the expression of *ILDat1.2 VSG* caused a rapid reduction in the endogenous *VSG221* mRNA. However, the high amounts of *ILDat2.1 VSG* did not lead to the downregulation of the endogenous *VSG221* transcript. This was surprising as a high-level expression of ectopic *VSG* not resulting in *VSG* mRNA balancing had not been observed. Therefore, ILDat2.1 VSG might have lacked an essential feature present in the other VSGs. As the ILDat2.1 protein was not translated, the first candidate to be investigated was the ER import signal. Translation and posttranslational modifications of VSGs occur in the ER, and as such, ER import signals are essential in the biosynthesis of VSGs. No experimental data has confirmed whether BSF trypanosomes rely on the SRP-dependent or independent pathways for translocating VSGs into the ER (Manna et al., 2014). However, it has been shown that procyclic *T. brucei* parasites and *S. cerevisiae* seems to rely on the SRP-independent pathway for translocation of GPI anchored proteins, whereas the SRP-dependent one is utilized for translocation of polytopic membrane proteins (Ast et al., 2013). Given that the same cytosolic factors involved in this process are conserved in *S. cerevisiae* and *T. brucei,* it is likely that they are functionally similar and could be essential in the regulation of *VSG* mRNA translocation. Different studies have arrived at contrasting conclusions regarding whether VSG protein is produced in just sufficient amounts or more than needed (Field et al., 2010; Manna et al., 2014; Tiengwe et al., 2016).

Overexpression of ILDat2.1 VSG reporter in which the native ER import signal was replaced with a bona fide *T. brucei* VSG ER import signal resulted in a rapid decrease in the endogenous *VSG* mRNA. This experiment led to three crucial insights: (i) that the mRNA balancing is not specific to *T. brucei* VSGs, (ii) mRNA balancing was independent of translation since ILDat2.1 protein was still not produced even with a functional ER import signal, and (iii) for the mRNA transcripts to be balanced, they need to be targeted to the ER. By inducibly expressing ER-targeted and cytosolic GFP reporters coupled to the VSG 3′UTR from the rDNA spacer, Batram showed that expression of high levels of cytosolic GFP reporter did not affect the *VSG* mRNA levels (Batram, 2013). However, when the GFP reporter was targeted to the ER by adding the EP1 ER import signal, the endogenous *VSG* mRNA was balanced. Due to cell death occasioned possibly by high-level production of the GFP reporter by the overexpression strategy used, a GFP reporter containing the *GFP* ORF and a *VSG* 3′UTR was constitutively expressed by integration downstream of the active VSG221 BES. This reporter cell line expressed high but non-toxic levels of GFP, and the

cells had no growth impairment. Consistent with the results of the GFP reporter overexpression experiments, the *VSG* mRNA balancing response was only elicited when the GFP reporter was targeted to the ER. These findings underpin the requirement of ER-targeting for the regulation of *VSG* mRNA amounts. Secondly, the data shows that the *VSG* mRNA regulation response is neither dependent on the VSG ORF nor the VSG ER import signal. Therefore, it appears that expression of high levels of ER-targeted non-VSG genes can elicit the *VSG* mRNA regulation response and that this regulation operates in *trans*.

The role of the conserved 16-mer motif in the regulation of *VSG* mRNA abundance is well documented (Berberof et al., 1995; Ridewood et al., 2017). An F-Box mRNA binding protein CFB2, which recruits a stabilization complex and the cap-binding translation initiation complex, has now been identified as the RBP associating with the 16-mer (do Nascimento et al., 2020). Thus, CFB2 has been proposed to be the limiting post-transcriptional factor that regulates the total *VSG* mRNA amounts in a cell. However, the data presented in this thesis suggests otherwise; it was possible to express wild type levels of ectopic *VSG121* and *ILDat2.1 VSG* reporters containing an apparently dysfunctional ILDat2.1 VSG ER import signal and the complete 3′UTR, which included the intact 16-mer motif, of a *T. brucei* VSG. The high levels of expression of these ectopic *VSG* reporters did not affect the endogenous *VSG* transcript levels. Therefore, the wild type *VSG* mRNA levels were highly exceeded in these cell lines. These findings imply that, in principle, the available pool of CFB2 protein is sufficient to stabilize *VSG* mRNAs above the wild type levels. This is consistent with a recent report that *T. brucei* has the capacity for the expression of *VSG* mRNA above the wild-type levels (Maudlin et al., 2021). Thus, though CFB2 is essential for stabilizing the *VSG* mRNA, it appears not to be the limiting factor regulating the steady-state *VSG* mRNA amount in the cell.

Maudlin et al. propose a model in which a pathway functions to regulate *VSG* mRNA in response to the folding efficiency of VSG protein. When a small percentage of the VSG protein folds efficiently, a large amount of the mRNA is produced, and the reverse occurs when a large amount of the VSG protein folds correctly. In situations where two VSGs with different folding efficiencies are expressed, the mRNA of the VSG with the highest folding efficiency is produced at low levels. Further, by introducing PTCs in the ectopic *VSG* ORF, it is shown that the endogenous VSG was only regulated if the ectopic VSG was GPI anchored (Maudlin et al., 2021). This data is inconsistent with the experiments conducted in this thesis, as it implies that balancing of the VSG transcripts is triggered at a point

downstream of ER targeting. However, Maudlin et al. expressed the ectopic VSG221 following integration upstream of the actively expressed VSG in the VSG121 BES, .used the ectopic VSG221 from the VSG121 BES, which led to the expression of different amounts of the endogenous and ectopic *VSG* transcripts. Additionally, the introduction of PTCs caused an overall increase in VSG transcripts above the wild type levels.

In summary, it appears that *VSG* mRNA balancing is not dependent on the *VSG* ORF and its ER import signal but is dependent on ER targeting of highly expressed genes. Given that the regulation of gene expression in trypanosomes is post-transcriptional, it makes sense for the parasites to have robust mechanisms to regulate the levels of the abundant surface protein transcripts. For example, mRNA balancing could be essential during antigenic exchange when the parasites transiently express two VSG antigens. The balancing of *VSG* transcripts can be likened to transcript buffering, a phenomenon that has been reported in yeast and mammalian cells (Abernathy et al., 2015; Haimovich et al., 2013; Singh et al., 2019; Sun et al., 2012). Transcript buffering involves crosstalk between transcription in the nucleus, and mRNA decay in the cytoplasm that ensures the maintenance of steady-state mRNA levels in a cell (Pérez-Ortín et al., 2012; Shalem et al., 2008). Transcription initiation factors, such as TFIID and TFIIH, and decay factors such as Xrn1 and Ccr4-NOT are involved in this transcript buffering (Timmers & Tora, 2018). However, the signal for crosstalk between transcription and mRNA decay and its directionality is not well known. As gene expression is regulated post-transcriptionally in trypanosomes (Clayton, 2002), the mechanistic aspects of transcript buffering are likely different from yeast and mammals. It is reported that complete buffering takes around 45 min and that models with mRNA half-lives exceeding the 12 min of *S. cerevisiae* mRNAs should be used to study transcript buffering (Timmers & Tora, 2018). Therefore, the trypanosome system with the dominant *VSG* mRNA species of a longer half-life (~2 hours) provides a versatile model for studying transcript buffering.

4.2 ILDat1.2 VSG expression-induced growth-arrest and stumpy development

Two distinct growth phenotypes were observed after inducing the expression of wild type *T. vivax* ILDat1.2 and ILDat2.1 VSG reporters in *T. brucei*: the ILDat1.2 overexpression cell line (221^{ES}.ILDat1.2^{tet}) ceased to proliferate whereas the ILDat2.1 cell line (221^{ES}.ILDat1.2^{tet}) proliferated normally. In the 221^{ES}.ILDat1.2^{tet} cell line, growth slowed down significantly within 12 hours of induction of ILDat1.2 VSG overexpression, followed by a complete cessation of growth. Cell cycle analysis showed a dual growth arrest, characterized by a stagnation of the parasites at the G1-phase and accumulation of precytokinesis parasites. It appears that the parasites initiated this precytokinesis cell cycle

arrest to avoid further dilution of the VSG coat since the ectopic VSG protein was not sufficiently expressed, and at the same time the endogenous *VSG221* mRNA was significantly reduced. This data agrees with a previous study which showed that ablation of the endogenous *VSG* transcripts by RNAi causes a precise precytokinesis arrest (Sheader et al., 2005). Until recently, it was unclear what aspect of the VSG (mRNA or protein) triggered this precise cell cycle checkpoint. However, a study by Ooi et al. suggests that it is levels of VSG protein or VSG protein synthesis that is "sensed" to activate the checkpoint response rather than the *VSG* mRNA; blocking VSG synthesis using Morpholino antisense oligonucleotides, that prevent translation without affecting the *VSG* mRNA level, led to the precise precytokinesis arrest (Ooi et al., 2018). In a separate study, ablation of the *GPI8* gene that encodes the catalytic subunit of the enzyme GPI transamidase — required for proper GPI-anchoring of proteins including VSGs — also causes a precytokinesis cell cycle arrest (Lillico et al., 2003). However, contrary to the precise *VSG* RNAi-induced precytokinesis arrest, *GPI8* RNAi-induced arrest was not tightly regulated as the cells re-initiated the S-phase with several rounds of mitosis resulting in multi-nucleated 'monster' cells (Lillico et al., 2003). This is then, in actual sense, not a precytokinesis cell cycle arrest. Thus, the precytokinesis cell cycle arrest observed after the induction of ILDat1.2 expression is similar to the *VSG* RNAi-induced precytokinesis phenotype.

Since the ILDat1.2 expressing parasites were arrested at the precytokinesis stage of the cell cycle, a decrease in the G1-phase pool was expected as the cells would go through the cell cycle and before arresting at the precytokinesis stage. However, the G1-phase pool appeared to remain constant (~55% of the population) in the course of ILDat1.2 expression. It was previously shown that overexpression of VSG118 in a background of VSG221 expressing monomorphic *T. brucei* cells caused a stalling in growth for approximately 5 days, after which normal proliferation was restored in some clones while some just grew slightly slower than the uninduced cells. In both cases VSG118 was produced and the endogenous VSG221 was downregulated (Henning, 2012). It was later reported that this growth phenotype was due to a prolongation of the G1-phase and accumulation of cells at this stage of the cell cycle (Specht, 2013). Similar growth phenotypes were observed when VSG121 was overexpressed in the same way in monomorphic parasites (Batram et al., 2014). Further, overexpression of VSG121 and VSG118 in pleomorphic parasites endogenously expressing Antat1.1 VSG (Zimmermann et al., 2017) corroborated the previous findings. However, this growth phenotype is not observed for the expression of all VSGs (Henning, 2012) indicating that this phenotype may be dependent on the VSG open reading frame.

A characteristic feature of the cells with a prolongated G1-phase was the VSG121 overexpression-induced BES-attenuation (Batram et al., 2014; Zimmermann et al., 2017). It is known that for a new VSG to be expressed, the BES-resident VSG must be replaced either by recombination events or the activation of a new BES. Two mechanisms control the switching of BES: the active BES can be silenced, followed by activation of the second BES (Aresta-Branco et al., 2016; Borst & Ulbert, 2001) or a second BES can be activated followed by the inactivation of the first BES (Batram et al., 2014; Zimmermann et al., 2017). By integrating a cytosolic *GFP* reporter downstream of the VSG221 BES promoter or just upstream of VSG221 in the active expression site, it was shown that the BES was directionally attenuated in the cells displaying a VSG overexpression-induced G1-phase prolongation, starting from the telomere end.

Integration of the GFP reporter in the active VSG221 promoter proximal region of the ILDat1.2 overexpression cell line, and subsequent quantification of the GFP intensity after 24, 48 and 72 h showed that the GFP intensity was unaffected. This was an indication that the activity of the BES was maintained at these timepoints. However, when the GFP reporter was expressed in a similar manner as above and VSG121 overexpressed, there was a rapid decrease in the GFP transcripts at earlier timepoints after induction, implying that attenuation of the active BES was initiated earlier on (Batram et al., 2014). Thus, analyzing the activity of the BES 24 h after inducing ectopic ILDat1.2 VSG expression could possibly not capture such early events. It is reported that the transient BES attenuation and the prolongation of the G1-phase potentially give the parasite a time window to probe for the functionality of the new BES. If the new BES is faulty, then the old BES is reactivated (Batram et al., 2014). Therefore, it is possible that inducing the expression of ILDat1.2 VSG caused attenuation of the BES, which was lifted since enough of the ectopic ILDat1.2 protein could not be produced. Thus, the activity of the BES appeared to be maintained.

Another feature of the G0/G1-phase prolongation is the activation of developmental competence marked with the expression of PAD1— a phenomenon observed in both monomorphic and pleomorphic cell lines (Batram et al., 2014; Zimmermann et al., 2017). These findings highlighted the existence of an alternative pathway of slender BSF differentiation to the stumpy form, distinct from the canonical SIF-dependent stumpy development pathway (Rojas et al., 2019; Vassella et al., 1997). To investigate whether the ILDat1.2 VSG overexpression derived G1-phase arrested cells acquired stumpy characteristics, an ILDat1.2 overexpression cell line which expressed the GFP:PAD1$_{UTR}$ reporter was generated. Induction of ILDat1.2 VSG expression caused the expression of

PAD1 in most of the cells. Further analysis of the cell cycle showed that almost 70% of the G1-phase arrested cells expressed PAD1 within 48 hours. Unexpectedly, more than half of the precytokinesis arrested cells also expressed PAD1. Since PAD1 is stumpy stage-specific, possibly the observed PAD1 positive precytokinesis arrested cells arise from a sub-population of G1-phase-arrested 'escapees' that were already committed to going through the cell cycle when transcription of the PAD1 was activated. Indeed, MacGregor et al. reported that the PAD1 mRNA is normally detected first in the intermediate cells (MacGregor et al., 2011), while the protein is expressed only expressed in fully differentiated stumpy forms (Dean et al., 2009). Therefore, the overexpression of a *T. vivax* VSG (ILDat1.2) in *T. brucei* can be interpreted as an unsuccessful switching event that has the potential to trigger the developmental transition from long slender BSF to short stumpy forms in monomorphic parasites. However, further characterization of the ILDat1.2 overexpression cell line is required to ascertain whether other features of stumpy parasites are developed upon the induction of ILDat1.2 VSG overexpression.

Batram et al. showed that the PAD1 expressing G1-phase arrested cells responded to treatment with 6 mM *cis*-aconitate and expressed the EP1 procyclin on their surface, indicating that the PAD1 was fully functional as a transporter on the cell surface (Batram et al., 2014). It would be worthwhile to determine whether the PAD1 expressed on induction of ILDat1.2 expression was a functional transporter and if treatment of these cells with *cis*-aconitate would trigger the replacement of the VSG coat with EP1. Other features of stumpy development that could be further investigated include the upregulation of mitochondrial lipoamide dehydrogenase (LipDH) protein levels (Tyler et al., 1997), and presence of a branched mitochondrion, which is an adaptation to metabolic requirements once the stumpy parasites are taken up by the tsetse fly (Brown et al., 1973).

4.3 *T. vivax* VSG GPI signals are not recognized in *T. brucei*

GPI anchored surface proteins are widely distributed in eukaryotes (Paulick & Bertozzi, 2008). Proteins of different sizes, ranging from just a few amino acids to 200 KDa, for example receptors, enzymes, adhesion proteins and complement activators can be GPI anchored (Kinoshita, 2020). The cell surface proteins of most kinetoplastids, including the African trypanosomes, are GPI-anchored. The cell surface of BSF *T. brucei, T. congolense* and *T. vivax* is mainly covered by GPI-anchored VSGs and a small percentage of GPI anchored receptors and invariant surface proteins of known and unknown functions. The *VSG* genes of African trypanosomes are inherited from the same ancestral repertoire with the A- and B-type VSG lineages being inherited by *T. brucei* and *T. vivax* while only the

B-type VSG lineage is present in *T. congolense*. Additionally, *T. vivax* appears to have inherited additional surface protein lineages absent in *T. brucei* and *T. congolense* (Jackson et al., 2013).

Proteins that are destined to be GPI anchored contain GPI attachment signal peptides at their C-terminal ends. Eukaryotic GPI signals are less conserved, but the general structure consists of small amino acids the ω position, where the preassembled GPI anchor is attached, followed by 2 – 3 additional small amino acids. The mid region is made up of a polar core of between 5 – 10 residues, which is followed by a hydrophobic core of 15 – 20 residues (Ferguson et al., 2015). *T. brucei* VSG GPI signals conform to the general structure of eukaryotic GPI attachment signal peptides (Böhme & Cross, 2002). Additionally, it has been shown that the serine and lysine residues at the ω+2 and ω+7 positions, respectively, are conserved in *T. brucei* VSG GPI signal peptides. These conserved residues were expendable with no effect on VSG GPI-anchoring. However, deletion of the hydrophobic region of the GPI signal peptide resulted in the accumulation of unanchored VSGs in the ER (Böhme & Cross, 2002). The observation that the predicted *T. vivax* ILDat1.2 and ILDat2.1 GPI signal peptides do not harbor the conserved elements present in *T. brucei* VSG GPI signal peptides was surprising. Analysis of the predicted ILDat1.2 and ILDat2.1 GPI signal peptides showed that what would be the hydrophobic regions contained a mixture of both hydrophobic and hydrophilic residues. This difference in architecture of the *T. vivax* ILDat1.2 and ILDat2.VSG GPI signal peptides might have caused inefficient GPI attachment, leading to the production of low amounts of the VSG protein in *T. brucei*.

Replacement of ILDat1.2 GPI signal peptide with a bona fide *T. brucei* VSG GPI signal peptide led to the expression of high levels of ILDat1.2 VSG protein in *T. brucei*. This confirmed that the ILDat1.2 GPI signal was indeed not recognized by the *T. brucei* GPI anchoring machinery. It has been shown that the addition of GPI-anchors to protein polypeptides in the ER acts as a positive trafficking signal, with unanchored proteins being retained in the ER (Field et al., 1994; McDowell et al., 1998). It has further been reported that the expression of a BSF VSG reporter lacking the GPI signal sequence in procyclic *T. brucei* parasites caused reduced VSG transport in the ER. Though the soluble VSG is finally secreted, the slow transportation causes an accumulation of the VSG in the ER (McDowell et al., 1998). However, in BSF parasites, VSGs lacking the GPI signal are not secreted but misdirected to the lysosomes where they are degraded (Triggs & Bangs, 2003). This degradation of VSGs lacking the GPI anchor could possibly be the reason why only low levels of ILDat1.2 VSG protein were detected. In-depth mutagenesis analysis of the

native ILDat1.2 GPI signal should be carried out to determine the sequence elements that hinder its recognition in *T. brucei.*

Though ILDat1.2 protein could be sufficiently expressed by replacement of its GPI anchor signal by one from a *T. brucei* VSG, the expression of ILDat2.1 VSG could not be rescued with the same modification. This was unexpected given that the ILDat2.1 ER import signal was additionally replaced with a *T. brucei* VSG ER import signal to ensure correct ER targeting. Why then was the ILDat2.1 VSG not expressed in *T. brucei*? Possibly the lack of glycosylation of ILDat2.1 (Gardiner et al., 1996) affected its steady-state expression. We know that the majority of *T. brucei* VSGs are glycosylated (Schwede et al., 2015). Although whether this glycosylation influences VSG folding and expression is still debatable, some studies suggest this likelihood; it was possible to remove the individual glycans of VSG118 but not all the three of them. Again, one of the 2 glycans in VSG060 could be removed with no effect on expression, but not the other. However, glycan free VSG117 and VSG121 mutants have been successfully expressed in *T. brucei* (Hartel et al., 2016).

The tertiary structure of *T. brucei* VSGs is highly conserved despite the high sequence variations. It appears that for successful expression and GPI-anchoring, the VSG protein must acquire a specific fold (Wang et al., 2003). The three-helix core structure of the VSG is shared with other GPI-anchored surface proteins of BSF *T. brucei*, including the SRA protein, HpHbR, and the TFR (Higgins et al., 2013; Trevor et al., 2019; Zoll et al., 2018). Whether it is inefficient folding, lack of glycosylation or a different sequence feature that hindered the production of ILDat2.1 protein is not known. It appears ILDat2.1 VSG does not meet the threshold of a functional *T. brucei* VSG, underpinning the existence of a robust quality control system for the expression, processing and trafficking of VSGs in *T. brucei*.

Expression of a *T. congolense* VSG was possible in *T. brucei* without any modifications of its native ER and GPI signal sequences (Nicola Jones, unpublished). However, a *T. vivax* VSG could only be sufficiently expressed in *T. brucei* after replacing its GPI signal sequence with that of a *T. brucei* VSG. This difference shows that the GPI biosynthetic and anchoring machinery may be conserved in *T. brucei* and *T. congolense*, but different in *T. vivax*. The possibility of VSGs from *T. congolense* and *T. vivax* to form coats on *T. brucei* further underscores the close evolutionary relatedness of these African trypanosome species and their surface coats. However, only one *T. congolense* and two *T. vivax* VSGs, have been studied for expression in *T. brucei*. Therefore, additional *T. congolense* and *T. vivax* VSGs should be analyzed to ascertain these findings.

4.4 The expression of chimeric *T. brucei* and *T. vivax* VSGs in *T. brucei*

Antigenic variation drives the establishment of persistent bacterial and protozoan infections in their different hosts (Palmer & Brayton, 2007). These pathogens either consist of small- or large-genomes, meaning the size of their variable surface antigen repertoires could be different. To continuously evade the host's immune responses, these pathogens rely on segmental gene conversion events that generate a vast repertoire of potential new antigens (Brayton et al., 2001). Segmental gene conversion can occur by combinatorial events between open reading frames of functional genes or sections of pseudogenes to introduce a variation on the expressed antigens. In *T. brucei, VSGs* can undergo segmental gene conversions to generate so-called mosaic VSGs (Marcello & Barry, 2007; Taylor & Rudenko, 2006). Since the *T. brucei* genome encodes for ~1500 *VSG* genes and pseudogenes (which is not a small archive), combinatorial events between these genes further enrich the VSG archive. The mosaic VSGs tend to appear late in infections, and mosaicism is the major contributor to chronic infection (Hall et al., 2013).

It has been shown that different combinations might occur between the various *T. brucei* VSG NTD and CTD domain types (Hutchinson et al., 2003). Banking on the successful expression of ILDat1.2 in *T. brucei*, I sought to determine whether VSG chimeras composed of *T. vivax* and *T. brucei* domains could form a coat on *T. brucei* parasites. Chimeric VSG genes consisting of the predicted ILDat1.2 NTD and the VSG221 CTD were generated in this thesis. The mature polypeptide of the chimeric VSG is 378 amino acids long (ILDat1.2 NTD = 307; VSG221 CTD = 71) and could be easily distinguished from the mature VSG221 (433 amino acids) by size on Western blots. The 221^{KO}.ILD:CTD cell line, expressing the chimeric VSG had reduced growth, doubling after every 9.4 hours, which considerably slower than the 6.4 hours doubling time of the parental cell line. It was also possible to express a chimeric VSG consisting of the ILDat1.2 NTD and a truncated VSG221 CTD, lacking the structured region (This truncation was carried out by Nicola Jones). The truncated VSG221 CTD consists of 46 amino acids, and together with the 307-amino acid ILDat1.2 NTD, the resulting chimeric VSG polypeptide is 353 amino acids long. The 221^{KO}.ILD:ML cell line expressing this VSG chimera doubled every 10.6 hours. I report here that shorter VSGs (in terms of the primary structure) than the known *T. brucei* VSGs could be sufficiently expressed and form a coat on *T. brucei*. Although the structured CTD domain of VSG221 was expendable, it came with a fitness cost as the parasites stably expressing a chimeric VSG with this truncation had a slowed growth.

It has been shown that *T. brucei* VSGs with C-terminal domain truncations can be expressed. VSG221 lacking the structured carboxy-subdomain could be expressed, while deleting the linker region just after the NTD in addition to the structured region was lethal. Additionally, deleting both structured regions in the VSG121 CTD was tolerated, if the linkers were left intact (Henning, 2014). These experiments showed that, at least, a linker with a minimal size requirement was required in the C-terminal for adequate expression of VSG. In cases where the truncations were tolerated and the mutant VSG proteins expressed, cell growth was differently impacted.

Parasites expressing different antigen types have varying growth rates. This growth rate may be ES-dependent since the different ES vary in size and distribution of the ESAGs (Becker et al., 2004). A recent study suggested that shorter expressed VSG antigens are synonymous with faster growths (Liu et al., 2018). However, contrary to the findings of Liu et al., our results showed that the 221^{KO}.ILD:ML cell line expressing the shortest VSG had the slowest growth. In our experiments, the ectopic *VSG* genes were expressed from the same BES, and so there was no difference as far as the BES size and distribution of ESAGs were concerned. Therefore, it is likely that other than the VSG antigen size, other aspects of VSG processing such as VSG folding could influence growth.

4.5 Structural analysis of a *T. vivax* VSG

The available data on trypanosome VSG architecture is based on the structures of only *T. brucei* VSGs. As more *T. brucei* VSG structures become unraveled, it is increasingly becoming evident that although VSG NTDs share a conserved central core structure, there are considerable architectural variations in other regions of the VSG molecule. Here, the structure of a *T. vivax* VSG was studied. Given that BSF *T. vivax* parasites cannot be cultivated in axenic culture, the tractable *T. brucei* system was utilized to express *T. vivax* ILDat1.2 VSG. The native ILDat1.2 VSG GPI signal sequence was replaced with a *T. brucei* VSG GPI signal and cell lines stably expressing only ILDat1.2 VSG were generated. It was possible to purify high amounts of the ILDat1.2 VSG protein from the transgenic *T. brucei* cell lines, using same methods established for *T. brucei* VSGs, as they could be grown to high densities in culture media.

The ILDat1.2 structure was studied using X-ray crystallography which has atomic scale resolution and has been used to solve the NTD structures of *T. brucei* VSGs. ILDat1.2 crystals diffracted well, however attempts to solve the phase problem using molecular replacement based on the known VSG structures were unsuccessful. This was not surprising as most of the recent *T. brucei* VSG structures could not be solved in this way. Therefore,

phasing was done by the isomorphous replacement. ILDat1.2 has the conserved three-helix bundle scaffold of *T. brucei* and *T. congolense* surface proteins. The unstructured surface loops form the variable head region of the VSG molecule. Two disulfide bridges, formed by cysteine residues at positions 13 and 144 and positions 122 and 191, stabilize the head region. This cysteine residue pattern is conserved in the A-type *T. brucei* VSGs (Carrington et al., 1991) indicating that the ILDat1.2 has an A-type NTD.

Previous studies suggested that *T. brucei* VSGs were only N-glycosylated, and that this modification occurred at the membrane proximal regions of the NTD (Blum et al., 1993; Freymann et al., 1990). However, the recently solved *T. brucei* VSGsur and MITat1.13 structures reveal that VSGs can also be N-glycosylated closer to the head region (Zeelen et al., 2021). Additionally, the discovery of an *O*-glycan in the head region of MITat1.3 (Pinger et al., 2018) proved that *T. brucei* VSGs can be modified by other glycans not only N-glycosylated. ILDat1.2 has a single N-linked glycan on Asn59 positioned at the membrane-proximal end of the coiled-coil. The role of this N-linked glycan in the ILDat1.2 NTD remains to be determined.

T. vivax VSGs are the smallest characterized trypanosome VSGs reported to date, ranging from 80 – 90 kDa (Gardiner et al., 1996; Gardiner & Wilson, 1987). The mature ILDat1.2 VSG polypeptide is 356 residues long, 77 residues shorter than the mature MITat1.2 VSG polypeptide with 433 residues. The NTD domain of MITat1.2 is 362 residues, six residues more than the complete ILDat1.2 VSG. It was therefore conceivable that the ILDat1.2 VSGs could be shorter compared to the *T. brucei* VSGs. However, we find that the solved structure of ILDat1.2 NTD (262 residues) is 9.35 nm in length, approximately equal to ILTat1.24 (358 residues, 9.45 nm), with MITat1.2 (362 residues, 9.89 nm) being slightly longer due to the presence of a C-terminal helix that links the NTD to the CTD (Blum et al., 1993; Freymann et al., 1990).

During *T. brucei* VSG crystallization experiments, the CTD is normally cleaved presumably by endogenous peptidases (Freymann et al., 1984). As such only the NTD of *T. brucei* VSGs have been solved using X-ray crystallography. Therefore, it was not surprising that *T. vivax* ILDat1.2 was also cleaved. However, this cleavage was different from that of *T. brucei* VSGs as it occurred above what would be the hip region in *T. brucei* VSGs. In total, 94 resides were cleaved form the ILDat1.2 molecule. Studies on the *T. brucei* VSG C-terminal domains show that a linker of approximately 30 residues is required in the C-terminal for expression of the VSG (Jones et al., manuscript in preparation). Therefore, if an assumption is made that the ILDat1.2 C-terminal was at least a 30-residue linker, only 64 residues would

remain to make the complete ILDat1.2 NTD. Since the NTD of ILDat1.2, MITat1.2, and ILTat1.24 VSG are approximately the same height, it is unlikely that the 64 residues of ILDat1.2 would be sufficient to form the extensions similar to those at the base of MITat1.2, which is 40 residues more. Therefore, it is probable that the ILDat1.2 NTD base is free of the extensions present in *T. brucei* VSG NTDs. Superposition of the ILDat1.2 NTD monomer with the GPI-anchored TFR monomer of *T. brucei* (ESAG6) showed good structural similarity at the base of the NTD. The mature TFR polypeptide is almost the same size as the mature ILDat1.2 polypeptide, comprising 353 residues, just three less than the mature ILDat1.2 polypeptide. Thirty residues at the C-terminal end of the receptor were not solved, and it is suggested that this polypeptide links the receptor to the GPI-anchor (Trevor et al., 2019). Therefore, the membrane proximal region of the ILDat1.2 VSG could be more similar to that of the GPI-anchored ESAG6 monomer of the TFR molecule than that of the *T. brucei* VSGs.

A better understanding of the ILDat1.2 VSG architecture necessitates solving its complete structure. The full MITat1.1 VSG structure was solved by the combination of X-ray crystallography, NMR, and SAXS techniques (Bartossek et al., 2017). The NTD domain was solved by X-ray crystallography. Given that the cleaved MITat1.1 CTD does not crystallize, it was heterologously expressed in bacteria, and the solution structure was solved using NMR. The high-resolution structures were fitted into the low-resolution structures obtained with SAXS to build a low-resolution envelope of the whole VSG molecule. These experiments enabled the determination of the compact and relaxed conformations of the MITat1.1 VSG CTD. It is also possible to build low-resolution protein models from SAXS data without *a priori* high-resolution data from complementary methods (Mertens & Svergun, 2010). SAXS does not require crystals. Therefore, a low-resolution model of the complete ILDat1.2 VSG can be built from SAXS data. Recent advances in the field of cryo-EM technology have improved the resolution of cryo-EM to the atomic level. Since this technology also eliminates the requirement to crystallize the sample, it could possibly be utilized to solve the complete structure of ILDat1.2 VSG.

The ILDat1.2 NTD structure presented here is the first non-*T. brucei* VSG NTD and *T. vivax* surface protein structure to be solved. The ILDat1.2 VSG fold and the conservation pattern of the cysteine residues indicate common ancestry of the *T. vivax* and *T. brucei* VSG genes. This data is consistent with the finding that despite the early divergence of these two trypanosome species, the ancestral VSG gene lineages have been maintained (Jackson et al., 2012). However, it is likely that the *T. vivax* ILDat1.2 CTD differs from the known *T. brucei*

VSG CTDs (Schwede et al., 2015). A flexible VSG CTD is essential for the VSG shielding function and maintains high mobility of the VSG at different densities (Bartossek et al., 2017) . It would be interesting to determine the implications of this potential deviation in the CTD architecture on the shielding function and mobility of ILDat1.2 VSG on the trypanosome surface.

Transcriptomic and proteomic data have shown that *T. vivax* parasites have an expanded repertoire of cell surface proteins (Greif et al., 2013; Jackson et al., 2015). This has led to the suggestion that the *T. vivax* VSG coat may not be as tightly packed as the *T. brucei* one, implying that the invariant surface proteins of *T. vivax* may be easily accessible to the host immune factors. Indeed, vaccinating mice with a recombinant subunit vaccine comprising a conserved invariant flagellar protein (IFX), that localizes to the flagellar membrane in *T. vivax*, confers long-term protection (Autheman et al., 2021). Together, these studies show that the *T. vivax* and *T. brucei* surface coats may be differently constituted. Further structural characterization of other *T. vivax* GPI-anchored surface proteins other than the VSGs will be necessary to shed light on the architecture and function of the *T. vivax* surface.

Conclusion

Here, it is shown that BSF *T. brucei* has a robust mechanism to regulate the mRNA levels of the abundant VSGs. Apart from the involvement of the ER in the translocation and post-translational modifications of proteins, the trypanosome ER is shown to be further involved in regulating *VSG* mRNA amounts. The *VSG* mRNA levels were balanced only when the VSGs had bona fide ER import signals and thus efficiently targeted to the ER. Interestingly, this balancing is independent of the *VSG* ORF and the VSG ER import signal. Though the mechanistic details of this regulation are not known, it might involve a feedback loop that is elicited at the ER to keep the amount of cargo that gets in at an optimum.

The tractable *T. brucei* system was successfully utilized to study *T. vivax* VSGs. Expression of sufficient amounts of a *T. vivax* ILDat1.2 VSG protein in *T. brucei* was only possible when the VSG C-terminal GPI anchor signal was replaced with one of *T. brucei* VSG. This indicates that despite the common ancestry of VSG gene lineages of African trypanosomes, the VSG GPI attachment machineries of *T. brucei* and *T. vivax* differ. The successful expression of the *T. vivax* ILDat1.2 as the only produced VSG in transgenic *T. brucei* cells and its crystal structure further shows the relatedness of these parasites and their surface coats. Therefore the *T. brucei* system presents an avenue to study surface proteins of related trypanosome species such as *T. vivax* that cannot be propagated in axenic culture and lack forward and reverse genetic manipulation tools.

5.0 References

Abernathy, E., Gilbertson, S., Alla, R., & Glaunsinger, B. (2015). Viral Nucleases Induce an mRNA Degradation-Transcription Feedback Loop in Mammalian Cells. *Cell Host & Microbe*, *18*, 243–253.

Abola, E., Kuhn, P., Earnest, T., & Stevens, R. C. (2000). Automation of X-ray crystallography. *Nature Structural Biology*, *7 Suppl*, 973–977.

Acosta-Serrano, A., Vassella, E., Liniger, M., Kunz Renggli, C., 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 USA*, *98*, 1513–1518.

Al-Qahtani, A., Teilhet, M., & Mensa-Wilmot, K. (1998). Species-specificity in endoplasmic reticulum signal peptide utilization revealed by proteins from *Trypanosoma brucei* and Leishmania. *The Biochemical Journal*, *331*, 521–529.

Aregawi, W. G., Agga, G. E., Abdi, R. D., & Büscher, P. (2019). Systematic review and meta-analysis on the global distribution, host range, and prevalence of *Trypanosoma evansi*. *Parasites & Vectors*, *12*, 67.

Aresta-Branco, F., Pimenta, S., & Figueiredo, L. M. (2016). A transcription-independent epigenetic mechanism is associated with antigenic switching in *Trypanosoma brucei*. *Nucleic Acids Research*, *44*, 3131–3146.

Ast, T., Cohen, G., & Schuldiner, M. (2013). A network of cytosolic factors targets SRP-independent proteins to the endoplasmic reticulum. *Cell*, *152*, 1134–1145.

Autheman, D., Crosnier, C., Clare, S., Goulding, D. A., Brandt, C., Harcourt, K., Tolley, C., Galaway, F., Khushu, M., Ong, H., Ramirez, A. R., Duffy, C. W., Jackson, A. P., & Wright, G. J. (2021). An invariant *Trypanosoma vivax* vaccine antigen inducing protective immunity. *BioRxiv,* 10.1101/2021.02.10.430711.

Bangs, J. D., Crain, P. F., Hashizume, T., McCloskey, J. A., & Boothroyd, J. C. (1992). Mass spectrometry of mRNA cap 4 from trypanosomatids reveals two novel nucleosides. *The Journal of Biological Chemistry*, *267*, 9805–9815.

Barnwell, E. M., van Deursen, F. J., Jeacock, L., Smith, K. A., Maizels, R. M., Acosta-Serrano, A., & Matthews, K. (2010). Developmental regulation and extracellular release of a VSG expression-site-associated gene product from *Trypanosoma brucei* bloodstream forms. *Journal of Cell Science*, *123*, 3401–3411.

Bartossek, T., Jones, N. G., Schäfer, C., Cvitković, M., Glogger, M., Mott, H. R., Kuper, J., Brennich, M., Carrington, M., Smith, A.-S., Fenz, S., Kisker, C., & Engstler, M. (2017). Structural basis for the shielding function of the dynamic trypanosome variant surface glycoprotein coat. *Nature Microbiology*, *2*, 1523–1532.

Bartossek, T. (2017). Structural and functional analysis of trypanosomal variant surface glycoprotein using x-ray scattering techniques and fluorescence microscopy. Doctoral dissertation, Julius-Maximilians-Universität Würzburg.

Baruch, D. I., Pasloske, B. L., Singh, H. B., Bi, X., Ma, X. C., Feldman, M., Taraschi, T. F., & Howard, R. J. (1995). Cloning the *P. falciparum* gene encoding PfEMP1, a malarial variant antigen and adherence receptor on the surface of parasitized human erythrocytes. *Cell*, *82*, 77–87.

Batram, C., Jones, N. G., Janzen, C. J., Markert, S. M., & Engstler, M. (2014). Expression site attenuation mechanistically links antigenic variation and development in *Trypanosoma brucei*. *ELife*, *3*, e02324.

Batram, C. (2013). Die Kontrolle der monoallelen Expression, antigenen Variation und Entwicklung in *Trypanosoma brucei*. Doctoral dissertation, Julius-Maximilians-Universität Würzburg.

Beck, T., Krasauskas, A., Gruene, T., & Sheldrick, G. M. (2008). A magic triangle for experimental phasing of macromolecules. *Acta Crystallographica. Section D, Biological Crystallography*, *64*, 1179–1182.

Becker, M., Aitcheson, N., Byles, E., Wickstead, B., Louis, E., & Rudenko, G. (2004). Isolation of the repertoire of VSG expression site containing telomeres of *Trypanosoma brucei* 427 using transformation-associated recombination in yeast. *Genome Research*, *14*, 2319–2329.

Benz, C., Lo, W., Fathallah, N., Connor-Guscott, A., Benns, H. J., & Urbaniak, M. D. (2018). Dynamic regulation of the *Trypanosoma brucei* transferrin receptor in response to iron starvation is mediated via the 3'UTR. *PloS One*, *13*, e0206332.

Berberof, M., Vanhamme, L., Tebabi, P., Pays, A., Jefferies, D., Welburn, S., & Pays, E. (1995). The 3'-terminal region of the mRNAs for VSG and procyclin can confer stage specificity to gene expression in *Trypanosoma brucei*. *The EMBO Journal*, *14*, 2925–2934.

Berriman, M., Ghedin, E., Hertz-Fowler, C., Blandin, G., Renauld, H., Bartholomeu, D. C., Lennard, N. J., Caler, E., Hamlin, N. E., Haas, B., Böhme, U., Hannick, L., Aslett, M. A., Shallom, J., Marcello, L., Hou, L., Wickstead, B., Alsmark, U. C. M., Arrowsmith, C., . . . El-Sayed, N. M. (2005). The genome of the African trypanosome *Trypanosoma brucei*. *Science*, *309*, 416–422.

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

Böhme, U., & Cross, G. A. M. (2002). Mutational analysis of the variant surface glycoprotein GPI-anchor signal sequence in *Trypanosoma brucei*. *Journal of Cell Science*, *115*, 805–816.

Boothroyd, C. E., Dreesen, O., Leonova, T., Ly, K. I., Figueiredo, L. M., Cross, G. A. M., & Papavasiliou, F. N. (2009). A yeast-endonuclease-generated DNA break induces antigenic switching in *Trypanosoma brucei*. *Nature*, *459*, 278–281.

Boothroyd, J. C., Paynter, C. A., Cross, G. A., Bernards, A., & Borst, P. (1981). Variant surface glycoproteins of *Trypanosoma brucei* are synthesised with cleavable hydrophobic sequences at the carboxy and amino termini. *Nucleic Acids Research*, *9*, 4735–4743.

Borst, P. (2002). Antigenic Variation and Allelic Exclusion. *Cell*, *109*, 5–8.

Borst, P., & Ulbert, S. (2001). Control of VSG gene expression sites. *Molecular and Biochemical Parasitology*, *114*, 17–27.

Brayton, K. A., Knowles, D. P., McGuire, T. C., & Palmer, G. H. (2001). Efficient use of a small genome to generate antigenic diversity in tick-borne ehrlichial pathogens. *Proceedings of the National Academy of Sciences USA*, *98*, 4130–4135.

Brown, R. C., Evans, D. A., & Vickerman, K. (1973). Changes in oxidative metabolism and ultrastructure accompanying differentiation of the mitochondrion in *Trypanosoma brucei*. *International Journal for Parasitology*, *3*, 691–704.

Bruce, D., D. N. Nabarro, et al. (1903). Further Report on Sleeping Sickness in Uganda. *Reports of the Sleeping Sickness commission of the Royal Society*. 4: 1–87

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

Budzak, J., Kerry, L. E., Aristodemou, A., Hall, B. S., Witmer, K., Kushwaha, M., Davies, C., Povelones, M. L., McDonald, J. R., Sur, A., Myler, P. J., & Rudenko, G. (2019). Dynamic colocalization of 2 simultaneously active VSG expression sites within a single expression-site body in *Trypanosoma brucei*. *Proceedings of the National Academy of Sciences USA*, *116*, 16561–16570.

Bülow, R., Overath, P., & Davoust, J. (1988). Rapid lateral diffusion of the variant surface glycoprotein in the coat of. *Biochemistry*, *27*, 2384–2388.

Bürdel, J. (2010). Functional analysis of the C-terminal of domain of *Trypanosoma brucei.* Diploma Thesis. Technische Universität Darmstadt.

Büscher, P., Cecchi, G., Jamonneau, V., & Priotto, G. (2017). Human African trypanosomiasis. *The Lancet*, *390*, 2397–2409.

Capewell, P., Cooper, A., Clucas, C., Weir, W., & MacLeod, A. (2015). A co-evolutionary arms race: Trypanosomes shaping the human genome, humans shaping the trypanosome genome. *Parasitology*, *142*, S108-19.

Carrington, M., Miller, N., Blum, M., Roditi, I., Wiley, D., & Turner, M. (1991). Variant specific glycoprotein of *Trypanosoma brucei* consists of two domains each having an independently conserved pattern of cysteine residues. *Journal of Molecular Biology*, *221*, 823–835.

Chamond, N., Cosson, A., Blom-Potar, M. C., Jouvion, G., D'Archivio, S., Medina, M., Droin-Bergère, S., Huerre, M., Goyard, S., & Minoprio, P. (2010). *Trypanosoma vivax* infections: Pushing ahead with mouse models for the study of Nagana. I. Parasitological, hematological and pathological parameters. *PLoS Neglected Tropical Diseases*, *4*, e792.

Chattopadhyay, A., Jones, N. G., Nietlispach, D., Nielsen, P. R., Voorheis, H. P., Mott, H. R., & Carrington, M. (2005). Structure of the C-terminal domain from *Trypanosoma brucei* variant surface glycoprotein MITat1.2. *The Journal of Biological Chemistry*, *280*, 7228–7235.

Clayton, C. (2019). Regulation of gene expression in trypanosomatids: Living with polycistronic transcription. *Open Biology*, *9*, 190072.

Clayton, C. E. (2002). Life without transcriptional control? From fly to man and back again. *The EMBO Journal*, *21*, 1881–1888.

Cordero, E. M., Cortez, C., Yoshida, N., & da Silveira, J. F. (2019). Signal peptide recognition in *Trypanosoma cruzi* GP82 adhesin relies on its localization at protein N-terminus. *Scientific Reports*, *9*, 7325.

Cross, G. A. (1975). Identification, purification and properties of clone-specific glycoprotein antigens constituting the surface coat of *Trypanosoma brucei*. *Parasitology*, *71*, 393–417.

Cross, G. A. (1984). Structure of the variant glycoproteins and surface coat of *Trypanosoma brucei*. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *307*, 3–12.

Cross, G. A. M., Kim, H.-S., & Wickstead, B. (2014). Capturing the variant surface glycoprotein repertoire (the VSGnome) of *Trypanosoma brucei* Lister 427. *Molecular and Biochemical Parasitology*, *195*, 59–73.

Dagnachew, S., Terefe, G., Abebe, G., Sirak, A., Bollo, E., Barry, D., & Goddeeris, B. (2015). Comparative clinico-pathological observations in young Zebu (*Bos indicus*) cattle experimentally infected with *Trypanosoma vivax* isolates from tsetse infested and non-tsetse areas of Northwest Ethiopia. *BMC Veterinary Research*, *11*, 307.

Dale, G. E., Oefner, C., & D'Arcy, A. (2003). The protein as a variable in protein crystallization. *Journal of Structural Biology*, *142*, 88–97.

Dean, S., Marchetti, R., Kirk, K., & Matthews, K. R. (2009). A surface transporter family conveys the trypanosome differentiation signal. *Nature*, *459*, 213–217.

Deitsch, K. W., Lukehart, S. A., & Stringer, J. R. (2009). Common strategies for antigenic variation by bacterial, fungal and protozoan pathogens. *Nature Reviews Microbiology*, *7*, 493–503.

Desquesnes, M., & Dia, M. L. (2004). Mechanical transmission of *Trypanosoma vivax* in cattle by the African tabanid *Atylotus fuscipes*. *Veterinary Parasitology*, *119*, 9–19.

Dessau, M. A., & Modis, Y. (2011). Protein crystallization for X-ray crystallography. *Journal of Visualized Experiments : JoVE*(47).

Di Tommaso, P., Moretti, S., Xenarios, I., Orobitg, M., Montanyola, A., Chang, J.-M., Taly, J.-F., & Notredame, C. (2011). T-Coffee: A web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. *Nucleic Acids Research*, *39*, W13-7.

Dickerson, R. E., Kendrew, J. C., & Strandberg, B. E. (1961). The crystal structure of myoglobin: Phase determination to a resolution of 2 Å by the method of isomorphous replacement. *Acta Crystallographica*, *14*, 1188–1195.

Dickie, E. A., Giordani, F., Gould, M. K., Mäser, P., Burri, C., Mottram, J. C., Rao, S. P. S., & Barrett, M. P. (2020). New Drugs for Human African Trypanosomiasis: A Twenty First Century Success Story. *Tropical Medicine and Infectious Disease*, *5*.

do Nascimento, L. M., Egler, F., Arnold, K., Papavisiliou, N., Clayton, C., & Erben, E. (2020). The trypanosome Variant Surface Glycoprotein mRNA is stabilized by an essential unconventional RNA-binding protein. *Bioarxiv,* 10.1101/2020.10.08.331769

Ducruix, A., & Giegé, R. (1992). Methods of crystallization. Crystallization of nucleic acids and proteins: A practical approach. Ducruix, A. & Giegé, R. (Eds.) 73–98. (IRL Press, Oxford).

Duffy, J., Patham, B., & Mensa-Wilmot, K. (2010). Discovery of functional motifs in h-regions of trypanosome signal sequences. *The Biochemical Journal*, *426*, 135–145.

El-Sayed, N. M., & Donelson, J. E. (1997). African trypanosomes have differentially expressed genes encoding homologues of the Leishmania GP63 surface protease. *The Journal of Biological Chemistry*, *272*, 26742–26748.

Emsley, P., Lohkamp, B., Scott, W. G., & Cowtan, K. (2010). Features and development of Coot. *Acta Crystallographica. Section D, Biological Crystallography*, *66*, 486–501.

Engstler, M., Pfohl, T., Herminghaus, S., Boshart, M., Wiegertjes, G., Heddergott, N., & Overath, P. (2007). Hydrodynamic flow-mediated protein sorting on the cell surface of trypanosomes. *Cell*, *131*, 505–515.

Engstler, M., Thilo, L., Weise, F., Grünfelder, C. G., Schwarz, H., Boshart, M., & Overath, P. (2004). Kinetics of endocytosis and recycling of the GPI-anchored variant surface glycoprotein in *Trypanosoma brucei*. *Journal of Cell Science*, *117*, 1105–1115.

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

Evans, P., & McCoy, A. (2008). An introduction to molecular replacement. *Acta Crystallographica. Section D, Biological Crystallography*, *64*, 1–10.

FAO, (2017). Report on the future of food and agriculture. Trends and challenges. http://www.fao.org/3/i6583e/i6583e.pdf

Fast, B., Kremp, K., Boshart, M., & Steverding, D. (1999). Iron-dependent regulation of transferrin receptor expression in *Trypanosoma brucei*. *The Biochemical Journal*, *342*, 691–696.

Fenn, K., & Matthews, K. R. (2007). The cell biology of *Trypanosoma brucei* differentiation. *Current Opinion in Microbiology*, *10*, 539–546.

Ferguson, M. A., Homans, S. W., Dwek, R. A., & Rademacher, T. W. (1988). Glycosyl-phosphatidylinositol moiety that anchors *Trypanosoma brucei* variant surface glycoprotein to the membrane. *Science*, *239*, 753–759.

Ferguson, M. A.J., Hart, G. W., & Kinoshita, T. (2015). *Essentials of Glycobiology: Glycosylphosphatidylinositol Anchors* (A. Varki, R. D. Cummings, J. D. Esko, P. Stanley, G. W. Hart, M. Aebi, A. G. Darvill, T. Kinoshita, N. H. Packer, J. H. Prestegard, R. L. Schnaar, & P. H. Seeberger, Eds.) (3rd).

Fetene, E., Leta, S., Regassa, F., & Büscher, P. (2021). Global distribution, host range and prevalence of *Trypanosoma vivax*: A systematic review and meta-analysis. *Parasites & Vectors*, *14*, 80.

Field, M. C., Moran, P., Li, W., Keller, G. A., & Caras, I. W. (1994). Retention and degradation of proteins containing an uncleaved glycosylphosphatidylinositol signal. *The Journal of Biological Chemistry*, *269*, 10830–10837.

Field, M. C., Sergeenko, T., Wang, Y.-N., Böhm, S., & Carrington, M. (2010). Chaperone requirements for biosynthesis of the trypanosome variant surface glycoprotein. *PloS One*, *5*, e8468.

Figueiredo, L. M., Janzen, C. J., & Cross, G. A. M. (2008). A histone methyltransferase modulates antigenic variation in African trypanosomes. *PLoS Biology*, *6*, e161.

Florent, I. C., Raibaud, A., & Eisen, H. (1991). A family of genes related to a new expression site-associated gene in *Trypanosoma equiperdum*. *Molecular and Cellular Biology*, *11*, 2180–2188.

Franco, J. R., Cecchi, G., Priotto, G., Paone, M., Diarra, A., Grout, L., Simarro, P. P., Zhao, W., & Argaw, D. (2020). Monitoring the elimination of human African trypanosomiasis at continental and country level: Update to 2018. *PLoS Neglected Tropical Diseases*, *14*, e0008261.

Freymann, D. M., Metcalf, P., Turner, M., & Wiley, D. C. (1984). 6 A-resolution X-ray structure of a variable surface glycoprotein from *Trypanosoma brucei*. *Nature*, *311*, 167–169.

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.

Ganot, P., Kallesøe, T., Reinhardt, R., Chourrout, D., & Thompson, E. M. (2004). Spliced-leader RNA trans splicing in a chordate, *Oikopleura dioica*, with a compact genome. *Molecular and Cellular Biology*, *24*, 7795–7805.

Gardiner, P. R., & Wilson, A. J. (1987). *Trypanosoma (Duttonella) vivax*. *Parasitology Today*, *3*, 49–52.

Gardiner, P. R., Nene, V., Barry, M. M., Thatthi, R., Burleigh, B., & Clarke, M. W. (1996). Characterization of a small variable surface glycoprotein from *Trypanosoma vivax*. *Molecular and Biochemical Parasitology*, *82*, 1–11.

Giordani, F., Morrison, L. J., Rowan, T. G., Koning, H. P. de, & Barrett, M. P. (2016). The animal trypanosomiases and their chemotherapy: A review. *Parasitology*, *143*, 1862–1889.

Glover, L., Alsford, S., & Horn, D. (2013). Dna break site at fragile subtelomeres determines probability and mechanism of antigenic variation in African trypanosomes. *PLoS Pathogens*, *9*, e1003260.

Goldshmidt, H., Sheiner, L., Bütikofer, P., Roditi, I., Uliel, S., Günzel, M., Engstler, M., & Michaeli, S. (2008). Role of protein translocation pathways across the endoplasmic reticulum in *Trypanosoma brucei*. *The Journal of Biological Chemistry*, *283*, 32085–32098.

Green, D.W., Vernom Martin Ingram, & Max Feedinand Perutz (1954). The structure of haemoglobin - IV. Sign determination by the isomorphous replacement method. *Proceedings of the Royal Society of London A.*, *225*, 287–307.

Greif, G., Ponce de Leon, M., Lamolle, G., Rodriguez, M., Piñeyro, D., Tavares-Marques, L. M., Reyna-Bello, A., Robello, C., & Alvarez-Valin, F. (2013). Transcriptome analysis of the bloodstream stage from the parasite *Trypanosoma vivax*. *BMC Genomics*, *14*, 149.

Grünfelder, C. G., Engstler, M., Weise, F., Schwarz, H., Stierhof, Y.-D., Boshart, M., & Overath, P. (2002). Accumulation of a GPI-anchored protein at the cell surface requires sorting at multiple intracellular levels. *Traffic*, *3*, 547–559.

Günzl, A., Bruderer, T., Laufer, G., Schimanski, B., Tu, L.-C., Chung, H.-M., Lee, P.-T., & Lee, M. G.-S. (2003). Rna polymerase I transcribes procyclin genes and variant surface glycoprotein gene expression sites in *Trypanosoma brucei*. *Eukaryotic Cell*, *2*, 542–551.

Haimovich, G., Medina, D. A., Causse, S. Z., Garber, M., Millán-Zambrano, G., Barkai, O., Chávez, S., Pérez-Ortín, J. E., Darzacq, X., & Choder, M. (2013). Gene expression is circular: Factors for mRNA degradation also foster mRNA synthesis. *Cell*, *153*, 1000–1011.

Hajduk, S. L., Moore, D. R., Vasudevacharya, J., Siqueira, H., Torri, A. F., Tytler, E. M., & Esko, J. D. (1989). Lysis of *Trypanosoma bruce*i by a toxic subspecies of human high density lipoprotein. *The Journal of Biological Chemistry*, *264*, 5210–5217.

Hall, B. S., Pal, A., Goulding, D., Acosta-Serrano, A., & Field, M. C. (2005). *Trypanosoma brucei*: Tbrab4 regulates membrane recycling and expression of surface proteins in procyclic forms. *Experimental Parasitology*, *111*, 160–171.

Hall, J. P. J., Wang, H., & Barry, J. D. (2013). Mosaic VSGs and the scale of *Trypanosoma brucei* antigenic variation. *PLoS Pathogens*, *9*, e1003502.

Hannaert, V., Bringaud, F., Opperdoes, F. R., & Michels, P. am (2003). Evolution of energy metabolism and its compartmentation in Kinetoplastida. *Kinetoplastid Biology and Disease*, *2*, 11.

Hartel, A. J. W., Glogger, M., Jones, N. G., Abuillan, W., Batram, C., Hermann, A., Fenz, S. F., Tanaka, M., & Engstler, M. (2016). N-glycosylation enables high lateral mobility of GPI-anchored proteins at a molecular crowding threshold. *Nature Communications*, *7*, 12870.

Heckman, K. L., & Pease, L. R. (2007). Gene splicing and mutagenesis by PCR-driven overlap extension. *Nature Protocols*, *2*, 924–932.

Heijne, G. von (1998). Life and death of a signal peptide. *Nature*, *396*, 111–113.

Heijne, G. von (1988). Transcending the impenetrable: How proteins come to terms with membranes. *Biochimica Et Biophysica Acta (BBA) - Reviews on Biomembranes*, *947*, 307–333.

Hendrickson, W. A., & Ogata, C. M. (1997) [28] Phase determination from multiwavelength anomalous diffraction measurements. In *Methods in Enzymology. Macromolecular Crystallography Part A* 276, 494–523.

Henikoff, S., & Henikoff, J. G. (1992). Amino acid substitution matrices from protein blocks. *Proceedings of the National Academy of Sciences USA*, *89*, 10915–10919.

Henning J. (2014). Mutagenesestudie zur fünktionallen Charakterisierung der C-terminalen Domäne der variablen Oberflachenglykoproteine Afrikanischer Trypanosomen. Master thesis, Julius-Maximilians-Universität Würzburg.

Hereld, D., Krakow, J. L., Bangs, J. D., Hart, G. W., & Englund, P. T. (1986). A phospholipase C from *Trypanosoma brucei* which selectively cleaves the glycolipid on the variant surface glycoprotein. *The Journal of Biological Chemistry*, *261*, 13813–13819.

Hertz-Fowler, C., Figueiredo, L. M., Quail, M. A., Becker, M., Jackson, A., Bason, N., Brooks, K., Churcher, C., Fahkro, S., Goodhead, I., Heath, P., Kartvelishvili, M., Mungall, K., Harris, D., Hauser, H., Sanders, M., Saunders, D., Seeger, K., Sharp, S., . . . Berriman, M. (2008). Telomeric expression sites are highly conserved in *Trypanosoma brucei*. *PloS One*, *3*, e3527.

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 USA*, *110*, 1905–1910.

Horn, D. (2008). Codon usage suggests that translational selection has a major impact on protein expression in trypanosomatids. *BMC Genomics*, *9*, 2.

Horn, D. (2014). Antigenic variation in African trypanosomes. *Molecular and Biochemical Parasitology*, *195*, 123–129.

Horn, D., & Cross, G. A.M. (1997). Analysis of *Trypanosoma brucei* vsg expression site switching in vitro. *Molecular and Biochemical Parasitology*, *84*, 189–201.

Huber, D., Boyd, D., Xia, Y., Olma, M. H., Gerstein, M., & Beckwith, J. (2005). Use of thioredoxin as a reporter to identify a subset of Escherichia coli signal sequences that promote signal recognition particle-dependent translocation. *Journal of Bacteriology*, *187*, 2983–2991.

Hutchinson, O. C., Smith, W., Jones, N. G., Chattopadhyay, A., Welburn, S. C., & Carrington, M. (2003). VSG structure: similar N-terminal domains can form functional VSGs with different types of C-terminal domain. *Molecular and Biochemical Parasitology*, *130*, 127–131.

Ibel, K., & Stuhrmann, H. B. (1975). Comparison of neutron and X-ray scattering of dilute myoglobin solutions. *Journal of Molecular Biology*, *93*, 255–265.

Ikezawa, H. (2002). Glycosylphosphatidylinositol (GPI)-anchored proteins. *Biological & Pharmaceutical Bulletin*, *25*, 409–417.

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 Neglected Tropical Diseases*, *7*, e2121.

Jackson, A. P., Berry, A., Aslett, M., Allison, H. C., Burton, P., Vavrova-Anderson, J., Brown, R., Browne, H., Corton, N., Hauser, H., Gamble, J., Gilderthorp, R., Marcello, L., McQuillan, J., Otto, T. D., Quail, M. A., Sanders, M. J., van Tonder, A., Ginger, M. L., . . . Berriman, M. (2012). Antigenic diversity is generated by distinct evolutionary mechanisms in African trypanosome species. *Proceedings of the National Academy of Sciences USA*, *109*, 3416–3421.

Jackson, A. P., Goyard, S., Xia, D., Foth, B. J., Sanders, M., Wastling, J. M., Minoprio, P., & Berriman, M. (2015). Global Gene Expression Profiling through the Complete Life Cycle of *Trypanosoma vivax*. *PLoS Neglected Tropical Diseases*, *9*, e0003975.

Jamonneau, V., Ilboudo, H., Kaboré, J., Kaba, D., Koffi, M., Solano, P., Garcia, A., Courtin, D., Laveissière, C., Lingue, K., Büscher, P., & Bucheton, B. (2012). Untreated human infections by *Trypanosoma brucei gambiense* are not 100% fatal. *PLoS Neglected Tropical Diseases*, *6*, e1691.

Janzen, C. J., Lander, F., Dreesen, O., & Cross, G. A. M. (2004). Telomere length regulation and transcriptional silencing in KU80-deficient *Trypanosoma brucei*. *Nucleic Acids Research*, *32*, 6575–6584.

Jauncey, G. E. (1924). The Scattering of X-Rays and Bragg's Law. *Proceedings of the National Academy of Sciences USA*, *10*, 57–60.

Jeacock, L., Faria, J., & Horn, D. (2018). Codon usage bias controls mRNA and protein abundance in trypanosomatids. *ELife*, *7*.

Jeffries, T. R., Morgan, G. W., & Field, M. C. (2001). A developmentally regulated Rab11 homologue in *Trypanosoma brucei* is involved in recycling processes. *Journal of Cell Science*, *114*, 2617–2626.

Jones, N. G., Nietlispach, D., Sharma, R., Burke, D. F., Eyres, I., Mues, M., Mott, H. R., & Carrington, M. (2008). Structure of a glycosylphosphatidylinositol-anchored domain from a trypanosome variant surface glycoprotein. *The Journal of Biological Chemistry*, *283*, 3584–3593.

Jones, T. W., & Dávila, A. M.R. (2001). *Trypanosoma vivax* – out of Africa. *Trends in Parasitology*, *17*, 99–101.

Joshi, P. P., Shegokar, V. R., Powar, R. M., Herder, S., Katti, R., Salkar, H. R., Dani, V. S., Bhargava, A., Jannin, J., & Truc, P. (2005). Human trypanosomiasis caused by *Trypanosoma evansi* in India: The first case report. *The American Journal of Tropical Medicine and Hygiene*, *73*, 491–495.

Jung, I. (2018). Overexpression of *Trypanosoma vivax* ILDat1.2 variable surface glycoprotein in *T. brucei.* Bachelor thesis, Julius-Maximilians-Universität Würzburg.

Ke, H. (1997) Overview of isomorphous replacement phasing. *Macromolecular Crystallography,* 276, 448–461).

Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., & Sternberg, M. J. E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols*, *10*, 845–858.

Kelly, R. B. (1985). Pathways of protein secretion in eukaryotes. *Science*, *230*, 25–32.

Kendrew, J. C., Bodo, G., Dintsiz, H. M., Parrisg, R. G., WyckoffF, H., & Philipps S, D. C. (1958). A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature*, *181*, 662–666.

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

Kinoshita, T. (2020). Biosynthesis and biology of mammalian GPI-anchored proteins. *Open Biology*, *10*, 190290.

Kovářová, J., Nagar, R., Faria, J., Ferguson, M. A. J., Barrett, M. P., & Horn, D. (2018). Gluconeogenesis using glycerol as a substrate in bloodstream-form *Trypanosoma brucei*. *PLoS Pathogens*, *14*, e1007475.

Kraemer, S. M., & Smith, J. D. (2006). A family affair: Var genes, PfEMP1 binding, and malaria disease. *Current Opinion in Microbiology*, *9*, 374–380.

Kraus, A. J., Brink, B. G., & Siegel, T. N. (2019). Efficient and specific oligo-based depletion of rRNA. *Scientific Reports*, *9*, 12281.

Kunze, M., & Berger, J. (2015). The similarity between N-terminal targeting signals for protein import into different organelles and its evolutionary relevance. *Frontiers in Physiology*, *6*, 259.

Küpper, W., & Wolters, M. (1983). Observation on drug resistance of *Trypanosoma (Nannomonas) congolense* and *Trypanosoma (Duttonella) vivax* in cattle at a feedlot in the Northern Ivory Coast. *Tropenmedizin Und Parasitologie*, *34*, 203–205.

Kyte, J., & Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology*, *157*, 105–132.

Lamour, N., Rivière, L., Coustou, V., Coombs, G. H., Barrett, M. P., & Bringaud, F. (2005). Proline metabolism in procyclic *Trypanosoma brucei* is down-regulated in the presence of glucose. *The Journal of Biological Chemistry*, *280*, 11902–11910.

Lane-Serff, H., Macgregor, P., Peacock, L., Macleod, O. J., Kay, C., Gibson, W., Higgins, M. K., & Carrington, M. (2016). Evolutionary diversification of the trypanosome haptoglobin-haemoglobin receptor from an ancestral haemoglobin receptor. *ELife*, *5*.

Laxman, S., Riechers, A., Sadilek, M., Schwede, F., & Beavo, J. A. (2006). Hydrolysis products of cAMP analogs cause transformation of *Trypanosoma brucei* from slender to stumpy-like forms. *Proceedings of the National Academy of Sciences USA*, *103*, 19194–19199.

Lehane, M., Alfaroukh, I., Bucheton, B., Camara, M., Harris, A., Kaba, D., Lumbala, C., Peka, M., Rayaisse, J.-B., Waiswa, C., Solano, P., & Torr, S. (2016). Tsetse Control and the Elimination of Gambian Sleeping Sickness. *PLoS Neglected Tropical Diseases*, *10*, e0004437.

Li, B. (2015). Dna double-strand breaks and telomeres play important roles in *Trypanosoma brucei* antigenic variation. *Eukaryotic Cell*, *14*, 196–205.

Lillico, S., Field, M. C., Blundell, P., Coombs, G. H., & Mottram, J. C. (2003). Essential roles for GPI-anchored proteins in African trypanosomes revealed using mutants deficient in GPI8. *Molecular Biology of the Cell*, *14*, 1182–1194.

Lindner, A. K., Lejon, V., Chappuis, F., Seixas, J., Kazumba, L., Barrett, M. P., Mwamba, E., Erphas, O., Akl, E. A., Villanueva, G., Bergman, H., Simarro, P., Kadima Ebeja, A., Priotto, G., & Franco, J. R. (2020). New WHO guidelines for treatment of gambiense human African trypanosomiasis including fexinidazole: substantial changes for clinical practice. *The Lancet Infectious Diseases*, *20*, e38-e46.

Liu, D., Albergante, L., Newman, T. J., & Horn, D. (2018). Faster growth with shorter antigens can explain a VSG hierarchy during African trypanosome infections: A feint attack by parasites. *Scientific Reports*, *8*, 10922.

MacGregor, P., & Matthews, K. R. (2012). Identification of the regulatory elements controlling the transmission stage-specific gene expression of PAD1 in *Trypanosoma brucei*. *Nucleic Acids Research*, *40*, 7705–7717.

MacGregor, P., Savill, N. J., Hall, D., & Matthews, K. R. (2011). Transmission stages dominate trypanosome within-host dynamics during chronic infections. *Cell Host & Microbe*, *9*, 310–318.

Mahlab, S., & Linial, M. (2014). Speed controls in translating secretory proteins in eukaryotes-an evolutionary perspective. *PLoS Computational Biology*, *10*, e1003294.

Manful, T., Fadda, A., & Clayton, C. (2011). The role of the 5'-3' exoribonuclease XRNA in transcriptome-wide mRNA degradation. *RNA*, *17*, 2039–2047.

Manna, P. T., Boehm, C., Leung, K. F., Natesan, S. K., & Field, M. C. (2014). Life and times: Synthesis, trafficking, and evolution of VSG. *Trends in Parasitology*, *30*, 251–258.

Marcello, L., & Barry, J. D. (2007). Analysis of the VSG gene silent archive in *Trypanosoma brucei* reveals that mosaic gene expression is prominent in antigenic variation and is favored by archive substructure. *Genome Research*, *17*, 1344–1352.

Matthews, K. R., Tschudi, C., & Ullu, E. (1994). A common pyrimidine-rich motif governs trans-splicing and polyadenylation of tubulin polycistronic pre-mRNA in trypanosomes. *Genes & Development*, *8*, 491–501.

Matthews, K. R., Ellis, J. R., & Paterou, A. (2004). Molecular regulation of the life cycle of African trypanosomes. *Trends in Parasitology*, *20*, 40–47.

Maudlin, I. E., Kelly, S., Schwede, A., & Carrington, M. (2021). Vsg mRNA levels are regulated by the production of functional VSG protein. *Molecular and Biochemical Parasitology*, *241*, 111348.

Mbwambo, H. A., Mella, P. N., & Lekaki, K. A. (1988). Berenil (diminazene aceturate)-resistant *Trypanosoma congolense* in cattle under natural tsetse challenge at Kibaha, Tanzania. *Acta Tropica*, *45*, 239–244.

McCulloch, R., & Barry, J. D. (1999). A role for RAD51 and homologous recombination in *Trypanosoma brucei* antigenic variation. *Genes & Development*, *13*, 2875–2888.

McDowell, M. A., Ransom, D. M., & Bangs, J. D. (1998). Glycosylphosphatidylinositol-dependent secretory transport in *Trypanosoma brucei*. *The Biochemical Journal*, *335*, 681–689.

McPherson, A. (1989). Macromolecular crystals. *Scientific American*, *260*, 62–69.

McPherson, A. (1998). *Crystallization of biological macromolecules*. Cold Spring Harbor Laboratory Press.

McPherson, A., & Gavira, J. A. (2014). Introduction to protein crystallization. *Acta Crystallographica. Section F, Structural Biology Communications*, *70*, 2–20.

Mehlert, A., Wormald, M. R., & Ferguson, M. A. J. (2012). Modeling of the N-glycosylated transferrin receptor suggests how transferrin binding can occur within the surface coat of *Trypanosoma brucei*. *PLoS Pathogens*, *8*, e1002618.

Mertens, H. D. T., & Svergun, D. I. (2010). Structural characterization of proteins and complexes using small-angle X-ray solution scattering. *Journal of Structural Biology*, *172*, 128–141.

Miller, W. H. (1839). Treatise on Crystallography, Cambridge, England,

Muñoz-Jordán, J. L., Davies, K. P., & Cross, G. A. (1996). Stable expression of mosaic coats of variant surface glycoproteins in *Trypanosoma brucei*. *Science*, *272*, 1795–1797.

Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F., & Vagin, A. A. (2011). Refmac5 for the refinement of macromolecular crystal structures. *Acta Crystallographica. Section D, Biological Crystallography*, *67*, 355–367.

Nagamune, K., Ohishi, K., Ashida, H., Hong, Y., Hino, J., Kangawa, K., Inoue, N., Maeda, Y., & Kinoshita, T. (2003). Gpi transamidase of *Trypanosoma brucei* has two previously uncharacterized (trypanosomatid transamidase 1 and 2) and three common subunits. *Proceedings of the National Academy of Sciences USA*, *100*, 10682–10687.

Nakane, T., Kotecha, A., Sente, A., McMullan, G., Masiulis, S., Brown, P. M.G.E., Grigoras, I. T., Malinauskaite, L., Malinauskas, T., Miehling, J., Yu, L., Karia, D., Pechnikova, E. V., Jong, E. de, Keizer, J., Bischoff, M., McCormack, J., Tiemeijer, P., Hardwick, S. W., . . . Scheres, S. H.W. (2020). *Single-particle cryo-EM at atomic resolution, 587,* 152–156.

Navarro, M., & Gull, K. (2001). A pol I transcriptional body associated with VSG mono-allelic expression in *Trypanosoma brucei*. *Nature*, *414*, 759–763.

Needleman, S. B., & Wunsch, C. D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology*, *48*, 443–453.

Nielsen, H. (2017). Predicting Secretory Proteins with SignalP. *Methods in Molecular Biology*, *1611*, 59–73.

Nilsson, D., Gunasekera, K., Mani, J., Osteras, M., Farinelli, L., Baerlocher, L., Roditi, I., & Ochsenreiter, T. (2010). Spliced leader trapping reveals widespread alternative splicing patterns in the highly dynamic transcriptome of *Trypanosoma brucei*. *PLoS Pathogens*, *6*, e1001037.

Nilsson, I., Lara, P., Hessa, T., Johnson, A. E., Heijne, G. von, & Karamyshev, A. L. (2015). The code for directing proteins for translocation across ER membrane: Srp cotranslationally recognizes specific features of a signal sequence. *Journal of Molecular Biology*, *427*, 1191–1201.

Ooi, C.-P., Smith, T. K., Gluenz, E., Wand, N. V., Vaughan, S., & Rudenko, G. (2018). Blocking variant surface glycoprotein synthesis alters endoplasmic reticulum exit sites/Golgi homeostasis in *Trypanosoma brucei*. *Traffic*, *19*, 391–405.

Osório, A. L. A. R., Madruga, C. R., Desquesnes, M., Soares, C. O., Ribeiro, L. R. R., & Costa, S. C. G. d. (2008). *Trypanosoma (Duttonella) vivax*: Its biology, epidemiology, pathogenesis, and introduction in the New World--a review. *Memorias Do Instituto Oswaldo Cruz*, *103*, 1–13.

Pal, A., Hall, B. S., Jeffries, T. R., & Field, M. C. (2003). Rab5 and Rab11 mediate transferrin and anti-variant surface glycoprotein antibody recycling in *Trypanosoma brucei*. *The Biochemical Journal*, *374*, 443–451.

Palenchar, J. B., & Bellofatto, V. (2006). Gene transcription in trypanosomes. *Molecular and Biochemical Parasitology*, *146*, 135–141.

Palmer, G. H., & Brayton, K. A. (2007). Gene conversion is a convergent strategy for pathogen antigenic variation. *Trends in Parasitology*, *23*, 408–413.

Paulick, M. G., & Bertozzi, C. R. (2008). The glycosylphosphatidylinositol anchor: A complex membrane-anchoring structure for proteins. *Biochemistry*, *47*, 6991–7000.

Pays, E. (1991). Genetics of antigenic variation in African trypanosomes. *Research in Microbiology*, *142*, 731–735.

Pechmann, S., Chartron, J. W., & Frydman, J. (2014). Local slowdown of translation by nonoptimal codons promotes nascent-chain recognition by SRP in vivo. *Nature Structural & Molecular Biology*, *21*, 1100–1105.

Pérez-Morga, D., Vanhollebeke, B., Paturiaux-Hanocq, F., Nolan, D. P., Lins, L., Homblé, F., Vanhamme, L., Tebabi, P., Pays, A., Poelvoorde, P., Jacquet, A., Brasseur, R., & Pays, E. (2005). Apolipoprotein L-I promotes trypanosome lysis by forming pores in lysosomal membranes. *Science*, *309*, 469–472.

Pérez-Ortín, J. E., Miguel-Jiménez, L. de, & Chávez, S. (2012). Genome-wide studies of mRNA synthesis and degradation in eukaryotes. *Biochimica Et Biophysica Acta*, *1819*, 604–615.

Petoukhov, M. V., & Svergun, D. I. (2006). Joint use of small-angle X-ray and neutron scattering to study biological macromolecules in solution. *European Biophysics Journal*, *35*, 567–576.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, *25*, 1605–1612.

Pinger, J., Nešić, D., Ali, L., Aresta-Branco, F., Lilic, M., Chowdhury, S., Kim, H.-S., Verdi, J., Raper, J., Ferguson, M. A. J., Papavasiliou, F. N., & Stebbins, C. E. (2018). African trypanosomes evade immune clearance by O-glycosylation of the VSG surface coat. *Nature Microbiology*, *3*, 932–938.

Quax, T. E. F., Claassens, N. J., Söll, D., & van der Oost, J. (2015). Codon Bias as a Means to Fine-Tune Gene Expression. *Molecular Cell*, *59*, 149–161.

Raper, J., Fung, R., Ghiso, J., Nussenzweig, V., & Tomlinson, S. (1999). Characterization of a novel trypanosome lytic factor from human serum. *Infection and Immunity*, *67*, 1910–1916.

Rapoport, T. A. (2007). Protein translocation across the eukaryotic endoplasmic reticulum and bacterial plasma membranes. *Nature*, *450*, 663–669.

Reuner, B., Vassella, E., Yutzy, B., & Boshart, M. (1997). Cell density triggers slender to stumpy differentiation of *Trypanosoma brucei* bloodstream forms in culture. *Molecular and Biochemical Parasitology*, *90*, 269–280.

Rico, E., Rojas, F., Mony, B. M., Szoor, B., Macgregor, P., & Matthews, K. R. (2013). Bloodstream form pre-adaptation to the tsetse fly in *Trypanosoma brucei*. *Frontiers in Cellular and Infection Microbiology*, *3*, 78.

Ridewood, S., Ooi, C.-P., Hall, B., Trenaman, A., Wand, N. V., Sioutas, G., Scherwitzl, I., & Rudenko, G. (2017). The role of genomic location and flanking 3'UTR in the generation of functional levels of variant surface glycoprotein in. *Molecular Microbiology*, *106*, 614–634.

Ridgway, P., & Almouzni, G. (2001). Chromatin assembly and organization. *Journal of Cell Science*, *114*, 2711–2712.

Rifkin, M. R. (1978). Identification of the trypanocidal factor in normal human serum: High density lipoprotein. *Proceedings of the National Academy of Sciences, USA.*, *75*, 3450–3454.

Roditi, I., Carrington, M., & Turner, M. (1987). Expression of a polypeptide containing a dipeptide repeat is confined to the insect stage of *Trypanosoma brucei*. *Nature*, *325*, 272–274.

Roditi, I., Schwarz, H., Pearson, T. W., Beecroft, R. P., Liu, M. K., Richardson, J. P., Bühring, H. J., Pleiss, J., Bülow, R., & Williams, R. O. (1989). Procyclin gene expression and loss of the variant surface glycoprotein during differentiation of *Trypanosoma brucei*. *The Journal of Cell Biology*, *108*, 737–746.

Rojas, F., Silvester, E., Young, J., Milne, R., Tettey, M., Houston, D. R., Walkinshaw, M. D., Pérez-Pi, I., Auer, M., Denton, H., Smith, T. K., Thompson, J., & Matthews, K. R. (2019). Oligopeptide Signaling through TbGPR89 Drives Trypanosome Quorum Sensing. *Cell*, *176*, 306-317.e16.

Rossmann, M. G. (1990). The molecular replacement method. *Acta Crystallographica. Section A, Foundations of Crystallography*, *46*, 73–82.

Rossmann, M. G. (2001). Molecular replacement--historical background. *Acta Crystallographica. Section D, Biological Crystallography*, *57*, 1360–1366.

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.

Salmon, D., Geuskens, M., Hanocq, F., Hanocq-Quertier, J., Nolan, D., Ruben, L., & Pays, E. (1994). A novel heterodimeric transferrin receptor encoded by a pair of VSG expression site-associated genes in *T. brucei*. *Cell*, *78*, 75–86.

Schell, D., Evers, R., Preis, D., Ziegelbauer, K., Kiefer, H., Lottspeich, F., Cornelissen, A. W., & Overath, P. (1991). A transferrin-binding protein of *Trypanosoma brucei* is encoded by one of the genes in the variant surface glycoprotein gene expression site. *The EMBO Journal*, *10*, 1061–1066.

Scherf, A., Lopez-Rubio, J. J., & Riviere, L. (2008). Antigenic variation in *Plasmodium falciparum*. *Annual Review of Microbiology*, *62*, 445–470.

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

Schuster, S., Krüger, T., Subota, I., Thusek, S., Rotureau, B., Beilhack, A., & Engstler, M. (2017). Developmental adaptations of trypanosome motility to the tsetse fly host environments unravel a multifaceted in vivo microswimmer system. *ELife*, *6*.

Schwede, A., Macleod, O. J. S., Macgregor, P., & Carrington, M. (2015). How Does the VSG Coat of Bloodstream Form African Trypanosomes Interact with External Proteins? *PLoS Pathogens*, *11*, e1005259.

Shalem, O., Dahan, O., Levo, M., Martinez, M. R., Furman, I., Segal, E., & Pilpel, Y. (2008). Transient transcriptional responses to stress are generated by opposing effects of mRNA production and degradation. *Molecular Systems Biology*, *4*, 223.

Shapiro, S. Z., Naessens, J., Liesegang, B., Moloo, S. K., & Magondu, J. (1984). Analysis by flow cytometry of DNA synthesis during the life cycle of African trypanosomes. *Acta Tropica*, *41*, 313–323.

Sheader, K., Vaughan, S., Minchin, J., Hughes, K., Gull, K., & Rudenko, G. (2005). Variant surface glycoprotein RNA interference triggers a precytokinesis cell cycle arrest in African trypanosomes. *Proc. Natl. Acad. Sci.*, *102*, 8716–8721.

Sheldrick, G. M. (2010). Experimental phasing with SHELXC/D/E: Combining chain tracing with density modification. *Acta Crystallographica. Section D, Biological Crystallography*, *66*, 479–485.

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D., & Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, *7*, 539.

Silverman, J. M., Chan, S. K., Robinson, D. P., Dwyer, D. M., Nandan, D., Foster, L. J., & Reiner, N. E. (2008). Proteomic analysis of the secretome of *Leishmania donovani*. *Genome Biology*, *9*, R35.

Silvester, E., Ivens, A., & Matthews, K. R. (2018). A gene expression comparison of *Trypanosoma brucei* and *Trypanosoma congolense* in the bloodstream of the mammalian host reveals species-specific adaptations to density-dependent development. *PLoS Neglected Tropical Diseases*, *12*, e0006863.

Simarro, P. P., Cecchi, G., Paone, M., Franco, J. R., Diarra, A., Ruiz, J. A., Fèvre, E. M., Courtin, F., Mattioli, R. C., & Jannin, J. G. (2010). The Atlas of human African trypanosomiasis: A contribution to global mapping of neglected tropical diseases. *International Journal of Health Geographics*, *9*, 57.

Singh, P., James, R. S., Mee, C. J., & Morozov, I. Y. (2019). Mrna levels are buffered upon knockdown of RNA decay and translation factors via adjustment of transcription rates in human HepG2 cells. *RNA Biology*, *16*, 1147–1155.

Smith, T. K., Bringaud, F., Nolan, D. P., & Figueiredo, L. M. (2017). Metabolic reprogramming during the *Trypanosoma brucei* life cycle. *F1000Research*, *6*.

Smith, T. K., Vasileva, N., Gluenz, E., Terry, S., Portman, N., Kramer, S., Carrington, M., Michaeli, S., Gull, K., & Rudenko, G. (2009). Blocking variant surface glycoprotein synthesis in *Trypanosoma brucei* triggers a general arrest in translation initiation. *PloS One*, *4*, e7532.

Solano, P., Torr, S. J., & Lehane, M. J. (2013). Is vector control needed to eliminate gambiense human African trypanosomiasis? *Frontiers in Cellular and Infection Microbiology*, *3*, 33.

Specht, A. (2013). Charakterisierung der Funktion der N-Glykane des variablen Oberflächenglykoproteins MITat1.5 von *Trypanosoma brucei*. Master thesis, Julius-Maximilians-Universität Würzburg.

Spieth, J., Brooke, G., Kuersten, S., Lea, K., & Blumenthal, T. (1993). Operons in *C. elegans*: Polycistronic mRNA precursors are processed by trans-splicing of SL2 to downstream coding regions. *Cell*, *73*, 521–532.

Stevens, J. R., & Gibson, W. C. (1999). The evolution of pathogenic trypanosomes. *Parasitology Today*, *15*, 432–437.

Steverding, D., Stierhof, Y. D., Chaudhri, M., Ligtenberg, M., Schell, D., Beck-Sickinger, A. G., & Overath, P. (1994). Esag 6 and 7 products of *Trypanosoma brucei* form a transferrin binding protein complex. *European Journal of Cell Biology*, *64*, 78–87.

Stockdale, C., Swiderski, M. R., Barry, J. D., & McCulloch, R. (2008). Antigenic variation in *Trypanosoma brucei*: Joining the DOTs. *PLoS Biology*, *6*, e185.

Su, X.-z., Heatwole, V. M., Wertheimer, S. P., Guinet, F., Herrfeldt, J. A., Peterson, D. S., Ravetch, J. A., & Wellems, T. E. (1995). The large diverse gene family var encodes proteins involved in cytoadherence and antigenic variation of *plasmodium falciparum*-infected erythrocytes. *Cell*, *82*, 89–100.

Sun, M., Schwalb, B., Schulz, D., Pirkl, N., Etzold, S., Larivière, L., Maier, K. C., Seizl, M., Tresch, A., & Cramer, P. (2012). Comparative dynamic transcriptome analysis (cDTA) reveals mutual feedback between mRNA synthesis and degradation. *Genome Research*, *22*, 1350–1359.

Svergun, D. I., & Koch, M. H. J. (2003). Small-angle scattering studies of biological macromolecules in solution. *Reports on Progress in Physics*, *66*, 1735–1782.

Taylor, G. L. (2010). Introduction to phasing. *Acta Crystallographica. Section D, Biological Crystallography*, *66*, 325–338.

Taylor, J. E., & Rudenko, G. (2006). Switching trypanosome coats: What's in the wardrobe? *Trends in Genetics*, *22*, 614–620.

Tiengwe, C., Muratore, K. A., & Bangs, J. D. (2016). Surface proteins, ERAD and antigenic variation in *Trypanosoma brucei*. *Cellular Microbiology*, *18*, 1673–1688.

Timmers, H. T. M., & Tora, L. (2018). Transcript Buffering: A Balancing Act between mRNA Synthesis and mRNA Degradation. *Molecular Cell*, *72*, 10–17.

Trevor, C. E., Gonzalez-Munoz, A. L., Macleod, O. J. S., Woodcock, P. G., Rust, S., Vaughan, T. J., Garman, E. F., Minter, R., Carrington, M., & Higgins, M. K. (2019). Structure of the trypanosome transferrin receptor reveals mechanisms of ligand recognition and immune evasion. *Nature Microbiology*, *4*, 2074–2081.

Triggs, V. P., & Bangs, J. D. (2003). Glycosylphosphatidylinositol-dependent protein trafficking in bloodstream stage *Trypanosoma brucei*. *Eukaryotic Cell*, *2*, 76–83.

Turner, C. M., Aslam, N., & Dye, C. (1995). Replication, differentiation, growth and the virulence of *Trypanosoma brucei* infections. *Parasitology*, *111*, 289–300.

Tyler, K. M., Matthews, K. R., & Gull, K. (1997). The bloodstream differentiation-division of *Trypanosoma brucei* studied using mitochondrial markers. *Proceedings of the royal society B, Biological Sciences*, *264*, 1481–1490.

Ullu, E., Matthews, K. R., & Tschudi, C. (1993). Temporal order of RNA-processing reactions in trypanosomes: Rapid trans splicing precedes polyadenylation of newly synthesized tubulin transcripts. *Molecular and Cellular Biology*, *13*, 720–725.

Umaer, K., Bush, P. J., & Bangs, J. D. (2018). Rab11 mediates selective recycling and endocytic trafficking in *Trypanosoma brucei*. *Traffic*, *19*, 406–420.

Uzureau, P., Uzureau, S., Lecordier, L., Fontaine, F., Tebabi, P., Homblé, F., Grélard, A., Zhendre, V., Nolan, D. P., Lins, L., Crowet, J.-M., Pays, A., Felu, C., Poelvoorde, P., Vanhollebeke, B., Moestrup, S. K., Lyngsø, J., Pedersen, J. S., Mottram, J. C., . . . Pays, E. (2013). Mechanism of *Trypanosoma brucei gambiense* resistance to human serum. *Nature*, *501*, 430–434.

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.

van Grinsven, K. W. A., van den Abbeele, J., van den Bossche, P., van Hellemond, J. J., & Tielens, A. G. M. (2009). Adaptations in the glucose metabolism of procyclic Trypanosoma brucei isolates from tsetse flies and during differentiation of bloodstream forms. *Eukaryotic Cell*, *8*, 1307–1311.

van Vinh Chau, N., Le Buu Chau, Desquesnes, M., Herder, S., Phu Huong Lan, N., Campbell, J. I., van Cuong, N., Yimming, B., Chalermwong, P., Jittapalapong, S., Ramon Franco, J., Tri Tue, N., Rabaa, M. A., Carrique-Mas, J., Pham Thi Thanh, T., Tran Vu Thieu, N., Berto, A., Thi Hoa, N., van Minh Hoang, N., . . . Baker, S. (2016). A Clinical and Epidemiological Investigation of the First Reported Human Infection With the Zoonotic Parasite *Trypanosoma evansi* in Southeast Asia. *Clinical Infectious Diseases : An Official Publication of the Infectious Diseases Society of America*, *62*, 1002–1008.

Vanhamme, L., & Pays, E. (1995). Control of gene expression in trypanosomes. *Microbiological Reviews*, *59*, 223–240.

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

Vanhollebeke, B., Muylder, G. de, Nielsen, M. J., Pays, A., Tebabi, P., Dieu, M., Raes, M., Moestrup, S. K., & Pays, E. (2008). A haptoglobin-hemoglobin receptor conveys innate immunity to *Trypanosoma brucei* in humans. *Science*, *320*, 677–681.

Vanhollebeke, B., & Pays, E. (2010). The trypanolytic factor of human serum: Many ways to enter the parasite, a single way to kill. *Molecular Microbiology*, *76*, 806–814.

Vassella, E., Den Abbeele, J. V., Bütikofer, P., Renggli, C. K., Furger, A., Brun, R., & Roditi, I. (2000). A major surface glycoprotein of *Trypanosoma brucei* is expressed transiently during development and can be regulated post-transcriptionally by glycerol or hypoxia. *Genes & Development*, *14*, 615–626.

Vassella, E., Reuner, B., Yutzy, B., & Boshart, M. (1997). Differentiation of African trypanosomes is controlled by a density sensing mechanism which signals cell cycle arrest via the cAMP pathway. *Journal of Cell Science*, *110*, 2661–2671.

Vickerman, K. (1976). The diversity of the kinetoplastid flagellates. In: Lumsden WHR, Evans DA, editors. *Biology of the kinetoplastida*. London: Academic Press. p. 1–34.

Vickerman, K. (1969). On the surface coat and flagellar adhesion in trypanosomes. *Journal of Cell Science*, *5*, 163–193.

Vickerman, K. (1985). Developmental cycles and biology of pathogenic trypanosomes. *British Medical Bulletin*, *41*, 105–114.

Viegas, I. J., Macedo, J. P. de, Niz, M. de, Rodrigues, J. A., Aresta-Branco, F., Jaffrey, S. R., & Figueiredo, L. M. (2020). N6 -methyladenosine in poly(A) tails stabilize VSG transcripts. *BioRxiv* 10.1101/2020.01.30.925776

Vos, G. J., & Gardiner, P. R. (1990). Antigenic relatedness of stocks and clones of *Trypanosoma vivax* from east and west Africa. *Parasitology*, *100*, 101–106.

Walter, P., Gilmore, R., & Blobel, G. (1984). Protein translocation across the endoplasmic reticulum. *Cell*, *38*, 5–8.

Walter, P., & Johnson, A. E. (1994). Signal sequence recognition and protein targeting to the endoplasmic reticulum membrane. *Annual Review of Cell Biology*, *10*, 87–119.

Wang, J., Böhme, U., & Cross, G. A.M. (2003). Structural features affecting variant surface glycoprotein expression in *Trypanosoma brucei*. *Molecular and Biochemical Parasitology*, *128*, 135–145.

Weber, P. C. Overview of protein crystallization methods. In *Macromolecular Crystallography 1997*, 276, 13–22.

Wells, E. A. (1972). The importance of mechanical transmission in the epidemiology of nagana: A review. *Tropical Animal Health and Production*, *4*, 74–88.

Wickner, W., & Schekman, R. (2005). Protein translocation across biological membranes. *Science*, *310*, 1452–1456.

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. W., 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 Crystallographica. Section D, Biological Crystallography*, *67*, 235–242.

Wüthrich, K. (2001). The way to NMR structures of proteins. *Nature Structural Biology*, *8*, 923–925.

Yu, X., Jin, L., & Zhou, Z. H. (2008). 3.88 A structure of cytoplasmic polyhedrosis virus by cryo-electron microscopy. *Nature*, *453*, 415–419.

Yun, O., Priotto, G., Tong, J., Flevaud, L., & Chappuis, F. (2010). Nect is next: Implementing the new drug combination therapy for *Trypanosoma brucei gambiense* sleeping sickness. *PLoS Neglected Tropical Diseases*, *4*, e720.

Zeelen, J., van Straaten, M., Verdi, J., Hempelmann, A., Hashemi, H., Perez, K., Jeffrey, P. D., Hälg, S., Wiedemar, N., Mäser, P., Papavasiliou, F. N., & Stebbins, C. E. (2021). Structure of trypanosome coat protein VSGsur and function in suramin resistance. *Nature Microbiology*, *6*, 392–400.

Zíková, A., Verner, Z., Nenarokova, A., Michels, P. A. M., & Lukeš, J. (2017). A paradigm shift: The mitoproteomes of procyclic and bloodstream *Trypanosoma brucei* are comparably complex. *PLoS Pathogens*, *13*, e1006679.

Zimmermann, H., Subota, I., Batram, C., Kramer, S., Janzen, C. J., Jones, N. G., & Engstler, M. (2017). A quorum sensing-independent path to stumpy development in *Trypanosoma brucei*. *PLoS Pathogens*, *13*, e1006324.

Zoll, S., Lane-Serff, H., Mehmood, S., Schneider, J., Robinson, C. V., Carrington, M., & Higgins, M. K. (2018). The structure of serum resistance-associated protein and its implications for human African trypanosomiasis. *Nature Microbiology*, *3*, 295–301.

6.0 Appendix

6.1 Supplementary figures

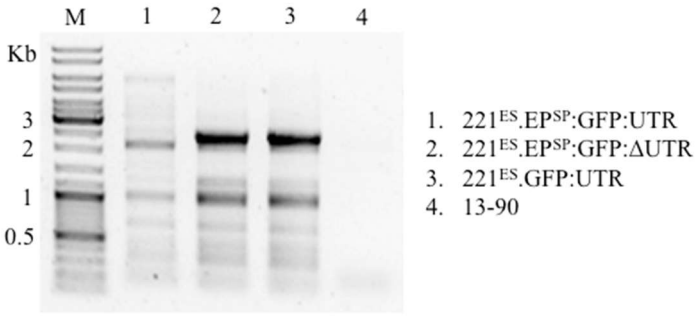

Figure 34: Integration PCR.
A 2368 bp fragment was amplified, confirming the correct integration of the GFP reporter constructs downstream of ES resident *VSG221*. The 13-90 cells and the 221ES.GFP:UTR cells served as the negative and positive controls, respectively.

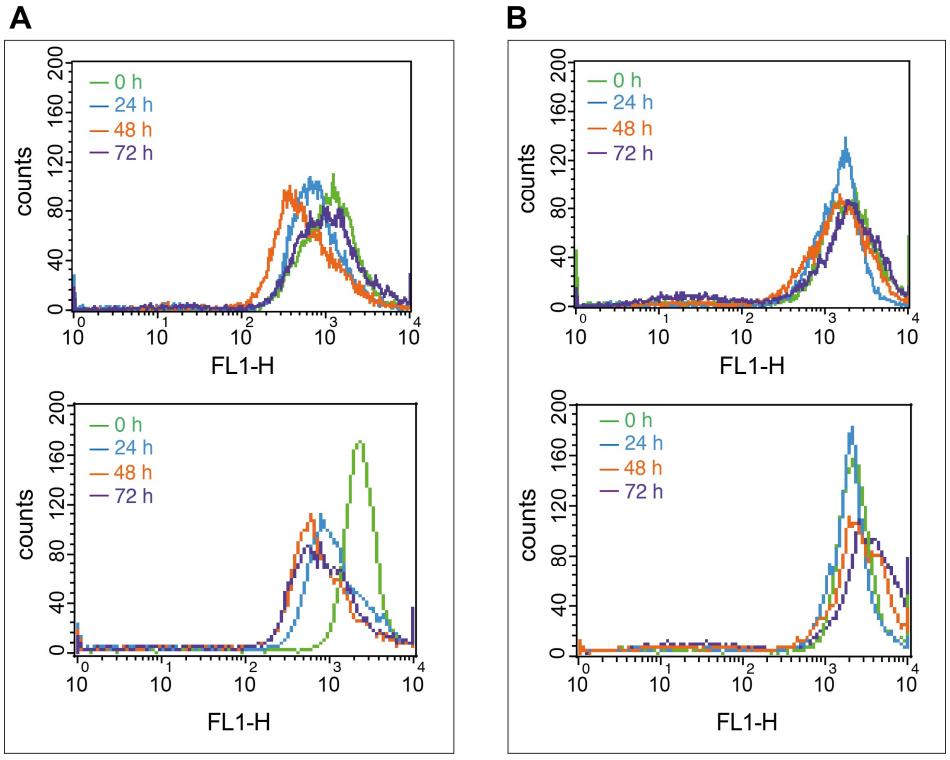

Figure 35: Expression site attenuation.
The expression site activity was measured before (0 h) and after inducing expression of the ectopic VSGs at different timepoints (24, 48,72 h). Intensity of the GFP reporter in **(A)** the GFP^{PRO}.221ES.121^{tet} control and **(B)**, GFP^{PRO}.221ES.ILDat1.2^{tet} cell lines were quantified by flow cytometry. The upper and lower panels represent two independent clones. 25,000 events were recorded at each time point.

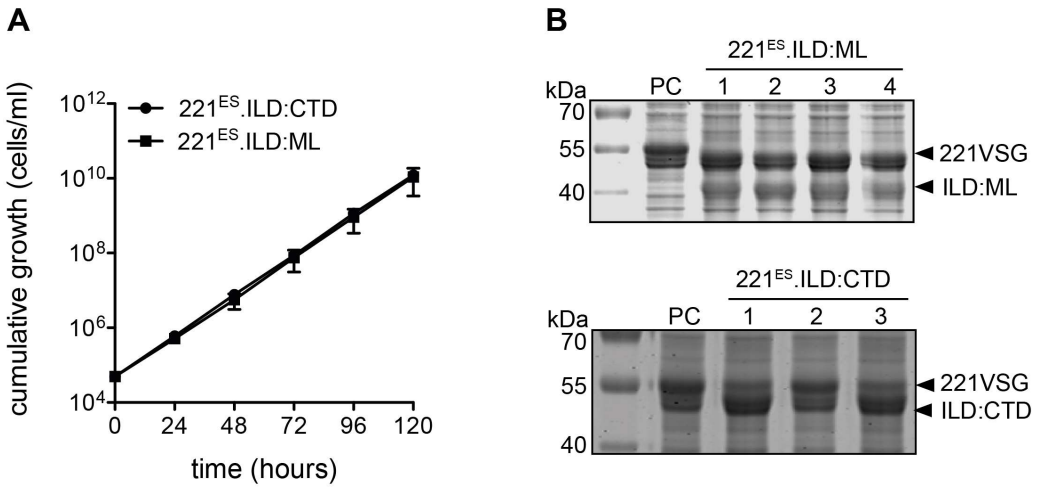

Figure 36: Expression of VSG chimeras in *T. brucei*.
(A) Growth curves of the 221ES.ILD:ML and 221ES.ILD:CTD cell lines. The values are averages of three independent clones and the error bars represent ± SD. **(B)** SDS-PAGE analysis of VSG protein expression in the 221ES.ILD:ML and **(C)** 221ES.ILD:CTD cell lines. Clone 2 of the 221ES.ILD:CTD cells appeared to be negative. The 13-90 parental cells (P) served as the VSG221 expression control.

6.2 Supplementary tables

Table 13: Fine-screen formulations for sitting drop diffusion

Table 1/2

Well	Salt	Buffer	Precipitant
A1	0.2 M NaCl	0.1 M HEPES pH 7.5	25% w/v PEG 4K
A2	0.2 M NaCl	0.1 M HEPES pH 7.5	25% w/v PEG 4K
A3	0.2 M NaCl	0.1 M HEPES pH 7.5	25% w/v PEG 4K
A4	0.2 M NaCl	0.1 M HEPES pH 7.5	23% w/v PEG 4K
A5	0.2 M NaCl	0.1 M HEPES pH 7.5	20% w/v PEG 4K
A6	0.2 M NaCl	0.1 M HEPES pH 7.5	18% w/v PEG 4K
A7	0.2 M NaCl	0.1 M HEPES pH 7.5	15% w/v PEG 4K
A8	0.2 M NaCl	0.1 M HEPES pH 7.5	25% w/v PEG 3.35K
A9	0.2 M NaCl	0.1 M HEPES pH 7.5	23% w/v PEG 3.35K
A10	0.2 M NaCl	0.1 M HEPES pH 7.5	20% w/v PEG 3.35K
A11	0.2 M NaCl	0.1 M HEPES pH 7.5	18% w/v PEG 3.35K
A12	0.2 M NaCl	0.1 M HEPES pH 7.5	15% w/v PEG 3.35K
B1	0.15 M NaCl	0.1 M HEPES pH 7.5	25% w/v PEG 4K
B2	0.2 M NaCl	0.1 M HEPES pH 7.6	25% w/v PEG 4K
B3	0.2 M NaCl	0.1 M HEPES pH 7.0	25% w/v PEG 4K
B4	0.2 M NaCl	0.1 M HEPES pH 7.6	20% w/v PEG 4K
B5	0.2 M NaCl	0.1 M HEPES pH 7.6	20% w/v PEG 3K
B6	0.2 M NaCl	0.1 M HEPES pH 7.6	18% w/v PEG 3K
B7	0.2 M NaCl	0.1 M HEPES pH 7.5	20% w/v PEG 3K
B8	0.2 M NaCl	0.1 M HEPES pH 7.5	20% w/v PEG 3K
B9	0.2 M NaCl	0.1 M HEPES pH 7.5	20% w/v PEG 3K
B10	0.2 M NaCl	0.1 M HEPES pH 7.5	20% w/v PEG 3K
B11	0.2 M NaCl	0.1 M HEPES pH 7.5	20% w/v PEG 3K
B12	0.2 M NaCl	0.1 M HEPES pH 7.5	25% w/v PEG 3K
C1	0.2 M NaCl	0.1 M HEPES pH 7.5	23% w/v PEG 3K
C2	0.2 M NaCl	0.1 M HEPES pH 7.5	18% w/v PEG 3K
C3	0.2 M NaCl	0.1 M HEPES pH 7.5	15% w/v PEG 3K
C4	0.2 M NaCl	0.1 M HEPES pH 7.0	20% w/v PEG 3K
C5	0.15 M NaCl	0.1 M HEPES pH 7.5	20% w/v PEG 3K
C6	0.10 M NaCl	0.1 M HEPES pH 7.5	20% w/v PEG 3K
C7		0.1 M HEPES pH 7.0	15% w/v PEG 20K
C8		0.1 M HEPES pH 7.5	15% w/v PEG 20K
C9	0.2 M NaCl	0.1 M HEPES pH 7.0	15% w/v PEG 20K
C10	0.2 M NaCl	0.1 M HEPES pH 7.0	15% w/v PEG 20K
C11	0.2 M NaCl	0.1 M HEPES pH 7.5	15% w/v PEG 20K
C12	0.2 M NaCl	0.1 M HEPES pH 7.5	15% w/v PEG 20K
D1	0.2 M KCl	0.1 M MES pH 6.5	25% w/v PEG 4K
D2	0.2 M KCl	0.1 M MES pH 6.5	23% w/v PEG 4K
D3	0.2 M KCl	0.1 M MES pH 6.5	20% w/v PEG 4K
D4	0.2 M KCl	0.1 M MES pH 6.5	18% w/v PEG 4K
D5	0.2 M KCl	0.1 M MES pH 6.5	15% w/v PEG 4K
D6	0.1 M KCl	0.1 M MES pH 6.5	25% w/v PEG 4K
D7	0.1 M KCl	0.1 M MES pH 6.5	23% w/v PEG 4K
D8	0.1 M KCl	0.1 M MES pH 6.5	20% w/v PEG 4K
D9	0.1 M KCl	0.1 M MES pH 6.5	18% w/v PEG 4K
D10	0.2 M KCl	0.1 M MES pH 6.0	15% w/v PEG 4K
D11	0.2 M KCl	0.1 M MES pH 6.0	23% w/v PEG 4K
D12	0.2 M KCl	0.1 M MES pH 6.0	20% w/v PEG 4K

Table 2/2

Well	Salt	Buffer	Precipitant
E1	0.2 M KCl	0.1 M MES pH 6.0	18% w/v PEG 4K
E2	0.2 M KCl	0.1 M MES pH 6.0	15% w/v PEG 4K
E3	0.1 M KCl	0.1 M MES pH 6.5	25% w/v PEG 4K
E4	0.1 M KCl	0.1 M MES pH 6.5	23% w/v PEG 4K
E5	0.1 M KCl	0.1 M MES pH 6.5	20% w/v PEG 4K
E6	0.1 M KCl	0.1 M MES pH 6.5	18% w/v PEG 4K
E7	0.1 M KCl	0.1 M MES pH 6.5	15% w/v PEG 4K
E8	0.5% v/v Tacsimate	0.1 M HEPES pH 7.0	10% w/v PEG MME 5K
E9	0.5% v/v Tacsimate	0.1 M HEPES pH 7.0	25% w/v PEG 4K
E10	0.5% v/v Tacsimate	0.1 M HEPES pH 7.0	23% w/v PEG 4K
E11	0.5% v/v Tacsimate	0.1 M HEPES pH 7.0	20% w/v PEG 4K
E12	0.5% v/v Tacsimate	0.1 M HEPES pH 7.0	18% w/v PEG 4K
F1	0.5% v/v Tacsimate	0.1 M HEPES pH 7.0	15% w/v PEG 4K
F2	0.5% v/v Tacsimate	0.1 M HEPES pH 7.5	10% w/v PEG MME 5K
F3	0.5% v/v Tacsimate	0.1 M HEPES pH 7.5	25% w/v PEG 4K
F4	0.5% v/v Tacsimate	0.1 M HEPES pH 7.5	23% w/v PEG 4K
F5	0.5% v/v Tacsimate	0.1 M HEPES pH 7.5	20% w/v PEG 4K
F6	0.5% v/v Tacsimate	0.1 M HEPES pH 7.5	18% w/v PEG 4K
F7	0.5% v/v Tacsimate	0.1 M HEPES pH 7.5	15% w/v PEG 4K
F8	0.2 M $MgCl_2$	0.1 M TRIS pH 8.5	25% w/v PEG 3.35K
F9	0.2 M $MgCl_2$	0.1 M TRIS pH 8.5	23% w/v PEG 3.35K
F10	0.2 M $MgCl_2$	0.1 M TRIS pH 8.5	20% w/v PEG 3.35K
F11	0.2 M $MgCl_2$	0.1 M TRIS pH 8.5	18% w/v PEG 3.35K
F12	0.2 M $MgCl_2$	0.1 M TRIS pH 8.5	15% w/v PEG 3.35K
G1	0.2 M $MgCl_2$	0.1 M TRIS pH 8.0	25% w/v PEG 3.35K
G2	0.2 M $MgCl_2$	0.1 M TRIS pH 8.0	23% w/v PEG 3.35K
G3	1 M LiCl		23% w/v PEG 4K
G4	1 M LiCl		20% w/v PEG 4K
G5	1 M LiCl		18% w/v PEG 4K
G6	1 M LiCl		15% w/v PEG 4K
G7	0.5 M LiCl		23% w/v PEG 4K
G8	0.5 M LiCl		23% w/v PEG 4K
G9	0.5 M LiCl		23% w/v PEG 4K
G10	1 M NH_4I	0.1 M BICINE pH 9.0	18% w/v PEG 4K
G11	1 M NH_4I	0.1 M BICINE pH 9.0	27% w/v PEG 4K
G12	1 M NH_4I	0.1 M BICINE pH 9.0	25% w/v PEG 4K
H1	1 M NH_4I	0.1 M BICINE pH 9.0	23% w/v PEG 4K
H2	1 M NH_4I	0.1 M BICINE pH 9.0	20% w/v PEG 4K
H3	1 M NH_4I	0.1 M BICINE pH 9.0	15% w/v PEG 4K
H4	0.4 M KNO_3	0.1 M MES pH 7.5	23% w/v PEG 4K
H5	0.4 M KNO_3	0.1 M MES pH 7.5	27% w/v PEG 4K
H6	0.4 M KNO_3	0.1 M MES pH 7.5	25% w/v PEG 4K
H7	0.4 M KNO_3	0.1 M MES pH 7.5	20% w/v PEG 4K
H8	0.4 M KNO_3	0.1 M MES pH 7.5	18% w/v PEG 4K
H9	0.4 M KNO_3	0.1 M MES pH 7.5	15% w/v PEG 4K
H10	0.4 M KNO_3	0.1 M MES pH 7.0	23% w/v PEG 4K
H11	0.2 M KCl	0.1 M MES pH 7.0	23% w/v PEG 4K
H12	0.2 M KCl	0.1 M MES pH 7.0	20% w/v PEG 4K

Table 14: Fine-screen formulations for hanging drop diffusion

Well	Salt	Buffer	Precipitant
A1	0.2 M KCl	0.1 M MES pH 6.5	25% PEG 4000
A2	0.2 M KCl	0.1 M MES pH 6.5	23% PEG 4000
A3	0.2 M KCl	0.1 M MES pH 6.5	20% PEG 4000
A4	0.2 M KCl	0.1 M MES pH 6.5	18% PEG 4000
A5	0.2 M KCl	0.1 M MES pH 6.0	30% PEG 4000
A6	0.2 M KCl	0.1 M MES pH 6.0	25% PEG 4000
A7	0.2 M KCl	0.1 M MES pH 6.0	23% PEG 4000
A8	0.2 M KCl	0.1 M MES pH 6.0	20% PEG 4000
A9	0.2 M KCl	0.1 M MES pH 5.5	30% PEG 4000
A10	0.2 M KCl	0.1 M MES pH 5.5	25% PEG 4000
A11	0.2 M KCl	0.1 M MES pH 5.5	23% PEG 4000
A12	0.2 M KCl	0.1 M MES pH 5.5	20% PEG 4000
B1	0.1 M KCl	0.1 M MES pH 6.5	25% PEG 4000
B2	0.1 M KCl	0.1 M MES pH 6.5	23% PEG 4000
B3	0.1 M KCl	0.1 M MES pH 6.5	25% PEG 3350
B4	0.1 M KCl	0.1 M MES pH 6.5	20% PEG 3350
B5	0.2 M NaCl	0.1 M HEPES pH 7.5	25% PEG 4000
B6	0.2 M NaCl	0.1 M HEPES pH 7.5	23% PEG 4000
B7	0.2 M NaCl	0.1 M HEPES pH 7.5	20% PEG 4000
B8	0.2 M NaCl	0.1 M HEPES pH 7.5	25% PEG 3350
B9	0.2 M NaCl	0.1 M HEPES pH 7.5	23% PEG 3350
B10	0.2 M NaCl	0.1 M HEPES pH 7.5	25% PEG 3000
B11	0.2 M NaCl	0.1 M HEPES pH 7.5	20% PEG 3000
B12	0.2 M NaCl	0.1 M HEPES pH 7.5	20% PEG 3000

6.3 List of abbreviations

3'UTR	3' untranslated region
5'UTR	5' untranslated region
µg	microgram
µl	microliter
µM	micromolar
aa	amino acid
AAT	animal African trypanosomiasis
BES	bloodstream expression site
BLE	phleomycin
BSA	bovine serum albumin
BSF	bloodstream form
bp	base pairs
cAMP	cyclic adenosine monophosphate
CTD	C-terminal domain
CIAP	calf intestinal alkaline phosphatase
DAPI	4′, 6-diamidino-2-phenylindole
ddH_20	double-distilled water
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
dNTP	deoxynucleotide triphosphate
DSB	double strand breaks
EGFP	enhanced green fluorescent protein
ESAG	expression site associated genes
ESB	expression site body
FACS	fluorescence assisted cell sorting
FCS	fetal calf serum
g	gram
g	acceleration due to gravity
G0/1	Gap 0/1 phase
GPI	glycosylphosphatidylinositol
HAT	human animal trypanosomiasis
h	hours
ISG	Invariant surface glycoprotein
kDa	kilodalton
l	liter
LB	Luria Bertani
M	molar
mM	millimolar
ml	milliliter

mRNA	messenger ribonucleic acid
NECT	nifurtimox-eflornithine combination therapy
NLS	nucleus localization signal
NTD	N-terminal domain
ORF	open reading frame
PAD	protein associated with differentiation
PAGE	polyacrylamide gel electrophoresis
PARP	procyclic acid repetitive protein
PEG	polyethylene glycol
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PDT	population doubling time
PFA	paraformaldehyde
PFR	paraflagellar rod
Pol I	polymerase I
PVDF	polyvinylidene difluoride
RBP	RNA binding protein
rDNA	ribosomal DNA
RNA	ribonucleic acid
RNAi	RNA interference
RT	room temperature
SD	standard deviation
SDS	sodium dodecyl sulfate
sec	second
SIF	stumpy induction factor
SRA	Serum resistance-associated
sVSG	Soluble Variant surface glycoprotein
TbHpHbR	*T. brucei* haptoglobin-hemoglobin receptor
*Tb*TfR	*T. brucei* Transferrin receptor
TDB	trypanosome dilution buffer
tet	tetracycline
TFR	transferrin receptor
TLF	trypanolytic factor
VSG	variant surface glycoprotein
v/v	volume per volume
WT	wild type
w/v	weight per volume

6.4 Affidavit

Erklärungen nach §4 Abs. 3 Satz 3, 5, 8 der Promotionsordnung der Fakultät für Biologie

I hereby declare that my thesis entitled: **"*Trans*-regulation of *Trypanosoma brucei* variant surface glycoproteins (VSG) mRNA and structural analysis of a *Trypanosoma vivax* VSG using X-ray crystallography"** is the result of my own work.
I did not receive any help or support from commercial consultants. All sources and/or materials applied are listed and specified in the thesis.
Furthermore, I verify that the thesis has not been submitted as part of another examination process neither in identical nor in similar form.

Eidesstattliche Erklärung

Hiermit erkläre ich an Eides statt, die Dissertation: **"*Trans*-regulation of *Trypanosoma brucei* variant surface glycoproteins (VSG) mRNA and structural analysis of a *Trypanosoma vivax* VSG using X-ray crystallography"**, eigenständig, d. h. insbesondere selbständig und ohne Hilfe eines kommerziellen Promotionsberaters, angefertigt und keine anderen als die von mir angegebenen Quellen und Hilfsmittel verwendet zu haben.
Ich erkläre außerdem, dass die Dissertation weder in gleicher noch in ähnlicher Form bereits in einem anderen Prüfungsverfahren vorgelegen hat.

6.5 Publications

Parts of this thesis are in preparation for submission to peer reviewed journals for publication.

1. Endoplasmic reticulum-targeting but not translation is required for mRNA balancing in trypanosomes. (Manuscript submitted in BioRxiv)
https://www.biorxiv.org/content/10.1101/2021.05.05.442555v1

2. Structure of a *Trypanosoma vivax* variant surface glycoprotein shows structural conservation of African trypanosome surface proteins.
(Manuscript in preparation)

6.6 Acknowledgements

I thank the almighty God for granting me good health and a sound mind during the entire period of my studies. I wish to express my deepest gratitude to Markus Engstler for giving me the immense opportunity to pursue my studies in his lab and for the exemplary supervision and mentorship. His funding beyond my scholarship period ensured that I completed this work. In a special way, I am very grateful to Nicola Jones for her ever-present support and guidance during this work, thank you for the training on cell biology techniques and the numerous discussions that we held. I thank Daniel Masiga for being part of my thesis committee.

My appreciation goes to the Deutsche Akademischer Austausch Dienst (DAAD) for awarding me an excellent Ph.D. scholarship that facilitated my research stay in Würzburg.

I must give thanks to the former and current members of the Cell- and Developmental Biology department. Thank you for being resourceful colleagues and friends in and out of the lab. I thank the technicians for ensuring the smooth running of the labs. I profoundly thank Uli and Manu for their unmatched support in the secretariat and for making sure that I had all the necessary documents for administrative purposes in time.

I also thank Jochen and Wolfgang, from the Kisker lab, for their support and technical knowledge in setting up protein crystallization experiments and analysis of X-ray diffraction data.

Blaise, Carina, Ezekiel, Irene, Majeed, Marie, and Obala; thank you for your friendship, encouragement, and for being part of this exciting journey. I am grateful to the entire “Kenyans in Würzburg” fraternity. Our monthly meetings and group activities gave Würzburg a feel of home away from home.

I acknowledge my parents for their unending love and for putting me through school against all odds. Thank you for your unwavering support and belief in me. I thank my brothers and sisters for their unfailing support, encouragement, and the deep family ties. Your constant messages of goodwill kept me going. To my fiancé Elizabeth, thank you for your great support, patience and all your love.