

Biomedical &  
Life Sciences

Lancaster  
University



**Characterisation of the *ESAG6* and *ESAG7* 3'UTRs involved  
in the iron starvation response in *Trypanosoma brucei***

Chloe Barnes, BSc (Hons)

35757681

Supervisors: Dr Mick Urbaniak & Dr Paul McKean

This thesis is submitted for the degree of MSc by Research Biomedical Science.

Lancaster University, Biomedical and Life Sciences

December 2021

## **Declaration**

I, Chloe Barnes, confirm that the work presented in this thesis is my own and has not been submitted for the award of a higher degree elsewhere.

## 1.0 Abstract

All organisms require iron for survival and the bloodstream form of the parasite *Trypanosoma brucei* has evolved a unique receptor that binds host transferrin to facilitate iron uptake. The *TbTfR* is encoded by two expression site associated genes, *ESAG6* and *ESAG7*, and is able to bind transferrin with a variable affinity from a wide range of host organisms. Upon iron starvation, the parasite is able to rapidly upregulate the expression of the transferrin receptor via a post-transcriptional mechanism mediated by the *ESAG6* 3'UTR. Here the *ESAG7* 3'UTR is defined, and truncations were made of the *ESAG6* and *ESAG7* 3'UTRs to identify motifs that are important for the upregulation of the receptor. The truncated sequences were ligated into a firefly luciferase reporter system and transfected into a tagged rRNA locus in a bloodstream form cell line. Luciferase assays were performed on the truncated cell lines to measure the upregulation of the reporter gene when iron starvation is induced. Normally, expression of the transferrin receptor is low under basal conditions and increases when the cells are incubated in media starved of iron. Under iron starvation conditions, it was observed that a number of the truncated cell lines were able to increase the expression of the reporter gene by a magnitude previously reported for the upregulation of the transferrin receptor. This response was only maintained when the 3' end of the 3'UTR remained undisrupted. From the 3'UTR a putative motif has been identified that may be responsible for mediating the upregulation of the transferrin receptor under iron starvation conditions.

# Table of Contents

<b>1.0 Abstract</b> .....	<b>3</b>
<b>2.0 Introduction</b> .....	<b>7</b>
<b>2.1 African Trypanosomiasis</b> .....	<b>7</b>
2.1.1 Clinical Presentation of African Sleeping Sickness .....	8
2.1.2 Epidemiology and Disease Control .....	9
2.1.3 Diagnosis of Human African Trypanosomiasis .....	11
2.1.4 Treatment of Human African Trypanosomiasis .....	12
2.1.5 Life Cycle .....	14
<b>2.2 <i>Trypanosoma brucei</i> Genetics</b> .....	<b>17</b>
2.2.1 RNA Polymerase I Transcription .....	17
2.2.2 Antigenic Variation .....	18
2.2.3 Monoallelic Exclusion .....	21
<b>2.3 The Importance of Iron</b> .....	<b>22</b>
<b>2.4. Structure and Function of the <i>Trypanosoma brucei</i> Transferrin Receptor</b> .....	<b>24</b>
2.4.1 Identification of the <i>Trypanosoma brucei</i> Transferrin Receptor .....	26
2.4.2 Genomic Organisation of the <i>Trypanosoma brucei</i> Transferrin Receptor .....	26
2.4.3 Localisation of the <i>Trypanosoma brucei</i> Transferrin Receptor .....	27
2.4.4 Ligand Binding .....	28
2.4.5 Processing of Transferrin.....	30
2.4.6 Conservation of the Transferrin Receptor in Trypanosomatids.....	30
<b>2.5 Iron Dependent Regulation of the <i>Trypanosoma brucei</i> Transferrin Receptor</b> .....	<b>32</b>
2.5.1. Iron Dependent Regulation of the Mammalian Transferrin Receptor.....	35
2.5.2 Importance of the 3' Untranslated Region of the <i>Trypanosoma brucei</i> Transferrin Receptor in the Iron Starvation Response .....	36
<b>2.6 Aims and Objectives</b> .....	<b>41</b>
<b>3.0 Materials and Methods</b> .....	<b>44</b>
<b>3.1. Materials</b> .....	<b>44</b>
3.1.1. Media.....	44
3.1.2. Antibiotics.....	45
3.1.3. Buffers .....	45
<b>3.2. Cell Culture</b> .....	<b>45</b>

<b>3.3 Agarose Gel Electrophoresis .....</b>	<b>45</b>
3.3.1. Gel Extraction .....	46
<b>3.4 Polymerase Chain Reaction (PCR).....</b>	<b>46</b>
3.4.1 PCR of <i>ESAG6</i> 3'UTR from Genomic DNA.....	46
3.4.2 PCR of <i>ESAG6</i> 3'UTR Truncations from Plasmid Template.....	47
3.4.3 PCR of <i>ESAG7</i> 3'UTR from cDNA.....	48
3.4.4 PCR of <i>ESAG7</i> 3'UTR Truncation from Plasmid Template .....	48
<b>3.5 Reverse Transcription.....</b>	<b>49</b>
<b>3.6 Cloning.....</b>	<b>50</b>
3.6.1 Ligation .....	50
3.6.2 Transformation .....	50
3.6.3 Isolation of Plasmid DNA .....	50
3.6.4 Analytical Restriction Enzyme Digest .....	51
3.6.5 Preparative RE Digest .....	51
<b>3.7 DNA sequencing .....</b>	<b>51</b>
<b>3.8 <i>Trypanosoma brucei</i> Transfection and Screening.....</b>	<b>52</b>
3.8.1 Linearising the DNA .....	52
3.8.2 Ethanol Precipitation .....	52
3.8.3 Amaxa Transfection.....	52
3.8.4 Screening for Transfected Clones.....	53
<b>3.9. Luciferase Activity Assay.....</b>	<b>54</b>
3.9.1 Induction of Iron Starvation .....	54
3.9.2 One Glo Luciferase Assay.....	54
<b>3.10 Data Analysis and Statistics.....</b>	<b>55</b>
<b>4.0 Results.....</b>	<b>56</b>
<b>4.1 Cloning and Truncation of the <i>ESAG6</i> and <i>ESAG7</i> 3' Untranslated Regions.....</b>	<b>56</b>
4.1.1 Truncation of the <i>ESAG6</i> 3'UTR.....	57
4.1.2 Identifying the <i>ESAG7</i> 3'UTR .....	61
4.1.3 Generation of the <i>ESAG7</i> 3'UTR Truncations.....	66
<b>4.2 Generation, Screening and Analysis of <i>Trypanosoma brucei</i> Cell Lines.....</b>	<b>68</b>
4.2.1 Characterisation of the <i>ESAG6</i> 3'UTR Response to Iron Starvation Conditions .....	72
4.2.2 Characterisation of the <i>ESAG7</i> 3'UTR Response to Iron Starvation Conditions .....	76
4.2.3 Summary of Luciferase Assay Results .....	79

<b>5.0 Discussion</b> .....	<b>82</b>
<b>5.1 Importance of Characterising the Iron Starvation Response</b> .....	<b>82</b>
<b>5.2 The Biological Role of the <i>Trypanosoma brucei</i> Transferrin Receptor</b> .....	<b>84</b>
<b>5.3 Analysis of Results</b> .....	<b>91</b>
5.3.1 Generation of the <i>ESAG6</i> 3'UTR and Truncations.....	91
5.3.2 Generation of the <i>ESAG7</i> 3'UTR and Truncations.....	94
5.3.3 Luciferase Assay Analysis.....	96
<b>5.4 Limitations</b> .....	<b>99</b>
<b>5.5 Post-Transcriptional Regulation in Trypanosomes</b> .....	<b>100</b>
<b>5.6 Recommendations for Future Research</b> .....	<b>104</b>
5.6.1 Further Analysis of the Putative Motif .....	104
5.6.2 Characterisation of the Iron Starvation Response of the Non-Expression Site Copy of <i>ESAG6</i> .....	106
<b>5.7 Conclusion</b> .....	<b>107</b>
<b>6.0 Acknowledgements</b> .....	<b>109</b>
<b>7.0 References</b> .....	<b>110</b>
<b>8.0 Appendix</b> .....	<b>122</b>

## 2.0 Introduction

### 2.1 African Trypanosomiasis

*Trypanosoma brucei* is the causative agent of the neglected tropical disease Human African Trypanosomiasis (HAT). This is a vector borne disease transmitted by the bite of an infected tsetse fly of the genus *Glossina* (Pays *et al.*, 2021). HAT is a substantial cause of morbidity and mortality in sub-Saharan with an estimated 65 million people at risk of contracting the disease in 36 endemic countries across an area referred to as the tsetse belt (WHO, 2021). There are two subspecies of *T. brucei* responsible for HAT, *T. b. rhodesiense* and *T. b. gambiense*. *T. b. rhodesiense* causes an acute infection and is present in 13 countries across Eastern and Southern Africa. However, they represent less than 5% of the total number of HAT cases reported annually (WHO, 2021). *T. b. gambiense* causes a chronic infection where people can be infected for months or years before they begin to show any symptoms. *T. b. gambiense* is widespread across 24 Western and Central African countries with 95% of all reported cases being *T. b. gambiense* (WHO, 2021). The two subspecies are morphologically indistinguishable but cause distinctive rates of disease progression and must be identified as different drugs are required to treat each subspecies (CDC, 2020).

Humans are regarded as the main reservoir for *T. b. gambiense*, however, animals are the major reservoir for *T. b. rhodesiense*, these can be domesticated animals such as cattle or wild animals such as those of conservation concern protected in game reserves (Wamwiri *et al.*, 2016). The transmission cycle of *T. b. gambiense* is considered to be human to tsetse fly and back to human, with animals playing little role in maintaining the transmission cycle. It is thought that due to the chronic infection caused in humans, that can be asymptomatic for long periods, this is enough to maintain the population of infected tsetse flies and continue the transmission cycle (Wamwiri *et al.*, 2016). From this it has been possible to eliminate infection with *T. b. gambiense* in local areas by targeting human treatment, however, animals can still be part of this transmission cycle so cannot be overlooked (Wamwiri *et al.*, 2016). The transmission cycle for *T. b. rhodesiense* relies on animal reservoirs and the cycle is typically animal to tsetse fly then back to animal or a human infection.

### 2.1.1 Clinical Presentation of African Sleeping Sickness

HAT occurs in two stages, stage I referred to as the haemolympathic stage and stage II the encephalitic stage (Kennedy, 2013). In stage I the parasite is actively dividing in the blood and the lymphatic system and travels to systemic organs such as the spleen and heart. Stage I is characterised by bouts of fever, headaches and itching at the site of the chancre (WHO, 2021). Symptoms in the early stage do not begin to show until 1-3 weeks after the tsetse bite and at this stage they are non-specific and can overlap with other diseases such as malaria, occasionally resulting in misdiagnosis (Kennedy, 2013). Progression to stage II occurs at different rates depending on the subspecies causing the infection. In *T. b. rhodesiense* infection, progression to stage II can occur in weeks post-infection, however, *T. b. gambiense* causes a chronic infection so progression to stage II can take much longer, often several months (Kennedy, 2013). To progress to stage II, the parasite crosses the blood brain barrier and has access to the central nervous system, at this stage the symptoms become more specific and more severe and the disease becomes more difficult to treat (Vincent *et al.*, 2016). Once a patient has progressed to stage II the clinical features of the disease include neurological dysfunction, depression and anxiety, psychotic episodes and disruption to the sleep-wake cycle (Vincent *et al.*, 2016). The disease is often referred to as 'African Sleeping Sickness' due to the perturbations to the sleep-wake cycle, these symptoms are present in 74% of infections and could present as reversal of the sleep-wake cycle, uncontrollable episodes of sleep and/or an alteration of the structure of sleep (Kennedy, 2013). More recently there has been growing evidence of asymptomatic *T. b. gambiense* infection, suggesting that some individuals are able to tolerate the infection without progression to stage II (Capewell *et al.*, 2019). Asymptomatic individuals are serologically positive, but parasites are not identified in the bloodstream by microscopy, so they are not diagnosed with HAT and are not offered treatment due to the toxicity of available drug treatments. There are no reported cases of asymptomatic *T. b. rhodesiense* infection. Unfortunately, if allowed to progress to stage II, human African trypanosomiasis is invariably fatal if untreated, with patients eventually falling into a coma and death (Vincent *et al.*, 2016).

### 2.1.2 Epidemiology and Disease Control

Over 65 million people are at risk of HAT where the tsetse fly is present, however, currently case numbers are very low with only 992 reported cases in 2019 (WHO, 2021). The risk of infection with *T. brucei* is determined by the occurrence of contacts between humans and the infected vector, this occurrence varies depending on socioeconomic factors such as livelihood and the range of the tsetse fly shown in **Figure 2.1** (Pays *et al.*, 2021). People living in rural communities have a higher exposure rate to the tsetse fly vector and therefore to parasites as they have a greater reliance on agriculture, fishing and hunting practices (WHO, 2021). Many of these affected rural populations have limited access to adequate healthcare making diagnosis and treatment of cases difficult. Despite this, the current level of infection is under control. However, even under control the disease causes major economic loss to communities affected. The Pan African Tsetse and Trypanosomiasis Eradication campaign reported in 2012 that when a person is infected with human African trypanosomiasis, the household loses 25% of its income. Due to the presence of the tsetse fly vector and the parasite infecting domestic animals, causing animal African trypanosomiasis (AAT) or nagana, it causes food insecurity in vast and fertile areas of the continent, and it has been stated that trypanosomiasis lies at the heart of Africa's struggle against poverty (PAAT, 2021). The estimated overall annual economic losses in human disease, infected livestock and the agricultural sector attributed to African trypanosomiasis is over \$4 billion (PATTEC, 2012).



**Figure 2.1: Map of Africa showing tsetse infested areas across the continent.** The areas shaded in red shows the range of the tsetse fly vector. Within this range, the population is at risk of human African trypanosomiasis due to close contact with the vector.

During the 1960s the prevalence of African trypanosomiasis was under a level of control with less than 5000 reported cases across the African continent (WHO, 2021). However due to relaxations in surveillance, the infection was allowed to re-emerge, reaching epidemic levels by the 1970s which lasted into the 1990s (WHO, 2021). In 1998, almost 40,000 cases were reported but the estimated number of undiagnosed and untreated cases was closer to 300,000 (Pays *et al.*, 2021). To combat the rising case levels and the public health crisis, National Control Programmes with the WHO and NGOs were set up to implement improved disease surveillance, clinical trials and increased availability of free drug treatment, resulting in a steady decline in cases. In 2009 the number of annual cases dropped below 10,000 for the first time and in 2019 only 992 cases were reported which is the lowest number since cases were recorded (WHO, 2021). Despite this the disease incidence varies between countries and even between villages within the same country, with over 70% of all cases in the last 10 years reported in the Democratic Republic of Congo (WHO, 2021). In some countries, including Botswana and Liberia, disease transmission appears to have stopped, as no new cases have been reported in over a decade. However, assessing the extent of cases across many African countries remains an issue due to unstable social circumstances and/or

difficult accessibility which hinder surveillance and diagnostic activities. Political and civil unrest is a major component in the breakdown in disease surveillance and treatment. Despite this, the World Health Organisation had targeted HAT for elimination as a public health problem by 2020 and interruption of transmission (no reported cases) by 2030. Strategies implemented to reach elimination targets include disease surveillance through active and passive case finding, rapid treatment of confirmed cases and vector control, to reduce the population of tsetse flies (CDC, 2020).

The identification of asymptomatic cases of *T. b. gambiense* infection has made the control and elimination of HAT more complicated. There is evidence that in asymptomatic cases, the parasites colonise the skin and parasites were identified in archived skin punctures from patients not previously diagnosed with HAT (Capewell *et al.*, 2019). Experiments conducted on mice model organisms showed they were able to harbour parasites in the skin despite no observable parasitaemia in the blood, however, the mice were still infectious to the tsetse vector. The presence of extravascular parasites has been known since HAT was first described, however this has been neglected in recent years. New evidence showing the occurrence of asymptomatic individuals with parasites present in the skin causes major implications for the elimination of the disease, as these individuals are able to continue the transmission cycle without being detected by routine screening by microscopy. Asymptomatic cases have also been linked to the re-emergence of HAT as they are a reservoir for the disease that has gone undetected for many years (Capewell *et al.*, 2019).

### 2.1.3 Diagnosis of Human African Trypanosomiasis

Due to the high mortality rate of late stage disease, rapid diagnosis and staging of the disease is important as different drugs are used to treat stage I and stage II infection, with the diagnostic procedure consisting of two successive steps, screening and staging (Bouteille *et al.*, 2012). Rapid diagnosis is also important for treatment as a stage I infection is much easier to treat than stage II because the drug does not need to be able to cross the blood brain barrier (Vincent *et al.*, 2016). Currently the WHO gold standard for diagnosis of human African trypanosomiasis is identification of parasites in blood or cerebrospinal fluid (CSF) samples taken from the patient (WHO, 2021). Clinical examinations are rare in the field due to the

non-specificity of symptoms associated with the infection (Bouteille *et al.*, 2012). Microscopy is not ideal in the field as it requires skilled technicians to be able to identify the parasites and also requires a lumbar puncture procedure to be performed to obtain the CSF. To tackle this a number of serological tests are currently being developed including ELISAs to detect the presence of antibodies or rapid diagnostic tests such as loop-mediated isothermal amplification of parasite DNA (Vincent *et al.*, 2016). However, whilst the tests must be rapid, they must also be suitable to be performed in the field in resource limited settings and still be able to stage the infection. Another serological test, the card agglutination test for trypanosomiasis (CATT), detects circulating antibodies in the patient from a blood sample taken from a finger-tip puncture with a recorded sensitivity of 87-93% (Bouteille *et al.*, 2012). The CATT test can be used to screen large numbers of the population, then those that are identified as CATT positive can have CSF samples drawn to confirm diagnosis and stage the disease.

#### 2.1.4 Treatment of Human African Trypanosomiasis

There are few drugs that can currently be used to treat human African trypanosomiasis and, of those that have been approved for use, three of them were discovered before the 1950s, highlighting the importance of research into drug discovery for the treatment of HAT (WHO, 2021). When a patient is diagnosed with African Sleeping Sickness it is essential that the subspecies of parasite is identified for treatment, which can be based on the geographical isolation of the two subspecies in endemic countries, and also that the progression of the infection is staged.

##### 2.1.4.1. Current Treatments

Pentamidine and suramin are used to treat stage I *T. b. gambiense* and *T. b. rhodesiense* infections, respectively. There are multiple side effects associated with these drugs, however, the more severe are reported when treating a *T. b. rhodesiense* infection with suramin and can include anaphylactic shock when there is a concomitant infection with filariasis, nephrotoxicity and haematuria (Bouteille *et al.*, 2012). When an infection has progressed to stage II the treatment is always administered in a hospital facility, due to the high toxicity

associated with the drugs and the logistical difficulties in their administration (Bouteille *et al.*, 2012). Melarsoprol was first discovered in 1949 and was originally used to treat stage II infections of all human African trypanosomiasis patients, however, due to the high toxicity of the drug, which is derived from arsenic, alternative therapies have been developed to treat stage II *T. b. gambiense* infection (WHO, 2021). There are numerous side effects reported when patients are treated with melarsoprol including, but not limited to, cardiotoxicity, hepatic dysfunction and myalgia. The most severe complication from treatment with melarsoprol is reactive encephalopathy which can be fatal in 3-10% of cases (WHO, 2021). Despite this, melarsoprol is still the front-line treatment for *T. b. rhodesiense* infection but is no longer used to treat stage I infections or *T. b. gambiense*. There are no other alternative treatments yet on the market that have been able to successfully treat stage II *T. b. rhodesiense* infections.

Current treatments for stage II *T. b. gambiense* infection include nifurtimox-eflornithine combination therapy (NECT) (WHO, 2021). Eflornithine was first registered in 1990 and was originally used as a monotherapy against *T. b. gambiense* infection, however, the treatment regimen was quite complex (Bouteille *et al.*, 2012). The introduction of nifurtimox in 2009 reduced the duration of the treatment by reducing the number of intravenous (IV) perfusions, as nifurtimox is administered orally. NECT is now one of the front-line treatments for *T. b. gambiense* infection, which is relatively safe but still requires treatment in a hospital facility due to logistical difficulties in the IV perfusions (Dickie *et al.*, 2020). In 2019, fexinidazole was recommended by the WHO as an oral treatment for *T. b. gambiense* infection and is currently used to treat stage I and non-severe stage II infections. It reduces the logistics of treatment as an oral treatment taken once a day for ten days, but it is important that patients stick to the regime to prevent relapse (Dickie *et al.*, 2020). The introduction of fexinidazole as an oral medication has simplified treatment against HAT and shows how important the production of safer and simpler drugs is to control and hopefully eliminate African Sleeping Sickness as a public health concern.

#### 2.1.4.2 New Developments

There are a number of other drugs currently in clinical trials for the treatment of HAT. Fexinidazole continues to undergo clinical trials for treatment of *T. b. rhodesiense* infection, with hopes to reduce the difficulty and side effects of treating *T. b. rhodesiense* infection (WHO, 2021). A new drug acoziborole is also in the clinical trials phase but early unpublished data is promising and indicates that the drug has demonstrated good efficacy and a good safety profile and can be used to treat stage II infections with a single oral dose, which has the potential to revolutionise treatment for human African trypanosomiasis (Dickie *et al.*, 2020).

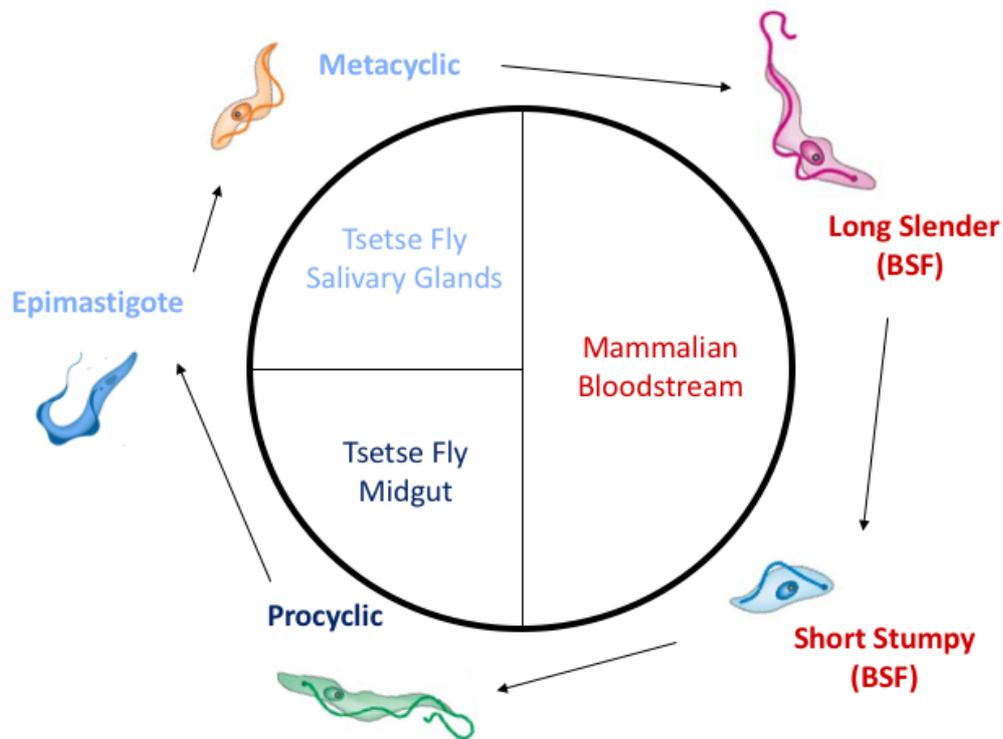
#### 2.1.4.3 Post-Treatment

Once an infection of HAT is treated, an 18 month to 2 year post-treatment follow up is necessary to declare a patient cured as relapse does occur occasionally (Bouteille *et al.*, 2021). These post-treatment check-ups prevent the parasite from being able to spread amongst the population again (Bouteille *et al.*, 2021). Unfortunately due to the antigenic response of the parasites, it is unlikely that a vaccine will ever be developed to prevent human African trypanosomiasis so having safe drugs that can be administered in areas with few resources is the next best thing, therefore the collaborations around the drug discovery programme for neglected tropical diseases is so important (Vincent *et al.*, 2016). To ensure that the current drugs available can be provided to those who need them the most, the WHO provides anti-trypanosome medications free of charge to endemic countries through public-private partnerships with Sanofi and Bayer Health care (WHO, 2021).

#### 2.1.5 Life Cycle

During the life cycle of *Trypanosoma brucei* the parasite passes from the tsetse fly vector to the mammalian host, and throughout this cycle the parasite must be able to adapt to changes in its environment to enable survival. In each environment the parasite is faced with new challenges, such as changes in nutrient availabilities and exposure to the host immune system

(Silvester *et al.*, 2017). The morphological changes observed during the complex life cycle of *T. brucei* shown in **Figure 2.2** are associated with changes in the surface coat that enhance survival of the parasite.



**Figure 2.2: Life cycle stages of *Trypanosoma brucei* in the mammalian host and the tsetse fly vector.** The stumpy bloodstream form (BSF) is pre-adapted for uptake into the arthropod vector and is taken up during a blood meal. In the midgut, the parasite proliferates before migrating to the salivary glands for release into the mammalian host. The life cycle stages that occur in the tsetse fly takes about 4 weeks. When injected into the mammalian host, the parasite proliferates as a long slender trypomastigote causing infection. When a density threshold is reached the parasite differentiates into a short-stumpy BSF cell. (Figure adapted from Lee *et al.*, 2007).

Whilst taking up a blood meal from an infected mammalian host, the tsetse fly also ingests pre-adapted stumpy bloodstream form (BSF) trypomastigotes, which rapidly colonise the midgut of the fly (CDC, 2021). Once in the tsetse fly midgut, the BSF cells transform into procyclic form cells (PCF). The uptake into the fly initiates a change in the surface coat of the parasite, from a protective variant surface glycoprotein layer (VSG) to a less dense procyclin coat made up of EP and GPEET (Matthews, 2005). In the midgut of the tsetse fly the parasite

divides by binary fission to establish an infection. There are a number of life cycle stages in the tsetse fly that occur in the proventriculus, characterised by changes in morphology but are not described here. The cells then arrest and migrate to the salivary glands where they differentiate into an epimastigote cell (CDC, 2021). The epimastigotes are no longer cell cycle arrested and proliferate, attaching to the surface of the salivary glands by their flagellum. At this stage the parasite once again arrests and differentiates into a metacyclic cell, which becomes pre-adapted to infect the mammalian host by reacquiring the VSG surface coat. The parasite is released into the lumen of the salivary glands and is injected into the mammalian host when the vector next takes up a blood meal. The stages of the life cycle that occur in the tsetse fly take around 3 weeks (CDC, 2021).

When the parasite enters the lymphatic system and passes into the bloodstream it transforms into a slender BSF trypanomastigote that proliferates in the bloodstream to establish an infection. The slender BSF cells release a molecule called stumpy-induction factor (SIF) which accumulates in the bloodstream (Zimmerman *et al.*, 2017). Once a threshold value of SIF is reached, the slender cells differentiate into stumpy cells. When transitioning to stumpy cells, they arrest in G1 phase of the cell cycle, ready to re-enter the cell cycle when taken up by the arthropod vector (Matthews, 2005). Components of the SIF transduction pathway were identified by genome-wide RNAi screening describing gene expression regulators that control stumpy formation (Mony *et al.*, 2014). More recently, Rojas *et al.* (2019) identified a G-protein coupled receptor analogue *TbGPR89*, which was shown to be surface expressed in only slender BSF. The *TbGPR* protein was identified as essential for the differentiation of stumpy cells by inducible expression in pleomorphic cell lines resulting in cell cycle arrest in the G1 phase, and the morphology of the cells differentiated to become stumpy (Rojas *et al.*, 2019). To confirm that this was via the SIF pathway, RNAi was used to knockdown RBP7, a predicted RNA-binding protein (RBP) required for SIF-induced stumpy formation (Rojas *et al.*, 2019). Trypanosomes lack conventional proton-coupled oligopeptide transporters (POT) however experiments with *E. coli* YjdL showed that *TbGPR89* is able to transport oligopeptides that can induce stumpy formation, suggesting that SIF may be an oligopeptide (Rojas *et al.*, 2019). Trypanosomes secrete prolyl oligopeptidase (*TbPOP*) and pyroglutamyl peptidase (*TbPGP*), and it is likely that these degrade numerous host substrates to produce oligopeptides that

can be transported by *TbGPR89* and induce stumpy formation by paracrine quorum sensing (Rojas *et al.*, 2019).

The stumpy cells maintain the VSG coat, so they are not rapidly cleared from the mammalian host. In the mammalian host longevity is essential to allow transmission to the vector and continue the transmission cycle. If the parasite were allowed to divide continuously the host would become overwhelmed and would succumb to the infection quite rapidly. A chronic infection can be established by the density dependent switch from slender BSFs to stumpy BSFs which are cell cycle arrested, preventing the parasite from proliferating unchecked, allowing host survival whilst also pre-adapting the parasite for uptake by the vector continuing the life cycle. This is also the cause of the waves of parasitaemia observed in *T. brucei* infection where variants of the parasites proliferate until reaching a threshold density and stumpy formation is induced (Mugnier *et al.*, 2016). The parasites are then cleared by the host immune response resulting in a decrease in parasitaemia. The infection is maintained by a few individuals which switched VSG and continue to proliferate to create a new wave of parasitaemia.

## **2.2 *Trypanosoma brucei* Genetics**

### 2.2.1 RNA Polymerase I Transcription

*Expression Site Associated Genes (ESAGs)* and *VSG* are arranged in polycistronic units with a single promoter located 67 bp upstream of the transcription initiation site (Gunzl *et al.*, 2015). The polycistronic units at the bloodstream expression sites (BESs) are unique to trypanosomes and they are transcribed exclusively by RNA polymerase I (RNA pol I). In most eukaryotes, RNA pol I is used to transcribe rRNA in the nucleolus (Gunzl *et al.*, 2015). *T. brucei* RNA pol I is able to transcribe 45S rRNA and transcribe mRNA from genes encoded outside of the nucleolus, such as procyclin and genes expressed at the BESs (Das *et al.*, 2008). In other eukaryotes RNA pol I does not make protein coding genes as it is unable to make the methyl cap, *T. brucei* RNA pol I is also unable to make this. Instead a 35-nucleotide leader derived from nuclear RNA, termed the spliced leader, is trans-spliced onto the 5' end of the mRNA (Gunzl *et al.*, 2010). The spliced leader contains a 7-methylguanosine cap and four methylated

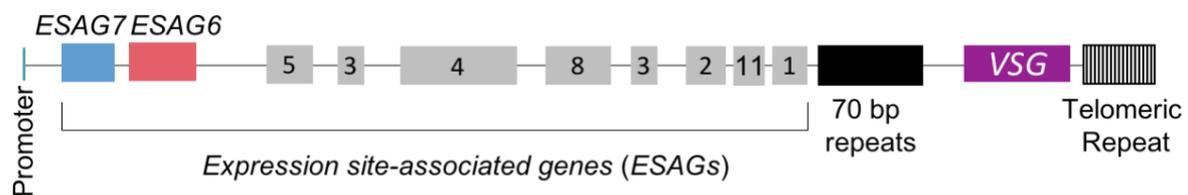
nucleotides at the beginning of the sequence to generate mature mRNA when the polycistronic unit is trans-spliced and polyadenylated to generate separate functional mRNAs encoding each of the genes in the polycistronic unit (Gunzl *et al.*, 2010). In other eukaryotes mRNA is synthesised by RNA polymerase II where the capping enzyme interacts with the C terminus of RNA pol II, so mRNA is capped co-transcriptionally rather than post-transcriptionally like in *T. brucei* (Gunzl *et al.*, 2003). At the BES RNA pol I is thought to allow for the high rates of transcription that occur at the active expression site to generate VSG mRNA (Alsford *et al.*, 2012). Due to the polycistronic arrangement of trypanosome genes, it is known that they are controlled post-transcriptionally, for example on the cell surface VSG is more abundant than other ESAG proteins.

### 2.2.2 Antigenic Variation

During the complex life cycle of *T. brucei*, in the mammalian host the parasite lives extracellularly, so is constantly exposed to the host immune system. To be able to survive in the mammalian host and develop a chronic infection observed in patients with *T. b. gambiense* the parasite is able to evade the host immune system by the expression of a dense VSG coat. The VSG coat is a monolayer of  $5 \times 10^6$  copies of the VSG dimer that cover the entire cell surface of the parasite (Hertz-Fowler *et al.*, 2008). VSGs are homo-dimeric proteins composed of two 50-60 kDa subunits, GPI anchored to the outer leaf of the plasma membrane (Hertz-Fowler *et al.*, 2008; Bartossek *et al.*, 2017). VSG molecules display lateral movement in the membrane and adopt two conformations to protect the cell surface. The key conformation consists of a compact arrangement where VSG is elevated above invariant plasma membrane proteins such as the transferrin receptor, to protect them from the host immune system (Bartossek *et al.*, 2017). Despite this, VSG is highly immunogenic so elicits an antibody response in the host and the antibodies can bind to the VSG surface coat (Vanhamme *et al.*, 2000). The parasite has evolved mechanisms to utilise the VSG coat to continue to evade the host immune response and prevent complement-mediated lysis even once an immune response has been initiated (Vanhamme *et al.*, 2000). The flagellar pocket is the only site of endocytosis and exocytosis on the surface of the parasite making up a surface area of only about 5% (Matthews, 2005). At low levels of anti-VSG antibodies, trypanosomes are able to internalise the VSG-antibody complex at the flagellar pocket, degrade the

antibodies and recycle the VSG dimer back to the cell surface (Zimmerman *et al.*, 2017). The parasite is able to turn over the VSG pool in approximately 12 minutes by the process of hydrodynamic flow, clearing the antibodies from the cell surface (Engstler *et al.*, 2004; Engstler *et al.*, 2007).

At high levels of anti-VSG antibodies, the parasite becomes overwhelmed and can no longer rely on kinetics to clear the antibodies from the surface (Zimmerman *et al.*, 2017). When the antibody response is sufficiently high to overwhelm the parasite, the cell is destroyed either by complement-mediated lysis or phagocytosis (Silvester *et al.*, 2017). To establish a chronic infection in an immunocompetent host, the parasite can continue to evade the immune cells by antigenic variation (Alsford *et al.*, 2012). Antigenic variation involves switching the expressed VSG surface coat for a new one that has not yet been exposed to the host immune system through phenotypic and clonal variation (Alsford *et al.*, 2012). A single trypanosome has an archive of up to 2000 VSG genes and pseudogenes to allow the parasite to develop a persistent infection in the host (Alsford *et al.*, 2012).



**Figure 2.3: General structure of a bloodstream expression site.** The *ESAG6* and *ESAG7* genes are located adjacent to the promoter, the furthest distance from the *VSG* gene and the telomere. The multiple BES have different arrangements of the *ESAGs*, some have duplicates of the same *ESAGs* or pseudogenes present.

*VSG* genes are expressed from telomeric BESs (**Figure 2.3**) only active in BSF cells and silenced in PCF. The BESs consist of life cycle stage specific genes including *ESAGs*, most of which are known to be cell surface expressed proteins, and *VSG* located after the 70 bps repeats downstream of the *ESAGs*, adjacent to the telomeric repeats (Jackson *et al.*, 2013). Many gene families involved in virulence and pathogenicity are located at the telomeres, this is likely due to the high rates of recombination that allow for high plasticity and generation of genetic diversity to enhance infection (Hertz-Fowler *et al.*, 2008; Young *et al.*, 2008). There are

multiple BESs present in the trypanosome genome that express similar but not identical copies of the *ESAGs* and *VSG*; the *ESAG* repertoire varies between them with some of the BES being more important than others for the expression of *VSG* (Young *et al.*, 2008). In *T. brucei* Lister 427, there are about 20 BESs, with 14 showing high levels of conservation in the overall structure and order of the *ESAGs* (Hertz-Fowler *et al.*, 2008). From the larger repertoire of *VSG* genes and pseudogenes, the parasite is able to express an indefinite number of different *VSG* molecules to maintain the infection in the host. There are four mechanisms by which *T. brucei* is able to change the expressed *VSG* dimer; through an *in-situ* switch, telomere exchange, gene conversion and segmental gene conversion (Vanhamme *et al.*, 2000). An *in-situ* switch involves the activation of a new BES and telomere exchange is the homologous recombination of the *VSG* in the active expression site. These two mechanisms only have a limited potential for variation and are often deployed by the parasites in early infection (Vanhamme *et al.*, 2000). Gene conversion and segmental gene conversion involve the recombination of genes either from internal cassettes being moved to the telomere, or segments of genes from the repertoire of ~2000 genes and pseudogenes recombining to give a new *VSG* sequence (Vanhamme *et al.*, 2000, Alford *et al.*, 2012). These mechanisms have a much greater potential of generating diversity and allow the parasite to maintain an infection until host death, or until intervention with drug treatment. There is, however, the risk that during recombination a non-functional gene is assembled, if this is the case a defective *VSG* protein is encoded that cannot protect the cell resulting in cell death (Mugnier *et al.*, 2016). Due to the polycistronic transcription of the BES, when an *in-situ* switch occurs and the parasite changes ESs, the *ESAG* repertoire also changes, so similar but different variations of the cell surface proteins are expressed, which can also affect the survival of the parasite in the host if the new proteins are not as well suited to the host environment. Zimmerman *et al.* (2017) suggested that differentiation into stumpy forms from proliferative slender forms can be triggered by activation of an ES that does not provide a good complement of *ESAGs* as a 'rescue programme'. When a parasite has differentiated into a stumpy form, the cell loses the ability to switch *VSG* so is exposed to the antibody response. It has, however, been shown that stumpy cells are able to remove *VSG*-antibody complexes from the cell surface quicker than slender forms potentially due to the shorter cell length or the slightly larger flagellar pocket (Engstler *et al.*, 2007). Once the antibody response is sufficiently high enough, the stumpy cells will also be cleared from the host, seen clinically as

waves of parasitaemia. When the antibody response is high only the cells that have had a productive VSG switch will survive and divide to form a new population with the same VSG coat, and this process is repeated in waves (Silvester *et al.*, 2017). It is due to the vast repertoire of VSG genes and the resulting antigenic variation of the parasite that makes it unlikely that there will ever be a vaccine developed to prevent the spread of human African trypanosomiasis, making diagnosis and treatment the key to eradicating the disease.

### 2.2.3 Monoallelic Exclusion

Despite there being multiple BESs in the parasite genome and a repertoire of thousands of variations of VSG, only a single copy of VSG is expressed on the cell surface at any one time (Vanhamme *et al.*, 2000). This is achieved as only a single expression site is active at any given time through the process of mono-allelic exclusion (Silvester *et al.*, 2017). There have been multiple attempts to select for cells expressing two expression sites, however, it resulted in unstable cells suggesting that multiple VSGs cannot be expressed simultaneously (Glover *et al.*, 2016). Despite each cell only expressing a single VSG, there are other cells in the population that will express different VSG as not all cells will switch to the same VSG, resulting in different generations with different VSG surface coats (Navarro *et al.*, 2001). There have been a number of theories about how a single expression site is active, such as controls of transcription initiation based on cDNA analysis, however, it has been shown that transcription is initiated at all BES but progression is poor in silent expression sites (Vanhamme *et al.*, 2000; Alsford *et al.*, 2012). Research suggests that transcription initiation occurs at a high level, simultaneously across all BES but RNA elongation and processing only continues in the single active expression site, with most transcripts remaining bound to chromatin (Vanhamme *et al.*, 2000). RT-PCRs have been used by Vanhamme *et al.* (2000) to show that transcription is initiated at multiple expression sites but is aborted in all except one. This occurs at a similar level in each ES, so the active expression site is not favoured for transcription initiation over silent sites (Navarro and Gull, 2001). Due to the unique use of RNA pol I to transcribe the polycistronic units, Navarro and Gull (2001) suggested that, as RNA pol I is usually associated with rDNA expression in the nucleolus, compartmentalisation processes regulate VSG ESs. From their study they found that there is a higher order architecture they termed the expression site body (ESB). The involvement of the ESB in mono-allelic exclusion was

supported by its absence in procyclic cells where VSG is not expressed and its presence in BSF cells where exclusive expression of a single VSG molecule is observed.

To identify components involved in allelic exclusion associated with the ESB, Glover *et al.* (2016) generated an RNAi library to screen for defects in telomere-exclusion expression. They identified a gene, *VEX1* (VSG exclusion 1) which was found to be closely associated with the RNA pol I focus at the active expression site and this localisation was life cycle stage specific and dependent on transcription (Glover *et al.*, 2016). In *VEX1* knockdown silent VSGs are transcribed and expressed on the cell surface resulting in a moderate growth defect, when overexpressed it results in increased expression of ESAGs from multiple BESs (Glover *et al.*, 2016). A winner takes all model was put forward that VEX1 is able to positively and negatively regulate RNA pol I transcribed loci, whereby access to VEX1 increases RNA pol I transcription to establish an active expression site that then mediates homology-dependent silencing at other sites (Glover *et al.*, 2016). Further analysis identified a *VEX2* gene that assembles a VEX1-VEX2 complex, recruitment of the two proteins to the active expression site requires active RNA pol I transcription (Faria *et al.*, 2019). When RNA pol I is inhibited, VEX1 and VEX2 are redistributed across the nucleus and remain there when transcription is prevented (Faria *et al.*, 2019). Knockdown of the VEX complex does not induce VSG switching, but instead results in a breakdown in allelic exclusion as multiple VSG are expressed on the cell surface. When only *VEX2* is knocked down an increased rate of expression of the ESAGs at the active ES is observed suggesting that VEX2 mediates suppression (Faria *et al.*, 2019). Overexpression was used to show that VEX1 is able to mediate the abundance of VEX2, but VEX1 association with VSG is VEX2 dependent, which may be important for preventing the activation of multiple BESs (Faria *et al.*, 2019).

### **2.3 The Importance of Iron**

Iron is a transition metal and its chemical properties make it essential to all forms of life. In mammals iron is required in a wide variety of essential cellular processes including DNA synthesis, energy metabolism, cell proliferation and it is a major component of haemoglobin and myoglobin involved in oxygen transportation (Kwabata *et al.*, 1999). Iron is a constituent

of haem, iron-sulphur (Fe-S) clusters and other iron containing centres in many redox enzymes, including those involved in the mitochondrial respiratory chain and ribonucleotide reductase involved in deoxyribonucleotide synthesis (Ganz *et al.*, 2006). Many proteins rely on iron as a cofactor in redox reactions and ligand binding (Cheng *et al.*, 2004).

Iron exists in two states, as a ferrous ( $\text{Fe}^{2+}$ ) or ferric ( $\text{Fe}^{3+}$ ) form and it can readily move between the two states by oxidation and reduction reactions (Hentze *et al.*, 2004). However, the conversion between the two states is dangerous. In the ferrous state the Fenton reaction can occur where  $\text{Fe}^{2+}$  reacts with hydrogen peroxide ( $\text{H}_2\text{O}_2$ ), which results in the production of reactive oxygen species which can damage proteins, nucleic acids and lipid membranes in the cell (Cheng *et al.*, 2004). In the ferric state, iron is insoluble and cannot be utilised by the cells (Cheng *et al.*, 2004). Along with this, iron overload and iron deficiency can both cause cell death and complications at a systemic level. Due to all of these factors the levels of iron in the body must be tightly regulated (Ganz *et al.*, 2006).

In the body iron is recycled and conserved to avoid any toxic effects and allow iron uptake, and the availability of free iron in the body is minimised by numerous mechanisms. In a healthy adult, 25-30 mg of iron is required daily for protein synthesis and cellular regeneration (Zhang *et al.*, 2014). Of this, only 10% is obtained from the diet, this is absorbed by intestinal cells through the action of DMT1, a metal ion symporter (Hentze *et al.*, 2004). This uptake is to compensate for iron lost from the body such as by bleeding. The remaining 90% of the daily required iron is recycled from senescent erythrocytes by splenic macrophages (Zhang *et al.*, 2014). Within mammalian cells most iron is utilised by the mitochondria, but the mechanism by which the mitochondria acquire iron from the cell is controversial, although there are a number of theories surrounding this (Gao *et al.*, 2021). Once iron is internalised by the mitochondrion there are three primary metabolic pathways for iron utilisation, these are haem synthesis, Fe-S cluster biogenesis and mitochondrial iron storage (Abbaspour *et al.*, 2014; Hentze *et al.*, 2004). Haem is a major component of haemoglobin in erythrocytes produced in the bone marrow and is essential for oxygen transportation to sites of respiration around the body (Ajioka *et al.*, 2006). Fe-S clusters are critical for electron transport in proteins in which they are present (Gao *et al.*, 2021). Any iron not actively being utilised by the host is bound to transport or storage proteins in a non-toxic state. In normal humans,

extracellular iron circulates bound to transferrin (Hentze *et al.*, 2004). Transferrin is a major serum glycoprotein that has a high affinity for iron, and it is abundant in serum at a normal range of 2.5-3 mg/mL (Pupim *et al.*, 2013). Transferrin captures iron released into the plasma from intestinal enterocytes and macrophages and delivers it to cells (Abbaspour *et al.*, 2014). Iron is bound to transferrin as  $Fe^{3+}$  and is non-reactive, however, due to the high affinity of iron to transferrin it is difficult to dissociate back into free iron (Cheng *et al.*, 2004). Due to this there are numerous iron uptake mechanisms that are transferrin-dependent that mediate uptake of iron bound to transferrin. Intracellularly excess iron is stored by ferritin. Ferritin is a ubiquitous, highly conserved multimeric protein (Hentze *et al.*, 2004). Ferroxidase is able to convert  $Fe^{2+}$  to  $Fe^{3+}$  as iron is internalised to detoxify and store the free iron within the cell.

#### **2.4. Structure and Function of the *Trypanosoma brucei* Transferrin Receptor**

To be able to survive within the mammalian host, *T. brucei* must be able to obtain iron from the surrounding environment. There are a number of mechanisms the parasite employs to import iron including a glyceraldehyde-3-phosphate dehydrogenase (GAPDH) receptor for uptake of lactoferrin or via a high affinity haptoglobin-haemoglobin receptor (HpHBR) to obtain haem from the host (Tanaka *et al.*, 2004; Higgins *et al.*, 2013). The key route in BSFs, however, is uptake of transferrin via the *T. brucei* heterodimeric transferrin receptor (*TbTfR*). Lactoferrin is an iron binding protein that is secreted by mammary glands and neutrophils so is available to trypanosomes in the host serum (Tanaka *et al.*, 2004). However, it is unlikely that trypanosomes utilise lactoferrin as the primary source of iron due to a 260-fold increase in the iron binding strength of the protein compared to transferrin, making it more difficult to obtain free iron (Tanaka *et al.*, 2004). Haem is also not the major source of iron for *T. brucei* as the parasites lack haem oxygenase and ferrochelatase, enzymes required to extract iron from haem (Stijlemans *et al.*, 2015). It is, therefore, more likely that haem obtained by the parasite is incorporated into haem proteins not as a source of free iron (Stijlemans *et al.*, 2015). Due to the inability of *T. brucei* to easily obtain free iron from haem and lactoferrin, uptake of transferrin is essential as a source of iron for parasite survival.

Transferrin is a major serum 79 kDa glycoprotein with two homologous iron-binding domains (Giometto *et al.*, 1993). Each of these domains can bind an Fe<sup>3+</sup> molecule to safely transport iron around the body in the bloodstream to where it is required (Trevor *et al.*, 2019). When iron is bound to transferrin, the protein adopts a closed conformation and is referred to as holo-transferrin, when there is no iron bound, the binding domains adopt an open, iron-free conformation referred to as apo-transferrin (Trevor *et al.*, 2019). For iron to dissociate from transferrin, there must be a change in the surrounding pH which occurs within a cell (Giometto *et al.*, 1993). Uptake of transferrin into mammalian cells occurs via receptor-mediated endocytosis of the mammalian transferrin receptor (TfR), where the receptor-transferrin complexes are endocytosed and trafficked to the endosome, where acidification causes a change in the conformation of transferrin from holo- to apo-transferrin and iron is released into the cell (Trevor *et al.*, 2019).

The host not only sequesters iron within transferrin to safely transport it around the body, but also to reduce the availability of free iron to invasive pathogens (Trevor *et al.*, 2019). Despite this, *T. brucei* has developed a way to scavenge iron from the mammalian host through the evolution of the *TbTfR* that is able to bind host transferrin with a high affinity to allow survival in the host (Trevor *et al.*, 2019). The *TbTfR* is evolutionarily distinct from the mammalian transferrin receptor which is a homo-dimeric protein consisting of two identical subunits with a single transmembrane segment (Bitter *et al.*, 1998). The *TbTfR* shows no sequence similarity to mammalian TfRs but is structurally similar to VSG expressed across the surface of BSF cells (Kariuki *et al.*, 2019). The *TbTfR* is transcribed polycistronically upstream of *VSG* in the telomeric expression site, encoded by two expression site associated genes *ESAG6* and *ESAG7* (Figure 2.4) (Steverding, 2000).

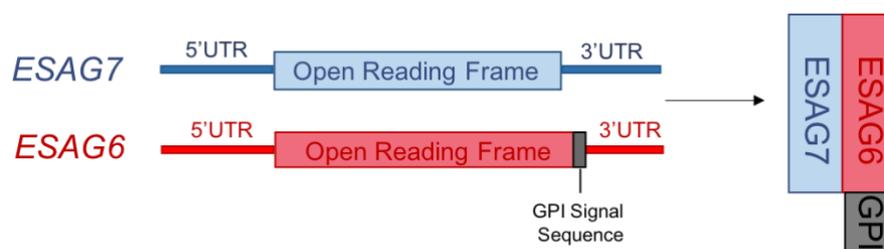


Figure 2.4: Schematic of the *TbTfR*, encoded by *ESAG6* and *ESAG7*.

#### 2.4.1 Identification of the *Trypanosoma brucei* Transferrin Receptor

The presence of a transferrin receptor to mediate the uptake of transferrin in trypanosomes was first hypothesised by Steverding *et al.* in 1995, where experiments using radiolabelled transferrin showed that it was cleared from the environment 200 × faster than would be possible by pinocytosis alone (Steverding *et al.*, 1995). Around this time, affinity chromatography was first used to isolate the transferrin receptor from BSF cells, and it was termed the transferrin-binding protein (Steverding *et al.*, 1995). *ESAG6* and *ESAG7* were identified as the components of the transferrin receptor.

#### 2.4.2 Genomic Organisation of the *Trypanosoma brucei* Transferrin Receptor

*ESAG6* and *ESAG7* are heterogeneously glycosylated proteins with approximately 400 amino acids and 350 amino acids respectively (Fast *et al.*, 1999; Mussman *et al.*, 2004). At the C-terminal domain of *ESAG6*, there is a signal sequence for a glycosylphosphatidylinositol (GPI) anchor to tether the receptor in the membrane (Bitter *et al.*, 1998). Due to the absence of the GPI anchor on *ESAG7*, the protein remains attached to the plasma membrane via interactions with *ESAG6* (Mussman *et al.*, 2004). Both *ESAG6* and *ESAG7* show significant amino acid sequence homology both with each other and *VSG*, despite *ESAG7* lacking the GPI anchor signal sequence (Bitter *et al.*, 1998). It has been suggested that the three proteins may have evolved from a common origin but have diverged to have their own distinct functions (Salmon *et al.*, 1997). The *TbTfR* is localised to the flagellar pocket unlike *VSG* which is present across the surface of the cell (Tiengwe *et al.*, 2016). Tiengwe *et al.* (2017) showed that there is evidence that the *TbTfR* can be detected on the cell surface when expression is increased. Like *VSG*, each expression site encodes a slightly different versions of the *TbTfR*, with up to 5% divergence in amino acid sequence (Mussman *et al.*, 2004).

The structure of the *TbTfR* revealed three long  $\alpha$ -helices in both *ESAG6* and *ESAG7* with the N-terminal helices mapping closely with the structure of *VSG* (Trevor *et al.*, 2019). The *TbTfR* is predicted to adopt the same fold as *VSG* due to the shared features including heptad repeats and conserved disulphides and glycine (Higgins *et al.*, 2014). The *ESAG6* and *ESAG7* subunits are approximately 25% smaller than *VSG* as they do not share the same C-terminal

domain, and it has been proposed that this is to shield the transferrin receptor from antibodies, as the receptor is held below the protective VSG coat (Steverding *et al.*, 2000). Not only could anti-receptor antibodies invoke an immune response against the parasite and cause lysis, but antibodies can also outcompete transferrin and inhibit transferrin uptake, restricting the parasite from obtaining iron (Steverding *et al.*, 2000). N-linked glycosylation of ESAG6 and ESAG7 has also been hypothesised to protect the receptor from immunoglobulins (Trevor *et al.*, 2019). N-glycans located where the receptor is narrowest are thought to prevent receptor crowding by VSG molecules to allow sufficient space for transferrin binding. Other N-glycans form a ring around the top of the receptor adjacent to the ligand-binding domain, also to prevent overcrowding but, more importantly, they cover a large section of the C-terminal domain not involved in ligand-binding, thus protecting the invariant receptor from the host immune response by reducing the likelihood of immunoglobulins binding (Trevor *et al.*, 2019). Saturation of the receptor with transferrin will also protect the receptor from immune recognition, because if host transferrin is bound the exposed area of the receptor is covered by something immunoglobulins will not deem as foreign (Trevor *et al.*, 2019).

#### 2.4.3 Localisation of the *Trypanosoma brucei* Transferrin Receptor

Under basal iron conditions, the transferrin receptor is localised to the flagellar pocket, anchored in the membrane by the GPI anchor, this is considered to be important for immune evasion by the parasite (Mussmann *et al.*, 2003). It has been suggested that at the flagellar pocket the accessibility for transferrin is limited by spatial constraints to protect the plasma membrane from the immune response (Melhert *et al.*, 2012). These authors predicted that the receptor could leave the flagellar pocket making sufficient space in the VSG coat to bind transferrin but also protect the invariant surface of the plasma membrane from the immune system due to the high number of N-glycosylation sites on the receptor compared to VSG. Previously Mussmann *et al.* (2003) had shown that there must be a retention mechanism to localise the TfR in flagellar pocket that was able to distinguish between the transferrin receptor and VSG despite their structural similarity. By overexpressing the transferrin receptor by transfection of additional *ESAG6* and *ESAG7* genes and inducing iron starvation, they showed that TfR in excess of basal levels spread across the surface of the cell like VSG

(Mussmann *et al.*, 2003). This occurs in response to the physiological stimulus of a reduction in available transferrin. The receptor was only shown to be retained at the flagellar pocket when the expression of the receptor increases 3-fold, any fold increase greater than this results in the receptor spreading across the surface of the cell. In 2004 Mussmann *et al.* showed a 5-fold increase in transferrin receptor expression and transferrin uptake demonstrating that the receptors are functional and there is a separate mechanism for internalisation of the receptor-transferrin complex and the localisation of the receptor at the flagellar pocket (Mussmann *et al.*, 2003; Mussmann *et al.*, 2004). However, Schwartz *et al.* (2005) also showed that the *TbTfR* is able to spill-over onto the cell surface yet these receptors were non-functional. When the cells were exposed to Tf-gold conjugates, electron microscopy showed gold particles in the flagellar pocket and nearby endosomal elements but no binding across the surface of the cell, this research looked at the GPI valence and went on to suggest that the TfR observed on the cell surface were actually ESAG6 homodimers, with two GPI anchors like VSG to hold the proteins in the membrane but unable to bind transferrin, and any ESAG7 homodimers would be degraded in the lysosome due to the lack of a GPI anchor (Schwartz *et al.*, 2005). This, however, does not explain the 5-fold increase in uptake of Tf demonstrated by Mussmann *et al.* (2004) when expression of the TfR can only increase 3-fold in the flagellar pocket, therefore, there must be more functional receptor across the cell surface.

#### 2.4.4 Ligand Binding

To be able to bind transferrin, the ESAG6 and ESAG7 monomers must be associated to form the *TbTfR* heterodimeric structure. *In vitro* ESAG6 and ESAG7 can form homodimers but lose the ability to bind transferrin (Salmon *et al.*, 1997). This, therefore, suggests that despite the sequences being almost identical, there are significant differences that are important to each monomer in the heterodimer necessary to generate the ligand-binding site. The membrane-distal loops encoded by the C-terminal domain of the two proteins, adopt different conformations which allows each of the subunits to contribute to the transferrin-binding domain (Trevor *et al.*, 2019). Heterodimeric *TbTfR* is stabilised by hydrogen bonds between the two proteins and slight differences in loop conformations near where the protein is anchored in the membrane (Trevor *et al.*, 2019). Only a single transferrin molecule is able to

bind to the *Tb*TfR and the receptor has affinity for both apo-transferrin and holo-transferrin which can be a limiting factor in iron uptake (Trevor *et al.*, 2019). The human homodimeric receptor has a much higher affinity for holo-transferrin over apo-transferrin with a 245,000-fold preference for holo-transferrin for TfR1 and 35-fold preference for holo-transferrin in TfR2, in comparison *Tb*TfR in BES17 only has a 3.4-fold preference for binding human holo-transferrin (Kleven *et al.*, 2018; Trevor *et al.* 2019). This is an important difference in infection as the human host is able to preferentially take up holo-transferrin and utilise the iron even in anaemic conditions when holo-transferrin is scarce. As the parasite has less of an ability to distinguish between the two, the parasite is more likely to take up apo-transferrin when holo-transferrin is depleted and will not obtain enough iron for survival.

The identification of the ligand-binding domain of the *T. brucei* transferrin receptor was originally achieved through sequence alignments between *ESAG6* and *ESAG7* and multiple *VSG* genes. The alignments showed some regions of variability, most likely the ligand-binding site, as it mapped with the most surface exposed loops of *VSG* in the N-terminal domain (Salmon *et al.*, 1997). There are multiple BES that encode similar but not identical transferrin receptors, these have varying affinities for transferrins from different mammalian hosts (Gerrits *et al.*, (2002). To prove this putative transferrin binding domain, Salmon *et al.* (1997) took advantage of the varying binding affinities by mutating a transferrin receptor with a low affinity for bovine transferrin, with a receptor with a high affinity for bovine transferrin. They performed site directed mutagenesis to replace amino acids in the putative surface loops with those from other expression sites to improve the affinity for bovine transferrin. The mutations made improved the binding of bovine transferrin by a 50-fold increase (Salmon *et al.*, 1997). They also created chimeric receptors where the N-terminal domain of *VSG* was fused to the C-terminal domain of either *ESAG6* or *ESAG7* (the ligand binding region), which were processed correctly to form membrane bound receptors with the ability to bind Tf (Salmon *et al.*, 1997). Changes in amino acids elsewhere in the receptor genes either were neutral or prevented the binding of transferrin, most likely due to changes in the folding of the receptor. This research, therefore, supports the hypothesis that the C-terminal surface exposed loops of the transferrin receptor subunits is the ligand binding domain (Salmon *et al.*, 1997). The crystal structure of the *Tb*TfR revealed that the receptor fits into a cleft in transferrin, binding most extensively with the C-lobe of the N-terminal domain (Trevor *et al.*, 2019). Comparisons

between the structure of mammalian and *T. brucei* transferrin receptors suggest that they bind transferrin with different structural features but they both bind to a similar site on the transferrin molecule (Trevor *et al.*, 2019).

#### 2.4.5 Processing of Transferrin

At the flagellar pocket, the receptor-ligand complex is internalised by clathrin-dependent endocytosis and it is hypothesised that the GPI anchor is cleaved, and the secondary messenger DAG is produced (Kariuki *et al.*, 2019). DAG activates a protein-tyrosine kinase phosphorylation cascade that results in the activation of components involved in the internalisation of the transferrin-receptor complex into the endosome. A decrease in pH in the endosome results in the release of iron from holo-transferrin and at this acidic pH, the apo-transferrin dissociates from the *TbTfR* and is transported to the lysosome for degradation, and the *TbTfR* is recycled back to the flagellar pocket (Kariuki *et al.*, 2019). Evidence for the proteolytic degradation of apo-transferrin was first presented in 1992 when Grab *et al.* showed, using immunofluorescence, that transferrin localises to the lysosome, rather than being recycled and exocytosed from the cell (Grab *et al.*, 1992). Iron released from holo-transferrin is reduced from  $Fe^{3+}$  to  $Fe^{2+}$  by ferric reductases and is imported into the cytoplasm where excess iron is transported to a storage compartment (Kariuki *et al.*, 2019).

#### 2.4.6 Conservation of the Transferrin Receptor in Trypanosomatids

It was previously unknown how procyclic *T. brucei* cells obtain iron in the insect host. Procyclin associated genes (PAGs) are present in the *T. brucei* genome, located at the beginning of the polycistronic unit (Haenni *et al.*, 2006). Within the PAG repertoire are *PAG1* and *PAG2*, the putative proteins of these genes are related to *ESAG6* and *ESAG7*, respectively (Haenni *et al.*, 2006). Their structure suggests that they should form a dimer similar to the BSF transferrin receptor, however, procyclic cells are unable to bind and take up transferrin (Mach *et al.*, 2013). It is also unlikely that procyclic cells make use of haem, like BSF cells, as no orthologue for haem oxidase has been identified in trypanosomatids and without it, iron cannot be released (Mach *et al.*, 2013). Mach *et al.*, showed that iron uptake in procyclic cells is an active process, similar to that utilised by *Saccharomyces cerevisiae* with both yeast and

procyclic cells exhibiting ferric reductase activity, as in *S. cerevisiae* reduction of ferric iron occurs mediated by reductase activity, before transport across the cell membrane. Similarly, procyclic *T. brucei* takes up iron in a two-step mechanism that involves the reduction of ferric iron to ferrous iron which is then transported, this allows acquisition of iron from a range of ferric complexes when other sources of iron are not available in the fly vector (Mach *et al.*, 2013). This route of iron acquisition has also been shown in species of *Leishmania* (Mach *et al.*, 2013).

The African trypanosome *Trypanosoma congolense* is closely related to *T. brucei*. It is also transmitted by the tsetse fly vector and is the major pathogen responsible for nagana in Africa. A group of transferrin receptor-like genes have been identified in two clades in the *T. congolense* genome (Jackson *et al.*, 2013). There are 45 genes in Fam15 that are homologous to *ESAG6* and *ESAG7* and 31 genes in Fam14 with sequences that match closely with *PAGs* expressed in procyclic cells (Jackson *et al.*, 2013). The *T. congolense* genes are more closely related to the genes in *T. brucei* than they are to each other showing that they are paraphyletic. The main difference between the genes in each of the separate species is their location in the genome, *ESAG6* and *ESAG7* are found almost exclusively at the telomeric expression sites, whereas the *T. congolense* orthologues are distributed throughout the sub-telomeres and are not usually close to the telomeres (Jackson *et al.*, 2013). Despite this, sequence comparisons suggest that they still function as a transferrin receptor and it has been shown that the genes are arranged in tandem pairs combining genes with a GPI anchor with ones without from the separate clades (Jackson *et al.*, 2013).

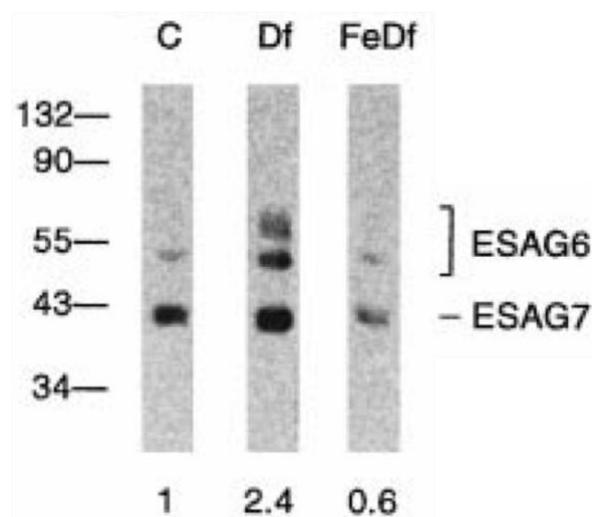
Another African trypanosome, *Trypanosoma vivax*, does contain *VSG*-like genes but they have a very distant relationship to both *T. brucei* *VSG* and *ESAG6* and *ESAG7*, they are not part of the TfR-like gene family of the two other trypanosome species suggesting the *T. vivax* does not have a transferrin receptor (Jackson *et al.*, 2013). It is likely that the three species evolved from a common ancestor, with *T. vivax* diverging first and *T. congolense* undergoing speciation after the evolution of procyclic expression sites, with an orthologous TfR present in both *T. brucei* and *T. congolense* but not *T. vivax* (Jackson *et al.*, 2013). The amino acid sequences in the binding domain of the TfR orthologues are conserved in *ESAG6* and *ESAG7* and *PAGs* and the *T. congolense* orthologues.

## 2.5 Iron Dependent Regulation of the *Trypanosoma brucei* Transferrin Receptor

The parasites have developed a mechanism to maintain iron uptake via the transferrin receptor when iron is scarce in the host environment. This has been studied by experimentally inducing iron starvation conditions in cell culture, which can be achieved by the addition of an iron chelator to the media such as deferoxamine, competition with apo-transferrin, addition of anti-TfR antibodies to the media or by serum switch (Benz *et al.*, 2018; Mussmann *et al.*, 2004). Serum switch induces iron starvation conditions when the media is supplemented with a serum containing transferrin that binds to the expressed TfR with a high affinity is switched with a medium supplemented with a serum with transferrin that binds to the expressed receptor with a low affinity. This is designed to mimic the conditions the parasite would be exposed to clinically when infecting a new mammalian host that may have a transferrin that binds to the receptor expressed in the active expression site with a lower affinity than in the previous mammalian host (Mussmann *et al.*, 2004). It is not an exact model as when the parasite infects a new mammalian host, metacyclic cells are injected from the tsetse fly that then differentiate into slender BSF cells.

Despite the polycistronic transcription of the transferrin receptor along with VSG, under basal conditions only  $3 \times 10^3$  transferrin receptors are expressed per cell compared to  $5 \times 10^6$  VSG homodimers (Salmon *et al.*, 1994; Jackson *et al.*, 1985). When iron starvation conditions are induced the parasite is able to sense and respond to the reduction in iron. A 2.5-5-fold increase in expression of the transferrin receptor is detected, with a corresponding increase in the uptake of transferrin under iron starvation conditions (Benz *et al.*, 2018). The upregulation of the transferrin receptor occurs when available transferrin is reduced but before the internal stores of iron have been depleted (Mussmann *et al.*, 2004). The observed upregulation of the transferrin receptor occurs equally at both an mRNA and protein level, however there is no increase in the level of VSG mRNA suggesting that the iron starvation response is controlled by a post-transcriptional regulation mechanism (Benz *et al.*, 2018).

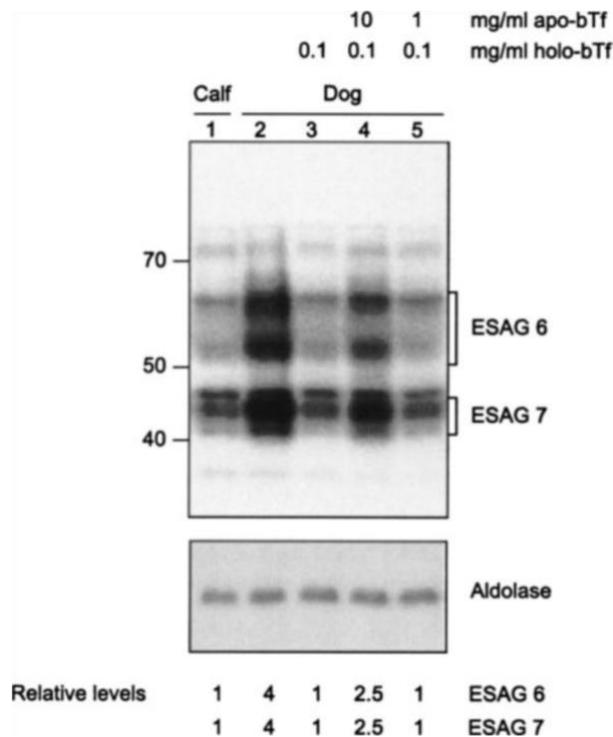
Initial experiments showing the iron-dependent regulation of the *Tb*TfR by Fast *et al.* (1999) utilised radiolabelled transferrin to demonstrate an increased uptake in iron depleted conditions. They observed a 3-fold increase in the uptake of transferrin when cell lines were incubated for 20 hours with deferoxamine, but also a 3-fold increase in the amount of TfR present and the amount of *ESAG6* mRNA (Fast *et al.*, 1999). Western blot analysis was used to observe the increase of the *Tb*TfR shown in **Figure 2.5**. Western blots are not the most accurate way to measure the upregulation of *Tb*TfR as there are multiple copies of the receptor across the different expression sites that can be transcribed even if the BES is not active and the antibodies are not specific enough to distinguish between the different copies of the receptor. They can also give smeared images due to the glycosylation of the TfR protein components, resulting in inaccurate bands that can be difficult to quantify. Fast *et al.* (1999) also went on to predict that the mechanism for the upregulation of the transferrin receptor under iron starvation conditions was to allow immediate adaptation to a new host environment. This was suggested as the amount of transferrin required by the parasite is low enough for there to always be an excess in the bloodstream of the host, and it is unlikely that anti-TfR antibodies are significant to starve the parasite of iron completely. This interpretation supports the theory that there are multiple variants of the *Tb*TfR in the BES to allow survival in multiple mammalian hosts with variable transferrins (Fast *et al.*, 1999).



**Figure 2.5:** Western blot analysis showing the expression of the *Tb*TfR under iron starvation conditions induced by the iron chelator deferoxamine (Fast *et al.*, 1999). Wild type MITat 1.4 BSF *T. brucei* cells were incubated for 20 hours with HMI-9 media (C), supplemented with

deferoxamine (Df), or with deferoxamine saturated with iron (FeDf). The receptor was then affinity purified and detected by immunoblotting with anti-ESAG6 and anti-ESAG7 antibodies. The resulting western blot was quantified by densitometric scanning, indicated beneath the lanes, which showed an approximate 3-fold increase in the presence of ESAG6 and ESAG7 under iron starvation conditions (Df) compared to the two controls (C, FeDf).

Mussmann *et al.* (2004) showed a 4-fold increase in the levels of the transferrin receptor by western blot analysis (**Figure 2.6**) when *T. brucei* 221 cells were switched from media containing 20% bovine serum to 20% dog serum and incubated for up to 8 hours. Under the experimental conditions, the cell population had less than doubled showing that the increase was not due to an increase cell density (Mussmann *et al.*, 2004). The addition of 0.1 mg bovine holo-transferrin prevented the increase in TfR levels, suggesting that the parasite sensed the change in transferrin when the serum was changed. This response could be reversed by the addition of excess apo-transferrin, the parasite does not preferentially take up holo-transferrin like the mammalian TfR. An increase in apo-transferrin starves the parasite, as when it is most abundant in the environment apo-transferrin outcompetes holo-transferrin and the cells do not receive enough iron. As a result, the parasite increases expression of the *TbTfR* to facilitate increased transferrin uptake and increase the chance of taking up holo-transferrin bound to iron (Mussmann *et al.*, 2004). The binding of apo-transferrin to the receptor showed that the ability of the trypanosomes to sense depleted iron in the media did not rely on the receptor being unoccupied (Mussmann *et al.*, 2004).



**Figure 2.6:** Western blot showing the upregulation of the *TbTfR* under iron starvation conditions induced by serum switch (Mussmann *et al.*, 2004). Lister 427 221a BSF *T. brucei* cells were transferred from media with 20% bovine serum (lane 1) for media with 20% canine serum (lane 1) supplemented with varying concentrations of bovine apo- and holo-transferrin, shown above the blot image, and incubated for 8 hours. Western blots were then performed on cell lysates of approximately  $1 \times 10^6$  with antibodies against the 221 transferrin receptor. A 4-fold increase in the expression of the TfR was observed when the cells were grown in canine serum as the 221 *TbTfR* has a lower affinity for canine transferrin compared to bovine transferrin. This could be reversed when the media was supplemented with bovine holo-transferrin. The cells remain iron starved when the media was supplemented with 10 mg/mL bovine apo-transferrin and only 0.1 mg/mL bovine holo-transferrin.

### 2.5.1. Iron Dependent Regulation of the Mammalian Transferrin Receptor

Regulation of the mammalian transferrin receptor is also controlled by a post-transcriptional mechanism which is distinct to that of trypanosomes (Benz *et al.*, 2018). Casey *et al.* (1988) showed that the 3' untranslated region (3'UTR) of the mammalian transferrin receptor is involved in the regulation of the receptor in response to changes in available iron. Biosynthesis of the transferrin receptor is reduced when iron is abundant and increased when

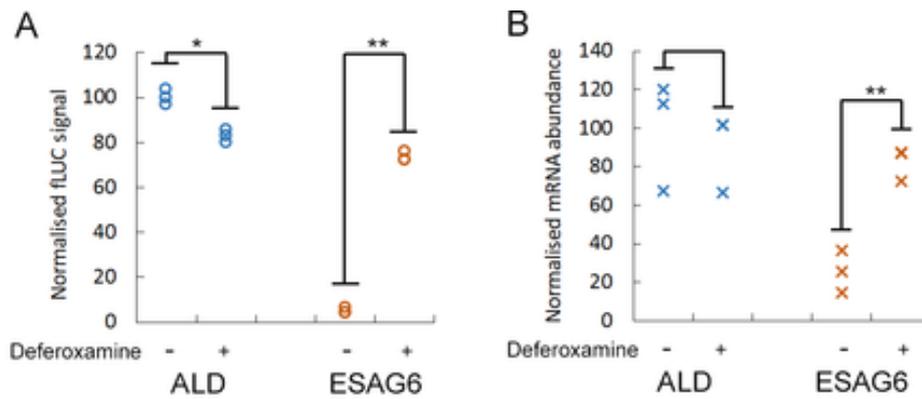
iron is limiting. The involvement of the 3'UTR was confirmed by deleting sequences resulting in elimination of iron-responsive regulation of transcript levels (Casey *et al.*, 1988). They also showed that inserting the TfR 3'UTR into the 3'UTR of the structural gene human growth hormone (hGH) and transfecting the chimeric receptor into murine cells resulted in regulation of the chimeric receptor resembling the regulation in wild type TfR in response to iron levels. Another test showed that no part of the open reading frame (ORF) of the transferrin receptor is involved in regulating the receptor in response to changes in iron levels by transforming cells with a TfR lacking the 3'UTR that could still bind transferrin but did not respond to changes in iron levels. They went on to identify iron responsive-like elements (IRE) by sequence comparison with IREs from the 5'UTR of ferritin, by placing these IRE-like elements into the 5'UTR of hGH. They were able to show an increase in the expression of hGH when iron increased, this is opposite to TfR regulation when iron is abundant showing that hGH was regulated like ferritin where expression increases with iron abundance as ferritin sequesters iron in the cytoplasm (Casey *et al.*, 1988). Fast *et al.* (1999) then went on to show that the mechanism for mammalian transferrin receptor regulation in response to iron levels is mediated by iron regulatory proteins (IRPs). Upon a drop in the iron levels, IRPs bind to iron responsive elements (IREs) present in the secondary structure of the 3'UTR of the mammalian transferrin receptor transcript. IRP binding increases the stability of the mRNA leading to an increase in expression of the receptor, however, they are not present in *T. brucei*, so this mechanism is not conserved (Fast *et al.*, 1999).

#### 2.5.2 Importance of the 3' Untranslated Region of the *Trypanosoma brucei* Transferrin Receptor in the Iron Starvation Response

The 3'UTR has been shown to be important for the regulation of the mammalian transferrin receptor. It also plays a key role in the regulation of the *TbTfR* but through a different mechanism. The involvement of the *ESAG6* 3'UTR in the upregulation of the receptor under iron starvation conditions was investigated by fusion of the *ESAG6* 3'UTR and *ALD* 3'UTR to a firefly luciferase (*fLUC*) gene (Benz *et al.*, 2018). The constructs were ligated into a specialised pRPaΔ vector downstream of an RRNA promoter. When transfected into 2T1 BSF cells, this vector integrates into a tagged *RRNA* locus to give uniform expression. Any

responses recorded in the luciferase assays are, therefore, due to the fused 3'UTR region rather than any positional effects as the location of the vector in the genome is known.

Luciferase assays performed by Benz *et al.* (2018) under normal conditions showed that there was a higher relative luminescence signal for cells expressing the *fLUC-ALD-3'UTR* vector compared to cells expressing the *fLUC-ESAG6-3'UTR* vector. This is because the *ALD* 3'UTR is not at all involved in the iron starvation response so is not regulated in the same way as the *ESAG6* 3'UTR, readings should therefore always be higher for the *ALD* cell lines as the *ALD-3'UTR* is not repressed and is a stable protein. Repeating the assay for the *ESAG6* 3'UTR cell line at varying cell densities showed that the relative luminescence increases with cell density showing that the assay is density dependent (Benz *et al.*, 2018). When iron starvation conditions were induced by serum switch and incubation with the iron chelator deferoxamine, Benz *et al.* (2018) showed an increase in the expression of the *TbTfR* at both an mRNA and protein level as shown in **Figure 2.7**. This shows that fusion of the *ESAG6* 3'UTR to a reporter gene demonstrates the dynamic regulation of the *TbTfR* at both an mRNA and protein level in response to iron starvation conditions, highlighting the importance of the *ESAG6* 3'UTR for regulation of the *TbTfR*. Treatment with deferoxamine for 5 hours significantly increased the luciferase activity of the *ESAG6* 3'UTR cell line approximately 10-fold when compared to the non-iron starved control cell line, and the same effect was confirmed at an mRNA level by RT-PCR (Benz *et al.*, 2018). A lower fold change was observed by serum switch to canine sera of around a 3.5-fold increase in luciferase activity of the *ESAG6* 3'UTR cell line.



**Figure 2.7: Luciferase activity assay showing increase in expression of the *fLUC-ESAG6-3'UTR* reporter under iron starvation conditions (Benz *et al.*, 2018).** A. Luciferase activity of *ESAG6* and *ALD* reporter cell lines incubated with or without 25  $\mu$ M deferoxamine for 5 hours, normalised to the untreated *ALD* signal. B. mRNA levels of *ESAG6* and *ALD* reporter cell lines normalised to untreated *ALD* mRNA levels. Both graphs show a significant increase in the level of the *fLUC-ESAG6-3'UTR* signal with the addition of deferoxamine, demonstrating that the 3'UTR is involved in the dynamic regulation of the transferrin receptor. The aldolase signal remains high. Points for three biological replicates is shown for each set of data. \* $p < 0.05$ , \*\* $p < 0.001$ .

This luciferase assay is a proxy for the expression of the transferrin receptor. The transferrin receptor is not being directly measured, as any alterations made to the receptor may be detrimental to the cells resulting in cell death. Despite this, the assay does mimic previous data showing a similar 3-6-fold change in the upregulation of the receptor under iron starvation conditions.

This post-transcriptional mechanism is distinct to that observed in iron-dependent regulation of the mammalian TfR through IRE, as knockout experiments of the *T. brucei* IRP-1 homologue aconitase has no effect on regulation of the *TbTfR* (Fast *et al.*, 1999). The results from the Benz *et al.* (2018) study supported other observed increases in the expression of the transferrin receptor, but first identified the importance of the *ESAG6* 3'UTR.

Research by Carbajo *et al.* (2021) also highlighted the importance of the 3'UTR in the iron starvation response of the *TbTfR*. Experiments were conducted to investigate the post-transcriptional regulation of the TfR in *T. brucei* through the addition of deferoxamine and

cycloheximide (CHX) to the cultures. Deferoxamine induced iron starvation and CHX blocked protein synthesis, the TfR was then quantified by immunoblotting, to show that expression of the TfR protein increased after the addition of deferoxamine, but the turnover rate remained unchanged, suggesting that the upregulation of the receptor is driven by mRNA stability when iron is scarce (Carbajo *et al.*, 2021). Tagging the *ESAG6* 3'UTR to an mNG reporter construct, showed that following deferoxamine treatment, there was an observed increase in *ESAG6* protein and mRNA in line with endogenous TfR. This led these authors to argue that TfR mRNA stability is reliant on putative conserved, cis-acting IRE in the 3'UTR, in an iron dependent manner for regulation, as first presented by Benz *et al.* (Carbajo *et al.*, 2021).

By sequencing the mRNA from cells treated with deferoxamine, Carbajo *et al.* (2021) identified a number of genes that were upregulated under iron starvation conditions, along with *ESAG6* and *ESAG7*. Aside from the TfR genes a putative RNA binding protein (RBP5) was identified that was most significantly upregulated, indicating that it is an iron regulated gene, that is co-regulated with the *TbTfR* (Carbajo *et al.*, 2021). qRT-PCR showed a 3-fold increase in *RBP5* mRNA and western blot analysis showed an 8-fold increase in the RBP5 protein when deferoxamine was added to the media. When  $\text{FeCl}_3$  was added to the media, along with pre-saturated deferoxamine, there was a 30% decrease in *RBP5* mRNA and a 90% downregulation of the protein, showing that RBP5 is regulated by iron availability at both an mRNA and a protein level, when iron is scarce they are upregulated and when iron is abundant, they are downregulated (Carbajo *et al.*, 2021). They also confirmed that this regulation is reliant on the 3'UTR of *RBP5* by replacing the *RBP5* 3'UTR with the 3'UTR of the paraflagellar rod protein which does not respond to changes in iron availability (Carbajo *et al.*, 2021). The loss of the *RBP5* 3'UTR resulted in the loss of the RBP5 protein ability to respond to changes in iron levels, showing that the *RBP5* 3'UTR is necessary for this response. It is likely that the regulation of RBP5 is controlled by the 3'UTR mediated by mRNA stability (Carbajo *et al.*, 2021). It is not yet clear what the role of RBP5 is, but data suggests that it does not bind to the *ESAG6* 3'UTR to directly regulate the transferrin receptor (Carbajo *et al.*, 2021). Yet, it is an essential protein as attempted elimination of *RBP5* through RNAi resulted in gene duplication, and RNA silencing reduced the mRNA level by 90% but had no effect on growth, indicating that RBP5

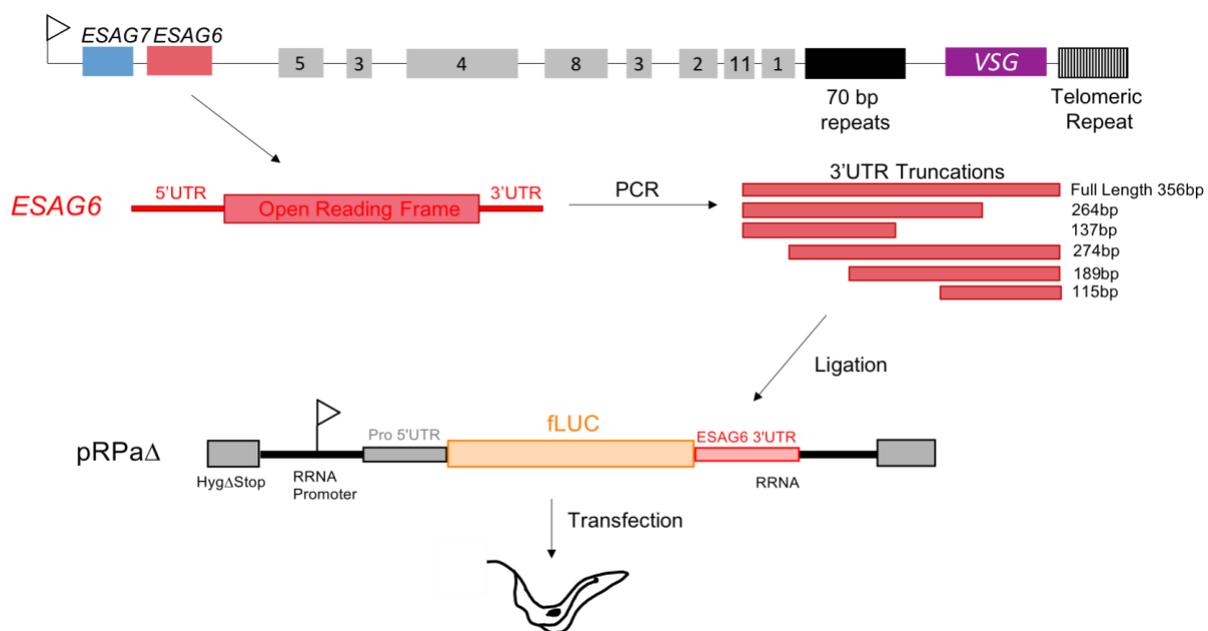
is essential but only low levels are required for parasite survival. Overexpression of RBP5 is toxic to the cells due to high RNA binding affinity (Carbajo *et al.*, 2021).

RBP5 stability is fairly atypical whereby the mRNA is highly stable, and the protein is unstable. Protein stability is not a key factor in the upregulation of the TfR (Carbajo *et al.*, 2021). This suggests that mRNA stability is essential for transferrin uptake in iron starvation conditions, as increased mRNA stability increases translation. It is likely that this stability is conferred by mRNA recruitment to translation machinery or through an RBP that binds to the 3'UTR (Carbajo *et al.*, 2021). Due to the similar kinetics in the *Tb*TfR and RBP5 responses to iron levels, they are likely regulated by a similar mechanism and potentially even by a common factor (Carbajo *et al.*, 2021).

The overall consensus is that when iron is scarce in the environment, trypanosomes upregulate expression of the transferrin receptor to increase transferrin and therefore iron uptake by receptor mediated endocytosis. This is undisputed and has been shown in multiple experiments through induction of iron starvation condition via a number of mechanisms including serum switch and the addition of deferoxamine (Benz *et al.*, 2018; Carbajo *et al.*, 2021; Fast *et al.*, 1999; Mussmann *et al.*, 2004). It has also been shown that this upregulation is via a post-transcriptional mechanism as other mRNA and proteins transcribed on the same polycistronic unit are not simultaneously upregulated (Benz *et al.*, 2018). What has yet to be experimentally demonstrated is how the receptor has been upregulated. Benz *et al.* (2018) showed that the *ESAG6* 3'UTR is important in this mechanism, as supported by research by Carbajo *et al.* (2021). It is likely that both the *ESAG6* and *ESAG7* 3'UTRs are responsible for regulating the transferrin receptor via a similar if not the same mechanism, due to the sequence similarity and their co-regulation for a functional receptor. It is likely that there is an RNA-binding domain present in the 3'UTRs for an RBP to bind and stabilise the mRNA increasing translation and expression of the *ESAG6* and *ESAG7* proteins. The evidence from Carbajo *et al.* (2021) that there is an identified RBP that is regulated by iron levels supports this hypothesis.

## 2.6 Aims and Objectives

The *Trypanosoma brucei* transferrin receptor is able to sense and respond to fluctuations in iron levels by a previously unknown post-transcriptional mechanism. The *ESAG6* 3'UTR has been demonstrated to mediate a specific response to iron starvation conditions in *T. brucei* Lister 427 through fusion of the 3'UTR to an *fLUC* reporter gene, at both an mRNA and a protein level with a magnitude previously observed for *TbTfR* upregulation (Benz *et al.*, 2018). Yet, it is unknown whether a certain motif in the 3'UTR is important for the upregulation of the *TbTfR* under iron starvation conditions or if the response requires the full sequence. It has also not yet been determined whether the *ESAG7* 3'UTR is involved in iron-dependent regulation of the receptor or not. The overall aim of this project is to characterise the *ESAG7* 3'UTR as the length is currently unknown and identify any important motifs of the *ESAG6* and *ESAG7* 3'UTRs that are involved in this regulation (**Figure 2.8**).



**Figure 2.8:** Experimental procedure to generate 2T1 BSF cell lines containing the truncated *ESAG6* 3'UTR sequences in the *fLUC* reporter system. The length of the *ESAG6* 3'UTR was identified by Benz *et al.* (2018) as 335 bp in length, primers were designed to amplify a fragment 356bp in length to include all identified polyadenylation sites. To identify any important motifs within this 356 bp sequence, N and C terminal truncations were made of the *ESAG6* 3'UTR at hypervariable regions between the BES and were ligated into a specialised

pRPaΔ vector containing a firefly luciferase gene. The truncated sequences in the pRPaΔ-*fLUC<sup>mut</sup>-ESAG6-3'UTR* were then transfected into 2T1 BSF cells and luciferase activity assays were performed to monitor the upregulation of the *TbTfR* under iron starvation conditions compared to the full length *ESAG6 3'UTR*.

The upregulation of the *TbTfR* when iron is scarce has been shown utilising multiple mechanisms to induce iron starvation conditions. In this research the iron starvation response is induced by switching the cultured cell lines from media supplemented with transferrin that binds to the receptor in BES1 with a high affinity for a lower affinity transferrin. The 2T1 cell line is monomorphic, so is stable with BES1 as the active expression site, expressing the associated ESAGs and VSG 221. The switching mechanism is still functional and occurs in the background, however, loss of VSG 221 is unfavourable, so any cells that had switched expression site are outgrown and are not able to establish a population (Liu *et al.*, 2018). In all experiments, the cells are not grown for longer than 5 hours under iron starvation conditions, and therefore have not been left in culture long enough to induce switching, so it is known that the cells are all expressing the *TbTfR* from BES1.

As the mechanism for regulation of the *TbTfR* is distinct from that used by mammalian cells involving IREs, RNA binding proteins (RBPs) could be important in trypanosomes. RBPs have been shown to be involved in modulating mRNA stability in trypanosomes, so this mechanism may be utilised in the regulation of the transferrin receptor in response to iron, as under iron starvation conditions an RBP could bind to TfR mRNA to increase stability and translation so more receptors are expressed on the cell surface (Carbajo *et al.*, 2021). The overall aim of this research is to identify a region of the *ESAG6* or *ESAG7 3'UTR* that is important for the iron starvation response. This may be a sequence that could be recognised by RBPs. Other RBP motifs identified in *T. brucei* have been short sequences of 16 nucleotides in *VSG 3'UTR* and 34 nucleotides in the *ESAG9 3'UTR* (Ridewood *et al.*, 2017; Monk *et al.*, 2013). By truncating the *ESAG6* and *ESAG7 3'UTRs* to shorter overlapping sequence an important sequence may be identified that could be further truncated to confirm a potential RBP binding motif.

There are three main predicted outcomes from this research. These include the truncations failing to respond to iron starvation suggesting that an upregulation control element has been

removed when the truncation was generated. Alternatively, the truncations may have a higher background signal, suggesting that a repressive control element has been removed. Finally, there could also be a combination of both of these responses observed in the luciferase activity assays suggesting that the iron starvation response is under more than one level of control. The secondary structure of the 3'UTRs were not taken into consideration when the truncations were designed and if important conformations have been disrupted there may not be an observed response.

## 3.0 Materials and Methods

### 3.1. Materials

#### 3.1.1. Media

##### 3.1.1.1 HMI11-T

2T1 BSF *T. brucei* cells were grown in HMI11-T media. This is HMI9 without serum plus with 1-Thioglycerol replacing  $\beta$ -mercaptoethanol and Glutamax replacing L-Glutamine (Hirumi and Hirumi, 1989). Media was made up in 5 L batches and stored as 450 mL sterile filtered aliquots at 4°C in the dark. FBS and Glutamax were added under sterile conditions prior to use.

**Table 3.1: Ingredients and amounts to make 5 L HMI11-T *T. brucei* BSF media.**

Ingredient	Product	5 L	500 mL
HMI-9 powder	Invitrogen	1 pk in 4.5L	
Sodium bicarbonate	Sigma S5761, RT	10 g	
pH adjusted to 7.3 with NaOH, sterile filter			
Foetal bovine serum	Labtech, -20 °C	-	50 mL
Glutamax 1, 100x	Labtech, -20 °C	-	5 mL

##### 3.1.1.2 LB Broth

500 mL LB broth was made up by dissolving 12.5 g LB broth powder (Melford) in 500 mL ddH<sub>2</sub>O and pH adjusted to 7.2 and autoclaved at 121°C for 1 hour.

##### 3.1.1.3 LB Agar

To make up 500 mL of LB agar, 20 agar capsules (Melford) were added to 500 mL H<sub>2</sub>O and autoclaved at 121°C for 1 hour. Once cooled carbenicillin (100  $\mu$ g/mL) was added and plates were poured in a sterile environment.

### 3.1.2. Antibiotics

**Table 3.2: Antibiotic uses and concentrations**

Antibiotic	Cells	Stock	Concentration Used	Manufacturer
Carbenicillin	<i>E. coli</i>	50 mg/mL	100 µg/mL	Melford
Puromycin	<i>T. brucei</i> Lister 427 2T1	2 mg/mL	0.2 µg/mL (10,000 ×)	Roche
Phleomycin	<i>T. brucei</i> Lister 427 2T1	10 mg/mL	0.5 µg/mL (20,000 ×)	Melford
Hygromycin	<i>T. brucei</i> Lister 427 2T1	50 mg/mL	2.5 µg/mL (20,000 ×)	Melford

### 3.1.3. Buffers

1L 1 × TAE running buffer was made up of 20 mL 50 × TAE (Fisher Scientific) and 980 mL ddH<sub>2</sub>O.

## 3.2. Cell Culture

2T1 (Lister 427) bloodstream form *T. brucei* cells (Alsford *et al.*, 2005) were maintained in 10mL HMI11-T media + puromycin + phleomycin (**Table 3.2**) in 25cm<sup>3</sup> filter cap flasks (SLS). Cells were incubated at 37°C with 5% CO<sub>2</sub> and 100% humidity. Cultures were split every two days to prevent overgrowth and cell death.

## 3.3 Agarose Gel Electrophoresis

Samples were run out on a 1.5% agarose gel with 10,000 × SYBR safe (Invitrogen) and 1 × TAE as running buffer, alongside 3 µL of 100 bp MW ladder (Promega) unless otherwise stated. 6 × purple loading dye (NEB) was added to colourless samples. Gels were imaged on a BioRad Gel Doc EZ imager.

### 3.3.1. Gel Extraction

DNA samples were analysed by gel electrophoresis. Bands of a desired MW were excised on a UV transilluminator (Syngene) and DNA was extracted using the Thermo Scientific GeneJet Gel Extraction kit as per manufacturer's instructions.

## 3.4 Polymerase Chain Reaction (PCR)

PCRs were performed in a BioRad T100 Thermal Cycler using thin walled 0.5 mL tubes with a total volume of 50  $\mu$ L per reaction. The primers and reaction conditions varied depending on the nature of the target fragment and are described below. The GoTaq polymerase system (Promega) was used for the PCRs unless otherwise stated, the components of a single reaction are shown in **Table 3.3**. For the GoTaq polymerase reactions, the thermocycler was programmed at 94°C for 2 min 45s to denature DNA, at a variable temperature for 1 min for primers to anneal and 72°C for 1 min for extension and repeated 34 times.

**Table 3.3: Components of a GoTaq polymerase reaction and the volumes used.**

PCR Component	Volume
5 × Green GoTaq buffer	10 $\mu$ L
MgCl <sub>2</sub> (25 mM)	5 $\mu$ L
Nucleotide dNTPs (10 mM)	1 $\mu$ L
Template DNA (variable)	0.1 $\mu$ L (unless otherwise stated)
Taq Polymerase	0.25 $\mu$ L
F-Primer (variable)	1 $\mu$ L
R-Primer (variable)	1 $\mu$ L
ddH <sub>2</sub> O	Variable, to make up total volume of 50 $\mu$ L

### 3.4.1 PCR of *ESAG6* 3'UTR from Genomic DNA

The *ESAG6* 3'UTR was amplified from 2T1 *T. brucei* genomic DNA using the OneTaq polymerase system (NEB) with a forward primer specific to the BES1 *ESAG6* ORF (5' -

GCAGTACATTTGAGTCTTT - 3') and a reverse primer complimentary to the end of the BES1 *ESAG6* 3'UTR (5'- AATAGGGCCCAGTAGAATTAGTCTAGTTT - 3'). Each reaction contained 10  $\mu$ L 5  $\times$  OneTaq standard reaction buffer, 1  $\mu$ L 10mM dNTPs, 0.25  $\mu$ L OneTaq Hot Start polymerase, 0.1  $\mu$ L template DNA, 1  $\mu$ L forward primer and 1  $\mu$ L reverse primer with the final volume made up to 50  $\mu$ L with ddH<sub>2</sub>O. The thermocycler was programmed at 94°C for 1 min to denature DNA, 55°C for 45 sec for primers to anneal and 68°C for 45 sec for extension. This cycle was repeated 34 times before a final extension step at 68°C for 10 min.

PCR products were analysed on an agarose gel (**Section 3.3**) to confirm amplification and the size of the band as ~350bp, and were gel extracted for ligation into pGEM-T easy (**Section 3.6**).

#### 3.4.2 PCR of *ESAG6* 3'UTR Truncations from Plasmid Template

Amplification of the *ESAG6* 3'UTR truncations was performed using the GoTaq polymerase system (Promega) (**Table 3.3**) with template DNA generated in 3.4.1 (pGEM-T-*ESAG6*-ORF-3'UTR) and an annealing temperature of 55°C. The primer sequences used to generate the truncations are listed below with the combinations and truncation lengths in **Table 3.4**.

Full length FP: 5'- AATAGAGGATCCGGGAAGGATGCGAC - 3'

FP84: 5'- TTATGGATCCCCTATTTCTTTTATTTGGGG - 3'

FP169: 5'- ATTAAGGATCCCTAAGGAAAGATGTGAGTTC - 3'

FP243: 5' - ATATTGGATCCGGAAAATGTAACAACCAAC - 3'

Full length RP: 5'- AATAGGGCCCAGTAGAATTAGTCTAGTTT - 3'

RP138: 5' - TATTAGGGCCCCACTGGAAACTTACGTTAC - 3'

RP265: 5'- AATATGGGCCCCATAGTTGGTTGTTACATTTTC - 3'

**Table 3.4: Truncations of the *ESAG6* 3'UTR of BES1 generated by PCR amplification to be ligated into the *fluc* reporter system and transfected into 2T1 BSF *T. brucei* cells.**

Truncation	Forward Primer	Reverse Primer	Size of Insert
A (+ve Control)	Full length FP	Full length RP	356bp

<b>B</b> (-ve Control)	-	Full length RP	-
<b>C</b>	Full length FP	RP265	264bp
<b>D</b>	Full length FP	RP138	137bp
<b>E</b>	FP84	Full length RP	274bp
<b>F</b>	FP169	Full length RP	189bp
<b>G</b>	FP243	Full length RP	115bp

Amplification was confirmed by gel electrophoresis and PCR products were purified using the GeneJet PCR purification kit (Thermo Scientific) as per manufacturer's instructions for ligation into pRPa $\Delta$ -*fLUC<sup>mut</sup>*-*ESAG6*-3'UTR.

#### 3.4.3 PCR of *ESAG7* 3'UTR from cDNA

From the reverse transcription reaction (**Section 3.5**), 5  $\mu$ L of cDNA was used as template DNA to amplify the *ESAG7* 3'UTR. The GoTaq polymerase PCR system was used with primers specific for the BES1 *ESAG7* ORF (5'- CAGCTTTACGAAHGAATTC - 3') and the adapter sequence primer (5'-CGCGTCGACTAGTAC - 3') with an annealing temperature of 53.1°C.

The PCR products were purified using a GeneJet PCR purification kit (Thermo Scientific) as per manufacturer's instructions and the resulting DNA was ligated into the pGEM-T easy vector system (**Section 3.6**). The resulting samples were then sent off for sequencing (**Section 3.7**) to identify the length of the *ESAG7* 3'UTR.

#### 3.4.4 PCR of *ESAG7* 3'UTR Truncation from Plasmid Template

The primers used to generate the *ESAG7* 3'UTR truncations are listed below, and the sizes of the inserts located in **Table 3.5**. The GoTaq polymerase system was used to generate the truncations (**Table 3.3**) with an annealing temperature of 60°C and template DNA generated in 3.4.3 (pGEMT-*ESAG7*-ORF-3'UTR).

Primer Sequences:

Full Length FP: 5' – TATAGGATCCAATGGAGTAAAGGCGAATT – 3'

FP147: 5' – TATTAGGATCCAAGGATGCGACAGAAG – 3'

Full Length RP: 5' – ATATTGGGCCCTTAGTCTAGTTTTCTTTATA – 3'

RP349: 5' – AATAGGGCCCCACACGTTTTTCGTAGC – 3'

**Table 3.5: Truncations generated of the *ESAG7* 3'UTR from BES1 by PCR amplification to be ligated into the *flUC* reporter system and transfected into 2T1 BSF *T. brucei* cells.**

Truncation	Forward Primer	Reverse Primer	Size of Insert
<i>ESAG7</i> 3'UTR	Full Length FP	Full Length RP	492bp
<b>1</b>	Full Length FP	RP349	364bp
<b>2</b>	FP147	Full Length RP	346bp
<b>3</b>	FP147	RP349	218bp

The presence of amplified DNA was confirmed by gel electrophoresis (**Section 3.3**) and the correct PCR products were purified using the Thermo Scientific PCR purification kit as per manufacturer's instructions. The resulting DNA truncations were ligated into the pRPaΔ-*flUC<sup>mut</sup>*-*ESAG6*-3'UTR vector (**Section 3.6**), to generate pRPaΔ-*flUC<sup>mut</sup>*-*ESAG7*-3'UTR variations.

### 3.5 Reverse Transcription

RNA was extracted from  $1 \times 10^6$  logarithmic phase 2T1 BSF *T. brucei* cells using a RNeasy plus mini kit (Qiagen) following the manufacturer's instructions, the amount of RNA extracted was quantified on a nanodrop (Thermo Scientific, 2000c Spectrophotometer). A reverse transcription reaction was set up using an Oligo-dT primer (5'-CGCGTCGACTAGTACTTTTTTTTTTTTTTTTTT-3') and Omniscript Reverse Transcriptase (Qiagen) to transcribe cDNA from 1 ug of the RNA template. A negative control reaction was set up excluding the reverse transcriptase enzyme from the reaction.

## 3.6 Cloning

The products from the PCRs described in section 3.4 were ligated into either the commercially available pGEM-T easy vector or into the pRPa $\Delta$ -*fluc*<sup>mut</sup> vector system (a kind gift from Sam Alford, LSHTM) to generate new variations of the vector with truncations of either the *ESAG6* 3'UTR or the *ESAG7* 3'UTR. The vector each product was ligated into is listed in the respective subsections of 3.4. The general cloning procedure is described below.

### 3.6.1 Ligation

Ligation reactions were set up containing 0.5  $\mu$ L T4 DNA ligase (Promega), 2.5  $\mu$ L 2  $\times$  rapid ligation buffer (Promega), 0.5  $\mu$ L vector (either pGEM-T easy (Promega) or pRPa $\Delta$ -*fluc*<sup>mut</sup>-*ESAG6*-3'UTR) and 1.5  $\mu$ L insert giving a total volume of 5  $\mu$ L. The ligation was left at room temperature for up to 1 hour.

### 3.6.2 Transformation

A 25  $\mu$ L aliquot of high efficiency *Escherichia coli* competent cells (C2987I) (NEB) was added to the ligation reaction and left on ice for up to 30 mins. Cells were heat-shocked at 42°C for 30 sec and then left on ice for 5 mins. Cells recovered in 150  $\mu$ L SOC (NEB) at 37°C with shaking for 45 mins. They were plated onto LB + Amp (100  $\mu$ g/mL) plates and incubated overnight at 37°C. If blue-white screening was necessary, 30  $\mu$ L 0.1 M IPTG (Melford) + 30  $\mu$ L 20  $\mu$ g/mL X-Gal was added to the plates before the cells were seeded.

### 3.6.3 Isolation of Plasmid DNA

Colonies were selected and grown overnight at 37°C with shaking in 5 mL LB broth + Amp (100  $\mu$ g/mL) in a 50 mL falcon tube with a loose lid.

Cell suspensions were centrifuged in a swing-bucket rotor at  $3500 \times g$  for 15 min to pellet transformed *E. coli* cells. Minipreps were then performed as per manufacturer's instructions (Thermo Scientific GeneJet Miniprep kit). Samples stored at  $-20^{\circ}\text{C}$  until further processing.

#### 3.6.4 Analytical Restriction Enzyme Digest

Analytical restriction enzyme (RE) digests were performed on the minipreps to screen for the correct sized insert. A total volume of  $10 \mu\text{L}$  was made up of  $0.5 \mu\text{L}$  restriction enzyme (*EcoRI* (NEB) for pGEM-T; *ApaI* (NEB) and *BamHI* (NEB) for pRPa $\Delta$ ),  $1 \mu\text{L}$   $10 \times$  cutsmart buffer (NEB),  $5 \mu\text{L}$  of miniprep and made up to  $10 \mu\text{L}$  with ddH<sub>2</sub>O. They were incubated at  $37^{\circ}\text{C}$  for 2 hours.

Digested samples were run adjacent to undigested minipreps in a gel. The size of the band was matched to the PCR band to confirm the correct insert.

#### 3.6.5 Preparative RE Digest

RE digests were set up with  $2 \mu\text{L}$  of each restriction enzyme,  $5 \mu\text{L}$  cutsmart buffer,  $35 \mu\text{L}$  PCR product/vector and made up to  $50 \mu\text{L}$  with ddH<sub>2</sub>O and incubated overnight at  $37^{\circ}\text{C}$  to prepare the DNA inserts for ligation.

### 3.7 DNA sequencing

Samples were sent off to DNA Sequencing and Services (Dundee University) at a DNA concentration and volume requested on the website, to confirm insert size and the correct sequence when aligned with *T. brucei* Lister 427 DNA sequences available online [<https://www.ncbi.nlm.nih.gov/nuccore/FM162566> GenBank: FM162566] (Hertz-Fowler *et al.*, 2008).

Samples ligated into pGEM-T easy were sequenced with a commercially available primer, complimentary to the T7 promoter in the pGEM-T easy vector backbone and constructs

ligated into pRPa $\Delta$ -*fLUC*<sup>mut</sup>-*ESAG6*-3'UTR were sent off for sequencing with the primer pRP3'UTRseq (5'-TAACCAAGACTCCAAAAGCC -3') complimentary to the vector backbone.

### **3.8 Trypanosoma brucei Transfection and Screening**

#### 3.8.1 Linearising the DNA

3  $\mu$ g of pRPa $\Delta$ -*fLUC*<sup>mut</sup>-*ESAG6*-3'UTR (full length *ESAG6* 3'UTR in the *fLUC* reporter system) was digested overnight at 37°C with 2  $\mu$ L of *Ascl* in a total volume of 50  $\mu$ L. Endonucleases were heat inactivated at 80°C for 20 min and the reaction transferred to ice.

#### 3.8.2 Ethanol Precipitation

5  $\mu$ L of 3M NaAc was added to the reaction on ice to give a final concentration of 300 mM NaAc, then 150  $\mu$ L of -20°C 100% EtOH was added to give a final concentration of 70% EtOH and the precipitation reaction was left at -20°C for up to 24 hours. The ethanol precipitation reaction was centrifuged at 18,000  $\times$  g at 4°C for 10 min. The supernatant was removed, and 1000  $\mu$ L -20°C 70% EtOH was added on ice and was centrifuged again at 18,000  $\times$  g at 4°C for 10 minutes. This step was then repeated. The supernatant was removed and replaced with 1000  $\mu$ L of -20°C 70% EtOH on ice and was centrifuged again for 5 mins at 18,000  $\times$  g at 4°C. The DNA sample was moved into the category II containment facility and the supernatant was removed allowing the pellet to air-dry for 10 mins in sterile conditions inside the biological safety cabinet. The DNA was then resuspended in 5  $\mu$ L of filter sterilised ddH<sub>2</sub>O.

#### 3.8.3 Amaxa Transfection

2T1 *T. brucei* BSF cells were grown to log phase ( $1.5 \times 10^6$  cells/mL) for the transfections. HMI11-T media without antibiotics was prewarmed to 37°C and nucleofector (+ supplement) (Lonza) was pre-warmed to room temperature.  $2 \times 10^7$  2T1 *T. brucei* BSF cells for one transfection and one control were centrifuged at 800  $\times$  g in a swing-bucket rotor at room temperature for 10 mins. The supernatant was removed, and the cells were resuspended in

200 µL nucleofector + supplement. 100 µL cells were transferred to a cuvette with 4 µL DNA and 100 µL cells were transferred to a no DNA control cuvette with 4 µL ddH<sub>2</sub>O. The Amaxa was set to programme X00-1 and the cuvettes were electroporated (Burkard *et al.*, 2011). The electroporated cells were then transferred to 25 mL warmed HMI11-T media without antibiotics and allowed to recover for 6 hours at 37°C with 5% CO<sub>2</sub> and 100% humidity. After 6 hours 25 mL HMI11-T media with 2 x phleomycin and 2 x hygromycin (**Table 3.2**) was added to the flasks. The transfected cell lines were then plated into 24 well plates with 1 mL media per well and allowed to grow at 37°C, 5% CO<sub>2</sub>, 100% humidity for up to 14 days.

#### 3.8.4 Screening for Transfected Clones

Once all the cells in the no DNA control had died the plates were disposed of, this occurred around day 4 post-transfection. From then on, each well in the test plates was screened daily by optical microscopy, with noticeable cell growth occurring around day 6 post-transfection. When a well was identified with living cells, it was marked and monitored until they had reached a high density that remained when agitated. The cells were then transferred to a flask in a 1 in 10 dilution into 10 mL HMI11-T + phleomycin and hygromycin (**Table 3.2**). The cells were then allowed to grow and were split regularly until the cells were dividing at an expected rate with the correct morphology.

##### 3.8.4.1 Drug Screen Assay for Correct Insertion

10 µL transfected and control cell lines ( $< 2 \times 10^6$  cells/mL) were added to 1 mL phleomycin and puromycin (**Table 3.2**) HMI11-T media and 1 mL phleomycin and hygromycin (**Table 3.2**) HMI11-T media in separate wells of a 24 well plate. Plates were left to incubate at 37°C with 5% CO<sub>2</sub> for up to 4 days. Wells were screened for living cells daily.

This transfection protocol was then repeated for the remaining five *ESAG6* 3'UTR truncations (C-G), the *ESAG7* 3'UTR full-length sequence and truncations (1-3), all in the *fLUC* reporter system.

### 3.9. Luciferase Activity Assay

#### 3.9.1 Induction of Iron Starvation

*ESAG6* 3'UTR truncated cell lines (C-G) and control cell lines (pRPaΔ-*fLUC*<sup>mut</sup>-*ESAG6*-3'UTR and pRPaΔ-*fLUC*<sup>mut</sup>-*ALD*-3'UTR) were grown overnight to log phase ( $5 \times 10^5$  cells/mL). Cells were counted and 1mL of each culture at  $5 \times 10^5$  cells/mL was then centrifuged in duplicate at 3500 × g for 10 minutes at room temperature. The supernatant was discarded, and the cells were resuspended in 1 mL HMI11-T media without GLX and FBS. The centrifugation step was then repeated, and the supernatant removed. One set of truncated and control cell lines was then resuspended in 1 mL 10% FBS HMI11-T media (normal conditions), and to induce iron starvation conditions, the other set of truncated and control cell lines was resuspended in 1 mL 10% canine HMI11-T media and were plated out into 24 well plates. The plates were incubated for 5 hours at 37°C with 5% CO<sub>2</sub> and 100% humidity.

Control cell lines contain the full length *ESAG6* 3'UTR fused to the *fLUC* reporter gene and the *ALD* 3'UTR fused to the *fLUC* reporter gene.

#### 3.9.2 One Glo Luciferase Assay

Cell counts were performed and recorded for all the cell lines under both conditions (iron-starved and non-iron starved). 50 µL of each sample was then plated out into white 96 well plates (Costar) in triplicate. 50 µL of OneGlo reagent (Promega) was added to each well and plates were read immediately on a plate reader (Fluoroskan Ascent FL, Thermo Electron Technologies). The plate reader was programmed to shake for 30 s, then to record luminescence from each well in 10 s increments. The reading was repeated until the signal peaked.

This luciferase assay protocol was then repeated for the *ESAG7* 3'UTR full-length sequence and truncations (1-3) against the same two control cell lines.

### 3.10 Data Analysis and Statistics

Averages were calculated from the triplicate readings and data was normalised to the *ALD* signal under basal conditions, by dividing the average reading for each cell line by the average *ALD*-FBS reading. The data was also normalised to a cell density of  $1 \times 10^6$  cells by dividing the luminescence by the cell count. An overall average was calculated from the three repeats for each cell line which was displayed on a graph as a bar, each individual repeat average was also displayed on the graph as a point. A paired two-tailed student T-test was performed on the data to identify any significant results where  $p < 0.05$ .

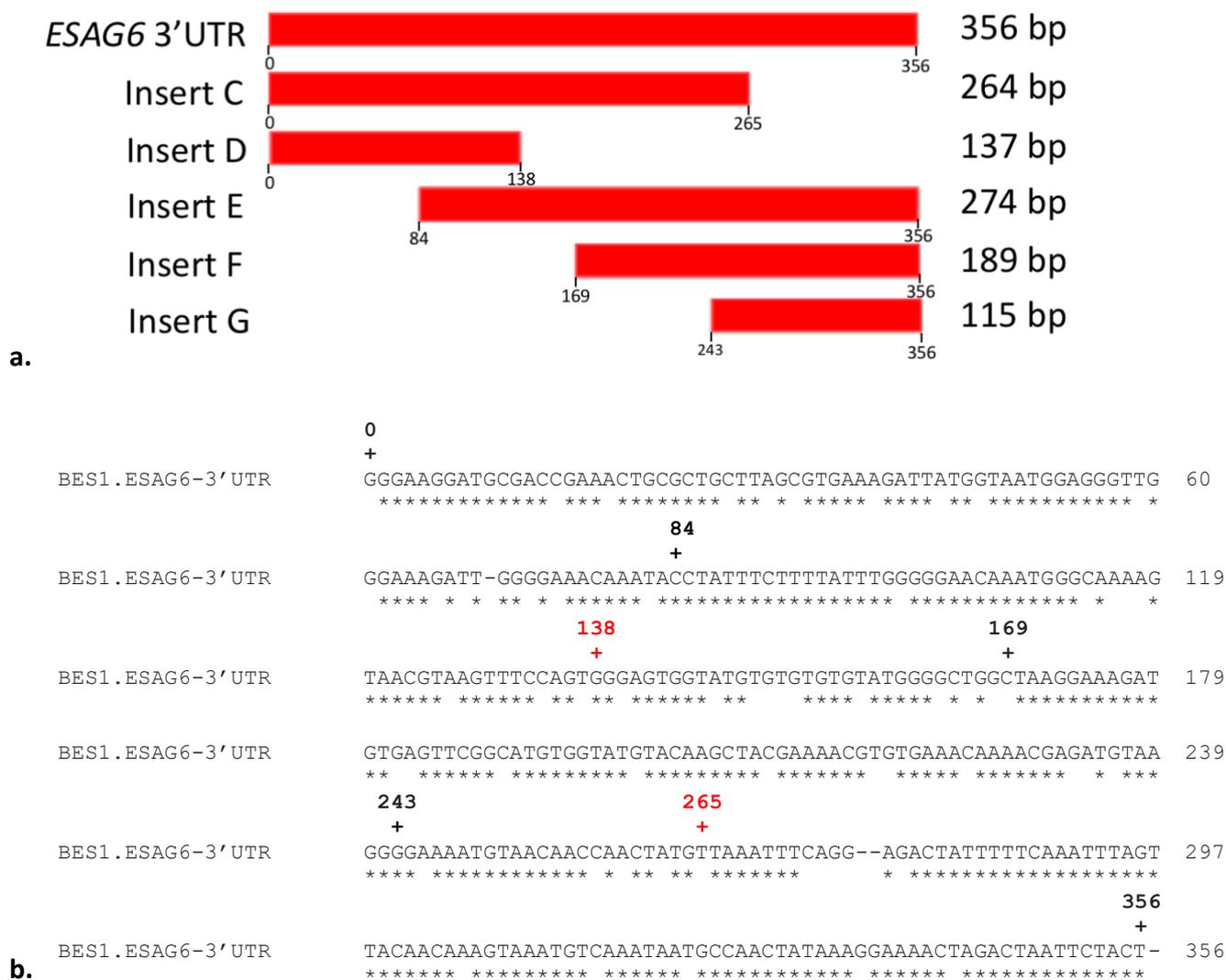
The fold-change was calculated by dividing the normalised average canine luminescence by the normalised average FBS luminescence for each cell line. This was plotted on a bar chart where the bars indicate the overall average fold-change and the points represent the fold-change for each biological replicate.

## 4.0 Results

### 4.1 Cloning and Truncation of the *ESAG6* and *ESAG7* 3' Untranslated Regions

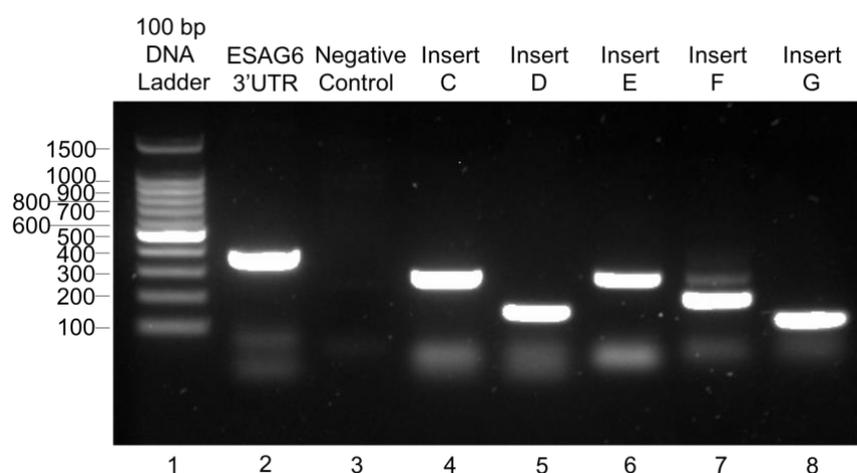
The length of the *ESAG6* 3'UTR was identified by Benz *et al.* (2018) as 335 bp in length. Primers were designed to amplify a full length *ESAG6* 3'UTR fragment 356 bp long to incorporate all identified polyadenylation sites in the sequence. The *ESAG6* 3'UTR has previously been demonstrated to mediate a specific response to iron starvation conditions in *T. brucei* Lister 427 through fusion of the 3'UTR to an *fLUC* or *GFP* reporter gene, at both an mRNA and a protein level with a magnitude previously observed for *TbTfR* upregulation (Benz *et al.*, 2018). Yet, it is unknown whether a certain motif in the 3'UTR is important for the upregulation of the *TbTfR* under iron starvation conditions or if the response is related to the secondary structure of the 3'UTR. It is assumed that the *ESAG7* 3'UTR will be involved in the iron starvation response via a similar mechanism as the *ESAG6* 3'UTR due to the co-regulation of the TfR protein components and the 97% similarity of the 3'UTRs across the multiple BESs, but it has not been experimentally demonstrated.

#### 4.1.1 Truncation of the *ESAG6* 3'UTR



**Figure 4.1: Truncations generated of the *ESAG6* 3'UTR. (a)** Shows the size of the 3' and 5' truncations in relation to the full length *ESAG6* 3'UTR. The diagram allows visualisation of where the sequences overlap and whether the sequence is truncated at the 5' or 3' end of the 3'UTR. **(b)** Alignment of the 13 copies of the *ESAG6* 3'UTR in the multiple BESs *ESAG6* is present in. Shows the location of the truncations in the 3'UTR sequences as indicated by the (+) and associated bp number, the two in red are the 3' truncations (Insert C and D). The alignment only shows the *ESAG6* sequence from BES1, but the conservation line represents the conservation across the 13 BESs.

To identify potential RBP motifs in the *ESAG6* 3'UTR that mediate the iron starvation response, the full length *ESAG6* 3'UTR sequence was used as template DNA to generate truncations of the 3'UTR. Primers were designed to truncate the 3'UTR from both the 5' and 3' ends at areas of hypervariability identified between the 13 copies of the *ESAG6* 3'UTR present in the 13 BES *ESAG6* is present in. The secondary structure was not taken into consideration. There are a number of short conserved sections in the *ESAG6* 3'UTR across the BESs that may be essential for the *TbTfR* response to iron starvation. The truncations were designed to study these conserved regions to identify if they are required for the upregulation of the receptor when iron is scarce, with their presence in some truncations and absence from others. **Figure 4.1** shows the alignment of the *ESAG6* 3'UTR and the location of the truncations in the sequence, allowing visualisation of the truncations and the conserved regions.



**Figure 4.2:** Agarose gel of the PCR amplified truncations of the *ESAG6* 3'UTR. 5' and 3' truncations of the full length *ESAG6* 3'UTR (356 bp); C- 264 bp, D- 137 bp, E- 274 bp, F- 189 bp, G- 115 bp. Gel shows that primers have annealed and successfully amplified fragments of an expected size compared to the 100 bp DNA ladder (Promega) run in lane 1. Lane 2 is the full-length fragment 356 bp in length, the negative control had only a reverse primer therefore no DNA was amplified.

The PCR reaction amplified fragments of an expected size for each of the five truncations of the *ESAG6* 3'UTR (**Figure 4.2**). To identify any important motifs in the *ESAG6* 3'UTRs the fragments were ligated into a specialised pRPa vector described by Benz *et al.* (2018) and the

plasmid DNA was sent off for sequencing to confirm the expected sequence had been successfully ligated. The pRPa $\Delta$ -*fLUC*<sup>mut</sup>-*ESAG6*-3'UTR vector is a modification of the pRPa<sup>GFP</sup>X vector from Alford *et al.* (2005) with the tetracycline operator removed so the insert is constitutively expressed. The firefly luciferase (*fLUC*) gene was incorporated into the vector from the pCRm-*LUC*-HYG vector from Yates (2014) to replace GFP. The *fLUC* sequence was mutated to remove an internal *Apal* site, replacing it with a *HindIII* site and incorporate a *BamHI* site at the end, described by Benz *et al.* (2018).

When the sequencing data was returned and aligned against the BES1 *ESAG6* 3'UTR sequence accessed from the published BES1 sequence on GenBank (FM162566), a number of mismatches were identified (data not shown) (Hertz-Fowler *et al.*, 2008). In each truncation, there was a number of polymorphisms ranging from 16 inconsistencies in Insert C and only a single inconsistency in Insert G. Further alignments of the sequenced truncations showed that the polymorphisms were located in the same positions along the full length *ESAG6* 3'UTR. To identify whether or not this was a coincidence, the template DNA used in the PCR reaction was also sequenced confirming that there were 18 mismatches in total, that had been introduced to the truncations from the template DNA. As the aim of this project is to identify important motifs involved in the upregulation of the *TbTfR* under iron starvation conditions, these truncations could not be used with any confidence. Any observed changes in the response of the *fLUC* reporter could be either due to the truncations or the polymorphisms in the sequence and there would be no way to distinguish which was the cause. Instead, new template DNA was generated.

#### 4.1.1.1 Cloning of the *ESAG6* 3'UTR

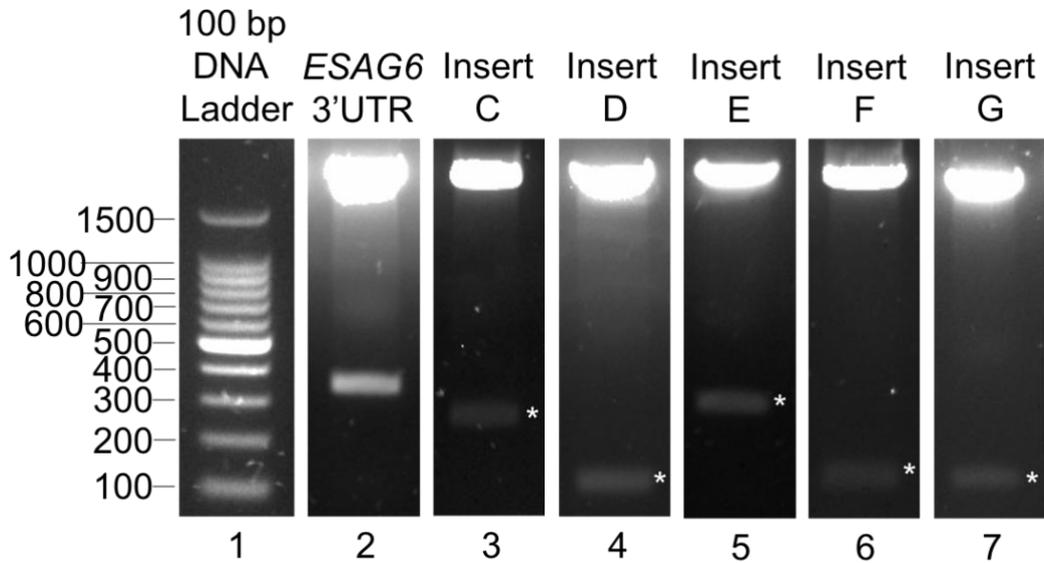
New problems arose whilst attempting to obtain new template DNA for the full length *ESAG6* 3'UTR. Initially genomic DNA was used as template DNA in a PCR with the full-length primers (listed in **Section 3.4.2**). When sequenced, however, there was an increase in the number of polymorphisms present in the sequences, in different locations to the initial mismatches identified. A reverse transcription reaction was then set up with RNA extracted from 2T1 BSF cells and the Oligo-dt primer (described in **Section 3.5**), the resulting cDNA was then used as template DNA in a PCR with the *ESAG6* 3'UTR full length forward primer and the adapter

primer. This, however, still resulted in polymorphisms in the DNA sequences generated. The source of this was eventually narrowed down to the primer sequences. Due to the similarity of the *ESAG6* 3'UTR between the multiple BES, the full length forward and reverse primers used in the above PCRs were complimentary elsewhere in the genome in other 'inactive' expression sites. A new primer was designed complimentary to a sequence of DNA near the end of the *ESAG6* open reading frame, specific to BES1 (5' - GCAGTACATTTGAGTCTTT - 3') and a PCR was performed using gDNA as template with this forward primer and the reverse primer complimentary to the end of the *ESAG6* 3'UTR. The use of the new forward primer confirmed that any 3'UTR sequences amplified were of the *ESAG6* gene expressed in BES1. A 422 bp sequence was amplified and ligated into the pGEM-T vector system, RE digests with *EcoRI* confirmed that a fragment of the correct size had been integrated into the vector and was sent off for sequencing. The sequencing data came back without a single polymorphism confirming that a correct *ESAG6* 3'UTR sequence had been obtained, identical to that available online [FM162566, GenBank] (Hertz-Fowler *et al.*, 2008). The sequenced plasmid was then used as template DNA to regenerate the five *ESAG6* 3'UTR truncations in **Figure 4.2**.

#### 4.1.1.2 Truncating the *ESAG6* 3'UTR

Originally to generate the truncations, PCR products were ligated into pGEM-T easy as an intermediate vector. Due the small size of the truncated fragments this was used to confirm integration into the vector as ligation into the commercially available vector has a higher success rate compared to the modified pRPaΔ vector. Due to time constraints of the project and the reproducibility of direct ligations, this protocol was modified so the truncations were cloned directly into pRPaΔ-*fluc*<sup>mut</sup>-*ESAG6*-3'UTR. Initially a PCR clean up kit (ThermoFisher) was used on the truncation PCR products and these were ligated into pRPaΔ-*fluc*<sup>mut</sup>-*ESAG6*-3'UTR. When analytical digests were performed with *Apal* and *BamHI* to screen the clones, larger inserts were being cut out, closer to 500 bp than 350 bp. An analytical digest was performed with *BamHI* and *HindIII*, which cuts the *fluc* gene out of the pRPaΔ vector backbone. Nothing was cut from the plasmid DNA in this digest which confirmed that some of the template plasmid DNA had survived the PCR purification and was subcloned directly into the *E. coli* cells giving an incorrectly high success rate in the cloning procedure. To avoid this in the future, the PCR products were gel purified so no template plasmid remained. The PCR

truncations were then ligated directly into pRPaΔ-*fLUC<sup>mut</sup>*-*ESAG6*-3'UTR, and successful integration was confirmed by analytical RE digests with *Apal* and *Bam*HI shown in **Figure 4.3**, and sequencing analysis. These constructs were now ready for transfection into the 2T1 BSF cells.



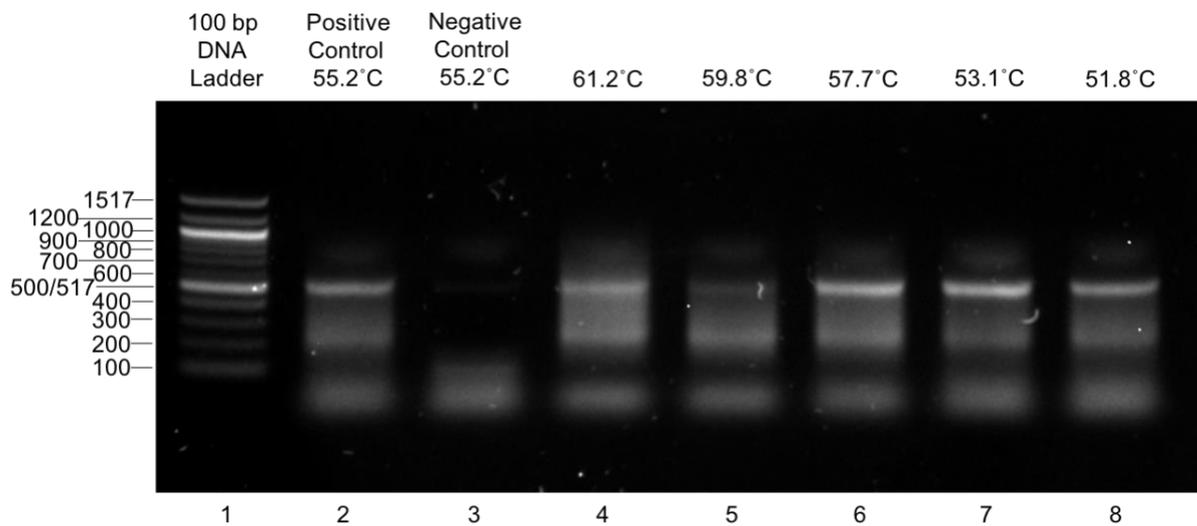
**Figure 4.3:** Agarose gel of analytical RE of the full length *ESAG6* 3'UTR and truncations cut out of the pRPaΔ-*fLUC<sup>mut</sup>*-*ESAG6*-3'UTR vector to confirm correct ligation. Inserts of an expected size (115-356 bp) were cut from the vector backbone in a double digest with *Apal* and *Bam*HI. The (\*) indicate the position of faint bands.

The restriction enzyme sites used to confirm integration were designed within the primers, the pRPaΔ-*fLUC<sup>mut</sup>*-*ESAG6*-3'UTR vector was mutated to remove an internal *Apal* site. The restriction enzyme digest is therefore specific, and DNA was only cut out from the plasmid DNA if there had been a successful ligation, this allowed identification of incorrect clones before sequencing.

#### 4.1.2 Identifying the *ESAG7* 3'UTR

The role of the *ESAG7* 3'UTR in the iron starvation response in *T. brucei* has not yet been studied. To investigate whether the *ESAG7* 3'UTR mediates a response via the same or a similar mechanism as the *ESAG6* 3'UTR, the size of the 3'UTR first needed to be identified. This was achieved by following a similar protocol to that used to identify the *ESAG6* 3'UTR by

Benz *et al.* (2018). A two-step RT-PCR reaction was carried out using an Oligo-dT primer sequence to transcribe mRNA into cDNA from RNA extracted from 2T1 BSF cells. The resulting cDNA was then used as template DNA in a gradient PCR (**Figure 4.4**) using primers specific to the BES1 *ESAG7* open reading frame, to confirm the DNA amplified was correct, and an adapter sequence.



**Figure 4.4: Agarose gel of gradient *ESAG7* 3'UTR RT-PCR reaction.** The gradient PCR reaction was set up to identify the optimal annealing temperature for the PCR reaction, which was determined at 53.1°C in lane 7. The PCR reaction shows strong bands around 500 bp in size, there are fainter bands around 200 bp. The negative control was set up excluding reverse transcriptase to confirm the reverse transcription reaction was successful.

The gradient PCR was used to find the optimum temperature for primers to anneal and for the *ESAG7* 3'UTR to be amplified. This was determined to be 53.1°C in Lane 7 of **Figure 4.4**. The RT-PCR was then repeated with this annealing temperature, the band excised and the DNA gel purified and ligated into pGEM-T and multiple clones were sent off for sequencing. Sequencing analysis identified that polyadenylation occurred at 453 bp, 477 bp and 492 bp downstream of the stop codon shown in **Figure 4.5**. The length of the *ESAG7* 3'UTR was taken as 492 bp in size. The fainter amplified bands around 200 bp were also sequenced however the resulting DNA could not be aligned with the *T. brucei* genome suggesting this was not part of the *ESAG7* 3'UTR.

```

BES1.ESAG7-3'UTR      AATGGAGTAAAGGCGAATTCAACTATACTGCAGAACCGGTCCGTGGACCTTTCACGGTAG 60
                        * * *** ***** ** ** * * * * * * * * * *
BES1.ESAG7-3'UTR      CGGGGTCCAACACAGTAGCAGTACATTTGAGTGTCTTACCCTGCACCTTGTCTTTTCAG 120
                        ***** ** * * * * * * * * * * * * * * * * * * * *
BES1.ESAG7-3'UTR      TTTTATTGTTGGGAGTGTCTGTGAAGGAAGGATGCGACAGAAGCTGCGCTGCTTAGCGTGA 180
                        ***** ***** * * * * * * * * * * * * * * * *
BES1.ESAG7-3'UTR      AAGATTATGGTAATGGAGGGTTGTGAAAGATT-GGGGAACAAAAACCTATTTCTTTTAT 239
                        * ***** ***** ** * * * * * * * * * * * * * * * *
BES1.ESAG7-3'UTR      TTGGGGGAACAAATGGGCAAAAGTAACGTAAGTTTCCAGTGGGAGTGGTATGTGTG--TG 297
                        *** ***** ** * * * * * * * * * * * * * * * * * *
BES1.ESAG7-3'UTR      TGTATGGGCTGGCTAAGGAAAGATGTGAGTTCGGCATGTGGTATGTACAAGCTACGAAA 357
                        ** * * * * * * * * * * * * * * * * * * * * * * * * * *
BES1.ESAG7-3'UTR      ACGTGTGAAACAAAACGAGATGTAAGGGGAAAATGTAACAACCAACTATGTTAAATTTCA 417
                        *** * * * * * * * * * * * * * * * * * * * * * * * * *
BES1.ESAG7-3'UTR      GG--AGACTATTTTCAAAATTTAGTTACAACAAAGTAAATGTCAAATAATGCCAACTATA 475
                        * * * * * * * * * * * * * * * * * * * * * * * * * *
BES1.ESAG7-3'UTR      +
                        +
AAGGAAACTAGACTAA----- 492
** ***** * **

```

**Figure 4.5: Identification of the *ESAG7* 3'UTR.** The crosses (+) on the BES1 sequence indicate the locations of three identified polyadenylation sites, the *ESAG7* 3'UTR was taken to be 492 bp in length. The alignment shows the conservation of the 13 *ESAG7* 3'UTRs taken from the BESs when 492 bp were aligned immediately downstream of the *ESAG7* stop codon using Clustal Omega. The '\*' show where the sequence is conserved across all 13 BESs. There was 98% conservation between the 13 3'UTRs showing a high level of sequence similarity, therefore, if an important motif is identified in the 3'UTR, it could be conserved across the other expression sites.

When the sequencing data was analysed a number of polymorphisms were identified in the sequences of the *ESAG7* 3'UTR, likely introduced by errors in the RT reaction or the PCR but may also have been due to the amplification of the *ESAG7* 3'UTR from other BESs. Unfortunately, all the sequences obtained at the length of 492 bp had one or more polymorphisms. In one of the 492 bp clones there was a single polymorphism located 69 bp downstream of the stop codon. A correct sequence 477 bp long had been obtained from a different clone so, rather than repeating the RT-PCR until a correct 492 bp sequence was confirmed, a new reverse primer (5' - TTTATAGTTGGCATTATTTGAC - 3') was designed to

amplify the remaining 15 bp using the 477 bp sequence as template DNA. The forward primer described in the methods (**Section 3.4.4**) was designed complimentary to the beginning of the *ESAG7* 3'UTR downstream of the stop codon and the longer reverse primer were used in a PCR to amplify the full length *ESAG7* 3'UTR using the correct 477 bp sequence in pGEM-T as template DNA. The resulting product was then gel purified and ligated into pGEM-T, integration was confirmed by an analytical RE digest with *Apal* and *BamHI* and sequencing data was used to confirm the correct 492 bp sequence for the *ESAG7* 3'UTR. To isolate this correct sequence, a preparative digest was performed with the same enzymes and the *ESAG7* 3'UTR was cut from pGEM-T, gel purified and ligated into the pRPa $\Delta$ -*fluc*<sup>mut</sup> vector system ready for transfection into the 2T1 BSF cells.

The observed length of 492 bp is consistent with expectations that the *ESAG7* 3'UTR would be longer than the *ESAG6* 3'UTR due to the lack of a GPI anchor signal sequence on the *ESAG7* gene. Sequence alignments of the end of the *ESAG6* open reading frame and 3'UTR and the *ESAG7* 3'UTR in **Figure 4.6**, show high conservation of the amino acid sequence.

BES1 . ESAG6-ORF-3' UTR	TTTAGGTGAAGAAGAGGAGACGATCCTGAAATCTAACTATACTGCAGAACCGGTCCGTGG	60
BES1 . ESAG7-3' UTR	-----AATGGAGTAAAGGCGAATTCAACTATACTGCAGAACCGGTCCGTGG	46
	*   *	
BES1 . ESAG6-ORF-3' UTR	ACCTTTCACGGTAGCGGGTCCAACACAGTAGCAGTACATTTGAGTCTTTTTACCGCTGC	120
BES1 . ESAG7-3' UTR	ACCTTTCACGGTAGCGGGTCCAACACAGTAGCAGTACATTTGAGTGTCTTCTACCGCTGC	106
	*****   *   *   *   *   *   *   *   *   *   *   *   *   *   *   *   *   *   *   *	
BES1 . ESAG6-ORF-3' UTR	ACTTTGTGTTTCAGCTTTATTGTTGGGAGTGCTGTGAGGGAAGGATGCGACCGAAACTGC	180
BES1 . ESAG7-3' UTR	ACTTTGTTTTTTCAGTTTTATTGTTGGGAGTGCTGTGAAAGGATGCGACAGAAGCTGC	166
	*****   *   *   *   *   *   *   *   *   *   *   *   *   *   *   *   *   *   *	
BES1 . ESAG6-ORF-3' UTR	GCTGCTTAGCGTGAAAGATTATGGTAATGGAGGGTTGGGAAAGATTGGGGAAACAAATAC	240
BES1 . ESAG7-3' UTR	GCTGCTTAGCGTGAAAGATTATGGTAATGGAGGGTTGTGAAAGATTGGGGAAACAAAAAC	226
	*****   *   *   *   *   *   *   *   *   *   *   *   *   *   *   *   *   *   *	
BES1 . ESAG6-ORF-3' UTR	CTATTTCTTTTATTTGGGGGAACAAATGGGCAAAAGTAACGTAAGTTTCCAGTGGGAGTG	300
BES1 . ESAG7-3' UTR	CTATTTCTTTTATTTGGGGGAACAAATGGGCAAAAGTAACGTAAGTTTCCAGTGGGAGTG	286
	*****   *   *   *   *   *   *   *   *   *   *   *   *   *   *   *   *   *   *	
BES1 . ESAG6-ORF-3' UTR	GTATGTGTGTGTGTATGGGCTGGCTAAGGAAAGATGTGAGTTCGGCATGTGGTATGTAC	360
BES1 . ESAG7-3' UTR	GTATGTGTGTGTGTATGGGCTGGCTAAGGAAAGATGTGAGTTCGGCATGTGGTATGTAC	346
	*****   *   *   *   *   *   *   *   *   *   *   *   *   *   *   *   *   *   *	
BES1 . ESAG6-ORF-3' UTR	AAGCTACGAAAACGTGTGAAAACAAAACGAGATGTAAGGGGAAAATGTAAACAACCAACTAT	420
BES1 . ESAG7-3' UTR	AAGCTACGAAAACGTGTGAAAACAAAACGAGATGTAAGGGGAAAATGTAAACAACCAACTAT	406
	*****   *   *   *   *   *   *   *   *   *   *   *   *   *   *   *   *   *   *	
BES1 . ESAG6-ORF-3' UTR	GTAAATTTTCAGGAGACTATTTTTCAAATTTAGTTACAACAAAGTAAATGTCAAATAATG	480
BES1 . ESAG7-3' UTR	GTAAATTTTCAGGAGACTATTTTTCAAATTTAGTTACAACAAAGTAAATGTCAAATAATG	466
	*****   *   *   *   *   *   *   *   *   *   *   *   *   *   *   *   *   *   *	
BES1 . ESAG6-ORF-3' UTR	CCAACTATAAAG-----	492
BES1 . ESAG7-3' UTR	CCAACTATAAAGGAAACTAGACTAA	492
	*****	

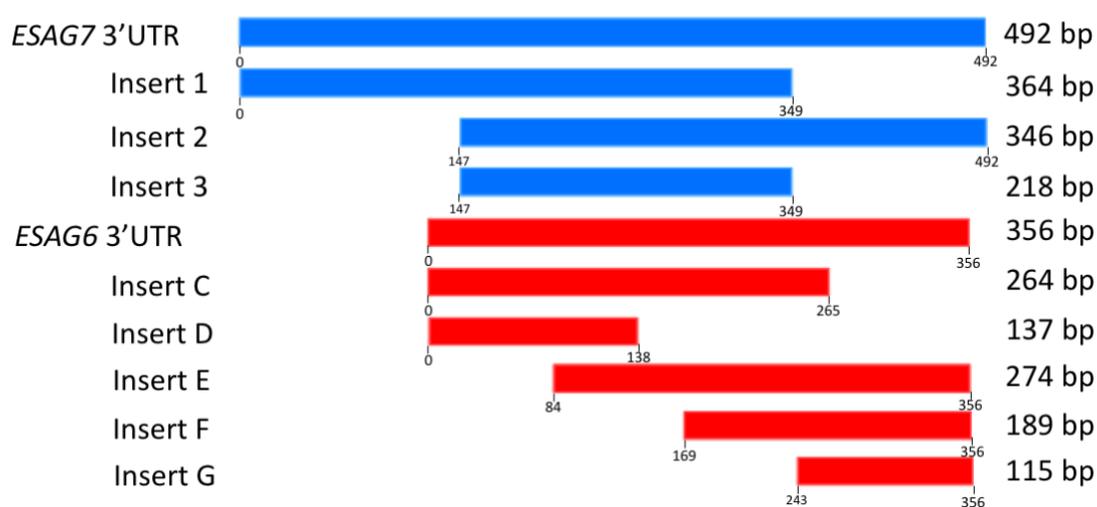
**Figure 4.6: The end of the *ESAG6* ORF and 3'UTR compared to the *ESAG7* 3'UTR of BES1.** The sequence is 157 bp of the end of the *ESAG6* ORF followed by the 335 bp 3'UTR, the highlighted nucleotide shows the start of the *ESAG6* 3'UTR. This sequence is the same length as the *ESAG7* 3'UTR (492 bp).

There are few mismatches between the 492 bp sequences of the *ESAG6* ORF 3'UTR and *ESAG7* 3'UTR of BES1 showing that the two are closely related (91% conservation). The beginning 143 bp section (the end of the *ESAG6* ORF) likely encodes the GPI signal sequence. This maintains close similarity to the beginning of the *ESAG7* 3'UTR supporting the theory that both proteins originally had a GPI signal sequence, but it has evolved away in *ESAG7* and the protein remains anchored to the membrane by associations with *ESAG6*. The alignment in **Figure 4.6** also shows that the *ESAG7* 3'UTR extends 14 bp beyond the 3' end of the *ESAG6* 3'UTR.

The identification of the *ESAG7* 3'UTR is important as the involvement of *ESAG7* 3'UTR in the regulation of the *TbTfR* under iron starvation conditions has not yet been experimentally demonstrated as it was excluded from the original research by Benz *et al.* (2018). The 3'UTR regions of *ESAG6* and *ESAG7* show high levels of similarity across the BES where they are present, at an average of 97% (**Appendix 7.1**). The aim of this project is to identify important motifs in the 3'UTRs of the *TbTfR* that control the upregulation when iron is scarce. With the high similarity between the 3'UTRs it is likely that any identified motifs may be conserved between *ESAG6* and *ESAG7* 3'UTRs and across the expression sites, as both proteins must be upregulated for a functional transferrin receptor and upregulation has been observed in *TbTfR* expressed in different ESs.

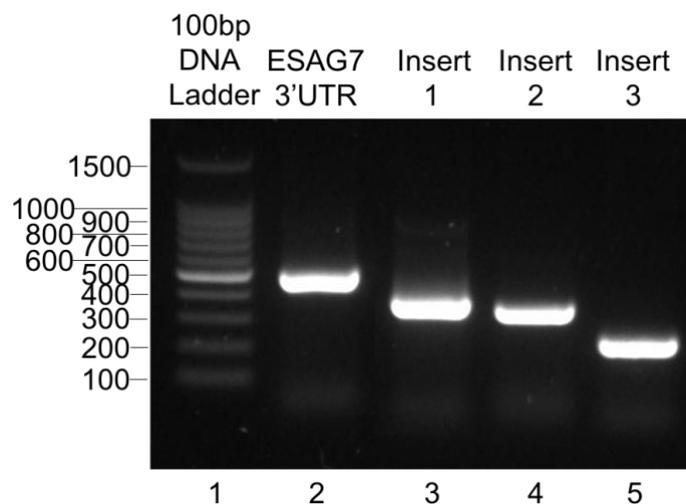
#### 4.1.3 Generation of the *ESAG7* 3'UTR Truncations

To further investigate if there are any important motifs in the *ESAG7* 3'UTR that are involved in the dynamic regulation of the *TbTfR*, primers were designed to 5' and 3' terminally truncate the *ESAG7* 3'UTR shown in **Figure 4.7**.



**Figure 4.7: Schematic diagram of the *ESAG6* and *ESAG7* 3'UTR truncated sequences.** The diagram shows the size of the 3' and 5' truncations in relation to the full length *ESAG6* and *ESAG7* 3'UTR. The diagram allows visualisation where the sequences overlap and whether the sequence is truncated at the 5', 3' or both ends of the 3'UTR.

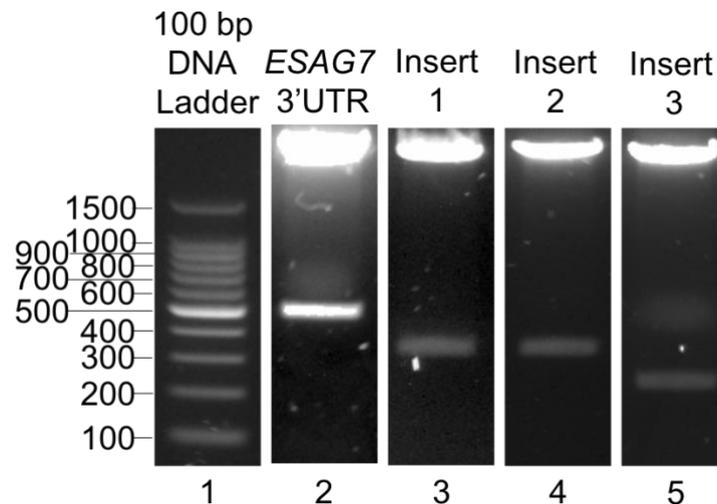
Two truncations were designed of a similar length to the *ESAG6* 3'UTR missing either the beginning or the end of the 3'UTR. A further truncation was made with a 5' and 3' terminal truncation to generate a sequence 218 bp long. As seen in **Figure 4.7**, Insert 1 is 364 bp in length, so similar in size to the *ESAG6* 3'UTR, however, as it is a 3' truncation there is a 122 bp sequence at the beginning of the truncation that is missing from the *ESAG6* 3'UTR. From the schematic, it is most likely that Insert 2 will respond in a similar way to the *ESAG6* 3'UTR as it is similar in length and is a 5' truncation missing the beginning of the *ESAG7* 3'UTR that aligns with the end of the *ESAG6* ORF and not the 3'UTR. Insert 3 is both 5' and 3' terminally truncated, due to the location of the truncations this insert is most similar to Insert C so, they may respond to iron starvation conditions in a similar way.



**Figure 4.8:** Agarose gel of PCR amplified truncations of the *ESAG7* 3'UTR. 5' and 3' terminal truncations of the *ESAG7* 3'UTR ranging from 218 bp to 364 bp in size, primers used to generate truncations listed in **Table 3.4**. Gel shows that primers have annealed and successfully amplified fragments of an expected size compared to the 100 bp DNA ladder (Promega) run in lane 1. PCR was performed using the GoTaq polymerase system with an annealing temperature of 60°C.

The PCR (**Figure 4.8**) amplified fragments of an expected size for the full length *ESAG7* 3'UTR and the truncations. These were then ligated into the pRPaΔ-*fluc*<sup>mut</sup> vector to generate new variations of a pRPaΔ-*fluc*<sup>mut</sup>-*ESAG7*-3'UTR vector. The isolated plasmid DNA was then

screened by analytical RE digests with *Apal* and *Bam*HI to confirm ligation of the PCR products cut from the vector backbone shown in **Figure 4.9**. Any isolated plasmids showing correct inserts cut from the pRPaΔ-*flUC<sup>mut</sup>* vector were sent off for DNA sequencing and aligned with the *ESAG7* 3'UTR of BES1 [FM162566, GenBank] (Hertz-Fowler *et al.*, 2008).



**Figure 4.9:** Agarose gel of analytical RE of the full length *ESAG7* 3'UTR and truncations cut out of the pRPaΔ-*flUC<sup>mut</sup>* vector backbone to confirm correct ligation. Inserts of an expected size (218-492 bp) were cut from the vector backbone in a double digest with *Apal* and *Bam*HI.

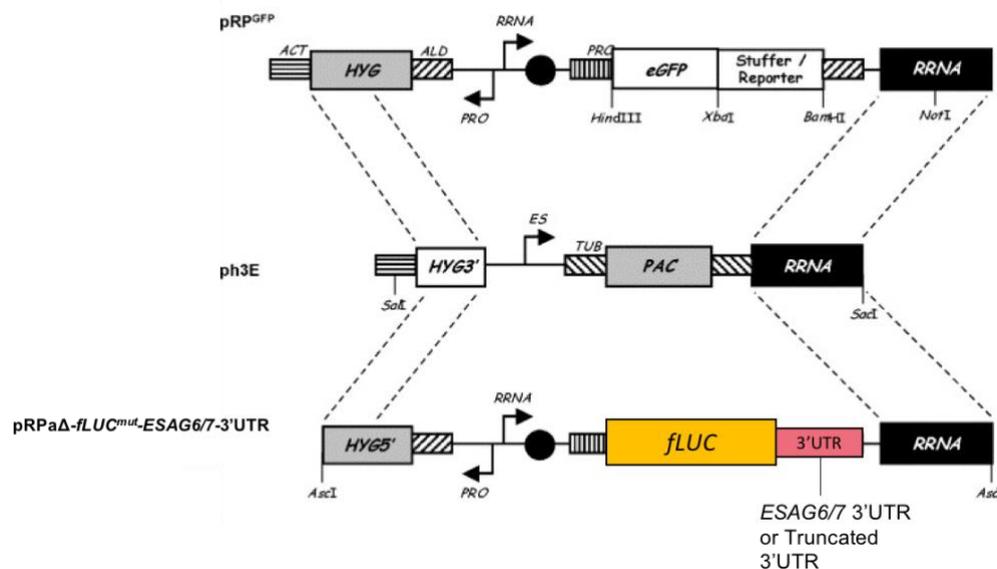
Once correct sequences, ligated into the pRPaΔ vector system, were obtained for all of the *ESAG6* and *ESAG7* 3'UTR full lengths and truncations, they were ready for further analysis. The DNA sequences were linearised and transfected into the 2T1 BSF *T. brucei* cells where luciferase activity assays were performed under normal and iron starvation conditions to analyse whether expression of the *TbTfR* changes with the truncations made (**Section 4.2**).

## 4.2 Generation, Screening and Analysis of *Trypanosoma brucei* Cell Lines

Once the *ESAG6* and *ESAG7* 3'UTR truncations had been successfully ligated into the *flUC* reporter system and the sequences had been confirmed, they were transfected into 2T1 BSF *T. brucei* cells. Luciferase activity assays were performed to measure luciferase expression and observe whether the reporter is still upregulated under iron starvation conditions or if the response is altered by the truncations made. This reporter system does not directly

measure TfR expression but acts as a proxy to measure the expression levels of the firefly luciferase reporter gene fused to sections of the receptor 3'UTR. Despite not directly measuring the *Tb*TfR, the *fLUC* gene fused to the full length 3'UTR has been shown to mimic previous data of the regulation of the *T. brucei* transferrin receptor, so is a good representation of how regulation of the receptor can be affected by different regions of the 3'UTRs (Benz *et al.*, 2018).

The firefly luciferase reporter gene incorporated into the pRPaΔ vector backbone is fused to the truncated 3'UTR sequences. To perform the luciferase activity assay OneGlo reagent (Promega) is added to the cells which contains a substrate that is broken down by the luciferase enzyme into a product (oxyfluoroluciferin) which emits luminescence that is recorded in the assay. The more luminescence emitted, the more luciferase enzyme present. Under iron starvation conditions the upregulation of the *fLUC* gene is directly proportional to the upregulation of the *Tb*TfR mediated by the 3'UTR.



**Figure 4.10: Schematic of the pRPaΔ-*fLUC*<sup>mut</sup>-ESAG6/7-3'UTR vector.** Figure adapted from Alsford *et al.* (2005) shows the location of the hygromycin resistance gene in the backbone of the vector and the *fLUC* gene fused to the truncated *ESAG6* or *ESAG7* 3'UTR downstream of the *RRNA* promoter. Before transfection, the vector is linearised by *Ascl* and integrates into the 'landing pad' *RRNA* locus in the 2T1 cell line. The 2T1 cell line is resistant to puromycin but when the DNA is transfected, it disrupts the puromycin resistance gene (*PAC*) and replaces

it with the hygromycin resistance gene (HYG) present in the pRPa $\Delta$ -*fLUC<sup>mut</sup>* vector so the cells acquire resistance to hygromycin which can be used to screen the transfected clones.

When the pRPa $\Delta$ -*fLUC<sup>mut</sup>*-*ESAG6*-3'UTR vector, containing the truncated sequences is transfected into the BSF cells, the vector is incorporated into a tagged *RRNA* locus in the 2T1 cell line, to avoid positional effects and give uniform expression, as the location of the construct is known (Alsford *et al.*, 2005). Therefore, any observed changes in the response of the cell lines to the induction of iron starvation conditions are due to the truncations in the *ESAG6* 3'UTR or *ESAG7* 3'UTR fused to the *fLUC* gene, rather than the location in the *RRNA* locus. When transfected into 2T1 cells, the vector insert disrupts a puromycin resistance gene and incorporates a hygromycin resistance gene present in the vector backbone (**Figure 4.10**). To confirm that the truncation-vector constructs had been successfully transfected into the correct location drug screen assays were performed. Where the cells survived in media containing hygromycin, and no longer survived in media containing puromycin, the transfection had been successful. If any clones survived in both, it was likely there was a mixed cell culture and the transfection was not successful, so these cell lines were discarded. The generated clones are listed in **Table 4.1**.

**Table 4.1: The number of clones generated for each of the full length and truncated *ESAG6* and *ESAG7* 3'UTRs.** There was a higher rate of transfection success for the *ESAG6* clones than for *ESAG7*, with more cells surviving the transfection with the correct drug selection, indicating the cells had taken up the truncation DNA.

<b>Truncation</b>	<b>Number of Clones Generated</b>	<b>% with Correct Drug Resistance (HYG)</b>	<b>Clones Available</b>
<b>ESAG6</b>	2	100%	6.1, 6.2
<b>Insert C</b>	2	100%	C1, C2
<b>Insert D</b>	4	75%	D1, D2, D4
<b>Insert E</b>	4	75%	E2, E3, E4
<b>Insert F</b>	5	100%	F1, F2, F3, F4, F5
<b>Insert G</b>	4	75%	G2, G3, G4
<b>ESAG7</b>	4	50%	7.1, 7.4
<b>Insert 1</b>	1	100%	1.1
<b>Insert 2</b>	2	50%	2.2
<b>Insert 3</b>	4	50%	3.1, 3.3

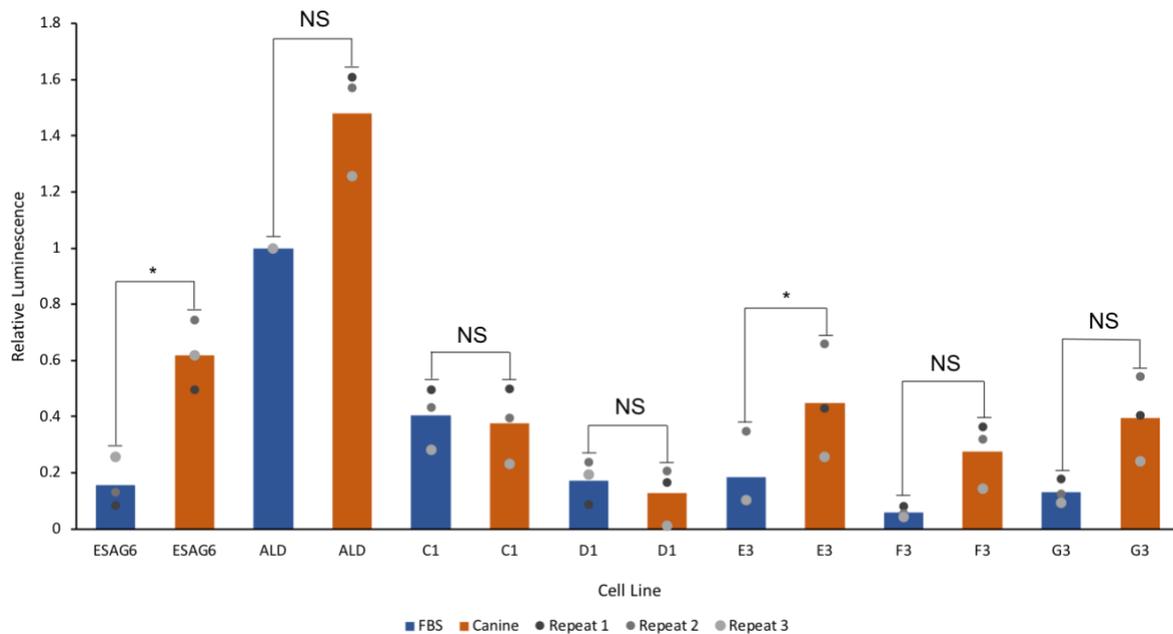
The 2T1 BSF strain is a culture adapted monomorphic cell line that is stably expressing *VSG* and the *ESAGs* from BES1, the cell line maintains the ability to switch BES but switching away from *VSG221* is unfavourable and cells that have switched are outgrown (Aitcheson *et al.*, 2005). In this experiment the cells are only left under iron starvation conditions for 5 hours which is not long enough to induce switching, therefore the majority of the cells are expressing the TfR from BES1. Due to the variable affinity of the *TbTfR* binding to different host transferrin, iron starvation conditions can be induced in the 2T1 cell lines by switching the media from a serum where the transferrin can bind with high affinity to a serum with transferrin that binds with a low affinity. It has been shown that the *TbTfR* expressed from BES1 has a high affinity for bovine transferrin and a low affinity for canine transferrin (Gerrits *et al.*, 2002). This has been taken advantage of in previous research to induce iron starvation conditions with a 2.5-5-fold increase in expression of the receptor (Fast *et al.*, 1999; Mussmann *et al.*, 2004). To induce the iron starvation response in the generated cell lines a serum switch was performed from media supplemented with 10% bovine serum to media

supplemented with 10 % canine serum. The cells were then incubated for 5 hours at 37°C before luciferase activity assays were performed.

The full length *ESAG6* 3'UTR was used as a control cell line to show how the transferrin receptor is normally upregulated under iron starvation conditions and allowed comparisons to be made between the normal response and the response of the truncated cell lines. A pRPaΔ-*fluc*-*ald*-3'UTR control cell line was also used which shows a linear increase in expression of the *Luciferase* with an increase in cell density but is unaffected by iron starvation (Benz *et al.*, 2018). Aldolase is highly expressed protein with a stable 3'UTR that is efficiently processed therefore the aldolase signal is always higher than the luciferase signal. Aldolase is not an iron regulated gene; therefore, any observed increases are due to an increase in cell density. The iron starved cell lines were compared to an untreated cell line to show how expression of luciferase changes under iron starvation conditions.

#### 4.2.1 Characterisation of the *ESAG6* 3'UTR Response to Iron Starvation Conditions

To induce iron starvation conditions into the generated *ESAG6* 3'UTR cell lines, they were centrifuged, washed and resuspended in 1 mL HMI11-T media supplemented with 10% canine serum or 10% foetal bovine serum. All cells had the same growth history to remove variation, as expression of the TfR is density dependent (Benz *et al.*, 2018). Cells were incubated for 5 hours then luminescence readings were taken in triplicate, with each assay repeated three times.



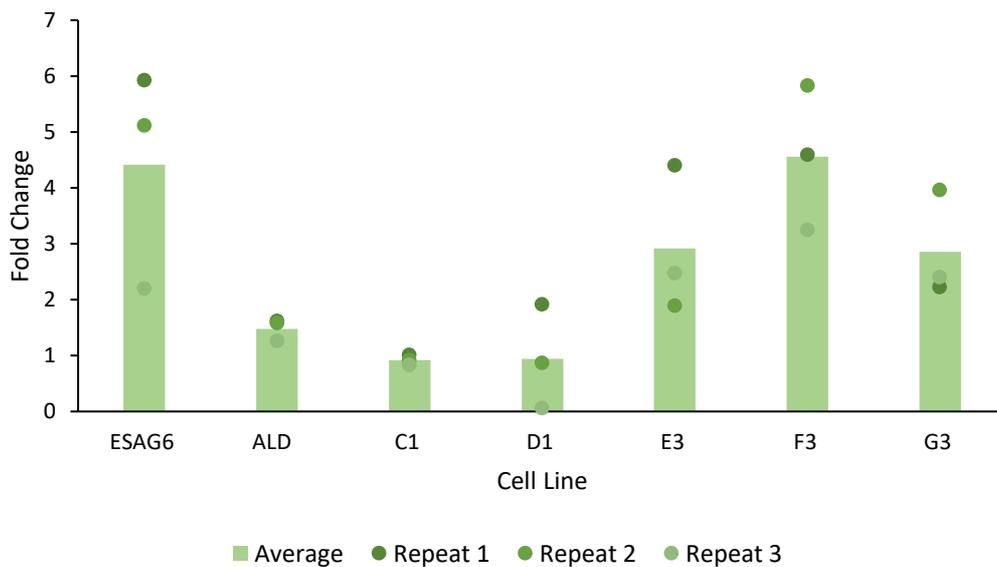
**Figure 4.11: Relative luminescence of luciferase expression in *ESAG6* 3'UTR truncated cell lines under normal and iron starvation conditions.** *ESAG6* 3'UTR truncated cell lines were incubated for 5 hours with HMI11-T supplemented with 10% canine serum (iron starved) or 10% FBS (non-iron starved). Luminescence was measured in triplicate for three biological replicates. Data was normalised to the ALD signal under non-iron starvation conditions and normalised to  $1 \times 10^6$  cells/mL. The bars on the graph indicate the overall average and the points represent the average for each biological replicate. Data was analysed using a paired two-tailed student T-test: NS  $p > 0.05$ , \*  $p < 0.05$ .

Data was normalised to the *ALD* signal in non-iron starvation conditions by dividing the average reading for each cell line by the average *ALD*-FBS reading as this was a consistently high luminescence reading. On the graph in **Figure 4.11**, this is indicated by a relative luminescence of 1. All readings aside from *ALD*-canine were lower than this. The data was also normalised to  $1 \times 10^6$  cells/mL by dividing the luminescence value by the cell count. The luciferase assay is density dependent so normalising the data by cell count showed a realistic representation of how the truncated cell lines were responding to iron starvation conditions without the data being skewed by variations in cell density. There are no error bars on the graph as there was little technical variation between the triplicate readings so calculated standard deviation values were very low. Repeating the assay on different days shows the biological variation between the cells.

The data in **Figure 4.11** show that inserts E3, F3 and G3 are all still responding to iron starvation conditions in a similar way to the full length *ESAG6* 3'UTR cell line, but only insert E3 is a statistically significant result ( $P < 0.05$ ). The lack of significant results in the assay is due to the variation in the biological replicates as indicated by the points in **Figure 4.11**. With more repeats, it is likely that more significant results would be obtained. This variation was expected as there is batch to batch variability in the different host serum, the OneGlo reagent is temperature dependent and the assay can be affected by cell growth, but the data has been normalised to this. The aldolase luminescence recorded throughout these experiments was high in canine serum (1.5) as the ALD cells grew at a much higher rate in this media compared to other batches of canine serum and FBS.

The relative luminescence for *ESAG6* and inserts E-G are all low ( $<0.2$ ) under non-iron starvation conditions and increase when the BSF cells are incubated with canine media. E3, F3 and G3 are all N-terminally truncated sequences, so the 3' end of the 3'UTR remains intact, whilst the beginning (5' end) of 3'UTR has been lost. This is interesting as it suggests that there may be an important element towards the end of the 3'UTR that is required for regulating the iron starvation response as when the end of the 3'UTR is lost, the cells lose the ability to respond to iron starvation conditions.

The relative luminescence recorded for inserts C1 and D1 does not increase when the cell lines are starved of iron unlike the full-length *ESAG6* control. They actually appear to decrease when the media is supplemented with canine serum, but this is only by a marginal amount. For Insert C1, the non-iron starved cells have quite a high average relative luminescence of about 0.4, compared to the other cell lines. This is about the same level as the C1 iron starved cells. As the C1 starved cells show a similar level of relative luminescence as the other cell lines in canine media that still appear to upregulate the transferrin receptor (*ESAG6*, E3, F3 and G3), it could suggest that the truncation for insert C1 may have lost a repressive element. Insert D1 does not appear to respond to iron starvation conditions by upregulation as the relative luminescence is low for cells supplemented with FBS and canine serum. The low level of variation between the relative luminescence under normal and iron starvation conditions suggests that the C1 and D1 cell lines are no longer able to respond to iron starvation conditions by upregulating expression of the *TbTfR* when iron is scarce in the environment.



**Figure 4.12:** Calculated fold-change in luciferase luminescence when *T. brucei* BSF cells are incubated in media supplemented with FBS or canine serum. The bars on the graph indicate the overall average fold-change and the points represent the average fold-change for each biological replicate. Data was normalised to  $1 \times 10^6$  cells/mL.

In the literature, the *T. brucei* transferrin receptor has been shown to be upregulated by a 2.5-5-fold increase under iron starvation conditions. The calculated fold-change in **Figure 4.12** is consistent with this, with the luciferase expression fused to the full-length *ESAG6* 3'UTR increasing 4.4-fold indicating an equal increase in the *TbTfR*. The fold-change was calculated by dividing the average canine luminescence by the average FBS luminescence for each cell line. The data was normalised to  $1 \times 10^6$  cells/mL before the fold-change was calculated to remove any variation due to cell density and give a realistic representation of how the cell lines responded to iron starvation.

**Figure 4.12** show the calculated fold-change is similar for *ESAG6* and Insert F3 with luciferase expression increasing 4.4-fold and 4.6-fold respectively. Inserts E3 and G3 also appear to be responding to iron starvation conditions with a 2.9 and 2.8-fold increase in luciferase expression respectively which is within the reported range for the upregulation of the *TbTfR* when iron is scarce. In the assay luciferase expression is proportional to expression of the transferrin receptor. Inserts C1 and D1 do not appear to be responding to the induction of

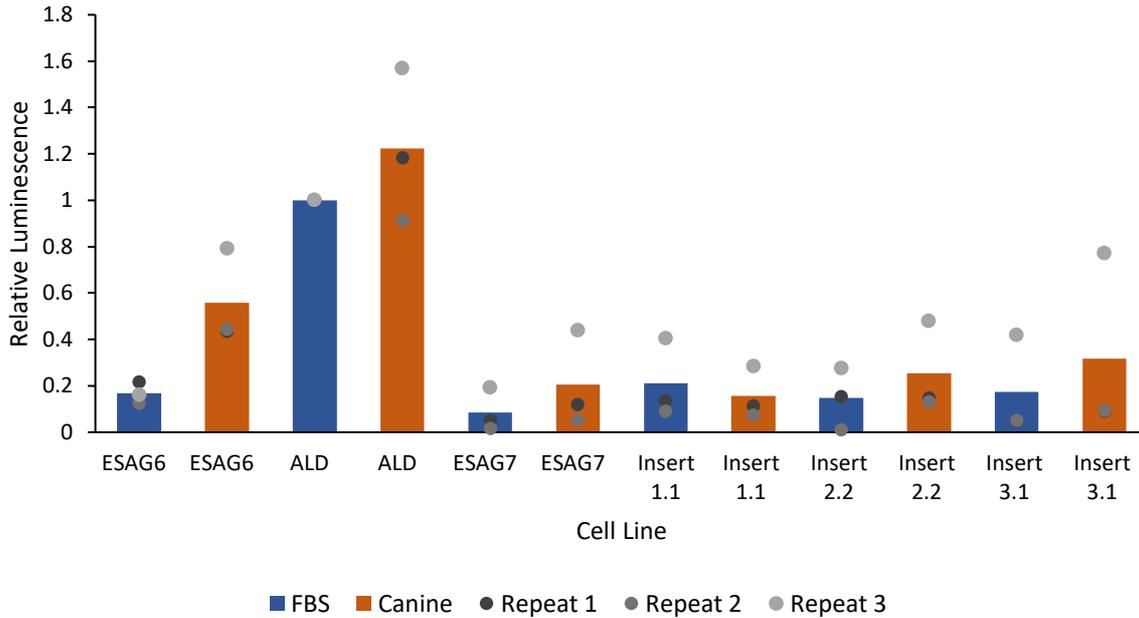
iron starvation, with a fold-change of less than 1. ALD is also not responding to iron starvation conditions with a fold-change of 1.5, this is expected for the ALD cell line as this is not an iron sensitive gene.

**Table 4.2: Summary table describing the *ESAG6* 3'UTR truncations and how the cell lines respond to iron starvation conditions.**

<b>Name</b>	<b>Size</b>	<b>5' or 3' Truncation</b>	<b>Responsive to Iron Starvation</b>
<b>Insert C</b>	264 bp	3'	Repressive?
<b>Insert D</b>	137 bp	3'	No Response
<b>Insert E</b>	274 bp	5'	Responsive
<b>Insert F</b>	189 bp	5'	Responsive
<b>Insert G</b>	115 bp	5'	Responsive

#### 4.2.2 Characterisation of the *ESAG7* 3'UTR Response to Iron Starvation Conditions

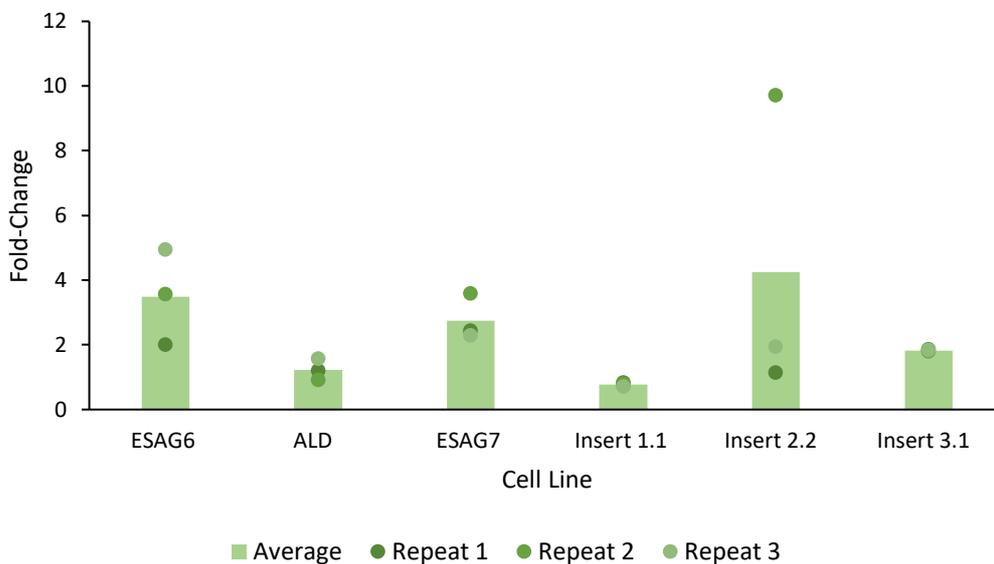
The role of the *ESAG7* 3'UTR in the iron starvation response has not yet been studied. It was assumed that due to the similarity between the 3'UTR of *ESAG6* and *ESAG7* that they are regulated via the same mechanism. This is supported by the evidence that the *TbTfR* is a heterodimer that can only bind transferrin when *ESAG6* and *ESAG7* are associated, therefore, they must be upregulated simultaneously at a post transcriptional level when iron is scarce. Due to this the reporter system generated for *ESAG6* was adapted by fusing the *ESAG7* 3'UTR and truncations to the *fLUC* reporter gene. The modified vector was then transfected into the 2T1 cells following the same protocol and screened for clones. Luciferase assays were then also performed on the *ESAG7* 3'UTR cell lines under normal and iron starvation conditions. The generated *ESAG7* 3'UTR cell lines were centrifuged, washed and resuspended in HMI11-T media supplemented with 10% FBS to media or 10% canine serum to induce iron starvation. Cells had the same growth history to remove variation and were incubated for 5 hours before relative luminescence was recorded in triplicate. The assay was repeated three times.



**Figure 4.13: Relative luminescence of luciferase expression in *ESAG7* 3'UTR truncated cell lines under normal and iron starvation conditions.** *ESAG7* 3'UTR truncated cell lines were incubated for 5 hours in media supplemented with 10% canine serum (iron starved) or with 10% FBS (non-iron starved). After 5 hours luminescence was recorded in triplicate for three biological replicates. Data was normalised to the ALD signal under non-iron starvation conditions and normalised to  $1 \times 10^6$  cells/mL. The bars on the graph indicate the overall average and the points represent the average for each biological replicate. Data was analysed using a paired two-tailed student T-test and all results were not significant with  $p > 0.05$ .

The relative luminescence values for the *ESAG7* 3'UTR full length and truncations (1.1, 2.2 and 3.1) shown in **Figure 4.13** were all quite low both under iron starvation conditions and normal conditions. All of the relative luminescence readings are lower than those obtained in **Figure 4.11**, including the *ESAG6* control which was 0.56 compared to 0.62 in **Figure 4.11**. The average relative luminescence for all of the cell lines under normal conditions (FBS), excluding ALD and *ESAG7*, are very similar around 0.2 (*ESAG6*- 0.17, 1.1- 0.21, 2.2- 0.15 and 3.1- 0.17). This shows that the cell lines have the same level of background expression of the luciferase gene when transferrin is readily available and therefore the same level of expression of the *TbTfR*. When the serum is switched to canine transferrin, the relative luminescence of the *ESAG6* control increases to 0.56, whereas none of the *ESAG7* cell lines increase above 0.34. This shows that the cells are still responding to the change in iron but not by the same

magnitude recorded for the *ESAG6* control cell line. Insert 1.1 shows a higher level or relative luminescence in non-iron starved media, similar to C1 and D1 in **Figure 4.11**. Repeat 3 had a much higher level of relative luminescence for each of the cell lines which will be affecting the averages and may be skewing the data to suggest that the cell lines are responding more than they actually are, however *ESAG6* is also higher for this repeat so the comparisons between the cell lines can still be made. From the data in **Figure 4.13**, it is difficult to confirm whether luciferase expression has been upregulated by iron starvation due to the very low readings for the relative luminescence. This may be because the assay has not been optimised for *ESAG7*.



**Figure 4.14:** Calculated fold-change in luciferase luminescence in *ESAG7* 3'UTR cell lines when *T. brucei* BSF cells are incubated in media supplemented with FBS or canine serum. The bars on the graph indicate the overall average fold-change and the points represent the average fold-change for each biological replicate. Data was normalised to  $1 \times 10^6$  cells/mL.

The calculated fold change shows the difference between the expression of the luciferase reporter gene under normal and iron starvation conditions, making it easier to visualise how much the expression of the luciferase gene and therefore the transferrin receptor has increased when iron is not readily available. In **Figure 4.14** luciferase expression in the *ESAG6* control cell line is increasing approximately 3.5-fold, within the reported range for the upregulation of the transferrin receptor under iron starvation conditions. It is however lower

than previously observed in these experiments in **Figure 4.12**. *ESAG7* and Insert 2.2 increase 2.7- and 4.3-fold respectively which is similar to the *ESAG6* cell line (3.5-fold), suggesting that they are both responding to iron starvation conditions by the same amount mediated by the 3'UTR. The individual repeats for *ESAG7* are fairly consistent so it is likely that the cell line is still responding to iron starvation conditions like *ESAG6* as was expected, despite the low values recorded for relative luminescence. Insert 2.2 however has a much greater value for Repeat 2 which would increase the average fold-change.

**Table 4.3: Summary table describing the *ESAG7* 3'UTR truncations and how the cell lines respond to iron starvation conditions.**

<b>Name</b>	<b>Size</b>	<b>5' or 3' Truncation</b>	<b>Responsive to Iron Starvation</b>
<b>ESAG7 3'UTR</b>	492 bp	n/a	Responsive
<b>Insert 1</b>	364 bp	3'	No Response
<b>Insert 2</b>	346 bp	5'	Responsive
<b>Insert 3</b>	218 bp	3' and 5'	No Response

#### 4.2.3 Summary of Luciferase Assay Results

Overall it appears that cell lines E3, F3, G3, *ESAG7* and 2.2 are responding to iron starvation conditions in a similar way to the control cell line *ESAG6*. The data shows that there is a low relative luminescence recorded for each of the listed cell lines under normal conditions and when the cells are incubated with canine serum, the relative luminescence increases. This is shown by the calculated fold change where there is an increase in luciferase expression proportional the *TbTfR* greater than 2.5-fold. For C1, it appears that the cell line may have lost a repressive element as there is a high relative luminescence recorded for the cell line under normal and iron starvation conditions. For D1, 1.1, and 3.1 there does not appear to be much of a response to iron starvation conditions when the fold-change is calculated suggesting that the cell lines have lost, at least partially, the ability to respond to iron starvation conditions. For the *ESAG7* truncations, this may be because the luminescence readings are too low because the assay has not been optimised for *ESAG7*. From the results above there are observed trends in the data that show some cell lines still responding to iron



The data collated from **Table 4.2**, **Table 4.3** and **Figure 4.15** show a distinct pattern. From the Figures throughout **Section 4.2** it is clear that some of the truncations have maintained the ability to respond to iron starvation conditions via upregulation of the luciferase gene and therefore upregulation of the *TbTfR* and some that have lost this ability. Of those that are still able to respond, they have all been truncated at the 5' end with the 3' end remaining intact (**Table 4.2** and **4.3**). It is therefore likely that there may be an important motif, located near the 3' end of the 3'UTR of *ESAG6* and *ESAG7* that is essential for upregulation of the *T. brucei* transferrin receptor under iron starvation conditions. When the 3'UTR sequences of *ESAG6* and *ESAG7* are aligned from the 13 BES they are present in, there is a 26-nucleotide conserved section located near the 3' terminus of the 3'UTR highlighted in **Figure 4.15**. When this sequence is searched against the truncations, it is only present in Inserts E, F, G and 2. All of these have been shown to respond to iron starvation by upregulation of the luciferase gene in a similar manner to the *ESAG6* 3'UTR (**Table 4.2** and **4.3**). This sequence could potentially be an RNA binding motif that binds an RBP to stabilise the mRNA and increase translation. This, however, is only speculation based on the results above and more specific luciferase assays would be required to confirm that this sequence is important for the upregulation of the *T. brucei* transferrin receptor under iron starvation conditions. There are other conserved regions between the expression sites, however they are not as long as this 26 bp sequence.

## 5.0 Discussion

### 5.1 Importance of Characterising the Iron Starvation Response

Understanding how trypanosomes are able to respond to changes in the availability of iron to enhance their survival could be important in aiding in the elimination of the disease. Clinically cells may experience iron starvation conditions in three circumstances. These include when the host has developed anaemia and there is an increased concentration of apo-transferrin available over holo-transferrin (Trevor *et al.*, 2019). The mammalian transferrin receptor has a higher affinity for holo-transferrin and is able to preferentially bind holo-transferrin over apo-transferrin by over 200,000-fold to obtain sufficient iron when apo-transferrin is most abundant (Trevor *et al.*, 2019; Kleven *et al.*, 2018). When the host develops anaemia, this is detrimental to the parasite as the *TbTfR* is not able to distinguish between holo- and apo-transferrin as efficiently as the mammalian receptor. When holo-transferrin is scarce, the parasite will take up more apo-transferrin, which has no iron bound, so the parasite will not be able to obtain enough iron.

Another circumstance where the cells may experience iron starvation conditions is when the parasite infects a new host organism. When the metacyclic cells differentiate into BSF cells they express the transferrin receptor and VSG from a single BES. If the parasite infects a new host and the expressed receptor does not bind the transferrin with a high affinity, initially, the parasite may not be acquiring enough iron from the environment (Mussmann *et al.*, 2004). Another time the cells may experience iron starvation conditions is when there is an immune response raised against the transferrin receptor. The transferrin receptor is shielded in the flagellar pocket by the VSG coat and the glycosylation of the proteins is thought to protect the receptor from anti-receptor antibodies (Gerrits *et al.*, 2002). However, the ligand-binding domain must be exposed to the external environment to allow transferrin to bind. Antibodies could bind to the ligand binding domain and block transferrin from binding, blocking iron uptake. When a low affinity receptor is expressed the antibodies are able to outcompete transferrin more successfully (Gerrits *et al.*, 2002). If the receptor is saturated

with transferrin, the antibodies cannot bind, however, as the immune response progresses, the concentration of antibodies will increase.

In relation to the amount of transferrin available in normal host serum, the parasite only requires a small amount of iron. Yet due to anaemia being a symptom of stage I *T. brucei* infection in humans and other mammalian hosts, in an infection the cells will face iron starvation conditions where holo-transferrin is less readily available, no matter how high affinity the receptor is or how much of an immune response has been raised. Iron is essential for many cellular processes in all eukaryotic cells, therefore, the upregulation of the transferrin receptor under iron starvation conditions is a crucial mechanism for parasite survival. If the signalling pathway that can detect low iron uptake or how the parasite is able to respond to this could be identified, components in the pathway could be used as new targets for drug development and could potentially aid in the elimination of this disease. The first step in identifying the components of this pathway is to isolate how the transferrin receptor is upregulated under iron starvation conditions.

Benz *et al.* (2018) showed that the upregulation of the transferrin receptor is regulated in part through the *ESAG6* 3'UTR. Fusion of the 3'UTR to a *firefly luciferase* or *GFP* reporter gene showed increased expression of the reporter genes when iron starvation conditions were induced through serum switch and incubation with the iron chelator deferoxamine (Benz *et al.*, 2018). Other trypanosome proteins such as *ESAG9* are also regulated via the 3'UTR where RNA-binding motifs have been identified (Monk *et al.*, 2013). It was hypothesised that there may be an RNA-binding domain located in the *ESAG6* 3'UTR and potentially the *ESAG7* 3'UTR due to their sequence similarity. It is likely that an RBP binds to an RNA binding domain in the 3'UTR which stabilises the mRNA to increase translation and expression of the receptor (Erben *et al.*, 2014; Carbajo *et al.*, 2021). It is known that the regulation must be post-transcriptional as there is no observed increase in VSG which is transcribed polycistronically from the same promoter (Benz *et al.*, 2018). To attempt to identify a potential binding domain, 5' and 3' truncations were made of the *ESAG6* and *ESAG7* 3'UTRs which were fused to the *fluc* reporter gene and transfected into 2T1 cells. Luciferase activity assays were performed on the cells with the truncated 3'UTRs under normal and iron starvation conditions compared to a full-length control cell line to see whether the upregulation of the TfR under

iron starvation was maintained or had been lost. If the cells still responded in the same way, it is likely a potential binding motif remained undisrupted in the truncation, if the response was lost and the cells showed no response a potential motif may have been lost or disrupted. However, it is unlikely that the response would be this simple as there may be repressive and up-regulatory elements that could bind to the 3'UTRs.

## 5.2 The Biological Role of the *Trypanosoma brucei* Transferrin Receptor

As previously described in Lister 427 *T. brucei* cells there are fifteen expression sites with *ESAG6* and *ESAG7* genes present that encode similar but not identical versions of the *TbTfR* (Bitter *et al.*, 1998). The sequences differ by less than 10% in the receptor (**Figure 5.1**), however, the variations have an effect on the affinity of the receptor to bind different host transferrin molecules (Gerrits *et al.*, 2002). There is some disagreement surrounding the reasoning for the variation between the *TbTfR* of the multiple BES. Two key arguments have emerged which suggest that the variation is to allow the parasite cope with the sequence diversity of host transferrin, enabling the parasite to adapt to a wide host range (Bitter *et al.*, 1998). The counter argument is that the variation is to enhance the parasites survival in the mammalian host by allowing the parasite to evade the host immune response, in a similar fashion to VSG switching. The two opposing viewpoints are discussed below.



**Figure 5.1:** Diagram showing the conserved regions of the *Trypanosoma brucei* transferrin receptor across the 15 bloodstream expression sites. The sequences are highly conserved, black sections represent conserved regions and the white sections represent variable regions.

It has been widely believed that the diversity in the receptors and their ability to bind transferrin has been driven by the parasite adapting to infect a broad host range, and that

being able to evolve a receptor that can bind with a high affinity to all host transferrins is advantageous (Bitter *et al.*, 1998). In the mammalian hosts the parasite is able to colonise, the amino acid sequences of the transferrin molecules differ by up to 30% (Gerrits *et al.*, 2002). The ability to switch between TfRs may allow the parasite to cope with this large variability in host transferrins. Without this adaptation the parasite could only bind some host transferrins with a high affinity, potentially making it difficult for the parasite to infect other hosts with low affinity binding (Bitter *et al.*, 1998).

Salmon *et al.* (2005), however, stated that the observed differences in parasite growth in different host sera is not due to the species specificity of transferrin (Salmon *et al.*, 2005). Transferrin is an essential growth factor for the parasite as it is a major source of iron. Therefore, disrupting transferrin uptake is detrimental to the cells. Growth variations have been observed when *T. brucei* BSF cells are grown in media supplemented with different serum which could be linked to the variations in the active expression site (Salmon *et al.*, 2005). To study any links between parasite growth and transferrin uptake, Salmon *et al.* (2005) recorded growth rate over a specific time period on clones expressing a VSG dominant in human infection (ETat 1.2R) or expressing a VSG dominant in murine infection (AnTat 1.3A). Excess bovine, human and murine transferrin was added to the clones and growth rate was recorded. The addition of various transferrins did not significantly alter the growth rate of any of the cell lines, suggesting that low and high affinity receptors are sufficient for transferrin uptake and a low affinity does not hinder growth (Salmon *et al.*, 2005). The results from this study indicate that the cells are able to remain viable in range of mammalian serum even where they encode a low affinity receptor, arguing that receptor affinity is not the driving evolutionary force behind the variation in TfR between BESs and that there must be other components influencing cell growth and survival. It is, however, known that when a low affinity transferrin is present the parasite will upregulate the expression of the *TbTfR* to obtain sufficient transferrin from the environment.

Research to support the theory that TfR variation enables adaptation to a wide host range involves switching the source of transferrin with high or low affinity for the *TbTfR* and analysing if there is a switch in the active expression sites of the culture. Experiments performed by Bitter *et al.* (1998) cultured trypanosomes expressing the VSG221 BES1

transferrin receptor with a very low affinity for canine transferrin in media supplemented with canine transferrin. They found that most of the cells stopped dividing, cells that outgrew the population were shown to have switched to the VO2 expression site where the transferrin receptor bind canine transferrin more efficiently (Bitter *et al.*, 1998). This hypothesis was furthered by Gerrits *et al.* (2002) where they suggested that a high affinity receptor ensures efficient transferrin uptake in the presence of anti-receptor antibodies (Gerrits *et al.*, 2002). This was also shown by looking at expression site switching and antibody presence in low and high affinity receptors when available transferrin is switched. These experiments, however, are conducted in an artificial environment and in an infection the amount of transferrin would be much higher and there may actually be a sufficient amount present for the trypanosome to acquire enough iron for survival even with a low affinity receptor when antibodies are present (Gerrits *et al.*, 2002).

Calculations based on binding affinities support the Gerrits *et al.* (2002) hypothesis that high-affinity receptors not only ensure efficient iron uptake, but that it is maintained in the presence of anti-receptor antibodies (Steverding, 2003). In an 8-hour time period a single trypanosome cell takes up approximately 85,000 Fe<sup>3+</sup> molecules, yet only requires 40,000 of these per generation doubling time. At this rate, only 50% of the transferrin receptors need to be saturated with transferrin for iron uptake to be sufficient (Steverding, 2003). In mammalian serum, there is a very high concentration of transferrin (30 µM), only a transferrin receptor with a K<sub>d</sub> value below this will be occupied with less than 50% transferrin. These calculations make it highly unlikely that BSF cells would be deprived of enough iron to induce switching to a higher affinity receptor (Steverding 2003). However, the cells may require a high affinity transferrin receptor *in vivo* when anti-receptor antibodies are present. Gerrits *et al.* (2002) showed that physiological concentrations of anti-receptor antibodies to a transferrin receptor with a K<sub>d</sub> value of 0.014 µM were able to inhibit the growth of the trypanosomes when cultured in media supplemented with canine serum (Steverding, 2003). This was enough to induce switching to a receptor that can bind canine transferrin with a higher affinity. This is because in an infection, anti-transferrin receptor antibodies can bind to the receptor and block transferrin binding. They compete with transferrin for access to the TbTfR and can outcompete transferrin that binds to a low affinity receptor more effectively. If the receptor can bind transferrin at a higher affinity it makes it less likely the receptor is

exposed to the anti-receptor antibodies (Gerrits *et al.*, 2002). The polyclonal antibodies used in this study however could have high concentrations of specific immunoglobulins because they were obtained from a hyperimmune serum, it remains to be shown whether high affinity antibodies are raised against the ligand binding domain of the transferrin receptor in a chronic infection (Steverding, 2003). There is also the point that transcription is initiated at inactive expression sites, resulting in 20% of expressed receptors not being transcribed from the active expression site (Steverding, 2003). These would have different binding affinities so if the TfR in the active ES is low affinity, there may also be high affinity receptors available.

More recently, Trevor *et al.* (2019) suggested that the variations of the transferrin receptor were not in fact to allow the parasite to adapt to a wide host range and bind different variations of transferrin with a high affinity, but instead to aid in the evasion of the host immune response (Trevor *et al.*, 2019). They argue that if variation is driven by a selection pressure to facilitate ligand binding in different hosts, the most variable region should be in the ligand binding domain that directly comes into contact with host transferrin (Trevor *et al.*, 2019). However, it was found that the most diverse region was the residues that came into contact with the immune system, with 68% of all polymorphisms occurring here. This therefore suggests that antigenic variation is the driving force of the diversification of the transferrin receptor. Despite this, of 16 identified residues do come directly into contact with transferrin, 5 of them (31%) are among the most polymorphic, indicating that the region that directly contacts transferrin is variable, to account for the variable binding affinity of the different host transferrins.

A key argument against the immune evasion theory is the existence of the resistance expression site (R-ES) present in *T. b. rhodesiense* cells. *T. b. rhodesiense* is a human infective subspecies of *T. brucei* that has evolved a mechanism that allows the parasite to survive in normal human serum (NHS) despite the presence of trypanolytic factor (TLF) (Zoll *et al.*, 2018). TLF1 and TLF2 contain two common components, haptoglobin-related protein (Hpr) and apolipoprotein L1 (ApoL1). Hpr is associated with high density lipoproteins which bind to haemoglobin that is taken up by the parasite via *TbHpHbR* also taking up ApoL1 (Pays *et al.*, 2006). ApoL1 is a pore forming component that is trafficked to the endosome where acidification results in ApoL1 inserting into the lysosomal membrane causing lysosomal

swelling and cell death (Zoll *et al.*, 2018). When studying *T. b. rhodesiense*, Xong *et al.* (1998) identified a strain of cells expressing the same VSG (ETat 1.10) that were resistant to normal human serum. Differential cDNA screening between resistant and sensitive strains identified a transcript present in only resistant cells, termed serum resistance-associated (SRA) which was found to encode a VSG-like protein encoded in an expression site (Xong *et al.*, 1998). Cells expressing SRA are able to survive in the human host as SRA binds to ApoL1 preventing the membrane-targeting domain from inserting into the lysosomal membrane (Pays *et al.*, 2006). Despite the requirement for ETat 1.10 for resistance to human serum, the VSG expressed was not important suggesting that SRA was only present in certain BESs and switching to this site would confer resistance to human serum, as long as antigenic variation was performed by gene conversion where VSG was replaced by a different VSG from the same expression site (Xong *et al.*, 1998). SRA was found to be to only be present in a single BES, termed the resistance expression site (R-ES), despite the 20 expression sites identified in *T. brucei*. ETat 1.10 is only activated when cells are exposed to NHS, in the absence of this selection pressure, the cells switch away from the R-ES as it lacks other essential ESAGs that are required for survival. Therefore, expression of SRA is essential for parasite survival in human serum but is also detrimental and when the selection pressure is removed the cells switch active expression site.

This is important in the argument against the hypothesis that expression site switching is driven by the transferrin receptor evasion of the host immune response as it shows that in this case expression site switching is driven by changes to host serum, not by changes in transferrin binding affinity, but it shows that the cells are able to respond to changes in host serum and components of that serum. This is also supported as when the *T. b. rhodesiense* parasites are no longer infecting a human host they switch away from the R-ES as it lacks important ESAGs showing that switching can be driven by the absence of certain ESAGs or the presence of detrimental ESAGs and if the TfR cannot obtain enough iron, this may be detrimental enough to the parasite to induce expression site switching. In the human host the cells are also able to survive with a single expressed transferrin receptor without being cleared by anti-receptor antibodies, suggesting that immune evasion is not essential for the variation of the receptor.

Trevor *et al.* (2019) also suggest that the receptors have not evolved to bind transferrin with a high affinity from a certain mammalian host but have actually reduced affinity for all host transferrins by evolutionary drift in the absence of the selection pressure to bind transferrin tightly (Trevor *et al.*, 2019). To support this theory an experiment was set up where cells expressing the TfR transcribed from BES1 were grown in media supplemented with four different host sera (horse, cow, rabbit and pig). Transferrin from these host animals has previously been shown to bind to the TfR in BES1 with varying affinities highest being obtained from foetal calf serum and the lowest being obtained from horse and pig serum. After 72 hours of growth, all cell lines survived in the different media and with sequencing analysis, showed that the predominant receptor in each of the cultures was still expressed from BES1, and other receptors were detected at low levels, showing the experiment was conducted for long enough to induce expression site switching. The results from this show that despite other transferrin receptors binding transferrin from some of the hosts with a higher affinity, they were no more likely to be expressed suggesting that there was not a selection pressure of transferrin affinity acting on the cells to induce expression site switch (Trevor *et al.*, 2019). Due to the low level of background switching, it is, however, likely that the detected TfRs were present at the start of the experiment also but have not become dominant as the parasite is stable expressing VSG221 (Liu *et al.*, 2018).

The way this experiment was conducted however raises some questions, as it is known that BES1 expresses VSG221, and it has been shown that switching away from this VSG to a different expression site can be detrimental to the cells (Liu *et al.*, 2018). Therefore, unless the cells were not receiving sufficient iron, switching away from the expressed transferrin receptor would be detrimental in terms of immune evasion. The experiment also does not actually measure whether the cells are experiencing iron starvation, as the expression of the *Tb*TfR was not analysed, therefore, it is not clear whether the cells were receiving sufficient iron from the expressed receptor which would not be inducible of expression site switching as the cells would not be under stress. It is also notable that dog serum was not used in this experiment alongside the other animal sera, as the BES1 TfR has been shown to have a very low affinity for canine transferrin and is used extensively in research as the serum of choice to induce iron starvation conditions and study the regulation of the transferrin receptor. If canine serum had been used the results may have been different with more switching

observed. Overall the results from this experiment suggest that expression of low affinity receptors does not induce expression site switching, however, the lack of dog serum and not measuring the expression of these receptors give an unreliable argument.

The mechanism by which antigenic variation occurs during an early infection supports the evidence that the presence of a low affinity receptor induces switching of expression sites, over protecting the receptor from the immune response. In early infection, the parasite is more likely to switch expression sites to change the expressed VSG and therefore also the expressed TfR (Vanhamme *et al.*, 2000). This is advantageous to the parasite in terms of scavenging for host transferrin as the parasite can switch expression sites until a suitable TfR variant has been expressed, as at this stage the parasite can continue to express the same transferrin receptor but remain protected from the host immune response by recombination of VSG from the repertoire of genes and pseudogenes. Evidence for this is shown by growing cultures of *T. brucei* expressing different variants of the *TbTfR* in media supplemented with different host transferrins. The parasites switched expression sites to take up transferrin with a greater efficiency in each of the separate cultures (Bitter *et al.*, 1998).

Overall it is likely that both of these hypotheses play a role in the variation of the transferrin receptor across the various BESs. In the original Bitter *et al.* (1998) paper they begin by stating that they thought the hypervariable region of the transferrin receptor has been selected for variability, possible as an immunodominant epitope of the receptor. But the initial experimental results led them onto the hypothesis that the variability may actually enable high affinity binding of multiple host transferrins (Bitter *et al.*, 1998). There is also the reasoning that expression site switching may not occur due to transferrin binding or immune evasion of the *TbTfR*, but instead there is selection of a pre-existing sub-population expressing VSG and the TfR from a different BES, as background switching occurs simultaneously in all cultures. There is clearly variation in the binding affinity of the transferrin receptor to various host transferrins, yet the parasite is still able to survive in the wide host range showing that the parasite is able to obtain sufficient iron from each host. This may be through switching to a BESs with a more suitable TfR or may just be through the upregulation of the receptor when the parasite is not taking up enough iron. The theory that receptor variation has actually evolved to aid in the avoidance of the host immune response has not yet become a widely

accepted theory. As previously discussed, the N-glycosylation of the TfR is thought to shield the receptor from antibodies preventing them from binding to the receptor, but also if the transferrin receptor binds transferrin with a high affinity, bound transferrin blocks the antibodies also protecting the parasite from the immune system favouring higher affinity receptors for optimum transferrin uptake and immune evasion (Steverding *et al.*, 2000; Gerrits *et al.*, 2002).

Despite the argument around whether transferrin binding affinity has driven the diversification of the variations of the *TbTfR* in the expression sites or not, this can be taken advantage of experimentally to induce iron starvation conditions in the cells. When the parasites colonise a new host, they could potentially enter an environment where the availability of iron is limited if the transferrin receptor expressed only binds transferrin with a low affinity (Fast *et al.*, 1999). Upon infection an immediate upregulation of the transferrin receptor would allow the parasite to obtain sufficient transferrin for survival and allow time for the selection of a suitable BES with a high affinity TfR to enhance the parasites survival in the new host environment (Fast *et al.*, 1999).

## 5.3 Analysis of Results

### 5.3.1 Generation of the *ESAG6* 3'UTR and Truncations

Generation of the of the correct truncated sequences for the *ESAG6* 3'UTR, and their ligation into the pRPa vector system, was quite a difficult process with a number of set-backs. The initial PCR was successful as shown in **Figure 4.2**, with correct sized bands amplified and gel purified. When it was attempted to ligate the inserts into pGEM-T easy there were a lot of unsuccessful transformants with a higher ratio of blue colonies grown on plates. When the white colonies were screened for the correct inserts, they mostly returned minipreps without any vector or insert DNA which appeared as a smear on the agarose gel. This was resolved when a new vector kit was purchased, it was likely the rapid ligation buffer had degraded due to repeated freezing and thawing. Once the PCR inserts were ligated into pGEM-T they were cut from the vector and ligated into the pRPa reporter system. When performing the RE

digests and purifications, it was easier with pGEM-T which is a pre-cut vector and has A-overhangs to prevent self-ligation. The ligations into pRPa $\Delta$  were more difficult as this was not a commercially available vector. There were a lot of problems with contamination in the laboratory due to Covid-19 safety procedures where Bunsen burners could not be used to maintain sterile conditions as the Covid-19 PPE was flammable and caused a health and safety hazard. Due to the contamination, many of the colonies grown on the transformed plates contained no plasmid DNA. To increase the efficiency of the ligation and reduce the background a number of minor alterations were made to the cloning protocol including increasing the concentration of the inserts with a SpeedVac (Savant ISS110), using new aliquots of DH5a cells, increasing the ligation volume to 5  $\mu$ L, and using new components of the ligation reaction (new rapid ligation buffer and new ligase). Eventually correct plasmid DNA was isolated for each of the inserts ligated into the pRPa-*fLUC*<sup>mut</sup> reporter system. These were then DNA sequenced to confirm the sequences prior to transfection into the 2T1 cell line.

When the DNA sequencing analysis was returned polymorphisms were identified in each of the inserts. These could be traced back to the template DNA that was used in the initial PCR. Due to the nature of the project, investigating the effect of alterations made to the 3'UTR on the iron starvation response of the transferrin receptor, the generated sequences could not be used for the luciferase assay as any changes to the iron starvation response of the reporter gene could not be attributed directly to the truncations made but may have been due to the variations in the sequences. Due to this, new template DNA was required. To attempt the generate template DNA of the full length *ESAG6* 3'UTR two procedures were attempted, a PCR with genomic DNA as template and the full length forward and reverse primers and an RT-PCR with extracted RNA as a template and the Oligo-dT adapter primer and the full length forward primer. When the resulting DNA was ligated into pGEM-T and sequenced, there was also a number of polymorphisms in the sequence, but in different locations than the initial mismatches identified. The source of the changes in the 3'UTR sequence were attributed to transcription of transferrin receptors from silent expression sites.

The expression site associated genes and VSG are transcribed from the bloodstream expression sites. In *T. brucei* there are 13 expression sites that encode *ESAG6* and *ESAG7* to

form a functional heterodimeric transferrin receptor. The different expression sites encode similar but not identical versions of the proteins that make up the receptor which alters the binding affinity of the receptors from different expression sites for different host transferrins. The 2T1 cell line is monomorphic so is stable expressing VSG221 from BES1 (Liu *et al.*, 2018). Spontaneous switching occurs in the background as shown by transcriptomics and proteomics but switching away from VSG221 is detrimental and any cells that have switched expression site are outgrown by the population expressing VSG221. Due to this the majority of the cells in the 2T1 cell line are expressing VSG221 and, therefore, the transferrin receptor from BES1 (Liu *et al.*, 2018). When the truncated cell lines were generated the 3'UTR from BES1 was used to design the primers to amplify the 3'UTR. Due to the similarity of the *ESAG6* 3'UTR across the BESs the forward and reverse primers designed were also complimentary to the beginning and the end of the *ESAG6* 3'UTR of other expression sites. When the PCR was performed, it is possible that gDNA or RNA from cells that had switched expression sites was present in the template DNA aliquot and therefore the sequences amplified may have been from a mixture of BES. So, when the DNA was sequenced it was not 100% conserved with the *ESAG6* 3'UTR of BES1 but may have contained sections of sequences from other *ESAG6* 3'UTRs in other expression sites. The sequences were aligned to the other *ESAG6* 3'UTRs available from the different expression sites, but the sequences did not match perfectly with any suggesting that there may be a mixture of sequences or that errors may have been introduced in the PCR, however, a high fidelity polymerase was used and the amplified sequence is short.

The polymorphisms in the sequence may not have only been introduced to the BES1 *ESAG6* 3'UTR sequences from DNA or RNA extracted from cells that had switched expression sites, but also from *ESAG6* transcribed from silent expression sites (Ansorge *et al.*, 1999). The cells have only a single active expression site through the process of monoallelic exclusion. Transcription, however, is initiated at all expression sites simultaneously but only continues at the active expression site due to the presence of the expression site body (Kassem *et al.*, 2014; Navarro and Gull, 2001). It has been shown that at inactive expression sites there is a repressive element at the telomere which prevents transcription, and as the distance from the telomere increases, the repression is less severe (Ansorge *et al.*, 1999). Transcription immediately downstream of the promoter in the silent expression sites can be detected by RT-PCR. *ESAG6* and *ESAG7* are located approximately 4 and 2 kb from the promoter,

respectively, so it had previously been assumed that a single cell only expresses a single transferrin receptor, in the same way that a single cell only expresses a single VSG (Ansorge *et al.*, 1999). However, experiments have shown that *ESAG6* RNA from other silent expression sites can be detected by RT-PCR with as much as 20% heterogeneity (Ansorge *et al.*, 1999). Single cell RT-PCR showed that several expression site transcripts can be identified with heterogeneity decreasing downstream of the promoter (Kassem *et al.*, 2014). It is, therefore, now known that a single cell can express multiple transferrin receptors from different expression sites but at a much lower level than the active expression site (Ansorge *et al.*, 1999). Due to this the 2T1 cell population has BES1 as the active expression site, but can also express a low level of different transferrin receptors from silent ESs, and this RNA could have been used as part of the template for the RT-PCR to give some of the mismatches detected when attempting to regenerate the *ESAG6* 3'UTR template DNA.

A new *ESAG6* 3'UTR template was generated eventually by designing a primer that was complimentary to the end of the *ESAG6* open reading frame, which is less conserved than the 3'UTR, so the primer was specific to BES1. With this primer it was known that the amplified 3'UTR was from the desired expression site. This PCR was performed with the same reverse primer using gDNA as the template. The resultant PCR product was ligated into pGEM-T easy and DNA sequencing revealed that a correct sequence for the *ESAG6* 3'UTR had been obtained. This was then used to regenerate the truncations. These were then ligated directly into the pRPa reporter system, bypassing the pGEM-T vector, which was used originally due to the small size of the fragments, as they were easier to ligate into the commercially available vector as an intermediate. Due to the reproducibility of direct ligations, this was not necessary, and the truncations were successfully ligated into pRPa without the intermediate vector and the sequences were confirmed by DNA sequencing. At this stage they could be transfected into the 2T1 cells.

### 5.3.2 Generation of the *ESAG7* 3'UTR and Truncations

The original research that demonstrated that fusion of the *ESAG6* 3'UTR to the *flUC* reporter gene increased luciferase expression proportional to the transferrin receptor, only focussed on the *ESAG6* 3'UTR (Benz *et al.*, 2018). It has, however, been shown that for a functional

transferrin receptor, ESAG6 and ESAG7 must be associated to form a heterodimer to allow transferrin binding and facilitate iron uptake with the receptor anchored in the plasma membrane (Salmon *et al.*, 1997). Benz *et al.*, identified that the *ESAG6* 3'UTR is 335 bp in length. The same method was utilised to identify the length of the *ESAG7* 3'UTR, by a two-step RT-PCR with primers complimentary to the end of the BES1 *ESAG7* ORF to ensure the correct 3'UTR was amplified and an Oligo-dT primer to identify sites of polyadenylation. Sequencing of the resulting clones identified that polyadenylation occurred at sites within a 40 bp range at locations of 453 bp, 477 bp and 492 bp downstream of the stop codon, which was expected based on the polyadenylation of the *ESAG6* 3'UTR (Benz *et al.*, 2018). The furthest polyadenylation site was located 492 bp downstream of the stop codon, so this was taken as the length of the *ESAG7* 3'UTR. It was expected that the length of the *ESAG7* 3'UTR would be longer than the *ESAG6* 3'UTR to account for the GPI anchor signal sequence located at the C terminus of the *ESAG6* open reading frame. When Trevor *et al.* (2019) identified the crystal structure of the transferrin receptor, there were 30 residues at the C-terminal end of ESAG6 that could not be resolved that most likely form a flexible polypeptide that links the receptor to the GPI anchor. Kabriri *et al.* (2021) confirmed that the 3' end of *ESAG6* codes for a 105 bp (35 amino acid) C-terminal GPI anchor signal sequence. When 157 bp of the *ESAG6* ORF and the 335 bp 3'UTR is aligned with the 492 bp *ESAG7* 3'UTR in **Figure 4.6**, the sequences are 91% conserved, showing that the end of the *ESAG6* ORF is closely related to the beginning of the *ESAG7* 3'UTR suggesting that ESAG7 may have previously been anchored in the membrane by a GPI anchor but due to the associations with anchored ESAG6, this has evolved away and is no longer required for transferrin receptor function. It has been shown that the *ESAG6* GPI anchor is essential for dimerisation and trafficking of the transferrin receptor and when the anchor motif is added to the end of *ESAG7*, it behaves as *ESAG6* (Biebinger *et al.*, 2003). Loss of the GPI anchor on *ESAG6* results in impaired dimerisation with *ESAG7* and is not trafficked to the membrane but likely degraded after lysosomal targeting (Biebinger *et al.*, 2003).

The identification of the *ESAG7* 3'UTR was essential for understanding the role of 3'UTRs of both proteins involved in transferrin uptake and how they can mediate expression of the receptor. The similarities between the 3'UTRs of both genes across the 13 ESs, suggests that in response to iron starvation, both genes are regulated via the same mechanism, but this

had not yet been experimentally demonstrated. It is likely that there is a conserved motif between the *ESAG6* and *ESAG7* 3'UTR that is a binding site, potentially for the same molecule or a homologue of the same protein which is required to stabilise the mRNA to increase gene expression and translation of the protein. This is also assumed, not just because of the similarity of the sequences but due to the co-regulation of the receptor, where each protein is required to form a functional dimer (Salmon *et al.*, 1997).

### 5.3.3 Luciferase Assay Analysis

Iron starvation conditions were experimentally induced by taking advantage of the variable binding affinity of the transferrin receptor. The receptor expressed from BES1, the source of the 3'UTR sequences, has been shown to be able to bind bovine transferrin with a much higher affinity than canine transferrin (canine transferrin has a disassociation constant > 1000 nM, whereas bovine transferrin has a  $K_d$  value of 80.1) (Mussmann *et al.*, 2003; Trevor *et al.*, 2019). Due to this, switching the host serum from 10% bovine to 10% canine in the HMI11-T media is enough to starve the cells of iron. The expression of the transferrin receptor in the luciferase assay was not directly measured, the relative luminescence relates to the expression of the *firefly luciferase* gene which is fused to the full length or truncated 3'UTRs which then mediate the upregulation or repression of the *luciferase* gene when iron starvation is induced. When cells containing the full length *ESAG6* 3'UTR were incubated in media supplemented with bovine transferrin the expression of the luciferase gene was low with an average recorded relative luminescence of 0.16 and 0.17 in **Figures 4.11** and **4.13**, respectively. When the *ESAG6* cells are incubated in media supplemented with 10% canine serum, so only canine transferrin is available to the cells, the average relative luminescence increases to 0.62 and 0.56 respectively, showing a 4.4 and 3.5 fold increase in the expression of the *luciferase* gene and therefore the expression of the transferrin receptor (**Figures 4.12** and **4.14**). The *ESAG6* control line shows how the wild type transferrin receptor responds to changes in iron availability without recording it directly through western blots, making it easier to measure receptor expression.

The *ESAG6* control cell line therefore shows a low level of background expression of the transferrin receptor under basal conditions, which then increases when iron availability

decreases by a magnitude of change previously reported for the upregulation of the transferrin receptor under iron starvation conditions (range of 2.5-5-fold increase in expression) (Benz *et al.*, 2018). Therefore, any of the truncated cell lines that have maintained the ability to respond to iron starvation conditions should show a similarly low level of *luciferase* expression under non-iron starvation conditions, which then increases between 2.5-5-fold when the cells are incubated in canine serum supplemented media. This was also observed for the *ESAG7* 3'UTR cell line, demonstrating that the *ESAG7* 3'UTR is also important for the transferrin receptor response to iron starvation conditions (**Figure 4.14**). There was, however, a much lower level of relative luminescence in both non-iron starvation conditions (0.09) and only increasing to 0.2 under iron starvation conditions (**Figure 4.13**). This is much lower than the *ESAG6* control but still gives a 2.8 increase in luciferase expression. This is on the lower side of the observed range of upregulation; however, this is most likely because the luciferase assay has not been optimised for *ESAG7* as the *ESAG6* reporter system had been adapted by replacing the *ESAG6* 3'UTR with the *ESAG7* 3'UTR. The plate reader may not be sensitive enough to measure the very low levels of luminescence, therefore, to increase sensitivity, the assay could be repeated with a higher cell density. At a higher cell density, the cells begin to experience iron starvation conditions in the non-iron starved control as the cells are competing with each other for available transferrin. To avoid this the assay could be repeated with an increased aliquot of OneGlo reagent and cells, so there is more substrate that can be broken down by luciferase to give off a higher level of luminescence that could be more accurately detected by the plate reader.

It was also not confirmed whether the *ESAG7* cell line was still expressing the transferrin receptor from BES1, although it was assumed, as switching away from VSG221 is detrimental to the cells (Liu *et al.*, 2018). This could be confirmed by RT-PCR of the *ESAG6* and *ESAG7* genes, if the cells had swapped expression sites away from a receptor with a low affinity for canine transferrin, the cells would not experience iron starvation conditions as strongly. To avoid this the assay could be repeated with media treated with deferoxamine to remove available transferrin from the environment, creating iron starvation conditions that would be experienced by all variations of the *TbTfR*. This was attempted initially, however, deferoxamine was toxic to the cells and high levels of cell death were being recorded. Overall, this suggests that the 3'UTR is important for the *ESAG7* response to iron starvation conditions

but does not prove that it is regulated via the same mechanism as *ESAG6* or that the same section of the 3'UTR is responsible for this upregulation of the luciferase gene.

There are also a number of truncations that also respond to iron starvation conditions in a similar manner to the *ESAG6* control cell line. These were Inserts E-G which are *ESAG6* truncations and the *ESAG7* truncation Insert 2. These showed an increase in *luciferase* expression between 2.9-4.6-fold as shown in **Figures 4.12** and **4.14**. The cell lines all show a low level of relative luminescence under basal conditions which increases when the cell lines are starved of iron in a similar way to the *ESAG6* control. Only Insert E3, however, showed a statistically significant increase in the expression of the transferrin receptor. This is likely because there was a lot of biological variation between the repeats of the luciferase assay and if more repeats had been obtained, it is likely there would have been more significant results. The assays also need to be repeated with other clones to confirm that the patterns identified are consistent when other copies of the cell lines are screened.

From the *ESAG6* and *ESAG7* truncations that maintain the ability to increase the expression of the transferrin receptor under iron starvation conditions, there was an observable pattern in the truncations made as shown in **Figure 4.7**. All of these truncations have been made at the 5' end so that the 3' end of the 3'UTR remained undisrupted. When visualised in **Figure 4.7**, they all show regions of overlap and for Insert 2, it has lost the front part of the 3'UTR that aligns with the *ESAG6* open reading frame, making the truncation most like the *ESAG6* 3'UTR. These results suggested that there may be a region towards the end of the *ESAG6* and *ESAG7* 3'UTR that is responsible for the response of the receptor when iron is less readily available in the environment. The conservation of the 3'UTR also strongly suggests that it may potentially be the same motif responsible for this response in both *ESAG6* and *ESAG7*, which was unclear previously. When the sequences are aligned in **Figure 4.15** a 26 bp sequence (5'-ACTATTTTTCAAATTTAGTTACAACA -3') was highlighted that is conserved across both 3'UTRs in all 13 BESs the two genes are present in. The presence of this sequence in only the truncations that have shown the upregulation to iron starvation conditions and absence from those that do not respond, make it feasible that this is an important motif necessary for the upregulation of the transferrin receptor under iron starvation conditions. This is yet to be experimentally proven.

It is also likely that the response of the receptor to iron starvation is more complicated. Insert C1 in **Figure 4.11**, shows a high relative luminescence under both non-iron starved (0.4) and iron starved conditions (0.38). Due to the high level of background luminescence, it suggests that the 3'UTR has potentially lost a repressive element that is responsible for the low level of background expression of the transferrin receptor. This may also be present at the 3' end of the 3'UTR. This however is not shown in Insert D1 which has a recorded relative luminescence of 0.17 under non-iron starved condition which decreases to 0.13 when the cells are in canine media indicating a loss of upregulation. Insert D is a shorter 3' truncation (137 bp) compared to Insert C1 (264 bp) so it would be presumed that Insert D1 would also have lost this repressive element as the 3' truncation is larger. As this is not the case, it is unclear whether there is a repressive element involved in regulating the transferrin receptor and more repeats are needed for Insert C with different clones.

## 5.4 Limitations

The biggest limitation of the study is the lack of statistically significant data caused by the biological variation between the repeats of the luciferase assays. It was expected that there would be some variation as the assay can be affected by a number of factors that can influence the relative luminescence, these include cell density, temperature of the OneGlo reagent and the batch to batch variability of canine media. To try and remove variation wherever possible, these factors were taken into consideration. The luciferase assay results were normalised to  $1 \times 10^6$  cells at the time of the luciferase assay. Most cells had grown to around this cell density in the 5-hour incubation so removed variation in the results as more cells would have a higher level of background luciferase expression. To remove the temperature variation, the OneGlo was defrosted and warmed to room temperature before the assays were carried out. All iron starvation conditions were induced with the same batch of canine media, so had the same concentration of canine transferrin available to the cells. Despite this, there was still a considerable amount of biological variation in the assay, resulting in few significant results. It would not be possible to remove all biological variation from the cells, so more repeats are needed for significance.

Despite the same batch of canine media being used throughout, different aliquots were used with different batches of HMI11-T which may have also introduced variation. The same batch of FBS was also used throughout the luciferase assays, however the components of FBS are less variable as it is more readily available. It was observed that with the batch of canine serum used, the ALD control cells were able to grow at a quicker rate than the ALD cells in FBS, it is unclear why this is the case and as the assay was normalised by cell count, this variation needs to be investigated.

Technical variation may have also been introduced into the luciferase assay, particularly at the time of the assay. When the cells were incubated for 5 hours in 24 well plates, the cells congregate around the edges, so some areas of the well are denser than others. To ensure that an equal number of cells are aliquoted, and the luminescence recorded, the wells are mixed by pipetting up and down before the assay and before taking a cell count. This reduces the likelihood that too few or too many cells will be assayed and counted, however, it does not completely prevent it.

To confirm the results reported in this study, the luciferase assays should be repeated at least once more to obtain significant results. The luciferase assays should also be repeated with other clones of the generated cell lines to confirm the patterns identified in **Figures 4.12 and 4.14**. Unfortunately, due to time constraints of the project and issues with cell growth and contamination this was not possible but will be repeated in the future.

## **5.5 Post-Transcriptional Regulation in Trypanosomes**

Benz *et al.* (2018) first demonstrated that the *ESAG6* 3'UTR mediates the regulation of the *TbTfR* in response to iron starvation, with an increase in mRNA and protein levels consistent with levels previously reported. It was suggested that this increase was driven by an increase in mRNA stability and/or increased translation efficiency, likely mediated by an RBP that could recognise either structural elements or sequence motifs, but this had not yet been demonstrated (Benz *et al.*, 2018; Carbajo *et al.*, 2021). Due to the co-regulation of *ESAG6* and

ESAG7 and the similarity of the 3'UTRs, it was assumed that a regulatory motif would be conserved between the 3'UTRs. To identify a potential motif, the 3'UTRs were truncated at sites of hypervariability and fused to the reporter system for analysis. A 26 bp putative motif (5'-ACTATTTTTCAAATTTAGTTACAACA-3') was identified located near the 3' end of the 3'UTR. This motif was conserved between the 3'UTRs of *ESAG6* and 7 and between the 13 BESs. It was also only present in truncated cell lines that maintained the iron starvation response, suggesting that it may be required in this response. The truncations were made as it was likely only a short section of the 3'UTR would be required to mediate the iron starvation response.

The mammalian transferrin receptor is also able to rapidly increase expression under iron starvation conditions via a mechanism controlled by the 3'UTR. Mammalian transferrin receptors increase expression under iron starvation conditions through interactions of iron regulatory proteins (IRP-1 and IRP-2) and iron responsive elements (IREs) (Rupani *et al.*, 2016). The regulation of the mammalian transferrin receptor has been shown to be distinct to that of the *TbTfR* as knockout of the IRP-1 homologue (aconitase) has no effect on *TbTfR* regulation (Fast *et al.*, 1999). Despite this, there are some potential similarities between the regulation of the receptors. Both have been shown to be mediated by the 3'UTR with five IREs identified in the mammalian transferrin receptor mRNA 3'UTR (Rupani *et al.*, 2016). It has been shown that IRP binding to these regions impacts mRNA stability which is essential for the upregulation of the receptor when iron is less readily available. This has been hypothesised for the *TbTfR* where something potentially binds to the 3'UTR to stabilise the mRNA and increase translation. The important region identified in the mammalian 3'UTR is 345 nucleotides containing three of the five identified IRE, so the entire 3'UTR is not essential for mammalian transferrin receptor regulation as was also shown for *T. brucei* transferrin receptor regulation.

Post-transcriptional control of gene expression is an essential process to all eukaryotes and RBPs play a critical role in this (Erben *et al.*, 2014). RBPs interact with cytosolic mRNA, normally, but not always, with the 3'UTR to stabilise mRNA and/or increase the translational efficiency. In trypanosomes RBP are implicated in a range of different cellular processes including differentiation, development, the cell cycle, rRNA processing and the heat shock response (Lueong *et al.*, 2016). RNA binding proteins are therefore key factors in gene

expression regulation, and it has been predicted that over 100 RBPs exist in trypanosomes based on canonical RNA-binding domains (Erben *et al.*, 2014). Of the predicted proteins, only a few have been characterised and many of those control mRNA abundance by increasing the stability of the transcript. Due to this it is highly likely that the dynamic regulation of the *TbTfR*, known to be mediated by the 3'UTR, is upregulated by RBP which either stabilise the mRNA or by causing an increase in translation, as is the function of most of the characterised RBP in *T. brucei*.

There have also been some RNA binding domains identified in the 3'UTR of trypanosome genes present in the expression site. The 3'UTR of VSG has been shown to be essential for expression of VSG across the surface of the cell. A 16-nucleotide sequence has been identified in the VSG 3'UTR that is essential for stabilising the VSG transcript (Ridewood *et al.*, 2017). There is approximately 700-fold more VSG mRNA present in BSF cells than ESAG1 mRNA, despite the genes being transcribed on the same polycistronic unit (Cully *et al.*, 1985). VSG transcripts have been shown to have a much longer half-life, allowing increased translation, and this increased half-life has been linked to the 16-mer region in the 3'UTR, as when this is disrupted there is a dramatic reduction in VSG transcript stability (Ridewood *et al.*, 2017). It was predicted that binding to this 16 bp sequence could directly protect the mRNA transcript from degradation by blocking association with nucleases (Ridewood *et al.*, 2017). There is much less *ESAG6* and *ESAG7* mRNA present in BSF cells compared to VSG mRNA, as the receptor is less abundant on the cell surface. However, when the receptor is upregulated under iron starvation conditions, there is an increase in the concentration of receptor mRNA, without an increase in VSG mRNA which is already high. This increase is not due to increased transcription but may be due to an increased half-life of the *ESAG6* and *ESAG7* mRNA if something bound to the 3'UTR to protect the transcripts from degradation. It is yet unknown what if anything binds to either of these 3'UTRs but it could be hypothesised that they are regulated via a similar mechanism due to their localisation to the expression sites.

RNA binding domains have also been discovered in expression site associated genes. *ESAG9* genes are mostly located within polycistronic transcription units positioned at internal chromosomal and sub-telomeric regions (Monk *et al.*, 2013). The exact function of the protein is unknown but *ESAG9* is secreted by BSF cells and is developmentally expressed upon

transition towards stumpy cells (Rico *et al.*, 2017). The upregulation of *ESAG9* transcripts is one of the earliest events during stumpy differentiation and *ESAG9* mRNA is one of the highest expressed during the differentiation (Monk *et al.*, 2013). Analysis of the *ESAG9* 3'UTR identified a highly defined regulatory region responsible for mediating the developmental expression of *ESAG9* (Monk *et al.*, 2013). The identified motif was only 34 nucleotides long and exhibits both positive and negative regulatory potential depending on the life cycle stage of the parasite (Monk *et al.*, 2013). Like *ESAG6* and *ESAG7*, the regulatory element is shown to affect both mRNA and protein levels, but deletion and insertion experiments only had very limited effects (Monk *et al.*, 2013). A model of regulation was suggested where the motif is bound by a negative regulator in slender forms which is replaced by a positive regulator in stumpy forms (Monk *et al.*, 2013). It is unlikely that the putative motif identified in the *ESAG6* and *ESAG7* 3'UTRs functions in this manner as it would be expected that a negative regulator bound under non-iron starved condition and a positive regulator bound under iron starvation conditions. There would, therefore, be an increase in transferrin expression under non-iron starvation condition if the repressive regulatory motif was lost and no increase under iron starvation condition if the upregulatory motif was lost, however this was not the case. It is likely that the 26 bp putative motif identified in this study is only responsible for upregulation.

It has been predicted that the *TbTfR* 3'UTR has extensive structure suggesting that it may interact with multiple RNA binding proteins. When the truncations were made, this secondary structure was not taken into consideration, so it was unclear whether any of the truncations would respond to iron starvation at all. Further research investigating the upregulation of *ESAG9* to induce stumpy formation identified a negative regulator termed REG9.1 (Rico *et al.*, 2017). It was shown that depletion of REG9.1 promotes differentiation into stumpy cells, whereas upregulation promotes differentiation into procyclic cells, illustrating the importance of REG9.1 in transitioning through the parasite lifecycle (Rico *et al.*, 2019). REG 9.1 regulation of NeoR linked to the *ESAG9* 3'UTR was not lost upon deletion of the identified 34 nucleotide domain previously identified, demonstrating that regulation of *ESAG9* after REG9.1 has been depleted does not operate via this single motif alone (Rico *et al.*, 2017). Therefore the 3'UTR sequences that are involved in the regulation of *ESAG9* when REG9.1 is depleted are expected to be dispersed throughout the 3'UTR and function via a multifactorial mechanism (Rico *et al.*, 2017). This evidence makes it possible that there could be multiple

regions of the *ESAG6* and *ESAG7* 3'UTR that may be required for transferrin receptor regulation, and these may not be located in linear motifs, more truncations would need to be designed to further test this and identify other potential binding motifs and study the secondary structure.

To conclude, most sequences responsible for responding to environmental changes are present in the 3'UTR of genes, with the average length of a *T. brucei* 3'UTR being about 400 bp which is long enough to bind at least 14 different regulatory proteins (Clayton, 2019). The iron starvation response was shown to be mediated by the *ESAG6* and *ESAG7* 3'UTR and a 26 bp putative motif was identified that may be required for upregulation of the transferrin receptor under iron starvation conditions. There have been 100s of RBP identified in *T. brucei* but few of these have been characterised. Of those that have, their function is to stabilise mRNA or increase the translational efficiency, and it is likely that one of these mechanisms is how the parasite increases expression of the transferrin receptor when iron is scarce. RNA binding domains have been identified in genes present in the bloodstream expression site and other ESAGs as discussed above, with a 16 bp and 34 bp motif identified in the *VSG* 3'UTR and *ESAG9* 3'UTR respectively. Both of which have been implicated in the upregulation of gene expression by an RBP binding to the identified domain to stabilise the mRNA transcript.

## 5.6 Recommendations for Future Research

### 5.6.1 Further Analysis of the Putative Motif

Now that a potential motif has been identified that could be necessary for the upregulation of the transferrin receptor under iron starvation conditions, it needs to be confirmed. As explained above, the motif is conserved across the *ESAG6* and *ESAG7* 3'UTRS in the different expression sites and it has been shown that all variations of the transferrin receptor from the different expression sites are able to respond to iron starvation conditions, so a conserved region involved in the regulation was expected. From the results, this motif is present in the cell lines with 5' truncations of the 3'UTR and is only present in the cell lines that maintain

the ability to upregulate the expression of the *luciferase* gene when incubated in canine serum.

To confirm that this motif is required for the upregulation of the transferrin receptor under iron starvation conditions, the firefly luciferase reporter system can still be used. Firstly, the motif (5'-ACTATTTTTCAAATTTAGTTACAACA-3') could be disrupted by site directed mutagenesis, whilst the rest of the 3'UTR remains intact. The mutated 3'UTR would then be ligated into the pRPaΔ-*fluc*<sup>mut</sup> reporter system and transfected into 2T1 cells and the luciferase assay would be repeated. To confirm that this motif is important, when disrupted it would be expected that the cells would lose the ability to respond to iron starvation conditions. Therefore, in the mutated cell line, it would be expected that there would be a low background of luciferase expression under basal conditions. When the cells are incubated with media supplemented with canine serum, it would be expected that there would be a similarly low level of relative luminescence recorded as the cells should have lost the ability to respond to iron starvation conditions. If the relative luminescence increases in canine media, it is likely that the identified motif is not essential to the iron starvation response or there may be other regions of the 3'UTR that could also be important.

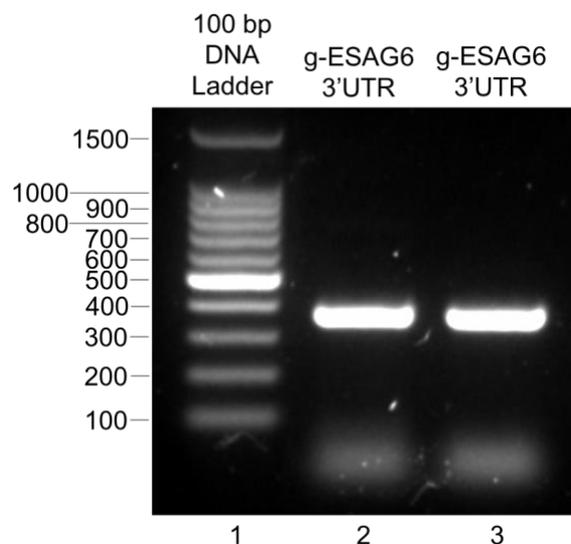
To further confirm that the motif (5'-ACTATTTTTCAAATTTAGTTACAACA-3') in the *ESAG6* and *ESAG7* 3'UTRs mediates the upregulation of the transferrin receptor under iron starvation conditions, the 26 bp sequence could be taken and fused to the *aldolase* 3'UTR. *Aldolase* is not an iron regulated gene so does not respond to changes in the availability of iron but has a stable 3'UTR so has a high level of relative luminescence recorded when fused to the *fluc* gene as shown in **Figure 4.11** and **Figure 4.13**. The 26 bp sequence could be added to the *ALD* 3'UTR to see if the additional motif alters the response of the aldolase cell line to iron starvation conditions. As the *ALD* relative luminescence is high under basal and iron starved conditions, the addition of the 26 bp motif to the *ALD* 3'UTR may cause the repression of *ALD*, down to the lower level of luciferase expression observed for the *ESAG6* control cell line under non-iron starved conditions, which would then increase when the cells are starved of iron. However, due to the high stability of the *ALD* 3'UTR, luciferase expression may remain high under non-iron starved conditions and increase by a similar magnitude as *ESAG6* when iron-starved. This may not be a clear fold-change, however, due to the already high expression of

the luciferase gene. For this reason the *ALD* control cell line may not be the most suitable 3'UTR for fusion of the putative 26 bp motif, due to the high level of background, as there may be other regions of the *ESAG6* and *ESAG7* 3'UTRs that bind repressive elements that result in the low background observed for the *ESAG6* control cell line. The 26 bp motif could be fused directly to the *flUC* reporter gene to identify how the cells respond to iron starvation conditions with the addition of the 26 bp motif alone, aiding in the understanding of how the motif may mediate the iron starvation response.

#### 5.6.2 Characterisation of the Iron Starvation Response of the Non-Expression Site Copy of *ESAG6*

In the *T. brucei* genome, there are several *ESAG* families that are located outside of the expression site despite their usual co-localisation with VSG on the polycistronic units transcribed by RNA polymerase I (Monk *et al.*, 2013). These non-expression site localised genes include *ESAG6* and they are often referred to as GRESAGs (gene related to ESAGs), however, the non-expression site *ESAG6* will be distinguished as *gESAG6*. The non-expression site localised ESAGs are transcribed as normal eukaryotic genes by RNA polymerase II which generates mRNA for all *T. brucei* protein coding genes that are not encoded in the expression sites (Monk *et al.*, 2013). It has been shown that these non-expression site related ESAGs can be functional, despite their positioning. Due to this, it would be interesting to confirm if the internal *ESAG6* and *ESAG7* genes are able to produce a receptor that is able to respond to iron starvation conditions. The internal *ESAG6* sequence is available from *T. brucei* Lister 927 on TriTrypDB [Tb927.9.15680]. It was assumed that 3'UTR would be 335 bp, the same as the expression site *ESAG6*, however the sequences did not align well. This may be because the sequence came from a different *T. brucei* strain, or because the genes and 3'UTRs had diverged. Primers were designed to amplify a 356 bp sequence of the internal *ESAG6* 3'UTR shown in **Figure 5.2**, which was ligated into the pRPa $\Delta$ -*flUC*<sup>mut</sup> reporter system and transfected into the 2T1 cell line. The sequence was generated with Lister 427 cells, so when the sequence in pRPa was sequenced, there were 4 polymorphisms identified (located in the same position) when aligned with the Lister 927 internal *ESAG6* 3'UTR, most likely differences between the strains. Unfortunately, the resulting clones became contaminated and were disposed of and there was not enough time to re-transfect. It therefore remains unknown

whether the internal *ESAG6* 3'UTR is able to respond to iron starvation. It is however, highly unlikely that the internal *ESAG6* would have maintained this response as there is very little conservation between the expression site *ESAG6* 3'UTR and the internal *ESAG6* 3'UTR and the putative motif identified above is not present in the internal *ESAG6* 3'UTR. It is also not clear if the non-expression site *ESAG6* copy is able to transcribe a functional protein that would be able to bind transferrin or respond to iron levels.



**Figure 5.2: Agarose gel of the internal copy of the *ESAG6* 3'UTR.** A band around 350 bp has been successfully amplified from Lister 427 BSF 2T1 *T. brucei* genomic DNA using primers designed to amplify a 356 bp sequence of the non-telomeric *ESAG6* 3'UTR using the GoTaq polymerase system (Promega) with an annealing temperature of 60°C.

## 5.7 Conclusion

Overall, truncation of the *ESAG6* and *ESAG7* 3'UTR does not result in complete loss of the iron starvation response. The truncations showed that the 3' end of the 3'UTR is most important for the upregulation of the transferrin receptor under iron starvation conditions, as when lost the expression of the luciferase reporter gene is similar to that under basal conditions. From the 3' end of the 3'UTR a 26 bp motif was identified through sequence alignments of *ESAG6* and *ESAG7* 3'UTRs from the 13 BESs. The 26 bp motif is conserved between the two 3'UTRs and between the different expression sites. Most importantly, the motif is only present in the truncations that maintained the ability to upregulate the expression of the luciferase gene

and therefore the transferrin receptor by a magnitude previously observed. RNA binding domains have been identified in the 3'UTR of a number of *T. brucei* genes that are responsible for upregulating expression of the gene in relation to changes in the environment. Most likely through RBP binding to the domain to stabilise the mRNA and/or increase translation efficiency, it is likely that this is how the transferrin receptor is upregulated. Of the domains identified, there is one present in the *VSG* 3'UTR and the *ESAG9* 3'UTR showing that genes in the expression site are regulated via RBPs. To confirm that this putative motif is an RNA binding domain, further experiments are required such as site directed mutagenesis or fusion of the motif to a different 3'UTR. It is also likely that this is not the only domain required for the *TbTfR* upregulation response. To confirm the identified patterns and improve significance the luciferase assays should be repeated with the current clones and others.

If the putative domain is recognised as essential for transferrin receptor upregulation and therefore parasite survival, it could be used to aid drug discovery. To make a successful drug that could be used to treat human African trypanosomiasis, it must be affordable. Drugs could target the 3'UTR by developing a peptide-mimetic that would directly mimic the motif to disrupt signalling, however, this would be too expensive to treat a neglected tropical disease. It would therefore be much more feasible to target components of the signalling pathway, but the components must first be identified. The domain would initially be used as a selection marker through the fusion of puromycin to the 3'UTR. This could then be used to identify components of the signalling pathway whereby the amount of drug resistance would be dependent on the iron starvation pathway via a genome wide RNAi screen (Baker *et al.*, 2011). This data could then be combined with existing genome-wide libraries to identify components of the signalling pathway. First, however, it must be shown that there is enough variation between high- and low-level expression to enable selection, if the motif is confirmed as essential.

Word Count: Approximately 34,000 words.

## **6.0 Acknowledgements**

I would like to acknowledge the guidance that I have received from my supervisors Dr Mick Urbaniak and Dr Paul McKean and the support provided by Dr Harsh Pawar within the laboratory. This support and guidance enabled me to successfully complete the research and write up for this study.

## 7.0 References

Abbaspour N., Hurrell R., Kelishadi R. (2014). Review on iron and its importance for human health. *Journal of Research in Medical Sciences*. V.19(2), pp.164-174.

Aitcheson N., Talbot S., Shapiro J., Hughes K., Adkin C., Butt T., Sheader K., Rudenko G. (2005). VSG switching in *Trypanosoma brucei*: antigenic variation analysed using RNAi in the absence of immune selection. *Molecular Microbiology*, v.57(6), pp.1608-1622.

Ajioka R.S., Phillups J.D., Kushner J.P. (2006). Biosynthesis of haem in mammals. *Biochimica et Biophysica Acta - Molecular Cell Research*, v.1763(7), pp.723-736.

Alsford S., Horn D. (2012). Cell-cycle regulation of VSG expression site silencing by histones and histone chaperones ASF1A and CAF-1b in *Trypanosoma brucei*. *Nucleic Acids Research*, v.40(20), pp.10150-10160.

Alsford S., Kawahara T., Glover L., Horn D. (2005). Tagging a *Trypanosoma brucei* rRNA locus improves stable transfection efficiency and circumvents inducible expression position effects. *Molecular and Biochemical Parasitology*, v.144(2), pp.142–8.

Ansorge I., Steverding D., Melville S., Hartmann C., Clayton C. (1999). Transcription of 'inactive' expression sites in African trypanosomes leads to expression of multiple transferrin receptor RNAs in bloodstream forms. *Molecular and Biochemical Parasitology*, v.101, pp. 81-94.

Baker N., Alsford S., Horn D. (2011). Genome-wide RNAi screens in African trypanosomes identify the nifurtimox activator NTR and the eflornithine transporter AAT6. *Molecular and Biochemical Parasitology*, v.176(1), pp.55-57.

Bartossek T., Jones N.G., Schafer C., Cvitkovic 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*, v.2, pp.1523-1532.

Benz C., Lo W., Fathallah N., Connor-Guscott A., Bennis 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*, v.13(12), e.020633.

Biebinger S., Helfert S., Steverding D., Ansorge I., Clayton C. (2003). Impaired dimerisation and trafficking of ESAG6 lacking a glycosyl-phosphatidylinositol anchor. *Molecular and Biochemical Parasitology*, v.132(2), pp.93-96.

Bitter W., Gerrits R., Kieft R., Borst P. (1998). The role of transferrin receptor variation in the host range of *Trypanosoma brucei*. *Letters to Nature*, v.391, pp.499-502.

Bouteille B., Buguet A. (2012). The detection and treatment of human African trypanosomiasis. *Research and Reports in Tropical Medicine*, v.3, pp.35-45.

Burkard G.S., Jutzi P., Roditi I. (2011). Genome-wide RNAi screens in bloodstream form trypanosomes identify drug transporters. *Molecular and Biochemical Parasitology*, v.175, pp.91-94.

Burri C., Yeramian P.D., Allen J.L., Merolle A., Serge K.K., *et al.* (2016). Efficacy, safety and dose of pafuramidine, a new oral drug for treatment of first stage sleeping sickness, in a phase 2a clinical study and phase 2b randomised clinical studies. *PLoS Neglected Tropical Diseases*, v. 10(2), e.0004362.

Capewell P., Atkins K., Weir W., Jamonneau V., Camara M., Clucas C., Swar N.R.K., Ngoyi D.M., Rotureau B., Garside P., Galvani A.P., Bucheton B., MacLeod. (2019). Resolving the apparent transmission paradox of African sleeping sickness. *PLoS Biology*, v.17(1), e.3000105.

Carabajo C.G., Cornell L.J., Madbouly Y., Lai Z., Yates P.A., Tinti M., Tiengwe C. (2021). Novel aspects of iron homeostasis in pathogenic bloodstream form *Trypanosoma brucei*. *PLoS Pathogens*, v.17(6), e.1009696.

Casey J.L., Hentze M.W., Koeller D.M., Caughman S.W., Rouault T.A., Klausner R.D., Harford J.B. (1988). Iron-responsive elements: regulatory RNA sequences that control mRNA levels and translation. *Science*, v.240, pp.924-928.

CDC. (2020). Parasites – African Trypanosomiasis. Online. [Accessed 7/7/21]. Available at: <https://www.cdc.gov/parasites/sleepingsickness/>

Cheng Y., Zak O., Aisen P., Harrison S.C., Walz T. (2004). Structure of the human transferrin receptor-transferrin complex. *Cell*. V.116(4), pp.565-576.

Clayton C. (2019). Regulation of gene expression in trypanosomatids: living with polycistronic transcription. *Open Biology*, v.9(6), e.190072.

Cully D.F., Ip H.S., Cross G.A. (1985). Coordinate transcription of variant surface glycoprotein genes and an expression site associated gene family in *Trypanosoma brucei*. *Cell*, v.42(1), pp. 173-182.

Das A., Banday M., Bellofatto V. (2008). RNA polymerase transcription machinery in trypanosomes. *Eukaryotic Cell*, v.7(3), pp.429-434.

Dickie E.A., Giordani F., Gould M.K., Maser 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*, v.5(1), doi.10.3390/tropicalmed5010029.

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*, v.131(3), pp.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*, v. 117(7), pp.1105-1115.

Erben E.D., Fadda A., Lueong s., Hoheisel J.D., Clayton C. (2014). A genome-wide tethering screen reveals novel potential post-transcriptional regulators in *Trypanosoma brucei*. *PLoS Pathogens*, v.10(6), e.1004178.

Faria J., Glover L., Hutchinson S., Boehm C., Field M.C., Horn D. (2019). Monoallelic expression and epigenetic inheritance sustained by *Trypanosoma brucei* variant surface glycoprotein exclusion complex. *Nature Communications*, v.10(1), e.3023.

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

Ganz T., Nemeth E. (2006). Regulation of iron acquisition and iron distribution in mammals. *Biochimica et Biophysica Acta - Molecular Cell Research*, v.1763(7), pp.690-699.

Gao J., Zhou Q., Wu D., Chen L. (2021). Mitochondrial iron metabolism and its role in diseases. *Clinica Chimica Acta*. V.513, pp.6-12.

Gerrits H., Mußmann R., Bitter W., Kieft R., Borst P. (2002). The physiological significance of transferrin receptor variations in *Trypanosoma brucei*. *Molecular and Biochemical Parasitology*, v.119, pp.237-247.

Giometti B., Gallo P., Tavolato B. (1993). Transferrin receptors in the central nervous system. *Methods in Neurosciences*, v.11, pp.122-134.

Glover L., Hutchinson S., Alsford S., Horn D. (2016). VEX1 controls the allelic exclusion required for antigenic variation in trypanosomes. *PNAS*, v.113(26), pp.7225-7230.

Grab D. J., Wells C.W., Shaw M.K., Webster P., Russo D.C.W. (1992). Endocytosed transferrin in African trypanosomes is delivered to lysosomes and may not be recycled. *European Journal of Cell Biology*, v.59, pp.398-404.

Gunzl A. (2010). The pre-mRNA splicing machinery of trypanosomes: complex or simplified?. *Eukaryotic Cell*, v.9(8), pp.1159-1170.

Gunzl A. Bruderer T., Laufer G., Schimanski B., Tu L.C., 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*, v.2(3), pp.542-551.

Gunzl A., Kirkham J.K., Nguyen T. N., Badjatia N., Park S. H. (2015). Mono-allelic VSG expression by RNA polymerase I in *Trypanosoma brucei*: Expression site control from both ends? *Gene*, v.556(1), pp.68-73.

Haenni S., Renggli C.K., Fragoso C.M., Oberle M., Roditi I. (2006). The *procyclin-associated genes* of *Trypanosoma brucei* are not essential for cyclical transmission by tsetse. *Molecular and Biochemical Parasitology*, v.150(2), pp.144-156.

Hentze M.W., Muckenthaler M.U., Andrews N.C. (2004). Balancing acts: molecular control of mammalian iron metabolism. *Cell*, v.117(3), pp.285-297.

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., Taylor J.E., Walker D., White B., Young R., Cross G.A.M., Rudenko G., Barry J.D., Louis E.J., Berriman M. (2008). Telomeric expression sites are highly conserved in *Trypanosoma brucei*. *PLoS ONE*, v.3(10), e.3527.

Higgins M.K., Carrington M. (2014). Sequence variation and structural conservation allows development of novel function and immune evasion in parasite surface families. *Protein Science*, v.23(4), pp.354-365.

Higgins M.K., Tkachenko O., Brown A., Reed J., Raper J., Carrington M. (2013). Structure of the trypanosome haptoglobin-haemoglobin receptor and implications for nutrient uptake and innate immunity. *PNAS*, v.110(5), pp.1905-1910.

Hirumi H., Hirumi K. (1989). Continuous cultivation of *Trypanosoma brucei* blood stream forms in a medium containing a low concentration of serum protein without feeder cell layers. *Journal of Parasitology*, v.75(6), pp.985-989.

Jackson A.P., Allison H.C., Barry J.B., Field M.C., Hertz-Fowler C., Berriman M. (2013). A cell-surface phylome for African trypanosomes. *PLoS Neglected Tropical Diseases*, v.7(3), e.2121.

Jackson D.G., Owen M.J., Voorheis H.P. (1985). A new method for the rapid purification of both the membrane-bound and released forms of the variant surface glycoprotein from *Trypanosoma brucei*. *Biochemical Journal*, v.230(1), pp.195-202.

Kabriri M., Steverding D. (2021). *Trypanosoma brucei* transferrin receptor: functional replacement of the GPI anchor with a transmembrane domain. *Molecular and Biochemical Parasitology*, v.242, e.111361.

Kariuki C. K., Stijlemans B., Magez S. (2019). The trypanosomal transferrin receptor of *Trypanosoma brucei* a review. *Tropical Medicine and Infectious Disease*, v.4(4), pp.126.

Kassem A., Pays E., Vanhamme L. (2014). Transcription is initiated on silent variant surface glycoprotein expression sites despite monoallelic expression in *Trypanosoma brucei*. *PNAS*, v.111(24), pp.8943-8948.

Kennedy P. G. E. (2013). Clinical features, diagnosis and treatment of human African trypanosomiasis (sleeping sickness). *The Lancet Neurology*, v.12(2), pp.186-194.

Kleven M.D., Jue S., Enns C.A. (2018). The transferrin receptors TfR1 and TfR2 bind transferrin through differing mechanisms. *Biochemistry*, v.57(9), pp.1552-1559.

Kwabata H., Yang R., Hirama T., Vuong P.T., Kwano S., Gombart A.F., Koeffler P. (1999). Molecular cloning of transferrin receptor 2: a new member of the transferrin receptor-like family. *Journal of Biological Chemistry*. V.274(30), pp.20826-20832.

Lee S.H., Stephens J.L., Englund P.T. (2007). A fatty-acid synthesis mechanism specialised for parasitism. *Nature Reviews Microbiology*, v.5, pp.287-297.

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. *Nature (Scientific Reports)*, v.8, e.10922.

Lueong S., Merce C., Fischer B., Hoheisel J.D., Erben E.D. (2016). Gene expression regulatory networks in *Trypanosoma brucei*: insights into the role of the mRNA-binding proteome. *Molecular Microbiology*, v.100(3), pp.457-471.

Mach J., Tachezy J., Sutak R. (2013). Efficient iron uptake via a reductive mechanism in procyclic *Trypanosoma brucei*. *Journal of Parasitology*, v.99(2), pp.363-364.

Matthews K.R. (2005). The developmental cell biology of *Trypanosoma brucei*. *Journal of Cell Science*, v.118(2), pp.283-290.

Melkert A., Wormald M.R., Ferguson M.A.J. (2012). Modelling of the N-glycosylated transferrin receptor suggests how transferrin binding can occur within the surface coat of *Trypanosoma brucei*. *PLoS Pathogens*, v.8(4), e.1002618.

Monk S.L., Simmonds P., Matthews K.R. (2013). A short bifunctional element operates to positively or negatively regulate *ESAG9* expression in different developmental forms of *Trypanosoma brucei*. *Journal of Cell Science*, v.126(10), pp.2294-2304.

Mony B.M., MacGregor P., Ivens A., Rojas F., Cowton A., Young J., Horn D., Matthews K. (2014). Genome-wide dissection of the quorum sensing signalling pathway in *Trypanosoma brucei*. *Nature*, v.505, pp.681-685.

Mugnier M.R., Stebbins C.E., Papavasiliou F.N. (2016). Masters of disguise: Antigenic variation and the VSG coat in *Trypanosoma brucei*. *PLoS Pathogens*, v.12(9), e.1005784.

Mussmann R., Engstler M., Gerrits H., Kieft R., et al. (2004). Factors affecting the level and localisation of the transferrin receptor in *Trypanosoma brucei*. *The Journal of Biological Chemistry*, v.279(39), pp.40690-40698.

Mussmann R., Janssen H., Calafat J., Engstler M., Ansorge I., Clayton C., Borst P. (2003). The expression level determines the surface distribution of the transferrin receptor in *Trypanosoma brucei*. *Molecular Microbiology*, v.47(1), pp.23-35.

Navarro M., Gull K. (2001). A poll transcriptional body associated with VSG mono-allelic expression in *Trypanosoma brucei*. *Nature*, v.414, pp.759-763.

Pan African Tsetse and Trypanosomiasis Eradication Campaign (PATTEC). (2012). Online. [Accessed: 6/7/21]. Available at: [https://au.int/sites/default/files/documents/32884-doc-how the pattec initiative can be supported.pdf](https://au.int/sites/default/files/documents/32884-doc-how%20the%20pattec%20initiative%20can%20be%20supported.pdf)

Pays E. (2006). The variant surface glycoprotein as a tool for adaptation in African trypanosomes. *Microbes and Infection*, v.8(3), pp.930-937.

Pays E., Nolan D.P. (2021). Genetic and immunological basis of human African trypanosomiasis. *Current Opinion in Immunology*, v.72, pp.13-20.

Programme Against African Trypanosomiasis (PAAT). (2021). Online. [Accessed 6/7/21]. Available at: <http://www.fao.org/paat/the-programme/our-structure/en/>

Pupim L.B., Martin C.J., Ikizler A. (2013). Nutritional Management of Renal Disease – Assessment of Protein and Energy Nutritional Status. Academic Press. Third Edition. pp.137-158.

Rico E., Ivens A., Glover L., Horn D., Matthews K.R. (2017). Genome-wide RNAi selection identifies a regulator of transmission site-enriched gene families and cell-type differentiation in *Trypanosoma brucei*. *PLoS Pathogens*, v.13(3), e. 1006279.

Ridewood S., Ooi C., 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 *Trypanosoma brucei*. *Molecular Microbiology*, v.106(4), pp.614-634.

Rojas F., Silvester E., Young J., Milne R., Tettey M., Houston D.R., Walkinshaw M.D., Perez-Pi I., Auer M., Denton H., Smith T.K., Thompson J., Matthews K.R. (2019). Oligopeptide signalling through *TbGPR89* drives trypanosome quorum sensing. *Cell*, v.176(2), pp.306-317.

Rupani D.N., Connell G.J. (2016). Transferrin receptor mRNA interactions contributing to iron homeostasis. *RNA*, v.22(8), pp.1271-1282.

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*, v.78(1), pp.75-86.

Salmon D., Hanocq-Quertier J., Paturiaux-Hanocq F., Pays A., Tebabi P., Nolan D.P., Michel A., Pays E. (1997). Characterisation of the ligand-binding site of the transferrin receptor in *Trypanosoma brucei* demonstrates a structural relationship with the N-terminal domain of the variant surface glycoprotein. *The EMBO Journal*, v.16(24), pp.7272-7278.

Salmon D., Paturiaux-Hanocq F., Poelvoorde P., Vanhamme L., Pays E. (2005). *Trypanosoma brucei*: growth differences in different mammalian sera are not due to the species-specificity of transferrin. *Experimental Parasitology*, v.109(3), pp.188-194.

Schwartz K.J., Peck R.F., Tazeh N.N., Bangs J.D. (2005). GPI valence and the fate of secretory membrane proteins in African trypanosomes. *Journal of Cell Science*, v.118(23), pp.5499-5511.

Silvester E., McWilliam K. R., Matthews K. R. (2017). The cytological events and molecular control of life cycle development of *Trypanosoma brucei* in the mammalian bloodstream. *Pathogens*, v.6(3), doi.10.3390/pathogens6030029.

Soysa R., Carter N.S., Yates P.A. (2014). A dual luciferase system for analysis of post-transcriptional regulation of gene expression in *Leishmania*. *Molecular and Biochemical Parasitology*, v.195(1), pp.1–5.

Steverding D. (2000). The transferrin receptor of *Trypanosoma brucei*. *Parasitology International*, v.48(3), pp.191-198.

Steverding D. (2003). The significance of transferrin receptor variation in *Trypanosoma brucei*. *Trends in Parasitology*, v.19(3), pp.125-127.

Steverding D., Stierhof Y.D., Fuchs H., Tauber R., Overath P. (1995). Transferrin-binding protein complex is the receptor for transferrin uptake in *Trypanosoma brucei*. *The Journal of Cell Biology*, v.131, pp.1173-1182.

Stijlemans B., Beschin A., Magez S., Van Ginderachter J. A., de Baetselier P. (2015). Iron homeostasis and *Trypanosoma brucei* associated immunopathogenicity development: a battle/quest for iron. *BioMed Research International*, v.2015, e.819389.

Tanaka T., Abe Y., Inoue N., Kim W.S., Kumura H., Nagasawa H., Igarashi I., Shimazaki K. (2004). The detection of bovine lactoferrin binding protein on *Trypanosoma brucei*. *The Journal of Veterinary Medical Science*, v.66(6), pp.619-625.

Tiengwe C., Muratore K.A., Bangs J.D. (2016). Variant surface glycoprotein, transferrin receptor and ERAD in *Trypanosoma brucei*. *Cell Microbiology*, v.18(11), pp.1673-1688.

Tiengwe C., Bush P.J., Bangs J.D. (2017). Controlling transferrin receptor trafficking with GPI-valence in bloodstream stage African trypanosomes. *PLoS Pathogens*, v.13(5), e.1006366.

Trevor C.E., Gonzalez-Munoz A.L., Macleod O.J.S., Woodcock P.G., Rust S., *et al.* (2019). Structure of the trypanosome transferrin receptor reveals mechanisms of ligand recognition and immune evasion. *Nature Microbiology*, v.4(12), pp.2074-2081.

Vanhamme L., Poelvoorde P., Pays A., Tebabi P., Xong H.V., Pays E. (2000). Differential RNA elongation controls the variant surface glycoprotein gene expression sites of *Trypanosoma brucei*. *Molecular Microbiology*, v.36(2), pp.328-340.

Vincent I. M., Daly R., Courtioux B., Cattanach A.M., Bieler S., Ndung'u J. M., Bisser S., Barrett M. P. (2016). Metabolomics identifies multiple candidate biomarkers to diagnose and stage human African Trypanosomiasis. *PLoS Neglected Tropical Disease*, v.10(12), e.0005140.

Wamwiri F.N., Changasi R.E. (2016). Tsetse flies (*Glossina*) as vectors of human African trypanosomiasis: a review. *BioMed Research International*, v.2016, e.6201350.

WHO. (2021). Trypanosomiasis, human African (sleeping sickness). [Accessed on 5/7/21]. Online. Available at: [https://www.who.int/news-room/fact-sheets/detail/trypanosomiasis-human-african-\(sleeping-sickness\)](https://www.who.int/news-room/fact-sheets/detail/trypanosomiasis-human-african-(sleeping-sickness))

Xong H.V., Vanhamme L., Chamekh M., Hamers R., De Baetselier P., Pays E. (1998). A VSG expression site-associated gene confers resistance to human serum in *Trypanosoma rhodesiense*. *Cell*, v.95(6), pp.839-846.

Young R., Taylor J.E., Kurioka A., Becker M., Louis E.J., Rudenko G. (2008). Isolation and analysis of the genetic diversity of repertoires of VSG expression site containing telomeres from *Trypanosoma brucei gambiense*, *T. b. brucei* and *T. equiperdum*. *BMC Genomics*, v.9, doi.10.1186/1471-2164-9-385.

Zhang D., Ghosh M.C., Rouault T.A. (2014). The physiological functions of iron regulatory proteins in iron homeostasis – an update. *Frontiers in Pharmacology*, v.5(124), doi.10.3389/fphar.2014.00124.

Zimmerman 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*, v.13(4), e.1006324.

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*, v.3, pp.295-301.

## 8.0 Appendix

BAD AVG GOOD

cons : 97

```
BES1 .ESAG6-3' UTR -----
BES2 .ESAG6-3' UTR -----
BES3 .ESAG6-3' UTR -----
BES4 .ESAG6-3' UTR -----
BES5 .ESAG6-3' UTR -----
BES7 .ESAG6-3' UTR -----
BES11 .ESAG6-3' UTR -----
BES12 .ESAG6-3' UTR -----
BES13 .ESAG6-3' UTR -----
BES14 .ESAG6-3' UTR -----
BES15 .ESAG6-3' UTR -----
BES17 .ESAG6-3' UTR -----
BES1 .ESAG7-3' UTR AATGGAGTAAAGGCGAATTCAACTATACTGCAGAACC GGTCCTGGACCTTTCACGGTAGCGG
BES2 .ESAG7-3' UTR AATGGAGTAAAAGCGAATTCAACTATACTGCAGAACC GGTCCTGGGGACATTTACGGGAGCGG
BES3 .ESAG7-3' UTR AATGGAGTAAAAACGAATCTAACTATACTGCAGAGCCTGTCCGTGGACCTTTAACGGGAGCGG
BES4 .ESAG7-3' UTR AATGGAGTAAAAGCGAATTCAACTATACTGCAGAACC GGTCCTGGGGACATTTACGGGAGCGG
BES5 .ESAG7-3' UTR AATGGAGTAAAAGCTAATTCAACTATACTGCAGAACC GGTCCTGGGGACATTTACGGGAGCGG
BES8 .ESAG7-3' UTR AATGGAGTAAAAGCGAATTCAACTATACTGCAGAACC GGTCCTGGGGACATTTACGGGAGCGG
BES10 .ESAG7-3' UTR AATGGAGTAAAATCGAATCTAACTATACTGCAGAGCCTGTCCGTGGACCTTTAACGGGAGCGG
BES11 .ESAG7-3' UTR -----AAGCAAATTCAACTATACTGTAGAGCCTGTCTGGGGACCTTTCACGGGAGCGG
BES12 .ESAG7-3' UTR AATGGAGTAAAAGCGAATTTAACTATACTGCAGAACC GTCCGTGGACCTTTCACGGGAGCGG
BES13 .ESAG7-3' UTR AATGGAGTAAAAGCGAATCTAACTATACTGTAGAACCTGTCCGTGGACCTTTCACGGGAGCGG
BES14 .ESAG7-3' UTR AATGAAGTAAAAGCGAATCTAACTATACTGCAGAACC GTCCGTGGACCTTTCACGGGAGCGG
BES15 .ESAG7-3' UTR AATGGAGTAAAAGCGAATTTAACTATACTGCAGAACC GTCCGTGGACCTTTCACGGGAGCGG
BES17 .ESAG7-3' UTR AATGGAGTAAAACGAATCTAACTATACTGCAGAGCCTGTCCGTGGACCTTTCACGGTAGCGG
```

cons

```
BES1 .ESAG6-3' UTR -----
BES2 .ESAG6-3' UTR -----
BES3 .ESAG6-3' UTR -----
BES4 .ESAG6-3' UTR -----
BES5 .ESAG6-3' UTR -----
BES7 .ESAG6-3' UTR -----
BES11 .ESAG6-3' UTR -----
BES12 .ESAG6-3' UTR -----
BES13 .ESAG6-3' UTR -----
BES14 .ESAG6-3' UTR -----
BES15 .ESAG6-3' UTR -----
BES17 .ESAG6-3' UTR -----
BES1 .ESAG7-3' UTR GGTCCAACACAGTAGCAGTACATTTGAGTGTTCCTACCGCTGCACTTTGTTTTTCAGTTTTAT
BES2 .ESAG7-3' UTR GGTCCAACGCGGCAGCAGTACATTTGAGTGTTCCTACCGCTGCACTTTGTTTTTCAGTTTTAT
BES3 .ESAG7-3' UTR GGGCCAACGCGGCAGCAATACATTTGAGTGTTCCTACCGCTGCACTTTGTCTGTTTCAGTTTTAT
BES4 .ESAG7-3' UTR GGTCCAACGCGGCAGCAGTACATTTGAGTGTTCCTACCGCTGCACTTTGTTTTTCAGTTTTAT
BES5 .ESAG7-3' UTR GGTCCAACGCGGCAGCAGTACATTTGAGTGTTCCTACCGCTGCACTTTGTTTTTCAGTTTTAT
BES8 .ESAG7-3' UTR GGTCCAACGCGGCAGCAGTACATTTGAGTGTTCCTACCGCTGCACTTTGTTTTTCAGTTTTAT
BES10 .ESAG7-3' UTR GGGCCAACGCGGCAGCAATACATTTGAGTGTTCCTACCGCTGCACTTTGTCTGTTTCAGTTTTAT
BES11 .ESAG7-3' UTR GGTCCAACGCGGCAGCAATACATTCGAGTGTATCTACCGCTGCACTCTGTTGTTTCAGTTTTAT
BES12 .ESAG7-3' UTR GGGCCAATGCGGCAGCAATACATTTGAGTGTTCCTACCGCTGCACTTTGTTGTTTCAGTTTTAT
BES13 .ESAG7-3' UTR GGTCCAACACAGTAGCAGTACATTTGAGTGTTCCTACCGCTGCACTCTGTTGTTTCAGTTTTAT
BES14 .ESAG7-3' UTR GGTCCAACACAGTAGCAGTACATTTGAGTGTTCCTACCGCTGCACTTTGTCTGTTTCAGTTTTAT
BES15 .ESAG7-3' UTR GGTCCAACACAGTAGCAGTACAGTTCGAGTGTTCCTACCGCTGCACTTTGTTGTTTCAGTTTTAT
BES17 .ESAG7-3' UTR GGGCCAACGCGGCAGCAATACATTTGAGTGTTCCTACCGCTGCACTTTGTCTGTTTCAGTTTTAT
```

cons

BES1 .ESAG6-3' UTR -----GGGAAGGATGCGACCGAAACTGCGCTGCTTAGCGTGAAAGATTATG  
 BES2 .ESAG6-3' UTR -----AGGAAGGATGCGACAGAAGCTGCGCTGCTTAGTGTGAAAGATTATG  
 BES3 .ESAG6-3' UTR -----AGGAAGGATGCGACCGAAACTGCGCTGCTTCGCGTGAAAGATTATG  
 BES4 .ESAG6-3' UTR -----AGGAAGGATGCGACAGAAGCTGCGCTGCTTAGCGTGAAAGATTATG  
 BES5 .ESAG6-3' UTR -----AGGAAGGATGCGACAGAAGCTGCGCTGCTTAGCGTGAAAGATTATG  
 BES7 .ESAG6-3' UTR -----GGGAAGGATGCGACCGAAACTGCGCTGTTTCGCGTGAATGATTATG  
 BES11 .ESAG6-3' UTR -----GGGAAGGATGCGACCGAAACTGCGCTGCTTCGCGTGAATGATTATG  
 BES12 .ESAG6-3' UTR -----GGGAAGGATGCGACCGAAACTGCGCTGCTTCGCGTGAAAGATTGTG  
 BES13 .ESAG6-3' UTR -----GGGAAGGATGCGACCGAAACTGCGCTGCTTAGCGTGAAAGATTATG  
 BES14 .ESAG6-3' UTR -----AGGAAGGATGCGACCGAAACTGCGCTGTTTCGCGTGAAAGATTATG  
 BES15 .ESAG6-3' UTR -----GGGAAGGATGCGACCGAAACTGCGCTGCTTAGCGTGAAAGATTATG  
 BES17 .ESAG6-3' UTR -----AGGAAGGATGCGACCGAAACTGCGCTGCTTCGCGTGAAAGATTATG  
 BES1 .ESAG7-3' UTR TGTTGGGAGTGCTGTGAAGGAAGGATGCGACAGAAGCTGCGCTGCTTAGCGTGAAAGATTATG  
 BES2 .ESAG7-3' UTR TGTTGGGAGTGCTGTGAAGGAAGGATGCGACAGAAGCTGCGCTGCTTAGCGTGAAAGATTATG  
 BES3 .ESAG7-3' UTR TGTTGGGAGTGCTGTGAAGGAAGGATGCGACCGAAACTGCGCTGCTTCGCGTGAAAGATTATG  
 BES4 .ESAG7-3' UTR TGTTGGGAGTGCTGTGAAGGAAGGATGCGACAGAAGCTGCGCTGCTTAGCGTGAAAGATTATG  
 BES5 .ESAG7-3' UTR TGTTGGGAGTGCTGTGAAGGAAGGATGCGACAGAAGCTGCGCTGCTTAGCGTGAAAGATTATG  
 BES8 .ESAG7-3' UTR TGTTGGGAGTGCTGTGAAGGAAGGATGCGACAGAAGCTGCGCTGCCTAGCGTGAAAGATTATG  
 BES10 .ESAG7-3' UTR TGTTGGGAGTGCTGTGAAGGAAGGATGCGACCGAAACTGCGCTGCTTCGCGTGAAAGATTATG  
 BES11 .ESAG7-3' UTR TGTTGGGAGTGCTGTGAGGGAAGGATGCGACCGAAACTGCGCTGCTTCGCGTGAATGATTATG  
 BES12 .ESAG7-3' UTR TGTTGGGAGTGCTGTGAGGGAAGGATGCGACCGAAACTGCGCTGCTTCGCGTGAAAGATTATG  
 BES13 .ESAG7-3' UTR TGTTGGGAGTGCTGTGAGGGAAGGATGCGACCGAAACTGCGCTGTTTCGCGTGAATGATTATG  
 BES14 .ESAG7-3' UTR TGTTGGGAGTGCTGTGAAGGAAGGATGCGACCGAAACTGCGCTGTTTCGCGTGAAAGATTATG  
 BES15 .ESAG7-3' UTR TGTTGGGAGTGCTGTGAGGGAAGGATGCGACAGAAGCTGCGCTGTTTCGCGTGAAAGATTATG  
 BES17 .ESAG7-3' UTR TGTTGGGAGTGCTGTGAAGGAAGGATGCGACCGAAACTGCGCTGCTTCGCGTGAAAGATTATG

cons \*\*\*\*\* \*\* \* \*\*\*\*\* \* \* \*\*\*\*\* \*\* \*\*

BES1 .ESAG6-3' UTR GTAATGGAGGGTTGGGA-AAG-ATTGGGGAAACAAATACCTATTTCTTTTATTTGGGGGAACA  
 BES2 .ESAG6-3' UTR GTAATGGAGGGTTGTGA-AAG-ATTGGGGGAACAAAAACCTATTTCTTTTATTTGGGGGAACA  
 BES3 .ESAG6-3' UTR GTAATGGAGGGTTGGGA-AAG-ATCGGTGAAACAAATACCTATTTCTTTTATTTGGGGGAACA  
 BES4 .ESAG6-3' UTR GTAATGGAGGGTTGTGA-AAG-ATTGGGGGAACAAAAACCTATTTCTTTTATTTGGGGGAACA  
 BES5 .ESAG6-3' UTR TTAATGGAGGGTTGTGA-AAG-ATTGGGGGAACAAAAACCTATTTCTTTTATTTGGGGGAACA  
 BES7 .ESAG6-3' UTR ATAATGGAGGGTTGGGA-AAG-ATTGGGGGAACAAATACCTATTTCTTTTATTTGTGGGAACA  
 BES11 .ESAG6-3' UTR ATAATGGAGGGTTGGGA-AAG-ATCGGGGAACAAATACCTATTTCTTTTATTTGGGGGAACA  
 BES12 .ESAG6-3' UTR GTAATGGAGGGTTCGGA-AAG-ATTGGGGGAACAAAAACCTATTTCTTTTATTTGTGGGAACA  
 BES13 .ESAG6-3' UTR GTAATGGAGGGTTGGGA-AAG-ATTGGGGGAACAAAAACCTATTTCTTTTATTTGGGGGAACA  
 BES14 .ESAG6-3' UTR GTAATGGAGGGTTGGGA-AAG-ATTGGGGGAACAAAAACCTATTTCTTTTATTTGGGGGAACA  
 BES15 .ESAG6-3' UTR GTAATGGAGGGTTGGGA-AAG-ATTGGGGGAACAAAAACCTATTTCTTTTATTTGGGGGAACA  
 BES17 .ESAG6-3' UTR GTAATGGAGGGTTGGGA-AAG-ATCGGTGAAACAAATACCTATTTCTTTTATTTGGGGGAACA  
 BES1 .ESAG7-3' UTR GTAATGGAGGGTTGTGA-AAG-ATTGGGGGAACAAAAACCTATTTCTTTTATTTGGGGGAACA  
 BES2 .ESAG7-3' UTR ATAATGGAGGGTTGGGA-AAG-ATCGGGGAACAAATACCTATTTCTTTTATTTGGGGGAACA  
 BES3 .ESAG7-3' UTR GTAATGGAGGGTTGGGA-AAG-ATCGGTGAAACAAATACCTATTTCTTTTATTTGGGGGAACA  
 BES4 .ESAG7-3' UTR GTAATGGAGGGTTGTGA-AAG-ATTGGGGGAAGAGATACCTATTTCTTTTACTTGGGGGAACA  
 BES5 .ESAG7-3' UTR GTAATGGAGGGTTGTGA-AAG-ATTGGGGGAACAAAAACCTATTTCTTTTATTTGGGGGAACA  
 BES8 .ESAG7-3' UTR GTAATGGAGGGTTGGGA-AAG-ATTGGGGAAACAGATACCTATTTCTTTTATTTGGGGGAACA  
 BES10 .ESAG7-3' UTR GTAATGGAGGGTTGGGA-AAG-ATCGGTGAAACAAATACCTATTTCTTTTATTTGGGGGAACA  
 BES11 .ESAG7-3' UTR ATAATGGAGGGTTGGGA-AAG-ATCGGGGAACAAATACCTATTTCTTTTATTTGGGGGAACA  
 BES12 .ESAG7-3' UTR GTAATGGAGGGTTGGGA-AAG-ATTGGGGGAACAAAAACCTATTTCTTTTATTTGTGGGAACA  
 BES13 .ESAG7-3' UTR ATAATGGAGGGTTGGGA-AAG-ATTGGGGGAACAAATACCTATTTCTTTTATTTGTGGGAACA  
 BES14 .ESAG7-3' UTR GTAATGGAGGGTTGGGA-AAG-ATTGGGGGAACAAAAACCTATTTCTTTTATTTGGGGGAACA  
 BES15 .ESAG7-3' UTR ATAATGGAGGGTTGGGA-AAG-ATTGGGGGAACAAAAACCTATTTCTTTTATTTGGGGGAACA  
 BES17 .ESAG7-3' UTR GTAATGGAGGGTTGGGA-AAG-ATCGGTGAAACAAATACCTATTTCTTTTATTTGGGGGAACA

cons \*\*\*\*\* \* \*\* \* \*\* \* \* \* \* \* \*\*\*\*\* \* \* \* \* \*\*\*\*\*

BES1 .ESAG6-3' UTR AATGGGC AAAAGTAACGTAAGTTTCCAGTGGGAGTGGTATG TGTGTGT--GTATGGGGCTGGC  
 BES2 .ESAG6-3' UTR AATGGGC AAAAGTAACGTAAGTTTCCAGTGGGAGTGGTATG TGTGTGT--GTATGGGGCTGGC  
 BES3 .ESAG6-3' UTR AATGGG AAAAGTAACGTAAGTTTCCAGCGGAGTGG--T--ATG TGTGTGTATGGGGCTGGC  
 BES4 .ESAG6-3' UTR AATGGGC AAAAGTAACGTAAGTTTCCAGTGGGAGTGGTATG TGTGTGT--GTATGGGGCTGGC  
 BES5 .ESAG6-3' UTR AATGGGC AAAAGTAACGTAAGTTTCCAGTGGGAGTGGTATG TGTGTGTGT--ATGGGGCTGGC  
 BES7 .ESAG6-3' UTR AATGGGC AAGTAACGTAAGTTTCTAGCGGAGTGG----TGTGTGT--CATGGGACGGGT  
 BES11 .ESAG6-3' UTR AATGGGC AAAAGTAACGTAAGTTTCCAGCGGTAGTGG----TGTGTGT--TATGGGGCTGGC  
 BES12 .ESAG6-3' UTR AATGGG AAAAGTAACGTAAGTTTCTAGCGGAGTGG--T--ATG TGTGTGTATGGGGCTGGC





BES17.ESAG7-3'UTR -----

cons



**Appendix 8.1: Alignment of the *ESAG6* and *ESAG7* 3'UTR from all BESs the genes are transcribed in.** The 335 bp *ESAG6* 3'UTR is aligned with the 492 bp *ESAG7* 3'UTR show a 97% conservation. Alignment created on TCoffee.