Characterisation and functional analysis of the developmentally regulated Expression Site Associated Gene 9 family in *Trypanosoma brucei*

Eleanor M. Barnwell

PhD – The University of Edinburgh – 2009

Abstract

Trypanosoma brucei is a protozoan parasite that is the causative agent of sleeping sickness in sub-Saharan Africa. *T. brucei* has a complex life cycle involving passage between a mammalian host and the tsetse fly. The parasite evades the mammalian immune system via expression of Variant Surface Glycoprotein (VSG) on the cell surface. VSG genes are expressed at telomeric expression sites and at these sites are a number of Expression Site Associated Genes (ESAGs). One unusual ESAG, ESAG9*,* is developmentally regulated: RNA for these genes accumulates during the transition from slender to stumpy cells in the mammalian bloodstream and cell-associated protein is only detected transiently in stumpy and differentiating cells. Transgenic cell lines were generated which ectopically express one or more members of the ESAG9 gene family. Biochemical and cytological analyses using these cell lines indicated that some members of this family are glycosylated and GPI-anchored, and also that one gene, ESAG9-K69, is secreted. ESAG9-K69 is also secreted by wild-type stumpy parasites. *In vivo* experiments with tsetse flies did not conclusively show whether ESAG9 proteins play a role in the establishment of a tsetse fly mid-gut infection by transgenic trypanosomes. However, *In vivo* and *ex vivo* experiments using the mouse model of trypanosomiasis indicated that expression of ESAG9 proteins may alter parasitaemia in the mouse and results in a significant decrease in the proportion of CD4+ T cells in the mouse spleen.

Declaration

The cDNA subtraction selection experiment and subsequent Northern Blot analysis was carried out by Professor Keith Matthews.

Subsequent to the above exception, I declare that the material presented in this thesis is my own work and has not been submitted to any other University or for any other degree or qualification.

Eleanor M. Barnwell, Edinburgh University, March 2009

Acknowledgements

First and foremost, I would like to thank Keith Matthews for his supervision and his continued support, for giving me such an interesting project to work on in the first place, and for having faith that the antibody was worth it. I would also like to thank everybody in the Matthews lab (past and present), who are too many to mention, for being great colleagues, and more importantly, great friends. They are proof of the theory that molecular biologists make excellent bakers.

Alvaro Acosta-Serrano was an important collaborator on this project and I would like to thank him for his many hours helping me dissect tsetse flies and for his help and advice with other assays. I would like to thank Mark Blaxter for being my second supervisor; Craig Lapsley, for helping me dissect the tsetse flies; Andy Bell for helping me to set up the tsetse fly qPCR assay; Deborah Hall, for doing the mouse work; and Rick Maizels and Simmi Mahajan, for a fruitful collaboration to do the FACs analysis.

I would like to thank Terry Preston and Vaughan Southgate for inspiring me to be a parasitologist, and Keith Gull and Michael Ginger for teaching me how to actually go about being one.

My time studying in Edinburgh has been both a scientific and a personal journey. For accompanying me on this journey I would like to thank the following people: everyone in Ashworth, especially my PhD cronies (you know who you are); all my flat-mates past and present; Ruth, who was one of the first people that I met, and who has been a great office mate; and Tabitha, Helen, Laura and Adel, just for being the very special people that they are.

I would like to thank my family, whom I love very dearly, and who believe in me more than I believe in myself.

Abbreviations

ANOVA	Analysis of variance
BARP	Brucei alanine rich protein
cAMP	Cyclic adenosine monophosphate
CCA	Citrate/cis-Aconitate
CD4	Cluster of differentiation 4
Con A	Concanavalin A
CRD	Cross-reacting domain
DAPI	4,6-diamidine-2, phenylindole
DiD1	Defective in differentiation cell line 1
DNA	Deoxyribonucleic acid
EATRO	East African Trypanosomiasis Research Organisation
ECL	Enhanced chemiluminescence
EDTA	Ethylenediaminetetraacetic acid
EP	Glutamic acid proline procyclin
ES	Expression site
ESAG	Expression site associated gene
FACS	Fluorescence activated cell sorting
FCS	Foetal calf serum
FSC	Forward scatter
FITC	Fluorescein isothiocyanate
g	Gravities
GmmRel	*Glossina morsitans morsitans* Relish
GPEET	Glycine–proline–glutamic acid–glutamic acid–threonine proclclin
GPI	glycosylphosphatidylinositol
GPI-PLC	glycosylphosphatidylinositol-specific phospholipase C
GPS	Guinea pig serum
Hp-Hb	Haptoglobin-Hemoglobin
HRP	Horseraddish peroxidase
iNOS	Inducible nitric oxide synthase
IFN-γ	Interferon gamma
IP	Immunoprecipitation
K	Thousand
kb	Kilobase-pair
kDa	Kilodaltons
kDNA	Kinetoplast DNA
LB	Luria Broth
KO	Knockout
MASP	Mucin-associated surface protein
MIF	Macrophage inhibitory factor
mm	Millimeter
mM	Millimolar
mRNA	Messenger RNA
NO	Nitric oxide
PAD	Protein(s) associated with differentiation
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PNGase F	Peptide N-Glycosidase F
PP2C	Protein phosphatase 2C
PSI-BLAST	Position-specific BLAST
qPCR	Quantitative polymerase chain reaction
RNA	Ribonucleic acid

RNAi	RNA interference
ROS	Reactive oxygen species
rpm	Revolutions *per* minute
rRNA	Ribosomal RNA
SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SIF	Stumpy induction factor
SRA	Serum resistance associated gene
SSC	Side scatter
TBS	Tris-buffered saline
TAE	Tris-acetate EDTA
TetR	Tetracycline repressor protein
Tf-R	Transferrin receptor
TLF	Trypanosome lytic factor
TLTF	Trypanosome-derived lymphocyte triggering factor
TNF	Tumour necrosis factor
TRITC	Tetramethylrhodamine isothiocyanate
UTR	Untranslated region
UV	Ultra violet
VSG	Variable surface glycoprotein
WHO	World health organisation
WT	Wild type
ZFK	Zinc finger kinase
ZFP	Zinc finger protein
ZFPM	Zimmerman postfusion buffer
ZFPMG	Zimmerman postfusion buffer with glucose

Contents

Chapter 1 Introduction

This PhD project is regarding the characterisation and functional analysis of an expression site associated gene 9 (ESAG9) gene family in *Trypanosoma brucei*. Following a general introduction to African trypanosomiasis, this introductory chapter covers the following areas in more detail: the life cycle of *T. brucei*, the variant surface glycoproteins of *T. brucei*, antigenic variation, expression site associated genes, and the immune responses to *T. brucei* in the tsetse fly and mammal hosts. This sets a framework for the experimental analyses to be described in the subsequent chapters.

1.1 General background

1.1.1 African trypanosomes

African trypanosomes, *Trypanosoma brucei ssp*., are extracellular protozoan parasites of the order Kinetoplastida. They are the causative agents of sleeping sickness in humans and Nagana in cattle in sub-Saharan Africa (Bruce & Nabarro, 1903). The term 'Nagana' also refers to disease caused in ruminants by *Trypanosoma congolense* (Welburn *et al*., 2006). Between 300,000 and 500,000 people are thought to be infected with sleeping sickness and the disease is lethal unless treated (WHO African trypanosomiasis factsheet http://www.who.int/mediacentre/factsheets /fs259/en). Attempts to control the disease in the long term have not been successful; in fact the area affected by the disease has increased, and the disease is currently epidemic in Sudan, Angola and The Democratic Republic of Congo, where it causes more deaths that HIV/AIDS (http://www.who.int/mediacentre/factsheets/fs259/en). In contrast, *T. evansi*, *T. congolense*, and *T. vivax* cause disease in livestock but not in humans.

1.1.2 Geographical range of African trypanosomiasis

There are three subspecies of the African trypanosome *Trypanosoma brucei*: *Trypanosoma brucei brucei*, *Trypanosoma brucei rhodesiense*, and *Trypanosoma brucei gambiense* (Hoare, 1972). *T. b. brucei* causes Ngana in cattle. *T. b. rhodesiense* causes human sleeping sickness in the East of sub-Saharan Africa; this sub-species is also infective to cattle and wild animals which are a zoonotic reservoir of the disease (Welburn *et al.,* 2006). *T. b. gambiense* causes human sleeping sickness in West and central sub-Saharan Africa and is only infective to humans (Gibson 1987). There are 36 countries in sub-Saharan African in which sleeping sickness is a threat to 60 million people. The areas affected are shown in Figure 1.1.

Figure 1.1: Map of the continent of Africa showing which regions are affected by sleeping sickness. Countries that are coloured in green are not currently affected, countries that are coloured in blue or brown experience endemic African trypanosmiasis, and countries coloured in red experience epidemic African trypanosomiasis (www.who.int).

1.1.3 Symptoms and treatment of the disease

The West African form of the disease, caused by *T. b. gambiense*, causes a chronic infection where the patient may survive for several months or years. The East African form of the disease, caused by *T. b. rhodesiense*, is more acute and usually kills the patient within weeks or months. The initial period of the disease is characterised by a fever and joint pain, during which stage the parasites are in the bloodstream and lymphatic system. Once the parasites have crossed the blood-brain barrier, sleep alterations occur, as do psychiatric and sensory imbalances (Lundkvist *et al.,* 2004). The early stage of the disease is treated by either Pentamidine isethionate or Suramine; the only treatments for the advanced stage of the disease are compounds based on arsenic, such as Melarsoprol, and the drug itself can have lethal side effects (Gull, 2002). There is evidence of resistance to both the early-stage and late-stage drugs in a number of field isolates (de Koning, 2008).

1.2 The life cycle of the African trypanosome

1.2.1 Overview of the life cycle

The life cycle of this parasite involves transmission between mammal hosts by an insect vector, which is the tsetse fly, of the genus *Glossina* (Bruce, 1909). Tsetse flies are bloodsucking insects; the metacyclic infective stages of the parasite reside in the salivary glands of the fly and are passed on when the fly bites a mammal host. The parasites then multiply rapidly in the bloodstream as 'slender' forms and, once a threshold density is reached, the cells begin to differentiate to 'stumpy' forms. The stumpy cells are cell cycle arrested and can only re-enter the cell cycle once they have been taken up in the fly blood meal and differentiated to the procyclic insect form. The life cycle of the parasite, and the morphology of the parasites at each stage, are summarised in Figure 1.2.

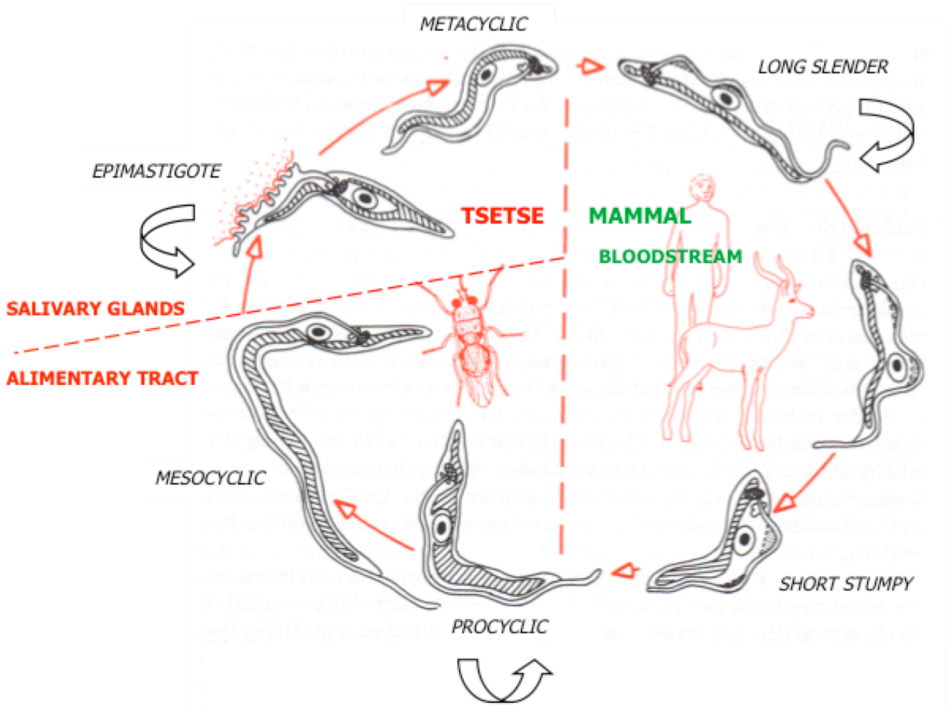


Figure 1.2: Summarised life cycle of the African trypanosome. African trypanosomes are zoonotic and are capable of invading a range of mammalian hosts. They are transmitted by the tsetse fly. The different cell morphologies of each stage of the life cycle are shown. This figure was made by Melanie Buhlmann using cell drawings by Keith Vickerman (Vickerman, 1985).

1.2.2 The mammalian stage of the life cycle

1.2.2.1 Slender form cells in the mammalian bloodstream

African trypanosome parasites are extracellular in the bloodstream of the mammal. They are covered with a dense surface coat of variant surface glycoprotein (VSG) and this will be discussed in detail in section 1.3. Bloodstream forms exist as a pleiomorphic population with two distinct morphologies: ‘slender’ forms and ‘stumpy’ forms (Robertson, 1912; Vickerman, 1965). Slender forms of the parasites are proliferative and divide rapidly in the bloodstream.

1.2.2.2 Differentiation of slender forms to stumpy forms

Slender form cells differentiate in a density-dependent manner to stumpy forms, as shown in Figure 1.3. It was noted as early as 1912 by Muriel Robertson that only the stumpy forms are capable of colonising the fly mid-gut (Robertson, 1912).

This differentiation process of slender to stumpy has been shown to be independent of cues from the mammal host (Mclintock *et al.,* 1990), and is in fact initiated by a quorum-sensing mechanism (Reuner *et al.,* 1997; Vassella *et al.,* 1997). The stumpy induction factor (SIF) is a small, soluble molecule released by the parasites which triggers the process of differentiation from slender to stumpy form (Vassella *et al.,* 1997); the identity of SIF has yet to be determined.

The molecule SIF was thought to initiate differentiation of slender cells to stumpy cells via the cyclic adenosine monophosphate (cAMP) signalling pathway (Vassella *et al.,* 1997); cAMP had previously been attributed a role in the life cycle of *T. brucei* as intracellular concentrations were shown to increase as the cells reached the peak of parasitaemia (Mancini & Patton, 1981). A comparison of the effects exposing slender cells to cAMP analogs, versus the hydrolysis products of cAMP analogs, indicated that it is in fact hydrolysed products of cAMP that are able to mediate differentiation to stumpy forms (Laxman *et al.,* 2006).

Additionally, other molecules apart from cAMP hydrolysis products have been implicated in slender to stumpy differentiation. A protein kinase called ZFK has been suggested to play a role as shown by the fact that a null mutant line displayed an increased rate of differentiation from slender forms to stumpy forms (Vassella *et al.,* 2001). However this effect was only seen in culture; the ZFK null mutants did not display an altered progression of parasitaemia in the mouse host, which would be expected if the parasites were differentiating more quickly to stumpy forms.

Other kinases have been implicated in differentiation of trypanosomatids, including the Mitogen-Activated Protein (MAP) kinase family (Rotureau *et al.,* 2009). MAP kinase enzymes are involved in signal cascades and can be activated by both extracellular stimuli and by stress (Theodosiou & Ashworth, 2002). A null mutant of

TbMAPK5 MAP kinase showed an increased rate of differentiation and this effect was seen both *in vitro* and *in vivo* (Domenicali Pfister *et al.,* 2006).

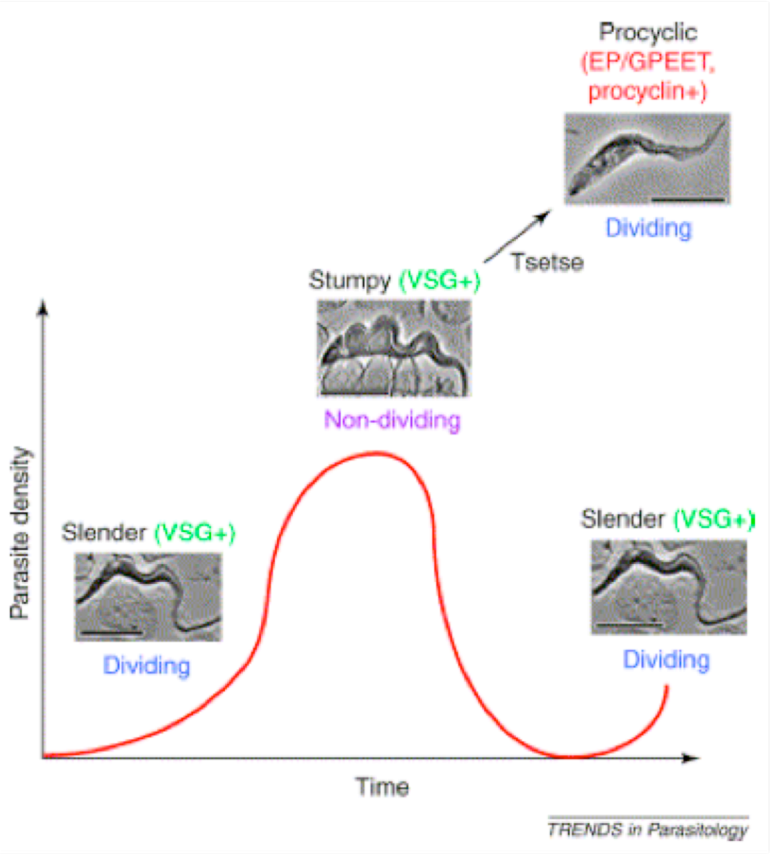


Figure 1.3: Density-dependent differentiation from slender to stumpy forms in the bloodstream of a mammal. Whilst the parasites are proliferating as slender forms in the bloodstream, they are secreting a molecule called Stumpy Induction Factor (SIF), which causes them to differentiate to the non-dividing stumpy form. Stumpy forms are competent for transmission to the tsetse fly vector. Stumpy forms that have not been transmitted will be destroyed by the host antibody response. Slender and stumpy forms express the VSG surface coat (VSG+), and procyclic forms express the Procyclin surface coat (EP/GPEET procyclin+) (Matthews *et al*., 2004).

1.2.2.3 Differentiation from stumpy forms to procyclic (insect) forms

The stumpy cells are arrested at the G_1/G_0 phase of the cell cycle (Shapiro *et al.,* 1984) and remain cell cycle arrested until they have been transmitted to a tsetse fly and received signals to differentiate to the insect stage of the life cycle, the procyclic form. The commitment to undergo differentiation is irreversible once the cells have become cell cycle arrested (Matthews & Gull, 1994).

Stumpy cells are pre-adapted for survival in the tsetse fly midgut in a number of ways. The energy source used by the cells changes from glucose, which is utilised in the bloodstream of the mammal (Michels et al. 2000), to the amino acid proline which is available from the blood meal in the tsetse fly midgut. Metabolic and structural modifications take place to facilitate this change, including reorganisation of the mitochondrial structure (Brown 1973) to allow proline metabolism to commence. The components required for the mitochondrial metabolism are encoded by both nuclear and mitochondrial genes, and complex processes must be involved in their coordinate regulation (Priest & Hajduk, 1994; Timms *et al.,* 2002).

Stumpy cells are competent to differentiate in a synchronous fashion over a short timescale to the procyclic form once the correct signals have been received (Ziegelbauer *et al.,* 1990). They are more resistant to the proteolytic environment of the midgut than slender forms (Sbicego *et al.,* 1999; Turner *et al.,* 1988b). Stumpy cells shed their VSG coat from four hours after the initiation of differentiation and gain a Procyclin coat over 4 to 16 hours (Matthews & Gull, 1994; Roditi *et al.,* 1989; Ziegelbauer *et al.,* 1990).

The process of differentiation can be triggered *in vitro* by a temperature drop to 27°C and addition of citrate/*cis*-aconitate (CCA) to the media (Brun & Schonenberger, 1981; Czichos *et al.,* 1986). The mechanisms via which synchronous differentiation could be triggered by the same signals *in vivo* have been explored. By selecting a cell line called DiD-1 (Tasker *et al.,* 2000), which was unable to undergo differentiation, the PAD family of proteins were identified as potential regulators of differentiation. PAD proteins are stumpy-specific, thermo-regulated carboxylate transporters

expressed on the surface of stumpy cells (Dean *et al.,* 2009). They are capable of transmitting the CCA signal that triggers stumpy cells to differentiate *in vitro*. Indeed citrate could also be the *in vivo* trigger given that, when combined with a reduction in temperature, it is effective at inducing stumpy to procyclic differentiation at concentrations *in vitro* that correspond to endogenous levels in the mammalian bloodstream and in the tsetse fly (Hunt *et al.,* 1994; Jacobs & Lee, 1964).

The reduction in temperature, or 'cold-shock', of stumpy parasites is also a trigger of differentiation. A drop in temperature of 15°C causes both slender and stumpy form parasites to express EP Procyclin protein, which is a surface molecule of procyclic form parasites (Engstler & Boshart, 2004). Fascinatingly, this protein only gains access to the surface of the cell in stumpy form parasites, indicating that surface expression of GPI-anchored proteins is differentially controlled in slender and stumpy forms. Cold-shock also increases the expression of at least one of the PAD family of proteins (Dean *et al.,* 2009). Temperature has been found to be an important factor in regulating the gene transcription and morphology of other protozoan parasites, including *Plasmodium falciparum* (Fang & McCutchan, 2002) and *Leishmania spp*. (Zilberstein & Shapira, 1994).

A protein tyrosine phosphatase has also been identified as a regulator of differentiation. The inhibition of a protein tyrosine phosphatase called TbPTP1 results in a small proportion of monomorphic cells spontaneously undergoing differentiation (Szoor *et al.,* 2006). Tyrosine dephosphorylation of proteins therefore prevents stumpy form cells from differentiating to procyclic forms, and so could be described as a molecular brake that prevents differentiation from occurring until the correct stimuli are received.

The stumpy cells can only re-enter the cell cycle once they have been taken up in the fly blood meal, otherwise they will eventually be killed by the host antibody response. However they are more resistant to antibody-mediated lysis than slender forms (McLintock *et al.,* 1993). The speed at which slender forms and stumpy forms clear host antibody from their surface has been investigated. Host antibody binds to the VSG molecules on the surface of the cell, and these complexes, which have been described as 'molecular sails', are then cleared from the surface of forward-

swimming trypanosomes by hydrodynamic forces (Engstler *et al.,* 2007). Stumpy cells are capable of clearing immune complexes from their surface in 20 seconds, whereas it takes slender cells 35-42 seconds; however the immune complexes are not moving at a higher speed across the surface of stumpy forms (Engstler *et al.,* 2007). The faster rate of clearance may therefore be because stumpy forms have a higher rate of endocytosis into the flagellar pocket at the posterior of the cell. The flagellar pocket is the only site of endo- and exocytosis in trypanosomes (Vickerman, 1969; Webster & Russell, 1993).

1.2.3 The insect stages of the life cycle

The tsetse fly insect vector of African trypanosmiasis is blood-feeding, and the parasites are transmitted to the vector in the blood meal. Stumpy-form cells are competent to undergo differentiation to the insect stage of the life cycle, procyclic forms, as discussed in section 1.2.2.3. Procyclic forms are characterised by a metabolism based on mitochondrial activity, the expression of the Procyclin coat, and the ability to evade the tsetse fly innate immune system and complete the insect stage of the life cycle over a course of several weeks (Aksoy *et al.,* 2003). Tsetse fly innate immunity against trypanosome invasion will be discussed in detail in section 1.5.1. However, it is not only the capability of the trypanosome versus the immune system of the fly that determines if an infection can progress. Female flies are typically found to have higher infection rates, as are older flies, and a warmer climate also correlates with higher infection rates (Aksoy *et al.,* 2003).

The tsetse fly blood meal is taken into the fly crop, the position of which is shown in a diagram of the tsetse fly in Figure 1.4. Once in the fly mid-gut, the trypanosomes are enclosed in a peritrophic matrix, which partitions the bloodmeal. During the following three days, the bloodmeal is digested by the fly and the trypanosomes progress to the posterior midgut until they eventually penetrate the peritrophic matrix (Ellis & Maudlin, 1985). The Procyclin coat molecule is predominantly GPEET Procyclin for the first week of infection and it then switches to predominantly EP Procyclin (Acosta-Serrano *et al.,* 2001). After approximately seven days, the trypanosomes halt cellular division and become proventricular forms which are very

elongated; these return to the gut lumen and then migrate to the salivary glands. Proventricular forms become epimastigotes, which are proliferative and divide in an asymmetric fashion (Sharma *et al.,* 2008; Van Den Abbeele *et al.,* 1999). These cells become attached to the lining of the salivary glands (Robertson, 1913) via an outgrowth of their flagellar membrane. Epimastigote forms express a novel BARP surface coat molecule, the function of which is not known (Urwyler *et al.,* 2007). A further developmental step results in metacyclic cells, which are motile and proliferative (Vickerman, 1985). Metacyclics are competent for transmission to the mammal, and express the VSG surface coat (Tetley *et al.,* 1987) from metacyclic expression sites (Ginger *et al.,* 2002). Unlike the VSG repertoire in slender bloodstream forms (which will be discussed in section 1.3), the metacyclic form parasites only have around 27 possible VSG genes to choose from (Turner *et al.,* 1988a).

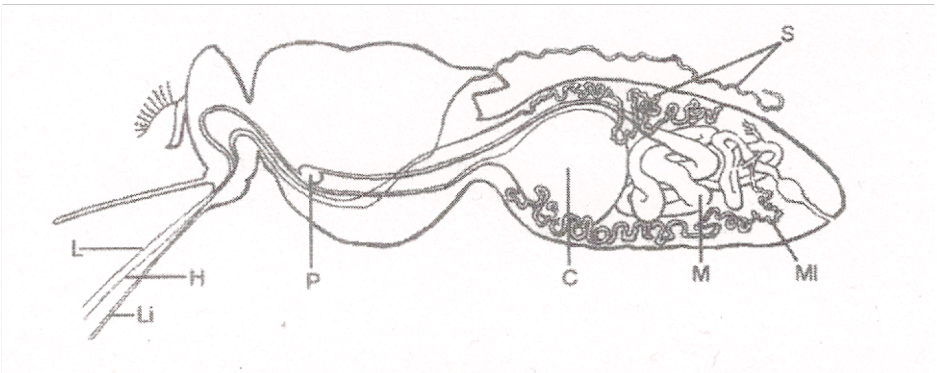


Figure 1.4: Diagram of a male tsetse fly. The body parts labeled are: C = crop; L = labrum; H = hypopharynx; Li = labium; M = midgut; MI = malpighian tubule; P = proventriculus; S = salivary glands (Aksoy *et al*., 2003).

When a fly bites a mammal host, the infective metacyclic cells are transmitted to the mammal where they initially divide at the site of the bite before progressing further into the bloodstream (Aksoy *et al.,* 2003). Interestingly, it has been suggested that tsetse fly saliva passed into the mammal bite results in an increased rate of

trypanosome infection (Caljon *et al.,* 2006). Tsetse flies appear to probe the mammal they are feeding upon more frequently when they are infected with trypanosomes (Jenni *et al.,* 1980). This may be because the presence of trypanosomes in the salivary glands interferes with the ability of the tsetse fly to detect the flow of blood. This could be a behaviour that the trypanosomes have evolved to enhance their transmission to the mammal host.

Trypanosomes are able to sexually reproduce whilst in the tsetse fly, though this is not an obligate stage of the life cycle (Gibson & Stevens, 1999; Jenni *et al.,* 1986). The genetic exchange is thought to occur in a Mendelian fashion by meiosis (MacLeod *et al.,* 2005; Turner *et al.,* 1990), after the parasites have migrated back to the salivary glands, but before they have become transmission-competent metacyclics (Tait *et al.,* 2007). A pleasing technique to further explore trypanosome mating has been developed by Gibson *et al*., whereby trypanosome lines expressing green fluorescent protein or red fluorescent protein are crossed in the fly, and the progeny of matings are distinguishable as they exhibit yellow fluorescence (Gibson *et al.,* 2008).

1.3 The variant surface glycoprotein coat

1.3.1 The structure of the VSG coat

Variant Surface Glycoproteins (VSGs) form a dense monolayer of around 10^7 molecules covering the entire trypanosome (Cross, 1975). These proteins are key to the immune evasion of bloodstream form African trypanosomes (Cross, 1978). The transmission electron micrograph in Figure 1.5 shows a cross-section through a cell and flagellum, and the VSG coat is visible as a very electron-dense covering over the whole surface.

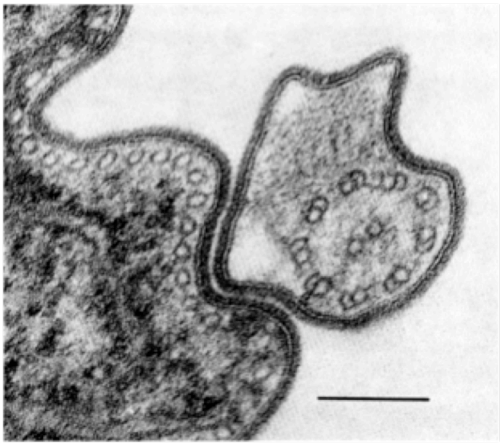


Figure 1.5: Transmission Electron Micrograph of *T. brucei* showing the VSG coat. In this cross section of a parasite, the VSG can be seen as the electron dense coat surrounding the parasite cell body and flagellum. Beneath the VSG coat is the plasma membrane. The sub-pellicular microtubules are visible as a series of hollow rings in the main body of the cell, and in the flagellum the microtubules form the 9+2 arrangement. Scale bar is 100nm (Cross, 1978).

Glycosylphosphatidylinositol (GPI) anchors attach proteins to cell membranes in eukaryotic cells (Ferguson, 1999). The ability of phosphatidylinositols to anchor proteins to lipid membranes was first realised when it was observed that a phospholipase C enzyme specific to phosphatidylinositol was capable of releasing alkaline phosphatase from cell membranes (Ikezawa *et al.,* 1976; Low & Finean, 1977). VSGs are dimeric proteins that are attached to the cell surface of bloodstream form African trypanosomes via two GPI anchors (Ferguson *et al.,* 1985a; Ferguson *et al.,* 1985b). The structure of a VSG dimer is shown in Figure 1.6, with each monomer having a GPI anchor which attaches it to the plasma membrane.

The C-terminal GPI anchor of VSG is hydrophobic, and can be cleaved off to release a soluble form of VSG by the action of either GPI phospholipase C (GPI-PLC) or metalloproteases. The GPI-PLC enzyme is only found in bloodstream-form parasites (Bulow *et al.,* 1989) and functions, alongside the zine metalloprotease Tb-MSPB, to remove the VSG coat from differentiating stumpy form trypanosomes (Grandgenett *et al.,* 2007; Gruszynski *et al.,* 2006). Bloodstream form and procyclic form parasites have stage-specific GPI anchors, and during differentiation the expression of the bloodstream form anchor is switched off, and the GPI-PLC resistant procyclic-form

anchor is switched on (Gruszynski *et al.,* 2006). The GPI-PLC enzyme is in fact non-essential, as null-mutants are still viable, though they are less virulent as infected mice survive longer when infected with null mutants than when infected with wild type parasites (Webb *et al.,* 1997). The facility of addition of GPI anchors to proteins is necessary for survival of bloodstream (though not procyclic) forms however (Lillico *et al.,* 2003).

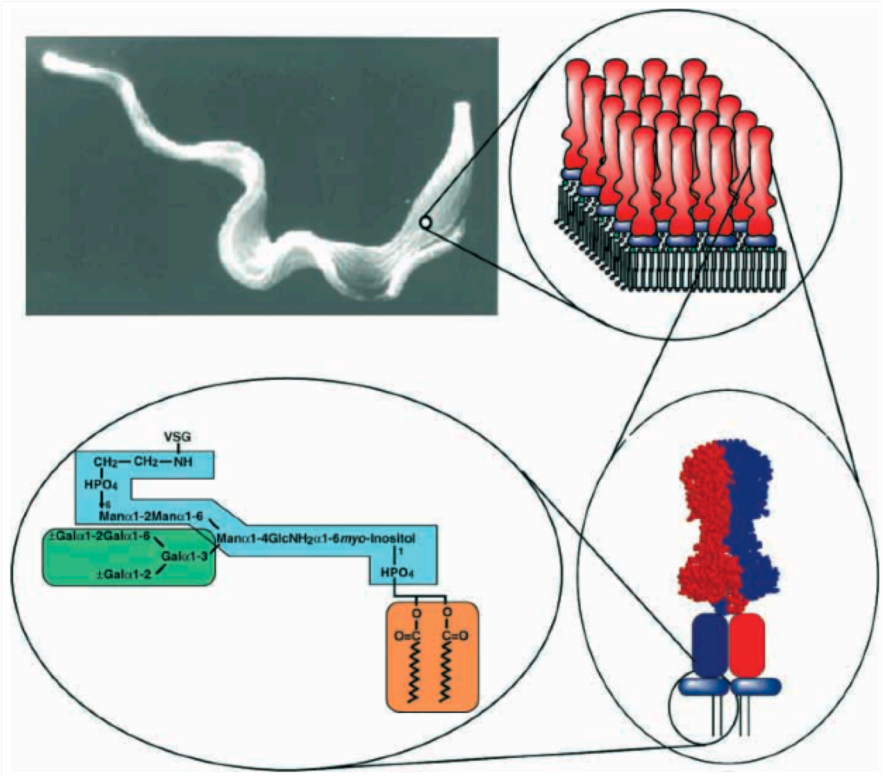


Figure 1.6: The structure of the VSG coat. Clockwise from top left: a scanning electron micrograph of a bloodstream form trypanosome; a diagram of the membrane lipid bi-layer with the VSG molecules protruding from the surface; a more detailed diagram of a VSG homodimer with two GPI anchors attached; the structure of a GPI anchor (Ferguson 1999).

1.3.2 Antigenic variation

The VSG surface coat of bloodstream form trypanosomes is highly immunogenic and results in the mammal raising an antibody-mediate response which kills all parasites with that particular surface coat. It was first noted by Ross and Thompson in 1910 that a patient with African trypanosomiasis experienced waves of parasitaemia (see Figure 1.7, panel A)(Ross & Thompson, 1910). Each wave of parasitaemia is characterised by the expression of a different surface antigen (VSG), in a process termed antigenic variation (Cross, 1978; Vickerman & Luckins, 1969). During the rise in parasitaemia, the majority of the parasites will express the same VSG coat as a monolayer over their entire surface. Antibodies will be raised against this VSG, which will then result in the majority of the parasites being destroyed by antibody-mediated lysis. However a few parasites will have switched to an alternative VSG gene and these parasites will evade the immune response to repopulate the host. This process is represented in a schematic fashion in Figure 1.7, panel B.

However it is important to note that it is not only the host antibody response and antigenic variation that cause the fluctuating parasitaemia typical of a trypanosome infection. As the parasitaemia in the host is increasing, slender form parasites are secreting SIF and a quorum-sensing mechanism results in them differentiating to cell-cycle arrested stumpy forms in a density-dependent fashion, as discussed in Section 1.2.2.2. Mathematical modelling indicates that this density-dependent differentiation is also capable of causing fluctuating parasitaemia in the host (Lythgoe *et al.,* 2007; Tyler *et al.,* 2001). At the peak of parasitaemia, the majority of cells will be stumpy form and have the same dominant VSG type as the slender population from which they differentiated (Matthews and Gull, 1994).

The combination of antigenic variation and density-dependent differentiation, and the fluctuating parasitaemia that they cause, result in a limitation of the parasite population size in the host. This allows a chronic infection to develop. It is beneficial for the parasite to increase host longevity as this will increase its chances of transmission to a tsetse fly vector.

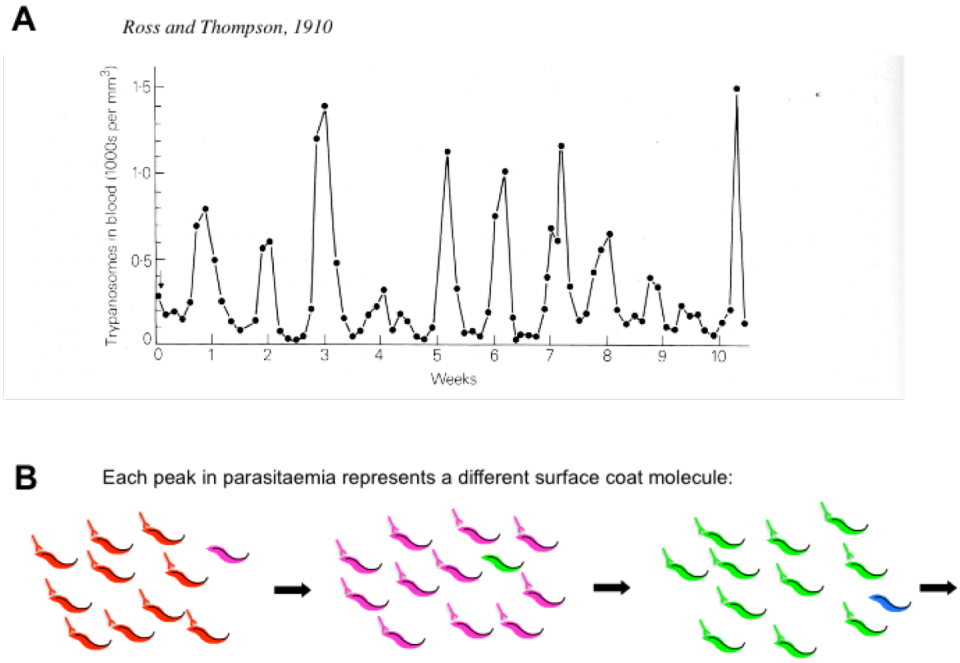


Figure 1.7: Antigenic variation during trypanosomiasis. Panel A shows the classic fluctuating parasitaemia profile in a patient with African trypanosomiasis (Ross and Thompson, 1910). Each peak in parasitaemia will be characterised by a different version of the VSG surface antigen. Panel B shows the majority of parasites expressing the same surface coat molecule (represented by a colour), against which antibodies will be raised (shown as a Y-shaped structure of the same colour). However a few trypanosomes will have switched their surface antigen to a slightly different version and hence escape the host antibody response.

1.3.3 Molecular mechanisms of antigenic variation

Complex molecular mechanisms underlie the process of antigenic variation. There are around one thousand VSG genes in the *T. b. brucei* genome (Berriman *et al.,* 2005; Van der Ploeg *et al.,* 1982); these are expressed from telomeric expression sites (ESs) (De Lange & Borst, 1982). There are around twenty VSG ESs and only one locus is actively transcribed at any one time by RNA polymerase I (pol I) (Gunzl *et al.,* 2003; Navarro & Gull, 2001); pol I only transcribes ribosomal RNAs in organisms outside the kinetoplastids (Lee & Van der Ploeg, 1997). The active expression site is located to a special expression site body (ESB) containing pol I (Navarro & Gull, 2001). The expression sites can be divided into metacyclic ESs, which are utilised in the salivary glands of the tsetse fly, and bloodstream ESs which take over after the first few days of infection into a mammalian host (Borst & Ulbert, 2001). The structure and contents of ESs will be discussed further in section 1.4.

There are complex molecular mechanisms underlying the phenotype of antigenic variation, in which only one VSG antigen is expressed at a time out of a large repertoire, and these are summarised in Figure 1.8. The mechanisms can be placed into four categories: first, and most frequent, is a process whereby the VSG gene in an active ES is replaced with a chromosome-internal VSG gene. This is the most common method for switching variant antigen gene (McCulloch, 2004), and allows the parasite to utilise the vast number of VSG genes that are not telomeric (Van der Ploeg *et al.,* 1982). A VSG gene can also be moved from another telomeric position by either telomere conversion or reciprocal exchange of telomeres (Borst *et al.,* 1998; Pays *et al.,* 1985). The trypanosome genome consists of 11 paired megabase chromosomes of sizes varying from 0.9 to 7 Mb as well as large numbers of intermediate and minichromosomes (Melville *et al*., 1999). There are VSG genes available at the telomeres of the megabase chromosomes and also at the telomeres of many minichromosomes (Van der Ploeg *et al.,* 1984; Weiden *et al.,* 1991). Finally, an *in situ* switch can occur, whereby the active ES is switched off, and a new ES is switched on.

The genome contains a vast repertoire of VSG genes, but interestingly a large majority of these genes do not have fully intact coding regions (Berriman *et al.,* 2005; Taylor & Rudenko, 2006). New VSG genes can be generated from these pseudogenes by recombination between homologous regions of these pseudogenes (Thon *et al.,* 1990).

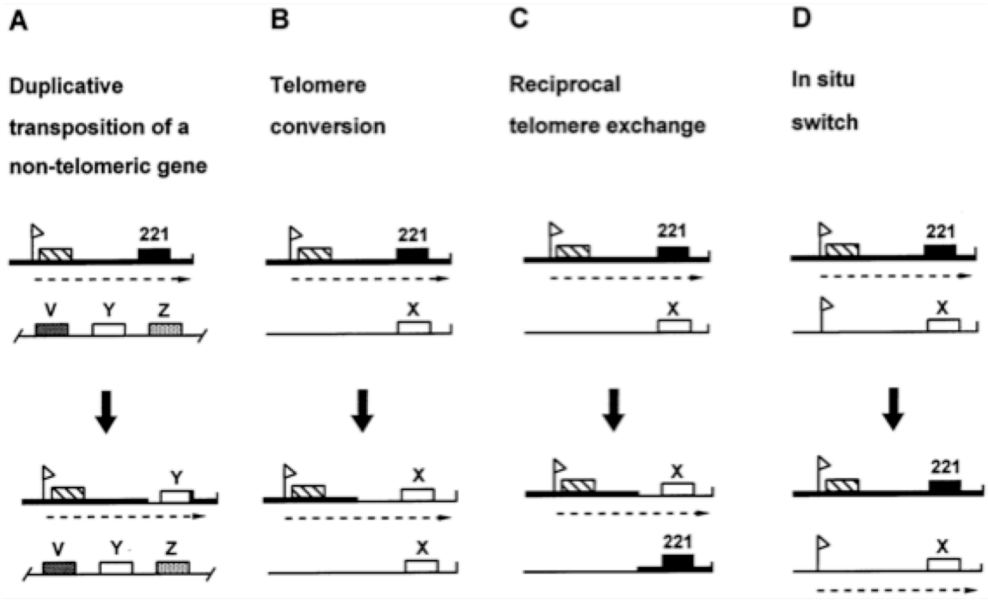


Figure 1.8: Mechanisms of switching expression to a new VSG surface coat molecule. There are four mechanisms for switching VSG antigen: A: gene conversion from a non-telomeric gene; B: telomere conversion (the same promoter is still used); C: telomere exchange (again the same promoter is still used); D: in situ activation of a different expression site (ES). The flags represent ES promoters, the dashed box represents a hygromycin resistance cassette used experimentally in distinguishing these mechanisms, the other boxes represent VSG genes, and the dashed arrow indicates which ES is being actively transcribed. This diagram is taken from a review by Borst *et al*. (Borst *et al.,* 1998).

1.4 Expression sites and Expression Site Associated Genes

It was discovered in 1985 that VSG genes are not the only genes transcribed from the telomeric expression sites. A number of mRNAs were identified which are co-transcribed within an active ES (Cully *et al.,* 1985b) by bloodstream form parasites. These genes were termed Expression Site Associated Genes, or ESAGs. ESAGs were subsequently demonstrated to be transcribed with VSG in a polycistronic fashion, from a promoter approximately 60 kb upstream of the telomeric VSG gene (Johnson *et al.,* 1987). This was the first report of polycistronic transcription in a eukaryotic cell.

It has since become clear that most trypanosomatid protein-coding genes are in fact transcribed in this fashion (Vanhamme and Pays, 1995), the one known exception being the metacyclic VSG genes expressed from metacyclic expression sites, which are transcribed monocistronically (Alarcon *et al.,* 1994; Ginger *et al.,* 2002). Metacyclic VSG ESs differ from those utilised by bloodstream form parasites in other ways. Metacyclic ESs contain only a small number of ESAG pseudogenes (Graham *et al.,* 1999), and not only are they transcribed in a monocistronic fashion, but they are also not subject to post-transcriptional control which is very unusual in trypanosomatids (Graham & Barry, 1995).

1.4.1 Genomic context of Expression Site Associate Genes

The structure of a typical *T. brucei brucei* ES is shown in Figure 1.9. There is a telomeric VSG gene, which is followed by a series of telomeric repeats, and then the end of the chromosome. Upstream of the VSG gene are 70 base-pair repeats, then a series of ESAG genes, and each ES is preceded by a promoter. The exact content of the ES varies in terms of which ESAGs are present. Indeed ESAGs, as is the case for VSG genes, are not found uniquely within expression sites. They are found both in sub-telomeric and chromosome-internal positions, and are usually found adjacent to or near other ESAG genes and/or other VSG genes (Berriman *et al.,* 2005; Hertz-

Fowler *et al.,* 2008). Unlike VSG genes, ESAGs that are outside telomeric ESs can be transcribed, as demonstrated by the fact that when a VSG is expressed from an ES containing no ESAGs, expression of certain ESAGs still occurs (Graham & Barry, 1991).

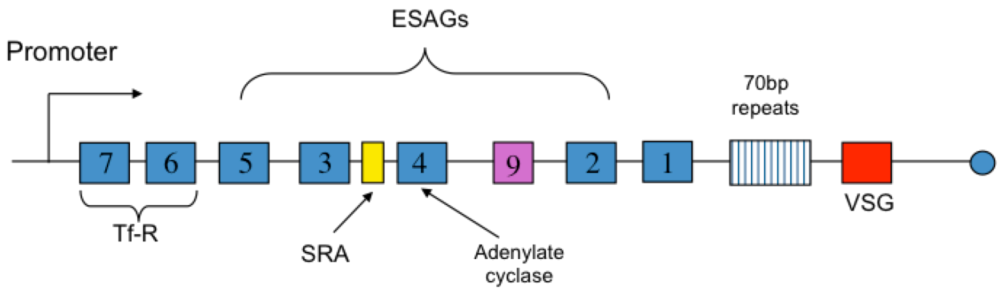


Figure 1.9: The structure of a typical bloodstream form expression site. Upstream of the telomeric VSG gene are a number of other genes, termed Expression Site Associated Genes (ESAGs). These are transcribed from the same promoter as VSG as a polycistronic transcription unit.

1.4.2 Diversity in Expression Sites

Thirteen ESAGs have been identified so far; ESAGs do not constitute one specific family of genes, instead many different types of genes can be ESAGs. Only ESAGs 6 and 7 are in all VSG bloodstream ESs that have been sequenced (Berriman *et al.,* 2002). It has been hypothesised that a function of the existence of different sets of ESAGs in different ESs could be to aid parasite survival in a range of mammalian hosts (Bitter *et al.,* 1998; Pays *et al.,* 2001). This is particularly relevant for ESAGs 6 and 7, and the Serum Resistance Associated gene (SRA), and these will be discussed further in sections 1.4.3 and 1.4.4. However the comparison of bloodstream form ESs from *T. b. brucei, T. b. gambiense*, and *T. equiperdum* did not provide any supportive evidence that the size of the host range (which varies between these species) correlates with the genetic diversity of the ESs (Young *et al.,* 2008). Although there is variation in the size of ESs and in the ESAGs present, the overall architecture of bloodstream form ESs is conserved (Hertz-Fowler *et al.,* 2008).

1.4.3 ESAGs 6 and 7: the Transferrin Receptor

ESAG 6 was initially ascribed as a membrane-associated, GPI-anchored transferrin-binding protein (Schell *et al.,* 1991b; Schell *et al.,* 1993). It was purified from membrane extracts by affinity chromatography using human transferrin. Treatment with GPI-PLC enzyme rendered ESAG6 hydrophilic which indicated that it is GPI anchored; the protein is specific to bloodstream-form parasites (Schell *et al.,* 1991b).

It was subsequently discovered that a heterodimer of ESAG6 and ESAG7 forms the transferrin receptor (Tf-R). While ESAG6 is GPI-anchored, ESAG7 is not, and both proteins are glycosylated (Salmon *et al.,* 1994; Steverding *et al.,* 1994). Immunogold labelling combined with scanning electron microscopy revealed that Tf-R is localised to the flagellar pocket lumen, the flagellar pocket membrane, and intracellular vesicles (Steverding *et al.,* 1994). Once transferrin is bound, the complex is internalised and trafficked to a lysosome where the transferrin is degraded to release the iron for use by the trypanosome and the receptor is subsequently recycled (Steverding *et al.,* 1995). Transferrin is therefore necessary for the growth of bloodstream form parasites (Schell *et al.,* 1991a). The transferrins available to trypanosomes in different hosts are not highly conserved. For example there is only a 70% amino acid sequence identity between bovine and human transferrin (Retzer *et al.,* 1996).

All the ESs in *T. b. brucei* contain ESAGs 6 and 7 and the genes in the different ESs vary in a hypervariable region. It was found that the Tf-Rs formed by ESAGs 6 and 7 from different ESs have stronger binding affinities for transferrin from different mammals. For example, the Tf-R from the 221 VSG expression site has a high binding affinity for bovine serum, but a low binding affinity for canine serum (Bitter *et al.,* 1998). When parasites in which the 221 VSG ES was active were grown in canine serum, this resulted in an outgrowth of parasites which had switched to an alternative ES containing a Tf-R with higher affinity for canine serum (Bitter *et al.,* 1998). Trypanosomes have also been found to be capable of switching the ESAG7 gene within an ES to a different one, rather than switching the active ES. When trypanosomes were forced to maintain expression from the 221 VSG ES via a

hygromycin resistance cassette, and grown in canine serum, some clones which had adapted to canine serum were found to retain the same ESAG6 gene in the 221 VSG ES, but replace the ESAG7 gene in the 221 VSG ES with another version from elsewhere in the genome (van Luenen *et al.,* 2005).

The ability of trypanosomes to switch Tf-R to maximise their ability to bind host transferrin has led to the hypothesis that ESAGs have evolved to facilitate the survival of trypanosomes in a range of mammal hosts (Bitter *et al.,* 1998; Pays *et al.,* 2001). However, there has been some debate as to whether the affinity of the Tf-R for transferrins from different mammals is actually physiologically relevant. When trypanosomes expressing Tf-R with a high affinity for human transferrin were grown in transferrin-depleted medium supplemented with transferrin from either bovine or human transferrin, the transferrins were equally well able to restore growth (Salmon *et al.,* 2005). The authors speculate that it was some other stress on the parasites resulting from growing them in canine serum that caused them to switch to another ES containing a different Tf-R in the experiments by Bitter *et al*. (Salmon *et al.,* 2005).

It has been hypothesised a high affinity of the Tf-R for the transferrin allows transferrin to successfully compete with anti-Tf-R antibodies for binding and internalisation. The host makes antibodies against Tf-R, which is exposed in the flagellar pocket (Bitter *et al.,* 1998). In *in vitro* assays, the incubation of anti-ESAG7 antibodies with trypanosomes significantly inhibited the uptake of transferrin (Salmon *et al.,* 1994) and polyclonal antibodies raised against the Tf-R were internalised by the parasites (Gerrits *et al.,* 2002). Indeed, growth of parasites which expressed the Tf-R from the 221 ES (which has a low affinity for canine serum) in canine serum and anti-221-Tf-R antibodies resulted in outgrowth of parasites which had switched to a different ES (with a Tf-R with a higher affinity for canine serum) occurring faster than it did if there were no antibodies present (Gerrits *et al.,* 2002). It is worth noting that this effect was only seen with certain anti-Tf-R antibodies and batches of canine serum.

However there has been evidence published that does not support the theory that competition by antibodies inhibits uptake of transferrin by the Tf-R. The study by

Gerrits *et al.* used rabbits that had been immunised with recombinant Tf-R to measure physiological antibody concentrations. When the titres of anti-Tf-R antibodies in mice chronically infected with trypanosomes (as opposed to in immunised rabbits) were measured they were found to be too low to significantly affect the ability of the Tf-R to take up transferrin (Steverding, 2006).

1.4.4 Serum Resistance Associated gene

It was noted that *T. b. rhodesiense* parasites displayed two phenotypes in terms of their susceptibility to lysis by human serum: sensitive and resistant, whilst *T. b. gambiense* parasites were always resistant (De Greef *et al.,* 1989). Resistance in *T. b. rhodesiense* parasites was found to correlate with the presence of a specific mRNA (De Greef *et al.,* 1989). Isolates were observed to switch from the resistant to the susceptible form and *vice versa* and all isolates were found to contain the gene for the transcript expressed by resistant forms. This gene was named the Serum Resistance-Associated gene (SRA) (De Greef & Hamers, 1994). The protein has sequence homology with VSG, especially at the C-termini (De Greef & Hamers, 1994) and was designated as having its evolutionary origins in a truncated VSG gene (Campillo & Carrington, 2003). The gene is an ESAG and is found within some VSG expression sites; cloning of the gene into the normally non human-infective *T. b. brucei* subspecies rendered these parasites resistant to lysis by human serum (Xong *et al.,* 1998). However SRA is not the only mechanism in the trypanosome's repertoire for resisting lysis by human serum; *T. b. gambiense* parasites are completely resistant, and *T. b. brucei* TRUE 927/4 parasites show an intermediate resistance phenotype (Vanhamme *et al.,* 2004); the mechanisms for these have not been elucidated but SRA is not responsible (Turner *et al.,* 2004).

Expression of SRA from a particular VSG ES is a definitive example of an ESAG playing a vital role increasing the host range of *T. b. rhodesiense*. The mechanism by which the SRA protein confers resistance to lysis by human serum will be discussed in section 1.5.2.4 of this chapter.

1.4.5 Characterisation of other ESAGs

Definitive functions have not been ascribed to the other ESAG genes. ESAG 1 is a membrane glycoprotein (Cully *et al.,* 1985a) of unknown function. ESAG4 appears to be an adenylate cyclase with transmembrane domains (Alexandre *et al.,* 1990; Paindavoine *et al.,* 1992; Pays *et al.,* 1989). ESAG8 is unusual in that it is a predominantly nucleolar protein which interacts with a Puf protein that is involved in mRNA stability (Hoek *et al.,* 2002). ESAG10 contains ten transmembrane domains and so is likely to be a membrane protein (Gottesdiener, 1994) and ESAG11 is predicted to be a glycosylated cell surface protein (Redpath *et al.,* 2000). Bioinformatic analysis of ESAG5 protein sequences indicates that they have similarity to a family of lipid transfer proteins, and predicts that ESAG5 proteins would be glycosylated and could be either membrane-bound or secreted (Barker *et al.,* 2008). A summary of what is known about the function and genomic location of the different ESAGs is given in Table 1.1.

ESAG number or name	Number of sequenced ESs in *T. b. brucei* Lister 427 which contain this ESAG	Number of non-telomeric copies in *T. b. brucei* TREU 927/4	Function or attributes	Core reference
1	12	11 genes 2 pseudogenes	Membrane glycoprotein of unknown function	Cully et al. 1985
2	12	10 genes 5 pseudogenes	None attributed	Hertz-Fowler et al. 2008
3	8; and 6 contain pseudogenes	22 genes 4 pseudogenes	Predicted to be membrane associated	Pays et al. 2001
4	9; and 2 contain a psueodgene	5 genes 4 pseudogenes	Putative adenylate cyclase	Pays et al. 1989
5	12; and 2 contain a pseudogene	7 genes	Predicted to be glycosylated and either membrane-associated or secreted	Barker et al. 2008
6	14	2 genes	Heterodimer with ESAG7 forms the transferrin receptor	Schell et al. 1993
7	14	1 gene	Heterodimer with ESAG6 forms the transferrin receptor	Steverding et al. 1994; Salmon et al. 1994
8	11	1 gene 1 pseudogene	Putative DNA-binding protein	Hoek et al. 2002
9	1	9 genes 10 pseudogenes	None attritubted	Florent et al. 1991
10	7 (these have an upstream promoter to the main ES)	3 genes	Putative transmembrane protein	Gottesdiener 1994
11	12 contain a pseudogene	6 genes 6 pseudogenes	Putative glycosylated cell surface protein	Redpath et al. 2000
12	5	None annotated	None attributed	Hertz-Fowler et al. 2008
SRA	None present in this strain	None present in this strain	Responsible for the resistance to lysis by human serum of *T. b. rhodesiense*	De Greef and Hammers 1994

Table 1.1: Summary table of ESAGs and their functions. The number of expression sites (ESs) in *T. b. brucei* Lister 427 which contain each ESAG are shown (Hertz-Fowler *et al*., 2008), as are the number of ESAGs in non-telomeric positions in *T. b. brucei* TREU 927/4 (Berriman *et al*., 2005). The function or putative function is given and a core reference for each ESAG.

1.5 African trypanosomes and immunity

1.5.1 The tsetse fly innate immune response

The tsetse fly mounts an immune response against trypanosomes to prevent both the establishment and the maturation of a trypanosome infection (Welburn & Maudlin, 1999). The effectiveness of this response is shown by the low level of trypanosome infections detected in field samples of tsetse flies in endemic areas, typically 10-12% (Masiga *et al.,* 1992; Morlais *et al.,* 1998a; Morlais *et al.,* 1998). The refractoriness of tsetse flies depends upon many factors and processes, involving the midgut, proventriculus, fat body and hemolymph (Lehane *et al.,* 2004).

The innate immune system has been more thoroughly studied in *Drosophila melanogaster*, and this has revealed similarities between well-characterised mammalian immune pathways such as the Tumor Necrosis Factor pathway, and the insect immune response (Tzou *et al.,* 2002). The occurrence of similar processes in tsetse flies has not been verified; however a number of innate immune molecules and functions have been characterised and these will be discussed.

1.5.1.1 Antimicrobial peptides

Four antimicrobial peptides have so far been conclusively identified in tsetse flies, and it is likely that there are more yet to be identified. Attacin, defensin, diptericin, and cecropin are all expressed in the fat body, and attacin and cecropin are also expressed in the proventriculus and haemolymph (Hao *et al.,* 2001; Hu & Aksoy, 2006; Kaaya *et al.,* 1987; Lehane *et al.,* 2008). These genes are differentially regulated depending on the microbial challenge. Exposing tsetse flies to *Escherichia coli* resulted in the transcription of attacin and defensin being strongly induced whereas procyclic trypanosomes elicited a lesser and temporary response, and bloodstream form trypanosomes did not induce the expression of defensin at all (Hao

et al., 2001). Diptericin however is always expressed regardless of the microbial challenge.

Exposure of procyclic and bloodstream form *T. b. brucei* to recombinant attacin protein (GmAttA1 gene) resulted in an inhibition of growth of both cell types (Hu & Aksoy, 2005). The recombinant protein did not have any activity against the tsetse fly gut endosymbiont *Sodalis glossinidius*. The expression of attacin is controlled by Relish (GmmRel), which is a transcriptional regulator of invertebrate humoral immune response genes that was originally identified in *D. melanogaster* (Hedengren *et al.,* 1999). Knockdown by RNA interference of GmmRel resulted in loss of 90% of the attacin transcript in tsetse flies (Hu & Aksoy, 2006). These flies experienced a two to three fold increase in parasite load in the midgut when exposed to trypanosomes compared to flies expressing normal levels of attacin.

1.5.1.2 Lectins

Exposure to lectins such as Con A is lethal to procyclic-form trypanosomes (Pearson *et al.,* 2000). The mechanism of killing involves the binding of the lectin to glycan chains on the Procyclin surface coat molecule. Mutants have been identified that express different EP surface protein isoforms, with only partial or no glycosylation, and are resistant to Con A (Acosta-Serrano *et al.,* 2000; Hwa *et al.,* 1999).

It is important to note that Con A is a plant lectin. However expressed sequence tag (EST) analysis of the *G. morsitans morsitans* midgut has revealed the presence of three putative lectin genes (Lehane *et al.,* 2003). A protein called TsetseEP has also been identified that is expressed in the midgut, though not the salivary glands, of *G. morsitans* (Chandra *et al.,* 2004). This protein has tentatively been classified as a lectin as it has structural similarity to known insect lectins, but it has not yet been shown to have lectin activity. Interestingly, the protein has some sequence similarity to the EP procyclin surface coat molecule of insect-form trypanosomes (Chandra *et al.,* 2004).

Circumstantial evidence for the involvement of lectins in the tsetse fly innate immune response to trypanosomes comes from the effect of feeding tsetse flies

sugars that inhibit lectin activity. Both D-glucosamine and N-acetyl-glucosamine block the binding of lectins to glycosylated macromolecules. The inclusion of these sugars with a tsetse fly feed results in the flies being more susceptible to trypanosome infection (Peacock *et al.,* 2006).

1.5.1.3 Proteases

Proteases are also found in the midgut of tsetse flies. The protease trypsin has been identified in the midgut as a lectin-trypsin complex (Abubakar *et al.,* 2006) and a number of enzymes with trypsin-like properties have also been found in the midgut (Liniger *et al.,* 2003). Exposure of pleiomorphic trypanosomes to trypsin showed that whilst slender forms are killed by these proteases, stumpy forms are resistant. In fact, interestingly, the trypsin stimulated the stumpy form parasites to undergo differentiation and express Procyclin (Sbicego *et al.,* 1999), so the presence of proteases in the midgut may actually be a signal to the invading parasites. One function of Procyclins on the surface of trypanosomes may be to protect the parasite from protease attack, and, indeed, the N-termini of the protein is cleaved in the tsetse midgut by proteolysis (Acosta-Serrano *et al.,* 2001). Although the Procyclin coat is not required for parasite survival in the fly gut, parasites lacking the coat do colonise the fly gut to a lesser degree (Guther *et al.,* 2006; Nagamune *et al.,* 2000; Ruepp *et al.,* 1997; Vassella *et al.,* 2009).

1.5.1.4 Reactive oxygen species

Reactive oxygen species (ROS) are known to be an important part of the mammalian immune system (this will be discussed in section 1.5.3) and they are also thought to play a role in invertebrate innate immunity. These ROS may be a method of signalling between different organs in the tsetse fly (Lehane *et al.,* 2004). The proventriculus of the tsetse fly produces nitric oxide (NO), hydrogen peroxide, and other reactive oxygen intermediates as a result of challenge with microbes (Hao *et al.,* 2003). Indirect evidence for a role for ROS comes from the fact that 'self-cured' flies (i.e. those that have successfully eliminated a trypanosome infection) have increased levels of catalase expression when compared to infected flies, and

increased levels of lambda crystallin when compared to uninfected flies. These antioxidant molecules could be involved in protecting the flies from reactive oxygen intermediates that they are producing (Lehane *et al.,* 2008). A total of 18 putative antioxidant genes were identified by EST analysis of the midgut of *G. morsitans morsitans* (Lehane *et al.,* 2003).

1.5.1.5 The role of endosymbionts

Tsetse flies have (at least) three important endosymbionts: *Wolbachia spp*., *Wigglesworthia glossinidia*, and *Sodalis glossinidius*. This last endosymbiont, *S. glossinidius*, has been found to be involved in the susceptibility of tsetse flies to trypanosome infection. It is not an obligate endosymbiont, and its elimination from tsetse flies using streptozotocin results in the flies having a lower rate of midgut infections (Dale & Welburn, 2001).

Refractoriness to trypanosome infection is inherited down the female line in tsetse flies (Moloo *et al.,* 1998). The endosymbionts of tsetse flies are also maternally inherited and they could be involved in susceptibility or resistance to infection. This phenotype of maternally-inherited refractoriness has been reported to only be seen in teneral flies (newly hatched unfed flies) however, and not those which have already received a blood feed (Welburn *et al.,* 1989).

An interesting model has been put forward which connects the endosymbiont populations of tsetse fly midguts with the susceptibility of teneral tsetse flies, via the action of lectins. When the fly is in the pupa, endosymbiotic bacteria digest chitin and release sugars which could be lectin-inhibitory (Peacock *et al.,* 2006; Welburn *et al.,* 1993); this may be a reason why younger flies are more susceptible to infection with trypanosomes.

1.5.2 Immune responses and immune evasion in the mammal stage of the life cycle

There is an evolutionary conflict between the host and the pathogen in terms of the host's ability to clear an infection and the pathogen's ability to avoid this clearance. The mammalian immune response and the complex immune evasion strategies of the trypanosome parasite will be discussed. The literature discussed in regards to mammalian immune responses will, unless specified, be referring to experiments using the murine model of trypanosomiasis.

1.5.2.1 Mammalian immune responses

During the course of trypanosomiasis, the majority of the parasites are periodically ablated through antibody-mediated cell destruction (Cross, 1978). This process is not mediated by complement, to which the parasites are not susceptible due to their VSG coat, but involves binding of antibody to the VSG surface antigen and subsequent phagocytosis of opsonised parasites by macrophages (Tomlinson and Raper, 1998).

Antibody-mediated parasite clearance is not the only tool that the mammalian immune system uses to counter trypanosomiasis however. Many cell types and cytokines are involved. There is an early interferon gamma (IFN-γ) response during trypanosomiasis – this type I cytokine is produced primarily during the first wave of infection and is involved in inflammatory immune responses and classical activation of macrophages (Baetselier *et al.,* 2001). These classically activated macrophages have dual functions in protecting the host. They secrete molecules that are toxic to trypanosomes, particularly nitric oxide (NO) and Tumor Necrosis Factor α (TNF-α). They also have the ability to phagocytose parasites that have been opsonised. Although the early production of IFN-γ is beneficial, over longer term infections the type I interferons α and β may cause a downregulation in IFN-γ which makes a mouse host more susceptible (Lopez *et al.,* 2008).

Nitric oxide (NO) is a reactive nitrogen intermediate with antimicrobial properties (Bogdan, 2001). NO results from the oxidation of L-arginine to L-citrulline by nitric oxide synthase (NOS). It is produced primarily, though not uniquely, by macrophages and can function to both stimulate and suppress the mammalian immune system (Bogdan, 2001). NO seems to have antagonistic properties during trypanosomiasis. It could benefit the host via its antimicrobial action and the depletion of the substrate L-arginine, which is required by trypanosomes, but it has also been reported to be a key element in immunosuppression caused by trypanosome infection (Beschin *et al.,* 1998). From day 4 of a trypanosome infection in mice, an increase in NO production by macrophages is seen when compared to uninfected mice, and this correlates with a decrease in proliferation of T cells in the spleen and a decrease in IFN-γ production (Beschin *et al.,* 1998). The inclusion of an NO synthesis inhibitor resulted in the proliferation of spleen cells being reduced to a lesser degree. So-called 'suppressor macrophages' have also been reported to produce prostaglandin E during trypanosomiasis, and this is thought to help mediate this decrease in proliferation of T cells (Schleifer & Mansfield, 1993).

Tumour Necrosis Factor α (TNF-α) is a cytokine that was originally discovered to have a role in tumour cell apoptosis, but has since also been found to have important roles in immune responses against pathogens (Pfeffer, 2003). TNF-α has both direct trypanolytic effects and indirect immunosuppressive effects during trypanosome infection. TNF-α lyses bloodstream-form trypanosomes *in vitro* (Daulouede *et al.,* 2001; Lucas *et al.,* 1994), and is endocytosed via the flagellar pocket (Magez *et al.,* 1997). After endocytosis, the TNF-α is trafficked to lysosome-like organelles, where it prevents proper osmoregulation by the parasites (Magez *et al.,* 1997). Interestingly, in pleiomorphic parasites, TNF-α only lysed cells if they were harvested at the peak of infection of a mouse and not early on in infection, which suggests that this cytokine has a more potent trypanolytic effect on stumpy cells. TNF-α knockout mice experience higher levels of parasitaemia, and a reduction in IFN-γ production (Magez *et al.,* 1999).

1.5.2.2 Protozoan parasite GPI-anchored proteins stimulate immune responses

Protozoan GPI-anchored proteins and purified GPIs have been shown to stimulate an immune response in their mammalian host. Effects have been seen in *Plasmodium spp*., *T. cruzi* and *T. brucei* (Magez *et al.,* 2002; Tachado *et al.,* 1997).

GPI anchored mucin glycoproteins from *T. cruzi* trypomastigotes were shown to stimulate production of nitric oxide (NO) by macrophages (Camargo *et al.,* 1997a; Camargo *et al.,* 1997b). This work was extended to compare the effect of GPI fractions from epimastigote (insect stage) and trypomastigote (extracellular bloodstream form stage) *T. cruzi* parasites (Almeida *et al.,* 2000). In *T. cruzi*, epimastigote GPI anchor fractions from mucins were not found to be bioactive, whereas very low concentrations of GPI fractions from trypomastigote mucins induced a macrophage response. GPI-mucins were able to bind to Cluster of Differentiation 1 d (CD1d) (Procopio *et al.,* 2002), which is a surface glycoprotein (Park & Bendelac, 2000) and is a ligand of NK1.1+ T cells (Bendelac *et al.,* 1995). However *T. cruzi* GPI-mucins but did not elicit an activation of NK1.1+ T cells (Procopio *et al.,* 2002).

GPI-anchored VSGs from *T. brucei* parasites have also been shown to have immunogenic properties (Coller *et al.,* 2003; Leppert *et al.,* 2007; Lopez *et al.,* 2008; Magez *et al.,* 1998; Schleifer & Mansfield, 1993). Soluble VSG (sVSG, cleaved from the surface of bloodstream form parasites via the action of GPI-PLC enzyme) binds to receptors on the surface of the macrophage (Leppert *et al.,* 2007). It induces IFN-γ dependent signals of macrophage activation, for example the production of NO (Coller *et al.,* 2003). This effect is thought to be important especially in the early stages of parasitemia in the mammal host.

1.5.2.3 Trypanosomes avoid and modulate the mammalian immune response

Trypanosomes avoid complete eradication by the mammalian immune system predominantly via the process of antigenic variation (as described in section 1.3.2), which allows them to evade obliteration by the antibody response mounted against VSG. However they also have other tactics for avoiding and suppressing the immune response of the host. Trypanosomiasis patients are more susceptible to secondary infections due to immunosuppression resulting from the parasite infection (Greenwood *et al.,* 1973).

Bloodstream form *T. brucei* parasites have been reported to secrete soluble factors that act as immunomodulators (Holzmuller *et al.,* 2008; Olsson *et al.,* 1991; Sternberg & Mabbott, 1996). *In vitro* culture of bloodstream form trypanosomes with either peritoneal macrophages or spleen cells resulted in an increase in production of NO by macrophages, and an inhibition of proliferation of T cells, respectively (Sternberg & Mabbott, 1996). This process was dependent on the presence of IFN-γ, and independent of cell contact between the trypanosomes and the macrophages. Procyclic forms did not elicit the same effects. Mortality in susceptible strains of mice has been linked with $CD4^+$ T cells producing large quantities of IFN-γ (Shi *et al.,* 2006), so it could be beneficial to the parasite to down-regulate T cell proliferation and hence increase the longevity of the host.

The secretomes of *T. brucei* strains that differ in virulence and pathogenicity, but are genetically similar, have been investigated. Trypanosomes secrete a large number of molecules, some of which are specific to strains that display different disease phenotypes (Holzmuller *et al.,* 2008). Medium containing the secreted molecules can activate arginase activity, which is involved in the alternative activation of macrophages (Holzmuller *et al.,* 2008).

A molecule termed Trypanosome-Derived Lymphocyte-Triggering Factor (TLTF; also called trypanin) was identified as a secreted immunomodulatory protein. TLTF was reported to be a 185 kDa protein, released by bloodstream form parasites, which caused CD8+ lymphoid cells to secrete IFN-γ (Olsson *et al.,* 1991; Olsson *et al.,* 1993). It was also suggested that the IFN-γ was taken up by the parasites and acted as a growth factor. The gene for TLTF was subsequently cloned and the recombinant protein found to have the same stimulatory effect on CD8+ T cells (Vaidya *et al.,* 1997). However further analysis of TLTF defined it as being a flagellar protein (Hill *et al.,* 2000), which would therefore not be either secreted or surface-associated. In the light of the localisation of the protein, and the fact that other laboratories have failed to replicate the stimulatory effects on CD8+ T cells, TLTF is considered to be an unlikely candidate immunomodulator (personal communication, Professor John Mansfield, University of Wisconsin-Madison).

B cells form part of the adaptive immune response to pathogens by creating antibodies against microbial antigens. Thus they are an important part of the immune response to trypanosomes, which is centred around antibody-mediated lysis. However, it appears that trypanosomes do not only use antigenic variation to avoid the host antibody response; they are also capable of manipulating the B cell populations responsible for creating antibodies. Experimental infection of mice with trypanosomes resulted in significant levels of B cell apoptosis, and mice were also then susceptible to challenge with trypanosomes expressing a VSG antigen to which they have already been exposed (Radwanska *et al.,* 2008).

1.5.2.4 The Trypanosome Lytic Factor in human serum

The Trypanosome Lytic Factor (TLF) in human serum causes lysis of susceptible strains of trypanosomes. It confers innate immunity against *T. b. brucei* parasites (Raper *et al.,* 2001).

An Apolipoprotein L-1 (ApoL-1) subfraction of the human high-density lipoprotein (which has long been known to be toxic to trypanosomes) has been isolated as the trypanosome lytic factor (TLF) (Hajduk *et al.,* 1989; Rifkin, 1978; Vanhamme *et al.,* 2003). In susceptible trypanosomes, the TLF is endocytosed via the flagellar pocket, and it forms a pore across the membrane of a late endosome or lysosome. This allows an influx of chloride ions, which causes the lysosome to swell and eventually results in cell death. However in resistant trypanosomes the TLF is trafficked to a vesicle containing the SRA (serum resistance associated) protein (Oli *et al.,* 2006). SRA is an ESAG, as discussed in section 1.4.4, and is responsible for resistance to lysis by TLF in resistant strains of *T. b. rhodesiense*. An alpha helix in the SRA protein has been shown to interact with an alpha helix in ApoL-1 (Vanhamme *et al.,* 2003) and via this interaction presumably prevents pore formation across the vesicle membrane. The mechanism for susceptibility or resistance to lysis is shown in Figure 1.10, which is taken from a Nature Reviews Microbiology review by Pays et al. (Pays *et al.,* 2006). Although the ApoL-1 is responsible for lysis, this lysis is much more efficient when the ApoL-1 is still in the human high density lipoprotein complex rather than when it is in isolation (Shiflett *et al.,* 2005).

The receptor responsible for taking up the trypanosome lytic factor is a Haptoglobin-Hemoglobin Receptor (TbHpHbR) (Vanhollebeke *et al.,* 2008). The prime function of this receptor is to take up Haptoglobin-Hemoglobin (Hp-Hb); the uptake of the Hp-Hb by TbHpHbR is associated with an increase in growth rate of bloodstream form trypanosomes because the heme is required by the cell (Vanhollebeke *et al.,* 2008). The heme appears to have a further potential function in bloodstream form trypanosomes. *In vivo* and *in vitro* analyses suggest that the uptake of Hp-Hb enables

parasites survive better when exposed to oxidative stress from macrophages; and this effect is negated when the oxidative stress is inhibited (Vanhollebeke *et al.,* 2008), for example by incubation with L-NAME, which is an inhibitor of NO synthase. The mechanism of this protection has not been elucidated.

However the TbHpHbR binds equally well to a Haptoglobin-related protein, which is associated with the human high density lipoprotein complex (which contains ApoL-1, the trypanolytic factor) (Vanhollebeke *et al.,* 2008; Vanhollebeke *et al.,* 2007). Therefore the receptor is beneficial to the parasite in all but human hosts (and some monkeys) as the Hp-Hb complex confers on the parasites both the ability to grow faster, and the ability to escape from oxidative stress induced by macrophages of the mammalian immune system. However, in human hosts this receptor becomes detrimental as it results in the uptake of ApoL-1 with the rest of the complex (Pays & Vanhollebeke, 2008). This effect has been evaded via the evolution of the SRA gene (an ESAG) in *T.b. rhodesiense*. *T. b. gambiense* is also immune to TLF but the mechanism is not known; resistance in this sub-species is not though to be associated with an ESAG (Pays *et al.,* 2006).

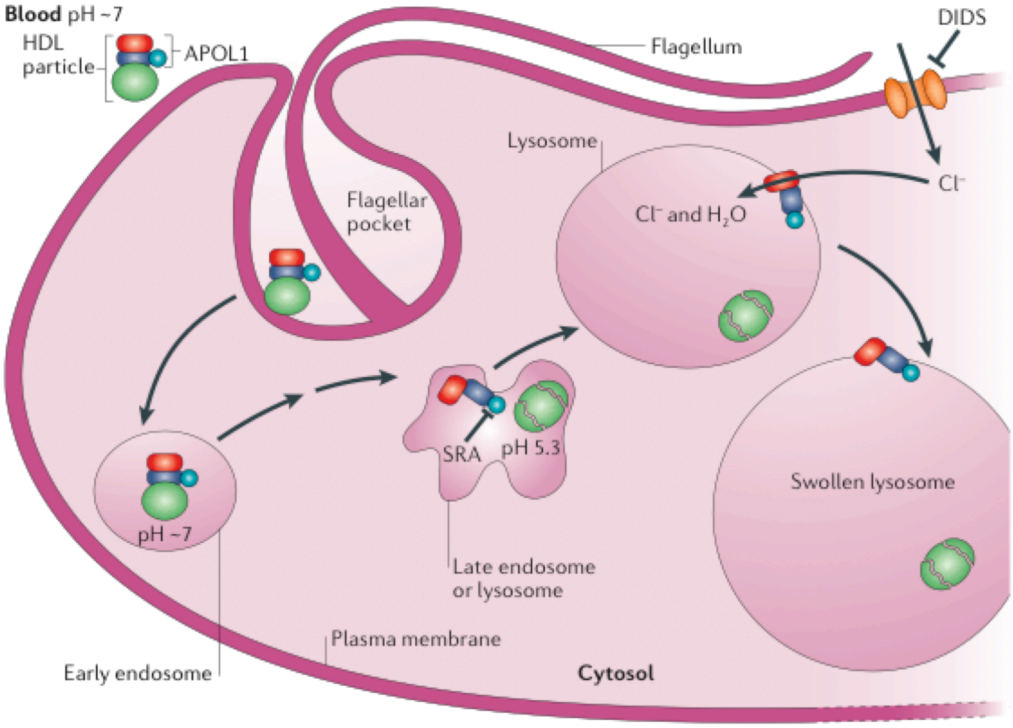


Figure 1.10: The mechanism of trypanosome lysis by ApoL-1. The human high density lipoprotein (HDL particle) is endocytosed via the flagellar pocket. It is enclosed in an endosome, wherein a sub-fraction of ApoL-1 is able to form a pore through the endosome plasma membrane. This results in an influx of chloride ions into the mature lysosome from the cytosol, followed subsequently by excessive swelling of lysosomes and hence trypanosome death. The SRA protein in some way prevents the ApoL-1 from making a pore across the plasma membrane of the maturing lysosome. The diagram is taken from a review in Nature Reviews Microbiology by Pays *et al.* (Pays *et al*., 2006).

1.6 Aims of this PhD project

Prior to the commencement of this PhD project, the extent of knowledge regarding the expression site associated gene 9 (ESAG9) gene family was limited to: (i) There are (at least) two ESAG9 genes in *Trypanosoma equiperdum*, one of which is in an expression site (Florent *et al.,* 1991); (ii) Sequencing of different strains of *T. brucei brucei* has revealed that ESAG9 genes are usually chromosome-internal and are rarely found in expression sites (Berriman *et al.,* 2005; Hertz-Fowler *et al.,* 2008); and (iii) the expression of ESAG9 genes is up-regulated in *T. b. brucei* stumpy form parasites (Keith Matthews, unpublished data).

We aimed to increase the depth of knowledge regarding this gene family in *T. b. brucei* in the following ways:

- By determining the protein expression profile of ESAG9 genes in slender, stumpy and differentiating parasites.
- By exploring, using a bioinformatic approach, the similarity of ESAG9 genes to any other characterised gene family, and the presence of any motifs of known function or predictions of post-translational modifications.
- By using transgenic cell lines, and antibodies raised against ESAG9 proteins, to explore the subcellular location and post-translational modifications of ESAG9 proteins.
- By using *in vitro*, *in vivo*, and *ex vivo* approaches to determine if ESAG9 genes have any function in host-parasite interactions, as is the case for some other ESAG genes.

Chapter 2 Materials and Methods

2.1 Trypanosomes

2.1.1 Strains

The 427-449 cell line was used for transfection in this study. The cell line originates from the *Trypanosoma brucei brucei* Lister 427 strain, which is monomorphic and culture-adapted. The isolation and history of this cell line can be found on the website of Professor George Cross (http://tryps.rockefeller.edu/). This cell line was stably transfected with a construct called pHD449 (Wirtz *et al.,* 1999), containing a tetracycline repressor gene to allow inducible ectopic expression, hence the name '427-449'. The pHD449 vector is described in section 2.2.3.5.1.

A pleiomorphic strain called EATRO 2340 was also used in this project. The term 'EATRO' refers to the East African Trypanosmiasis Research Organisation in Tororo, Uganda, where the strain was first isolated (Cunningham & Vickerman, 1962). This strain was thought until recently to be *T. b. rhodesiense*, but has recently been tentatively renamed as *T. b. brucei* (Young *et al.,* 2008).

Pleiomorphic cells are those that exhibit two morphologies in the bloodstream of the life cycle: slender and stumpy. They are competent to complete the life cycle via passage through tsetses flies. Monomorphic cells have been culture adapted to the extent that they no longer produce stumpy forms, and proliferate very rapidly in the mouse host causing a virulent infection (Fenn & Matthews, 2007).

2.1.2 Cell culture

Procyclic form cells were maintained at between 2×10^6 and 2×10^7 cells/ml in culture flasks and incubated at 27°C and passaged every 24 to 48 hours. SDM-79 media (Brun & Schonenberger, 1979; Cross & Manning, 1973) was used, with 10% v/v of foetal bovine serum (FBS, Gibco) and haemin to a final concentration of 2.5µg/ml.

Bloodstream form cells were maintained at between 1×10^5 and 2×10^6 cells/ml in culture flasks with filter lids to allow diffusion of CO_2, and passaged every 24 to 48 hours. They were incubated at 37°C with a CO_2 concentration of 5%. HMI-9 media was used (Hirumi & Hirumi, 1989) with 20% v/v FBS.

2.1.3 Purification from blood

Blood from trypanosome-infected mice was obtained by heart-puncture. The blood was then passed through a DEAE cellulose anion exchange column (Lanham & Godfrey, 1970) and the trypanosomes washed through the column with PSG (for all buffer recipes refer to Appendix A). The cell density was then counted using a Coulter Particle Count and Size Analyser (Beckman Coulter), the cells pelleted by centrifugation at 600 g for ten minutes, and the cells then resuspended in an appropriate volume of HMI-9 media. All mouse handling work was carried out by either Deborah Hall or Keith Matthews.

2.1.4 Differentiation

Stumpy cells extracted from blood were resuspended in pre-warmed HMI-9 media at a concentration of 2×10^6 cells per ml. They were incubated at 27°C with addition of *cis*-aconitate to a final concentration of 6mM to induce differentiation.

2.1.5 Transfection

For procyclic form transfections, 3×10^7 cells (per transfection) were harvested by centrifugation at 600 g for 10 minutes whilst in the log phase of growth. The cell pellets were washed in ice-cold ZPFM buffer (for recipe see Appendix A) and the resultant cell pellet resuspended in 500µl of ice-cold ZPFM buffer. Either 10µg linearised DNA, or an equivalent volume of sterile dH_2O for the negative control, was added to an electroporation cuvette (with a 4mm gap, from Molecular BioProducts) and then a 500µl aliquot of cells added to each cuvette. The cells were then transfected with a BTX 830 Electro Square Porator set to 3× 100µs pulses of 1700v with 200ms intervals. The transfected cells were immediately transferred to 10mls of pre-warmed SDM-79 media and recovered overnight in the incubator.

Successful transfectants were those which had integrated the linearised plasmid into their genome in the ribosomal RNA intergenic region. The constructs used contained drug resistance cassettes, which inferred drug resistance to the parasites. Successful transfectants were therefore selected with the drug hygromycin at 30µg/ml in culture flasks at a cell density of 2×10^6 cells/ml.

Bloodstream form cells were transfected as above, as far as the overnight recovery step, except that ZPFMG buffer was used instead of ZPFM, and the overnight recovery step was carried out in HMI-9 media. Successful transfectants were selected with either hygromycin (2µg/ml) or puromycin (0.5µg/ml) and selection was carried out in 96-well plates at a cell density of 1×10^5 cells/ml.

2.1.6 Cryopreservation of cell lines

Procyclic or bloodstream form cells were harvested by centrifugation at 600 g for 10 minutes whilst in the log phase of growth. The supernatant was poured off, leaving approximately 500µl of media, in which the cells were resuspended by flicking. An equal volume of either procyclic form freezing mix (SDM-79 with 14% glycerol) or bloodstream form freezing mix (HMI-9 with 14% glycerol) was then added and the cells transferred to cryotubes. The cells were stored at -80°C for 1-2 weeks, and then transferred to liquid nitrogen if longer-term storage was required.

2.2 Molecular biology

2.2.1 DNA gel electrophoresis

Agarose gels were made up in TAE buffer (for recipe see Appendix A) and the agarose was melted by heating in a microwave prior to addition of 3.5µl of Safeview (NBS Biologicals) for every 100ml of agarose gel. Generally a 1% gel was used for maximum separation in the relevant size range. Gels were cast in Biorad tanks and an appropriate volume of sample loaded along with 5µl of SmartLadder DNA ladder (Eurogentec). The samples were run at 95V until a sufficient degree of resolution was achieved, and the gel was then viewed using a UV light source.

2.2.2 PCR

The polymerase chain reaction (PCR) was carried out using either *T. b. brucei* EATRO 2340 genomic DNA or *Escherichia coli* purified plasmid DNA as a template. PCR reactions were set up in a final volume of 50µl with the following ingredients: 5µl dNTPs (2mM stock), 10µl of 5× PCR buffer (as supplied with the DNA polymerase), 6µl $MgCl_2$ (25mM), 2µl of forward and reverse primers (10mM stock), 1µl of DNA polymerase (either GoTaq® from Promega or Expand High Fidelity PCR system from Roche), 2µl DNA, and 22µl H_2O. The PCR was carried out using a PCR Sprint thermal cycler (Thermo Electron Corporation) with the following cycles: one cycle of 95°C for 4 minutes; 30 cycles of: 94°C 30 seconds, 55°C 45 seconds, 72°C 1 minute per kb; one cycle of 72°C for 2 minutes. 5µl of the resultant PCR was then run out in a 1% agarose gel. The sequences of primers used are given in Appendix B.

2.2.3 Cloning

2.2.3.1 Restriction enzyme digestion

Restriction enzyme digestions were typically carried out in a 30µl volume with between 1 and 6µl DNA, 1µl of each of the required enzyme(s), 3µl of 10× digestion buffer (as supplied with the enzyme), and the volume made up with dH_2O. The digestion reactions were incubated at 37°C for either 3 hours or overnight.

2.2.3.2 Ligation

Ligation of PCR products into P-Gem T-easy was carried out using the following ingredients: 0.5µl T4 DNA ligase, 0.5µl P-Gem T-easy vector, 1µl buffer, 8µl PCR product. Ligations were incubated either at room temperature for one hour, or at 4°C overnight.

Ligations of restriction-enzyme digested inserts into all other vectors was carried out with: 2µl plasmid DNA, beween 2 and 6µl insert DNA (dependent on concentration

of insert versus plasmid), 1µl of buffer, 0.5µl of T4 DNA ligase, and the volume made up to 10µl with dH_2O. These ligations were incubated at 4°C overnight.

2.2.3.3 Transformation of bacteria

Aliquots of chemically competent *Escherichia coli* (XL1-Blue from Stratagene) were thawed on ice and 3µl of ligation reaction added to 100µl of cells. The mixture was flicked gently and incubated for 20 minutes on ice. The cells were heat-shocked for 45 seconds at 42°C and then incubated on ice for 5 minutes. 150µl of L-broth was added and then the cells were incubated in a shaking incubator for one hour at 37°C. The cells were then plated out on LB agar containing 100µg/ml ampicillin and incubated overnight at 37°C.

2.2.3.4 Isolation and purification of DNA

DNA was purified from restriction enzyme digestions, or from agarose after gel electrophoresis, using a NucleoSpin Extract II kit (Machery-Nagel) according to the manufacturers instructions.

Small-scale preparations of plasmid DNA were achieved by inoculating 1.5ml of L-broth containing 100µg/ml ampicillin with a single colony from an agar plate and growing overnight at 37°C with shaking. The plasmid DNA was then isolated using solutions I, II and III (Birnboim & Doly, 1979). The bacterial suspension was pelleted by centrifugation in a table-top microcentrifuge at 13,000rpm for 5 minutes. The supernatant was removed, the pellet resuspended in 100µl of solution I, and incubated on ice for 5 minutes. 200µl solution II was added, followed by several inversions to mix and a 5 minute incubation on ice. 150µl of solution III was then added and the mixture inverted several times and incubated on ice for 5 minutes. The mixture was then centrifuged for 10 minutes at 13,000rpm in a table-top microcentrifuge. The supernatant was transferred to a new eppendorf, 900µl of 100% ethanol added, and the DNA then precipitated on ice for 15 minutes. The DNA was then pelleted by centrifugation for 10 minutes at 13,000rpm in a table-top microcentrifuge. The resultant DNA pellet was washed with 70% ethanol, then resuspended in 40µl of TE buffer and 2µl RNase (10mg/ml) and incubated at room temperature for five minutes. Samples were stored at -20°C.

Large-scale isolations of plasmid DNA were achieved by growing a 100ml bacterial culture and isolating the DNA using a QIAGEN Plasmid Midi kit according to manufacturer's instructions.

2.2.3.5 Plasmid constructs used

2.2.3.5.1 pHD 449 and pHD 451

The pHD 449 plasmid, designed by Biebinger *et al.* (Biebinger *et al.,* 1997), contains an open reading frame for the Tetracyclin repressor (TetR) bacterial protein, driven by the trypanosome procyclin promoter. The construct also contains a phleomycin resistance cassette for selection of successful transfectants. The construct is linearised with Not I enzyme and integrates into the ribosomal RNA spacer.

The pHD 451 plasmid, also designed by Biebinger *et al*. (Biebinger *et al.,* 1997), was used for tetracycline-inducible ectopic expression of proteins in both procyclic and bloodstream form parasites. The plasmid integrates into the ribosomal RNA spacer; a diagram of the plasmid is shown in Chapter 4 Figure 4.3.

2.2.3.5.2 pGEM T easy

This is a commercially-available plasmid (Promega) which was used to clone PCR products. A diagram of the plasmid is shown in Appendix C.

2.2.4 Sequencing reaction

Sequencing was carried out by the School of Biological Sciences Sequencing Service at Edinburgh University. For the sequencing reaction a 6µl sample was prepared containing 10-40 ng of plasmid DNA and 26 pmol of the required primer. The sequences were viewed using 4Peaks version 1.7.1 (© Mek&Tosj).

2.3 Western blotting

2.3.1 Preparation of protein samples

Between 1 and 4×10^7 cells were harvested by centrifugation for 10 minutes and 2,000rpm in a table-top centrifuge. The pellets were washed in PBS and the centrifuge step repeated. The resultant pellets were then resuspended in Laemmli (Laemmli, 1970) buffer to a concentration of 2×10^5 cells per 1µl of buffer. The samples were incubated at 100°C for 5 minutes and then vortexed and stored at -20°C until required.

2.3.2 SDS polyacrylamide gel electrophoresis

The polyacrylamide gels were prepared according to the recipes in Appendix A and poured into a Bio-Rad PROTEAN II casting apparatus. A resolving gel was poured first and allowed to set, and then a stacking gel was poured on top. The percentage of acrylamide used in the resolving gel was dependent upon the size of the protein being investigated; for ESAG9 proteins a 13% gel was utilised for maximum separation in the relevant size range.

Samples made from 2×10^6 cell equivalents were run in each lane, alongside a Benchmark™ protein ladder (Invitrogen). The gels were run in a Biorad apparatus at 150V in SDS Running Buffer for approximately 80 minutes, or until the dye front reached the bottom of the gel. Gels were stained in Coomasie's blue stain for 10 minutes if required, and then washed several times in Destain solution.

The gels were transferred to nitrocellulose transfer membrane (Whatman Protran membrane) at 25V for 30 minutes using a BioRad Trans-Blot semi-dry transfer cell. If required, the membranes were then stained with 0.4% ponceau solution for 10 minutes to visualise protein, and washed several times in dH_2O.

2.3.3 Antibody staining of the blot

Two systems were used for visualisation of proteins on blots: either chemiluminescence, or the LI-COR Odyssey system, which uses secondary antibodies conjugated to a fluorochrome. Membranes were blocked overnight at 4°C in either 5% Marvel (for chemiluminescent detection) or in 50% LI-COR Odyssey block, in PBS. Primary antibodies were applied (for concentrations see Appendix D) for one hour at room temperature with rocking, diluted in either 5% Marvel or 50% LI-COR Odyssey block in PBS. Membranes were then washed three times for 5 minutes in PBS. Secondary antibodies (either HRP-conjugated or LI-COR fluorochrome-conjugated) were applied for one hour at room temperature with rocking, diluted in either 5% Marvel or 50% LI-COR Odyssey block in PBS. Membranes were then washed three times for 5 minutes in PBS. In some instances more stringent washing was required, in which case the membranes were washed three times for 5 minutes in PBS, followed by once for 5 minutes in TBS-TNT, and once for 5 minutes in PBS.

For visualisation of a chemiluminescence signal, Enhanced chemiluminescence (ECL) substrate (GE Healthcare, Amersham) was applied for one minute. The blot was then exposed to an X ray film in a photographic cassette for between 1 minute and overnight depending on the primary antibody used, and exposed in a Compact X2 X-ograph.

For visualisation using the LI-COR Odyssey system, the blots were scanned using a LI-COR Odyssey Imager, which uses an infrared laser to detect the fluorochrome on the secondary antibody.

2.4 Immunoprecipitation with cell lysate or concentrated medium

2.4.1 Preparation of G-beads

For each 500µl volume IP, 180µl of G-beads (Sigma) were pelleted by centrifugation at maximum speed in a benchtop centrifuge for 1 minute. The supernant was removed and the beads were washed four times with IP lysis buffer (for recipe see

Appendix A) at 4°C. The beads were then resuspended in IP lysis buffer to a final concentration of 10% v/v.

2.4.2 Preparation of cell lysate for immunoprecipitation

Cells were cultured for 48 hours with tetracycline at a final concentration of 2µg/ml to induce ectopic gene expression. Between 2 and 5×10^8 cells were pelleted by centrifugation at 600 g for 10 minutes. The cell pellets were washed once in 50mls of PBS. The cell pellets were resuspended in 1ml of IP lysis buffer and the cells lysed by freeze-thawing in liquid nitrogen followed by 2 minutes incubation in a sonicating water bath (Geprüfte Sicherheit). A small amount of the cell lysate was then checked microscopically to ensure there were no intact cells. The cell lysate was stored at -80°C.

2.4.3 Preparation of concentrated medium for immunoprecipitation

Bloodstream form cells were grown to a concentration 6×10^6 cells/ml in a volume of 200mls over 48 hours with tetracycline at a final concentration of 2µg/ml. The cells were harvested by centrifugation at 600 g for 10 minutes. The cell pellet was saved for Western analysis. The medium (supernatant) was removed and filtered through 0.22µm filters. The medium was concentrated by centrifuging through 20ml VivaSpin (Invitrogen) columns at 4 000g at 4°C. The volume was reduced by a factor of 50. Roche EDTA-free protease inhibitor was added to the concentrated medium and it was stored at 4°C.

2.4.4 Immunoprecipitation

The cell lysate or concentrated medium was incubated for 30 minutes with a 1:10 dilution of G-beads to pre-block. Simultaneously, the antibody was incubated with 10µg of blocking peptide for 30 minutes, as a control to test the specificity of the immunoprecipitation, or with an equivalent volume of IP lysis buffer. Incubations were at 4°C on a rotating wheel. The supernatant was centrifuged at 10,000×g for 3 minutes at 4°C and removed from the blocking beads. The antibody, with or without pre-block, was incubated with the supernatant for 1 hour at 4°C on a rotating wheel

(for antibody concentrations see Appendix D). An equivalent volume of pre-washed G-beads were then added and incubated for a further hour.

The IP was pelleted by centrifugation at 10 000 g for 3 minutes at 4°C. The supernant was saved for Western analysis. The beads were then washed 6 times with IP lysis buffer. The protein was eluted from the washed beads by incubation at 100°C for 10 minutes in 50µl of Laemmli buffer. The beads were vortexed and pelleted by centrifugation for 3 minutes at 10 000 g and the supernant (IP) removed. The flow-through and washes 1 to 6 were boiled in an equal volume of Laemmli buffer for 5 minutes. Samples were then analysed by Western blotting.

2.5 Immunofluorescence

Immunofluorescence was carried out using a protocol modified from Field *et al*. (Field *et al.,* 2004). Transgenic cells were induced 48 hours prior to immunofluorescence by addition of tetracycline to a final concentration of 2µg/ml. 5×10^6 cells were harvested by centrifugation at 600 g for 10 minutes. The cell pellets were washed in 5ml PBS. The cell pellets were then resuspended in 100µl of 1% paraformaldehyde in PBS and incubated for 10 minutes at 4°C. The fixed cells were spread out on the surface of either a normal glass slide for procyclic cells, or a silanised glass slide (Polysine® slides from Thermo Scientific) for bloodstream form cells. The cells were allowed to settle for 15 minutes. If required, the cells were permeabilised with 0.5% Triton (Sigma) in PBS for 45 seconds. Pre-block solution of 1% bovine serum albumin (BSA, Sigma) in PBS was applied to the slides for 10 minutes. The slides were then incubated in the required antibody (diluted to an appropriate concentration, see Appendix D) in a humid chamber for one hour. Slides were washed three times for 5 minutes in PBS. Slides were incubated in a fluorescent-conjugated secondary antibody for one hour in a humid chamber and then washed three times for 5 minutes in PBS. The slides were stained for 2 minutes with 10µg/ml 4',6-diamidino-2-phenylindole (DAPI) to visualise the DNA, and washed once in PBS. Coverslips were adhered to the slide using 50µl of Mowiol®

glue with p-Phenylenediamine (PDA) to a final concentration of 1mg/ml. The slides were stored in the dark at 4°C.

2.6 N-glycosylation assay

Transgenic bloodstream form cells were induced by addition of tetracycline to a final concentration of 2µg/ml 48 hours prior to assay. 6×10^7 cells were harvested by centrifugation at 600 g for ten minutes, washed with PBS, and the resultant cell pellet incubated at 100°C for 10 minutes in 30µl denaturing buffer (in PNGase F kit from New England Biosciences). 3µl reaction buffer, 3µl 10% NP.40 and 1µl Peptide N-Glycosidase F (all from the NEB kit) were then added and the sample incubated for one hour at 37°C. A negative control with no enzyme added was also included. The samples were then incubated at 100°C for 5 mins in Laemmli sample buffer and analysed by Western blotting.

2.7 GPI-anchor addition assays

2.7.1 Hypotonic lysis

Hypotonic lysis was carried out according to an adapted protocol for small-scale VSG isolation (Jones *et al.,* 2005). Bloodstream form cells (approximately 1×10^8 cells) were harvested by centrifugation for 10 minutes at 600 g and washed twice in ice-cold PBSG (1.8g/litre glucose in PBS). The cell pellets were resuspended in 100µl of hypotonic lysis buffer (10mM NaH_2PO_4 with Roche Complete, EDTA-free protease inhibitor) and incubated at 37°C for 5 minutes. The samples were then centrifuged at 14,000 g for 5 minutes at 4°C. The resultant pellets and supernatants were incubated at 100°C in Laemmli buffer for 5 minutes and analysed by Western blotting.

2.7.2 Myristate labeling and immunoprecipitation

2.7.2.1 Myristate labeling

Transgenic bloodstream form trypanosomes were induced at a concentration of 1×10^5 cells/ml in 50mls for 48 hours with 2µg/ml tetracycline. When the cultures had reached a density of 4×10^6 cells/ml they were harvested by centrifugation at 600 g for ten minutes. The cell pellets were resuspended in 50ml of modified fatty acid-free RPMI medium (see recipe in Appendix A) (Buxbaum *et al.,* 1994), and incubated at 37°C for 30 minutes to allow equilibration. Cells were pelleted by centrifugation at 600 g for ten minutes and resuspended in 9ml of the modified fatty acid free RPMI medium. To this medium 500µl of modified RPMI medium containing 80 micro-curies of triatiated myristate was then added. The flasks were incubated at 37°C for 3 hours to allow for incorporation of the tritiated myristate into newly-synthesised GPI anchors.

After 3 hours, the cells were pelleted by centrifugation at 600 g for ten minutes. The resultant supernatant was set aside for immunoprecipitation anlaysis. The cell pellet was washed in 5ml of ice-cold PBS and then stored on dry ice for immunoprecipitation analysis.

2.7.2.2 Immunoprecipitation and analysis

The immunoprecipitations of the conditioned medium and cell lysate were carried out as described in section 2.4. The resultant protein samples (in laemmli buffer) were run out in a 10% B.T. Nu-PAGE pre-case gel (Invitrogen) in NuPAGE SDS running buffer (Invitrogen). The gel was run at 150V for approximately one hour at which point the dye front was 0.5cm from the bottom of the gel. The gel was then stained in Coomasie's stain as previously described. The gel was de-stained by four 15 minute incubations in destain buffer. The gel was then washed in Amplify Fluorographic Reagent (Amersham) for fifteen minutes. The gel was placed between a piece of filter paper and a piece of clingfilm and dried in a gel drier for one hour at 80°C. The dried gel was then exposed to an Amersham high-sensitivity autoradiography film for 5 weeks in at -80°C.

2.8 AlamarBlue growth assay

AlamarBlue™ (Serotec) is a chemical that enables the growth of cells in culture to be assayed via a change in redox potential. As the cells divide and deplete the culture medium, this results in a change in the redox potential and hence absorbance of alamarBlue.

Procyclic form or bloodstream form cells were grown in culture flasks to a concentration of between 1 and 3×10^6 cells/ml. Cells were plated out in a 96-well plate (100µl per well) and AlamarBlue and either untreated or heat-inactivated animal sera were both added to a final concentration of 10% v/v. Absorbance at 540 and 595nm was measured at a series of time points over a 24 hour period using a BioTek® Elx808 plate reader and Gen5 software. The percentage reduction of alamarBlue was then calculated.

2.9 Tsetse fly experiments

2.9.1 Tsetse fly infections

Tsetse fly pupae (delivered by post from the Department of Entomology, Slovak Academy of Science, Bratislava, Slovakia) were hatched at room temperature and kept in wire cages, being fed on horses blood every other day. The tsetse flies were fed in a dark room with blood placed on a warmed plate under a membrane to simulate probing through skin. Bloodstream form trypanosomes were cultured in HMI-9 medium, and procyclic form trypanosomes in SDM-79 medium, with or without tetracycline. Prior to feeding to tsetse flies, the trypanosome cells were washed twice in SDM-79 medium (for procyclic form cells) or HMI-9 medium (for bloodstream form cells) and resuspended in 10ml horses blood to a concentration of 1×10^6 cells/ml; tetracycline was added to a final concentration of 25µg/ml as required. The trypanosome-infected blood was placed under a membrane on a warm plate (37°C), the cages containing tsetse flies placed over the top, and the flies allowed to feed through the membrane for 10 minutes at room temperature in the

dark. The flies that had fed were then separated into individual 5ml Bijoux pots. Flies were then dissected and the mid-gut removed at 24, 48, or 72 hours or seven days post-infection.

2.9.2 Quantification by microscopy

For microscope counting, the mid-gut was broken up using dissection scissors and then a micro-pestle in PBS. The number of trypanosomes present in 10µl of the mid-gut solution was counted in a haemocytometer. The number of trypanosomes in the whole mid-gut was then estimated.

2.9.3 DNA extraction using phenol chloroform

The tsetse fly mid-guts were dissected out, cut up with dissection scissors, transferred to 500µl buffer, broken up with a micropestle, vortexed for ten seconds, and stored on ice until transferral to a -80°C freezer. The buffers used were either PBS, CTAB, or TNES-Urea, the recipes for which can be found in Appendix A. The DNA was then extracted from the fly guts using a phenol-chloroform extraction. An equal volume of phenol chloroform was added to the thawed gut mixture. The sample was vortexed for 20 seconds and spun for 5 minutes at full speed in a table-top microcentrifuge. The upper layer was transferred to a new 1.5ml polypropylene tube and the lower layer discarded. These steps were then repeated once. A half-volume of ice-cold isopropanol was added and the DNA precipitated by incubation at -20°C for one hour. The DNA was washed twice in 70% ethanol, eluted in 10µl sterile H2O, and incubated at 100°C for 5 minutes before storage at -20°C.

2.9.4 DNA extraction using Gentra kit

The tsetse fly mid-guts were excised and placed in a test tube without any buffer and placed on dry ice. The fly guts were then thawed and 200µl cell lysis solution added (from the Gentra systems Puregene® Genomic DNA Purification Kit). The fly gut and the cell lysis solution were then thoroughly mixed by pipetting, mashing with a micropestle, and then vortexing for 20 seconds. The DNA was then extracted according to Gentra kit protocol D-5000A, including an overnight digestion in

Proteinase K. The DNA was rehydrated in 20µl sterile H_2O and incubated at room temperature overnight. Samples were stored at -20°C.

2.9.5 Quantification by Real Time PCR

Quantitative PCR reactions were set up in 96-well plates. The primers and probes used recognise a VSG gene and were designed by Andy Bell and Rosie Allister. The reaction volume in each well was 25µl, which included forward primer 600nM (TAGCGGCCACGAAAATGA), reverse primer 600nM (CCAGTTCCCCTAGCTTGGTT), TaqMan probe® 40nM (6-FAM-CAGCAATAGAAAAGCTCA-MGB), half the final volume of TaqMan® Universal PCR Master Mix No AmpErase®, 2µl DNA, and the volume made up to 25µl with sterile dH_2O. The samples were run in an Applied Biosystems 7000 Sequence Detection System for 40 cycles according to standard settings. To make standards for quantification, the DNA from a known number of cultured trypanosomes was extracted in phenol-chloroform. The standards were diluted by factors of 1×10^1, 1×10^2, 1×10^3 and 1×10^4 and each standard was plated out in triplicate. The number of trypanosome genomes in each sample was then determined by comparison to the standards using Applied Biosystems 7000 software.

2.10 Mouse infections

Transgenic parasites were inoculated into MF1 mice; the starting inoculation was of 1×10^3 or 1×10^4 bloodstream form monomorphic parasites. Ectopic expression of proteins was induced in the parasites via administering doxycycline to the mice in their drinking water at 200µg/ml with 5% sucrose to disguise the flavour. The control mice received drinking water with 5% sucrose only. The parasitaemia was recorded daily by blood smears from a tail snip using the Herbert and Lumsden rapid-matching method(Herbert & Lumsden, 1976). The mice were culled if they showed signs of discomfort. If naïve mice were required then they were inoculated with an equivalent volume of HMI-9 medium alone.

2.11 Ex vivo FACS

2.11.1 Mouse inoculation and spleen harvest

Mice (MF1) were inoculated with 1×10^4 bloodstream form transgenic parasites, or an equivalent volume of medium containing no parasites for the naïve controls. On the fourth day after inoculation, the mice were culled and spleens dissected out and placed in 2mls RPMI medium (from Sigma with 5% FCS added). The spleens were placed between two sterile pieces of gauze and mashed with forceps. The samples were then centrifuged for 5 minutes at 600 g. The pellet was resuspended in Sigma red blood cell lysing buffer and incubated at room temperature for four minutes prior to washing with RPMI medium. The cell densities of each spleen sample were measured with a CASY®1 cell counter.

2.11.2 FACS analysis

The spleen cell samples were stained for FACS (fluorescent-activated cell sorting) analysis in V-bottomed 96-well plates. Cells were washed with PBS containing 2% FCS and blocked with a 1:10 dilution of 5mg/ml rat IgG for 10 minutes. Cells were incubated on ice for 20 minutes with a 1:100 dilution (in PBS with 2% FCS) of the appropriate primary antibody or isotype control, which were conjugated to FITC, alexafluor 488 or PE to allow detection by the FACS machine (for suppliers of antibodies see Appendix D). All staining was carried out in the dark. Cells were then washed three times in PBS with 2% FCS and transferred to FACS tubes. 50,000 cells from each sample were analysed using a FACScalibur machine (Beckton Dickinson) and the data analysed with FlowJo software (Tree Star).

For analysis of the cytokine interferon gamma, the spleen cells were incubated with the protein transport inhibitor Golgistop™ (BD Biosciences) according to manufacturers protocol for four hours at 37°C prior to staining.

2.12 Statistical analysis

Statistical analyses of data were carried out using Minitab software version 15. The test used was analysis of variance (ANOVA), which is a general linear model. An ANOVA assumes that the data are normally distributed so this was confirmed prior to carrying out the test. P-values of less than 0.05 were considered to be statistically significant. A P-value of 0.05 can be described as a 5% chance that the null hypothesis (that there is no difference between the means of the samples) will be rejected when it should have been accepted. The outcomes of the ANOVAs were expressed as '$F_{xy} = 5$, $p = 0.05$' where x is the degrees of freedom, y is the error degrees of freedom, F is the F-ratio, and p is the P-value.

Chapter 3 Expression Site Associated Gene 9 genes are developmentally regulated in *T. brucei* and are predicted to have post-translational modifications and potential mucin-like properties

3.1 Identification of stage-specific mRNAs in T. brucei

3.1.1 Introduction

There are currently no published data regarding genes which are up-regulated only in the bloodstream stumpy form stage of the *T. brucei* life cycle. Stumpy cells are cell-cycle arrested, are able to survive for a limited amount of time in both the mammalian bloodstream and tsetse fly mid-gut environments, and are competent to differentiate to procyclic forms if the correct signals are received (refer to Chapter 1 section 1.2.2.3). It is highly likely that some of the unique capabilities of stumpy cells are controlled by stage-specific gene expression. Therefore, prior to the commencement of this PhD project, a cDNA subtraction selection was carried out by members of the Matthews lab to identify stumpy-specific mRNAs in *T. brucei* and, hence, genes which are only expressed by stumpy cells.

3.1.2 A search for stage-specific mRNAs in *T. brucei rhodesiense* yielded two genes that are developmentally regulated

A cDNA subtraction selection was carried out to compare transcripts present in slender and stumpy cells. Mice (Balb/c) were infected with monomorphic slender bloodstream forms (*T. b. brucei* EATRO 2340, stabilate GUP 2965) or pleiomorphic slender forms (*T. b. brucei* EATRO 2340, stabilate GUP 2962). The monomorphic cells were harvested and mRNA extracted as the control 'slender' population, and the pleiomorphic cells were allowed to progress until homogenously in the stumpy stage of the life cycle prior to harvesting. The mRNA was reverse transcribed to cDNA and this was then used to perform the subtraction selection. In simple terms, this involves mixing a 'tester' cDNA population (in this case, the stumpy form cDNA), and a 'driver' population (slender) which is in ten fold excess. The cDNA is heated

to denature the molecules and render them single-stranded. Adaptors are ligated to the ends of the tester cDNA molecules. The tester and driver cDNA populations are then incubated to allow annealing. Any tester single-stranded cDNAs which have not annealed (and so in this experiment were unique to the stumpy stage) are enriched by PCR using primers that anneal to the adaptors on the ends of the tester cDNA molecules (Sagerstrom *et al.,* 1997). The unique tester cDNAs can then be cloned. The two clones that were isolated from this process most frequently were called 'K9' and 'K69'. To confirm the stage-specific expression of K9 and K69, a Northern blot was performed using probes for these two genes against monomorphic slender, stumpy, and procyclic total RNA. The 'Defective in Differentiation clone1' or DiD-1 cell line was also used as this cell line is unable to differentiate from stumpy to procyclic forms (Tasker *et al.,* 2000) and has been proposed to lack a characteristic required for differentiation. The Northern blot is shown in Figure 3.1 (Professor Keith Matthews, unpublished data). This confirms that the mRNA abundance for these genes is highly up-regulated in stumpy forms, with a very low level of the transcripts present in monomorphic slender forms, and none in procyclic forms. A feint band for ESAG9-K9 was visible in the DiD-1 sample, whereas there was no band for ESAG9-K69, but this may simply reflect the length of exposure of the respective blots.

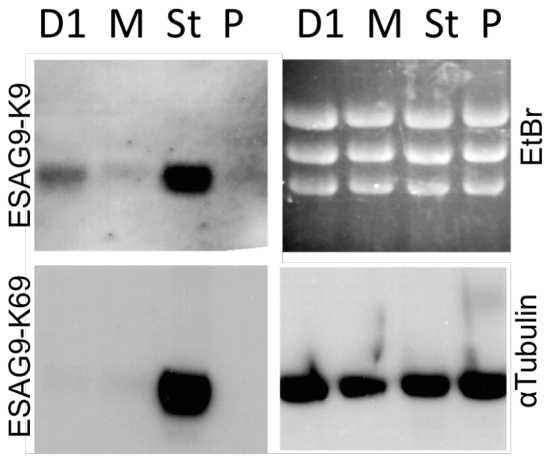


Figure 3.1: Northern blots showing that K9 and K69 are upregulated in Stumpy form cells. The lanes correspond to: Defective in Differentiation cell line 1 (**D1**) (Tasker *et al*., 2000); Monomorphic *T. b. brucei* EATRO 2340 (stabilate GUP 2965) (**M**); Stumpy form *T. b. brucei* EATRO 2340 (stabilate GUP 2962) (**St**); and procylic forms (**P**). The gels were stained with ethidium bromide to ensure equal loading (EtBr), and the Northern blots were carried out with probes against ESAG9-K9, ESAG9-K69, and α tubulin as a loading control.

3.1.3 These developmentally regulated genes are Expression Site Associated Gene 9 (ESAG9) genes

The genome of *T. b. brucei* strain TREU 927/4 (which will subsequently be referred to as *T. b. brucei* 927) has been sequenced and was published in 2005 (Berriman *et al*., 2005), and the annotated genome is available at GeneDB (www.genedb.org). The sequence of the clone K9 is present in the *T. b. brucei* 927 genome (accession number: Tb927.7.170), and is annotated as an Expression Site Associated Gene 9 (ESAG9). ESAG9 was first named in *T. equiperdum*: there are two characterised ESAG9 genes in this closely related species, ESAG9c being found within a VSG expression site and ESAG9u found outwith an expression site (Florent *et al.,* 1991). No function has been ascribed to these genes, but they are known to be transcribed. *T. b. brucei*, *T. b. gambiense*, and *T. b. rhodesiense* are very closely related sub-species that are morphologically indistinguishable, but have different disease-causing characteristics. *T. equiperdum* is a parasite of horses that is morphologically indistinguishable from *T. brucei*, but has a different geographic range and a truncated life cycle. It has been argued by some that *T. equiperdum* should in fact be designated as a sub-species of *T. brucei*, for example see Lai *et al.* (Lai *et al.,* 2008).

There are nine intact ESAG9 genes annotated in the *T. b. brucei* 927 genome, including ESAG9-K9, and ten pseudogenes. ESAG9-K69 is not present, but it can be termed an ESAG9 as it is as similar to the other ESAG9 genes as they are to each other (for an alignment of ESAG9 protein sequences, see Figure 3.9). All of the ESAG9 genes in *T. b. brucei* 927 are chromosome internal or sub-telomeric, and this reflects the fact that telomeres are not represented in the bacterial artitifical chromosome (BAC) library used for genome sequencing. However sequencing of the telomeres of *T. b. brucei* strain Lister 427 (which is more commonly used in laboratory experiments than 927) has been carried out with clones generated by transformation-associated recombination (TAR) cloning. Only one expression site in the 427 telomeres has been found to contain an ESAG9 gene (Hertz-Fowler *et al.,* 2008). Sequencing of *T. b. gambiense* strain DAL 927 is in progress and so far one ESAG9 gene has been annotated. An ESAG9 pseudogene has also been identified in a metacyclic expression site in *T. brucei* strain EATRO 795 (Graham *et al.,* 1999).

The precise genomic location of ESAG9-K69 in *T. b. brucei* strain EATRO 2340 is not known as this strain has not been sequenced. However, a 200-nucleotide cDNA clone from *T. b. rhodesiense* strain WRATat1.1 (NCBI accession W06557) has been annotated as an ESAG9 (Djikeng, Donelson and Majiwa, unpublished data) and bears 87% similarity to the ESAG9 Tb927.5.4620, and 59% similarity to ESAG9-K69.

3.1.4 A number of ESAG9 mRNAs are expressed in a stage-specific manner in *T. b. brucei*

To explore whether other ESAG9 genes were expressed in a stage-specific manner, Northern blots were performed with mRNA from two parasite strains: *T. b. brucei* EATRO 2340 (previously thought to be *T. b. rhodesiense* but tentatively renamed by the Rudenko group (Young et al. 2008) and *T. b. brucei* AnTat 1.1. Probes for K9 and K69 were used, as well as for three other ESAG9s: Tb927.5.4620, Tb09.v1.0330 and Tb927.3.5790 (an ESAG9 pseudogene which has a stop codon around two thirds of the way through the sequence). In addition, a control gene, ESAG2, was also assayed for its expression. This Northern blot is shown in Figure 3.2 (Professor Keith Matthews, unpublished data) and shows three interesting features of ESAG9 gene expression in EATRO 2340 parasites: (i) all ESAG9 transcripts are enriched in stumpy form parasites; (iii) both the strains express more than one ESAG9 gene at the same time; and (iii) the two strains express a different repertoire of ESAG9 genes. The differential patterns in the two strains is due to the presence or absence of these genes in the strains (Professor Keith Matthews, unpublished data), rather than the silencing of certain genes. ESAG2 was not expressed in a stumpy-specific manner, nor was another control gene, Pk2.7 (a protein kinase). AnTat 1.1 slender RNA was not used in this assay so it was not possible to confirm the enrichment of ESAG9 transcripts in stumpy forms in this strain.

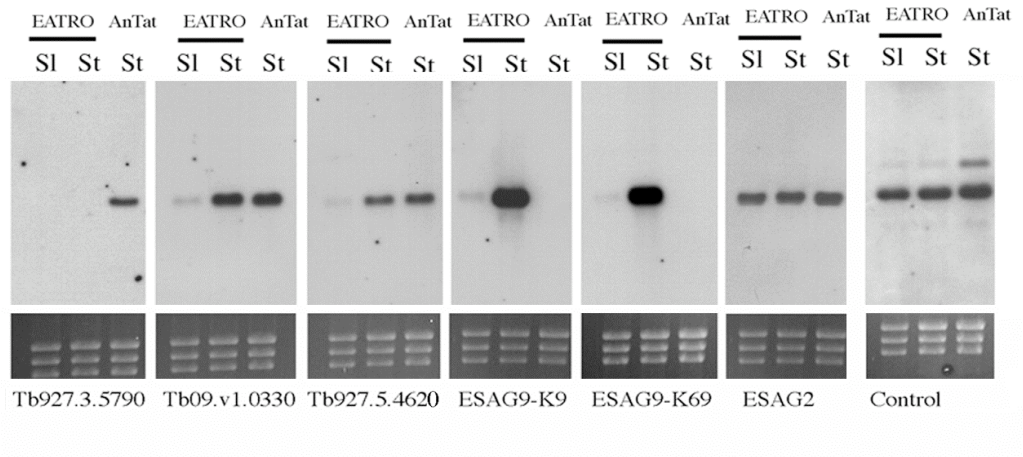


Figure 3.2: Northern blot showing ESAG9 mRNAs up-regulated in stumpy form cells. RNA was made from *T. b. brucei* EATRO 2340 slender and stumpy cells, and *T. b. brucei* AnTat 1.1 stumpy cells. Probes were used against ESAG9s: Tb927.3.5790 (pseudogene), Tb.v1.0330, Tb927.5.4620, ESAG9-K9, ESAG9-K69; ESAG2; and a control RNA (Pk2.7). The top panels are the Northern blots, and the bottom panels are ethidium bromide-stained agarose gels showing the ribosomal subunits to confirm equal loading of each lane.

3.1.5 A peptide antibody raised against ESAG9-K9 reveals that the protein is primarily expressed in differentiating stumpy form cells

An anti-peptide antibody was raised against ESAG9-K9 in rabbits using the following peptides, which are highlighted in Figure 3.3: C-TGPSKTVRRSNSVTS and QVHDGEQRDLEGRGC. The first step of this PhD project was to determine the ESAG9-K9 protein expression in stumpy form cells and during their synchronous differentiation to procyclic form cells using this anti-peptide antibody. Hence, stumpy form cells were harvested from a mouse and caused to initiate differentiation *in vitro* by a temperature drop to 27°C and addition of cis-aconitate (to a final concentration of 6mM), as shown in the schematic diagram in Figure 3.4.

```
MLKVAVTALL TLRSCAGPAE SRGTSLKVGT
GGVGVTVSTY SGCSESWSPD GRNMVCGSPS
TGPSKTVRRS NSVTSTPQKI QGQPVSGSIA
GAKRQEMQTA APSSGGAPHL TSVTLDNPRE
GKTQTKNTSD AVIRGPTERL TDLSSRGEPA
HAPVGNGKAT ERVKQVHDGE QRDLEGRGCK
DQSPARSEPA SGLMRDDGMD VSRPPARRTV
SGAEQESTRE LMTRNLSEQE SAENSTQEKK
SAANKGHAVM ISAALTLISF
```

Figure 3.3: The location in the K9 protein sequence of the peptides used to raise the K9 antipeptide antibody. The antibody was raised against both peptides in two rabbits and affinity purified.

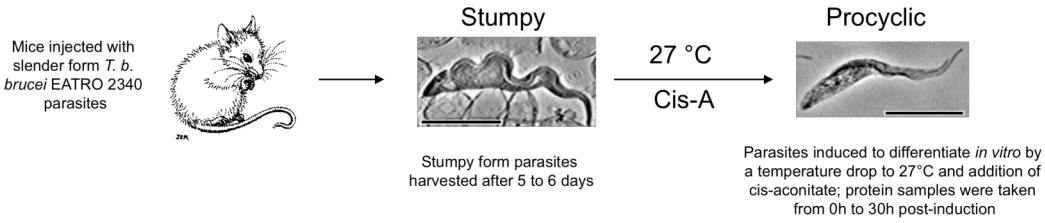


Figure 3.4: Schematic diagram of *in vitro* differentation of stumpy cells. Stumpy form parasites were harvested from a mouse and induced to differentiate to procyclic forms *in vitro* via the addition of 6mM cis-aconitate and a drop in temperature to 27°C. Protein samples were taken over a time-course from 0 hours to 30 hours.

Protein samples were prepared over a time-course of differentiation from 0h to 30h and analysed by Western blotting, as shown in Figure 5. This Western blot showed that the ESAG9-K9 had an interesting and unusual protein expression profile. The protein was expressed at a low level at 0h (i.e. in stumpy bloodstream form cells) but increased until expression peaked from 6 to 9 hours after the start of differentiation to procyclic form cells, after which the protein was reduced to undetectable levels. This profile was suggestive of ESAG9-K9 protein performing a function specifically during differentiation from stumpy to procyclic form. Comparison with the size markers indicated that ESAG9-K9 protein migrated at approximately 37kDa, which is 10kDa higher than its predicted protein mass of 27.1kDa as annotated in the GeneDB database. This could be due to post-translational modifications.

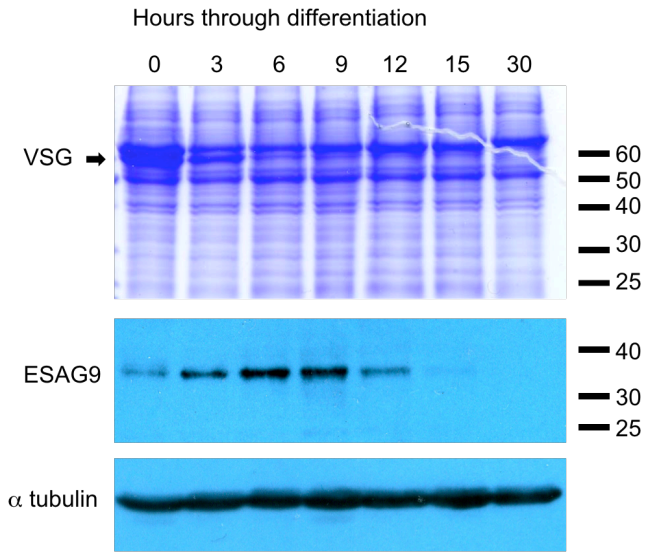


Figure 3.5: Western blot showing the expression of ESAG9-K9 protein during a 30-hour differentiation time course. The top panel shows a SDS-PAGE gel stained with Coomasie's stain with the band for VSG being indicated with an arrow. The middle panel is a Western blot generated using the ESAG9-K9 anti-peptide antibody. The bottom panel is a Western blot probed for α tubulin as a control for equal loading of lanes. The size markers indicated are in kilo-Daltons.

3.1.6 ESAG9-K9 protein is located intracellularly and sometimes in the flagella pocket

Using the ESAG9-K9 antibody, immunofluorescence was carried out on stumpy forms 6 hours into differentiation to procyclic forms (i.e. at the peak of ESAG9-K9 protein expression). Figure 3.6 panel A shows the signal from the ESAG9-K9 antibody when reacted with a mixed population of slender (non-differentiating) and stumpy (differentiating) cells. The cells were counter stained for EP procyclin expression to indicate those which had initiated differentiation to procyclic forms. It was clear that the cells staining positive for EP procyclin also had a much stronger signal with the ESAG9-K9 antibody. Although the EP-negative cells also showed a weak signal from ESAG9-K9, this was because this antibody gives a certain degree of background staining, as demonstrated when reacted against procyclic form cells (data not shown).

Figure 3.6 panel B shows two differentiating stumpy cells at a higher magnification. The ESAG9-K9 protein was located intracellularly, in a loosely punctate fashion with a halo around the nucleus. This could represent a signal from the endoplasmic reticulum, though the punctate spots within the cytoplasm could also be small vesicles such as lysosomes or glycosomes. Also a signal was sometimes observed at the flagellar pocket, which is the site of endo- and exocytosis in trypanosomes. EP procyclin showed a peripheral staining around the cell due to its surface location. Further analysis of the cellular distribution of ESAG9-K9 is detailed in Chapter 4.

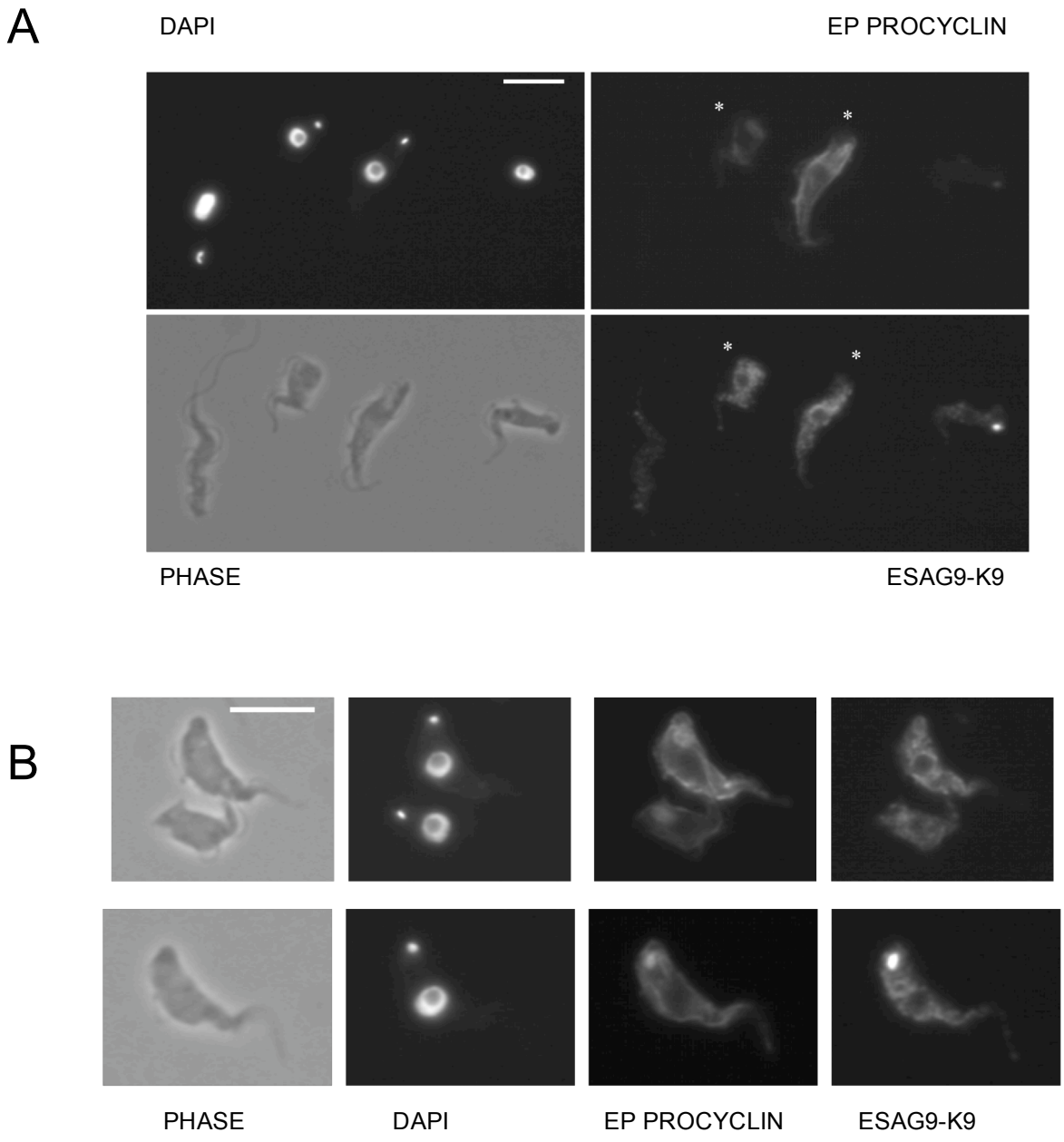


Figure 3.6: Immunofluorescent staining of stumpy form cells six hours into differentiation. *T. b. brucei* EATRO 2340 stumpy bloodstream-form cells were induced to differentiate with addition of cis-aconitate and a temperature drop to 27°C and stained with DAPI, to visualise the DNA, EP-procyclin antibody detecting the procyclic surface-coat molecule, and the ESAG9-K9 anti-peptide antibody. The phase images and the different channels are shown as labelled. Panel A shows four cells, two of which are procyclin-positive and differentiating, and these two cells are marked with an asterisk to distinguish them from the two cells which are not differentiating. Panel B shows differentiating, procyclin-positive cells, at a higher magnification. Scale bars represent 15 microns.

3.2 Bioinformatic analysis of ESAG9 sequences

3.2.1 Introduction

ESAG9 genes were shown to be transcribed, and translated in the case of ESAG9-K9, in a stage specific manner, with enrichment during differentiation. To date, no other genes have been identified with a similar expression profile in *T. brucei*.

The first step in the process of exploring the potential function of ESAG9 proteins was to analyse the DNA and protein sequences of the genes using bioinformatic comparison. Bioinformatic analysis is the process of carrying out *in silico* experiments on sequence data using algorithms to find or predict certain attributes of that data. Many algorithms have been written for this purpose and these are incorporated into tools that are freely available online and are straightforward to use. When utilised in conjunction with the genome database information which is also available online for many organisms, this represents a vast resource of information for the biologist. The databases and programmes used, and the web addresses at which they can be found, are summarised in Table 3.1.

Initially, the degree of conservation in the ESAG9 family was investigated. ESAG9 sequences were analysed for the presence of functional motifs, predicted post-translational modification sites, and their similarity to other proteins of known function. In addition, the 3'UTR sequences of ESAG9 mRNAs were searched for motifs that might be involved in their post-transcriptional regulation of gene expression. The analyses carried out, and the bioinformatics programmes used, will be discussed in more detail in the following sections.

Database or programme used	Brief description of purpose	Web address	Literature reference where available
GeneDB	Online genomic database for many organisms including *T. brucei*	www.genedb.org	N/A
ClustalW	Performs alignments of multiple DNA or protein sequences	www.ebi.ac.uk/Tools/clustalw2	Thompson *et al*., 1994, Larkin *et al*., 2007
BLAST at NCBI	Multi-organism BLAST and PSI-BLAST	http://blast.ncbi.nlm.nih.gov/Blast.cgi	Altschul *et al*., 1990 Altschul *et al*., 1997
Pfam	Protein families database	http://pfam.sanger.ac.uk/	Sonnhammer *et al*., 1997
Prosite	Protein families database, including pattern and profile information	http://expasy.org/prosite/	Hulo *et al*., 2007
SMART	Identifying protein domains and analysing domain architectures	http://smart.embl-heidelberg.de/	Letunic *et al*., 2006
TMHMM v2.0	Predicting transmembrane domains in proteins	http://www.cbs.dtu.dk/services/TMHMM-2.0/	Sonnhammer *et al*., 1998
SignalP v3.0	Predicting signal peptide cleavage sites	http://www.cbs.dtu.dk/services/SignalP/	Bendtsen *et al*., 2004
Big-PI	Prediction of GPI-anchor addition	http://mendel.imp.ac.at/gpi/gpi_server.html	Eisenhaber *et al*., 1999
GPI-SOM	Prediction of GPI-anchor addition	http://gpi.unibe.ch/	Unpublished
NetNGlyc	Prediction of N-glycosylation sites in human proteins	http://www.cbs.dtu.dk/services/NetNGlyc/	Unpublished
MEME	Motif identification tool	http://meme.sdsc.edu/meme/intro.html	Bailey and Elkan, 1995
S-fold	Statistical RNA folding	http://sfold.wadsworth.org/index.pl	Ding *et al*., 2004

Table 3.1: Bioinformatics tools used and the web addresses at which they can be found. The bioinformatics tools used in this chapter are listed, along with a brief description of their purpose, the web address at which they can be accessed and the literature reference if the work has been published.

3.2.2 The genomic context of ESAG9 genes

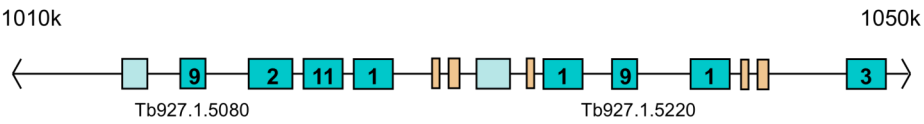
The diverse family of Expression Site Associated Genes, or ESAGs, are so called because of their genomic location (Cully *et al.,* 1986; Cully *et al.,* 1985b). They are located upstream of the single VSG gene found in telomeric expression sites and are co-transcribed by RNA polymerase I with VSG genes as part of a polycistronic unit in the active expression site. Many copies of ESAG genes are also found in sub-telomeric and chromosome internal positions, as is the case for VSG genes. The

ESAG9 genes in the *T. b. brucei* genome were named as such due to their similarity to two genes in *T. equiperdum*, one of which was found in a telomeric expression site, as discussed in section 1.1.2. Because the regions at or near the telomeres are poorly represented in the BAC sequencing, only ESAG9s within the chromosomes have so far been identified in the *T. b. brucei* strain 927 genome. The telomeres of chromosomes 4 and 6 from the genome reference strain are being sequenced but there are no reads available as yet. However, sixteen bloodstream form expression sites have been sequenced from *T. b. brucei* Lister 427 and of these only one contains a copy of ESAG9 (Becker *et al.,* 2004; Hertz-Fowler *et al.,* 2008).

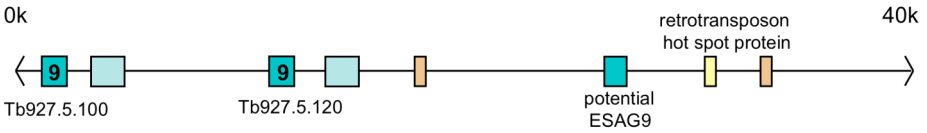
Although ESAG9 members are not ubiquitous ES components, the naming of the ESAG9 genes as ESAGs remains appropriate. This is because ESAG9s are always located in the genome near other VSG and ESAG genes and this is unlikely to be coincidental. This suggests that, even though most VSG expression sites that have been sequenced do not contain an ESAG9, there is some adaptive link between ESAG9 genes and other ESAGs and VSG genes.

Figure 3.7 shows the genomic context of each of the nine intact ESAG9 genes in the *T. b. brucei* 927 genome database. Figure 3.8 shows the position of the single ESAG9 gene annotated in the *T. b. brucei* strain Lister 427 genome, which is in BES1.22. In each case, an approximately 40 kilobase region is represented schematically, showing what other genes are near to each ESAG9. They are, in every case, found with VSG genes either as the adjacent gene or close by. Further, they are also often near other ESAGs, specifically ESAG1 (most frequently), ESAG2, ESAG3 and ESAG11. ESAG1 is a membranous glycoprotein (Cully *et al.,* 1986) of unknown function (Carruthers *et al.,* 1996); there is no published experimental data regarding the function of ESAGs 2 and 3; and ESAG11 is suggested to be a cell surface protein though this has not been demonstrated experimentally and its function is not known (Redpath *et al.,* 2000).

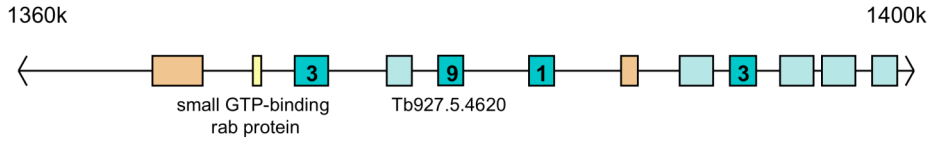
▸ **Tb927.1.5080 and Tb927.1.5220 location on chromosome 1:**



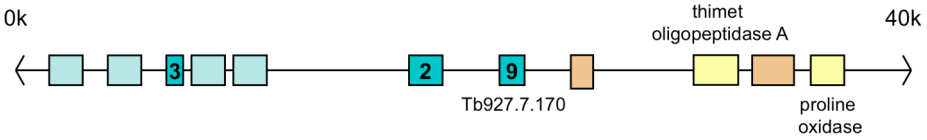
▸ **Tb927.5.120 and Tb927.5.100 (pseudogene) locations on chromosome 5:**



▸ **ESAG9-EQ (Tb927.5.4620) location on chromosome 5:**



▸ **ESAG9-K9 (Tb927.7.170) location on chromosome 7:**



▸ **Tb09.v1.0330, Tb09.160.5430, and Tb09.160.5400 locations on chromosome 9:**

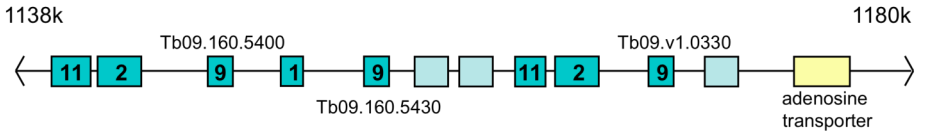
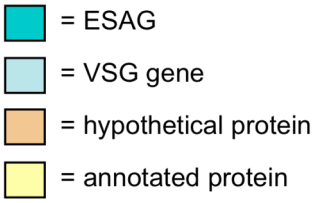


Figure 3.7: The genomic locations of ESAG9 genes in *T. b. brucei* strain TREU 927. The genomic locations of all ESAG9 genes and one pseudogene are shown, excepting Tb11.1000 which is on chromosome 11 but in an unordered contig so the precise location is not defined. The key indicates which colour relates to which type of gene. The diagrams are schematic and not exactly to scale.



- **Tb427.BES122.10 location in Bloodstream Form Expression Site:**

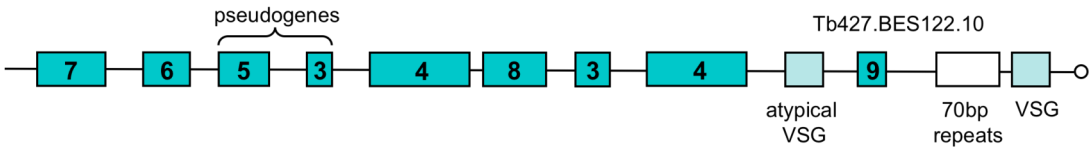


Figure 3.8: The genomic context of an ESAG9 gene in a bloodstream form expression site (BES122) in *T. b. brucei* strain 427. See Figure 3.7 for a key of the colours. The ESAG9 gene is the last gene before the 70 base-pair repeats, which are followed by the VSG gene and the chromosome end.

3.2.3 Alignment of ESAG9 gene family members

As detailed above, there are nine ESAG9 genes, and ten ESAG9 pseudogenes, in the *T. b. brucei* 927 genome. There is also one copy of ESAG9 in the Lister 427 expression sites, one so far sequenced in the *T. b. gambiense* genome, two copies in *T. equiperdum*, and the version ESAG9-K69 previously identified by members of the Matthews lab in *T. b. brucei* strain EATRO 2340. These genes combined constitute a group of 14 genes (excluding pseudogenes) in *T. b. brucei*, *T. b. gambiense* and *T. equiperdum*. *T. equiperdum* is very closely related to *T. brucei* and it has been argued that *T. equiperdum* should in fact be designated a subspecies of *T. brucei*, as phylogenetic analysis did not result in *T. equiperdum* sequences forming a monophyletic group (Lai *et al.,* 2008). A starting point for exploring the attributes of this protein family was to look at how similar the genes are to each other. This can be achieved using an alignment tool called Clustal W 2.0 (Larkin *et al.,* 2007; Thompson *et al.,* 1994). The Clustal W programme performs an alignment by carrying out three steps which consist of, in simple terms: aligning each possible pair of sequences to see how divergent they are from one another, building a phylogenetic

tree based on this information and aligning all the sequences in order based on how they branch in the tree.

The fourteen ESAG9 protein sequences were analysed using Clustal W and Figure 3.9 shows the phylogenetic tree generated by this programme. Phylogenetic trees are used to infer evolutionary relationships between sequences and species. This un-rooted tree indicates that Tb927.5.4620 is most closely related to the ESAG9c gene in *T. equiperdum*; this gene was therefore named ESAG9-EQ. The ESAG9c and ESAG9u genes are 67% similar to each other, and are 84% and 67% similar to ESAG9-EQ respectively. These three genes formed a cluster with Tb427.BES122.10 which is in a bloodstream form expression site in strain Lister 427. This tree also shows that the ESAG9 copy detected in *T. b. gambiense* is most closely related to Tb927.1.5080; these genes are 42% similar at the amino acid level.

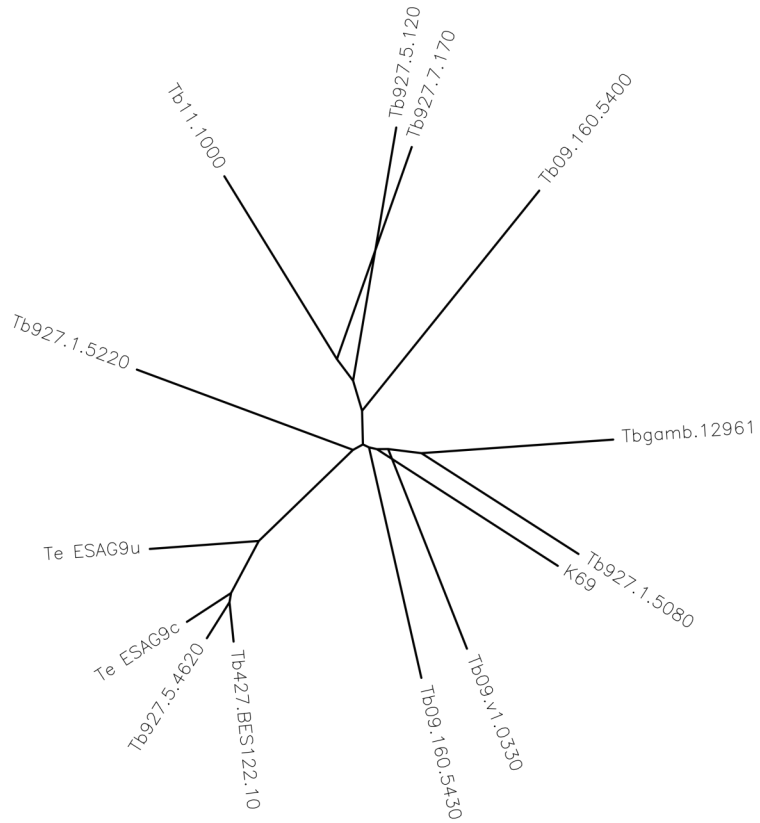


Figure 3.9: Phylogenetic tree of ESAG9 protein sequences. ESAG9 protein sequences from *T. b. brucei, T. b. gambiense,* and *T. equiperdum* were subjected to phylogenetic analysis using the ClustalW programme. This is an un-rooted tree where the length of the branches are proportional to the evolutionary distance between divergence events.

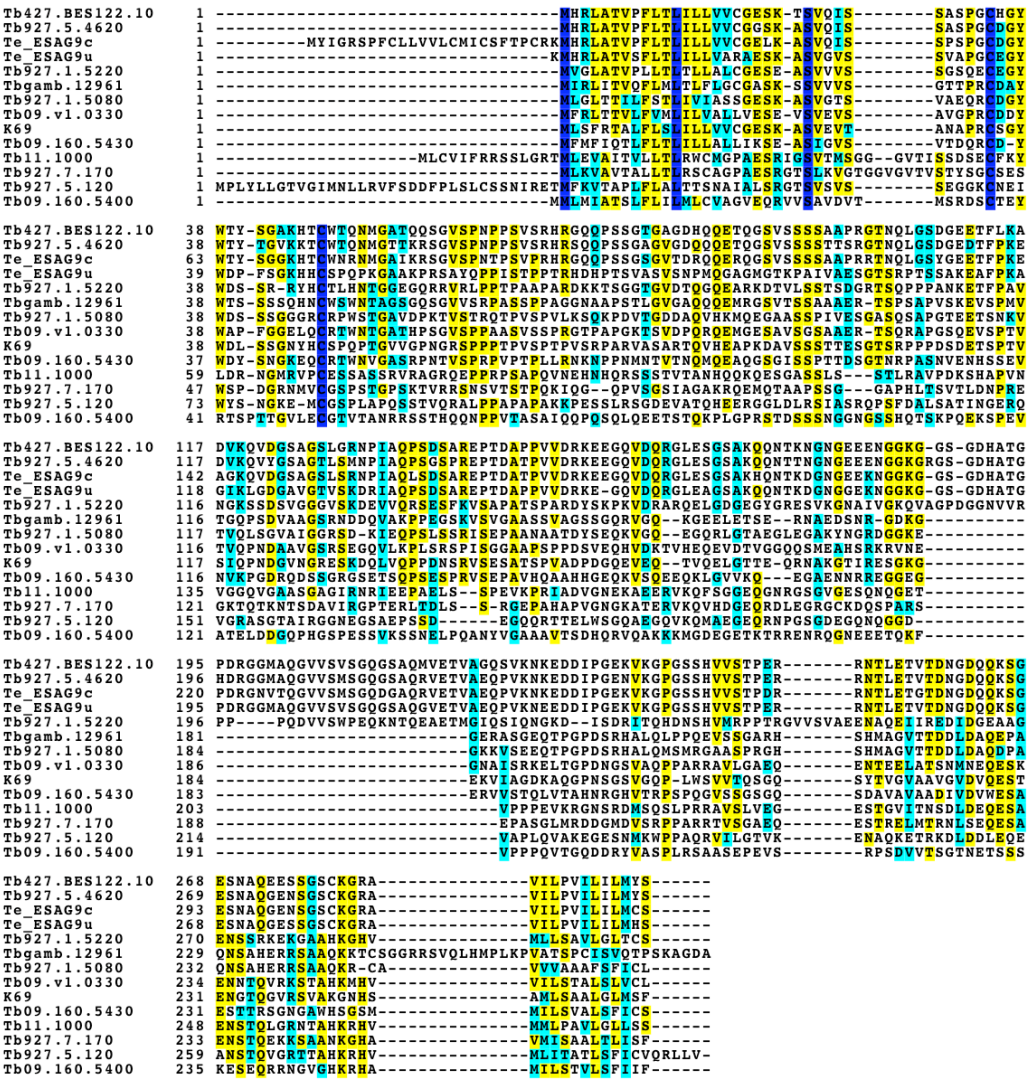


Figure 3.10: ClustalW alignment of the ESAG9 protein sequences in *T. b. brucei, T. b. gambiense* and *T. equiperdum*. A blue background indicates that the residue is conserved in all sequences; a yellow background indicates residues that are identical but not present in all sequences; and a turquoise background indicates residues that are similar (for example acidic residues).

The Clustal W alignment of the identified ESAG9 sequences is shown in Figure 3.10. Amino acid residues that were conserved in all sequences were coloured in blue, residues that were identical but were not in every sequence were coloured in yellow, and resides that were similar were coloured in turquoise. There is a large degree of divergence between the protein sequences. The sequences are more similar to each other at the N terminus, and to some degree at the C terminus, but in between there is little similarity, excepting *T. equiperdum* ESAG9c and ESAG9u, and *T. b. brucei* Tb927.5.4620, which are very closely related. The gene Tb927.5.4620 will subsequently be referred to as ESAG9-EQ in this thesis due to it being the ESAG9 gene in the *T. brucei* genome that is most similar to the ESAG9 in an expression site in the *T. equiperdum* genome.

The greater degree of similarity between the protein sequences at the N terminus could be explained by the presence of a signal peptide, and this will be discussed in section 3.2.9. However there are also five highly conserved residues that are not at the immediate N terminus, and these are: a serine (S, amino acid 25 in ESAG9-K9), a cysteine (C, aa43 in EASG-K9), a tyrosine (Y, aa46 in ESAG-K9), a tryptophan (W, aa47 in ESAG-K9), and another cysteine (aa56 in ESAG-K9). Not all sequences contain the conserved tyrosine and tryptophan but they are present in at least 12 out of 14 sequences.

3.2.4 A common motif in ESAG9 sequences is similar to a metal-binding domain in Protein Phosphatase 2 C enzyme

To explore whether the conserved residues discussed in section 3.2.3 are found as a motif in other proteins, a motif search of the *T. b. brucei* genome was carried out using the GeneDB website. The motifs used to search the genome were as follows, where X represents any amino acid: either SX_9, SX_{10}, SX_{15}, or SX_{17}, followed by either $CXYWX_8C$, CX_2YWX_8C or CX_3WX_8C.

Most of the combinations only recognised either other ESAG9 genes or pseudogenes, or hypothetical proteins of unknown function. One interesting exception to this is that the motif $SX_{15}CX_2YWX_8C$ is also found in a putative zinc-finger domain protein (accession: Tb927.3.4220). This series of 30 amino acids

overlaps with the cysteine-rich C3HC4 zinc-finger domain (Borden & Freemont, 1996) that is annotated in the database entry for this gene. However the ESAG9 sequences do not contain the other residues required for a zinc-finger so this hit may be coincidental.

More proteins were recognised if the serine residue was excluded from the motif. The motif CX_3WX_8C is found in a putative dynein heavy chain (Tb10.70.1720), a putative protein kinase (Tb11.01.2900), and in protein phosphatase 2C (PP2C; Tb10.70.2270). A literature search revealed that some of the conserved amino acids in this motif are vital for the activity and metal binding properties of mouse and human PP2C (Das *et al.,* 1996; Kusuda *et al.,* 1998), so an alignment of this domain in ESAG9s and in PP2C sequences from trypanosomatids, *Schizosaccharomyces pombe*, *Homo sapiens* and *Mus musculus* was performed, shown in Figure 3.10.

Amino acid substitution mutants have been made in *M. musculus* PP2C and the substitution of the first aspartic acid (D) in the consensus sequence, the tryptophan (W), and the second aspartic acid were shown to significantly decrease or completely ablate the activity of the enzyme (Kusuda et al., 1998). These residues are marked with black asterisks in Figure 3.10. The first aspartic acid, marked with a red asterisk in Figure 3.10, is predicted to bind Magnesium or Manganese ions in the *H. sapiens* protein phosphatase 1A (PPM1A; formerly named PP2C), and the glycine also marked with a red asterisk, is predicted to assist in this binding by maintaining the tertiary peptide structure (Das et al., 1996). However these residues are only present in some of the ESAG9 sequences, for example they are both present in ESAG9-EQ (Tb927.5.4620), but ESAG9-K9 (Tb927.7.170) and ESAG9-K69 both lack the aspartic acid. Due to the lack of overall homology between the ESAG9 and PP2C sequences it is not feasible that the ESAG9s are enzymically active, but it is perhaps possible that those ESAG9s with the relevant aspartic acid could bind Magnesium or Manganese ions. However it is worth noting that other parts of the PP2C protein are involved in creating this metal-binding site. Figure 3.11 shows that in the *H. sapiens* protein the Manganese ions interact not only with the ASP 239, which is found in the short conserved motif also present in ESAG9 proteins, but also with other parts of the PPM1A (formerly PP2C) protein which are not conserved in ESAG9 proteins.

```
Tb427.BES122.10  -SSASPGCHGYWTYSGAKHTCWTQNMGATQQS
Tb927.5.4620     -SSASPGCDGYWTYTGVKKTCWTQNMGTTKRS
Te_ESAG9c        -SSPSPGCDGYWTYSGGKHTCWNRNMGAIKRS
Tb09.160.5430    -SVTDQRCD-YWDYSNGKEQCRTWNVGASRPN
Tb927.1.5080     -SVAEQRCDGYWDSSSGGGRCRPWSTGAVDPK
Tb09.v1.0330     -SAVGPRCDDYWAPFGGELQCRTWNTGATHPS
Tbgamb.12961     -SGTTPRCDAYWTSSSSQHNCWSWNTAGSGQS
Te_ESAG9u        -SSVAPGCEGYWDPFSGKHHCSPQPKGAAKPR
K69              -TANAPRCSGYWDLSSGNYHCSPQPTGVVGPN
Tb927.1.5220     -SSGSQECEGYWD-SSRRYHCTLHNTGGEGQR
Tb927.7.170      VSTYSG-CSESWSPDGRNMVCGSPSTGPSKTV
Tb927.5.120      -SSEGGKCNEIWYSNGKEM-CGSPLAPQSSTV
Human_PPM1A      -QFIILACDGIWDVMGNEELCDFVRSRLEVTD
Mouse_PP2C       -EFVVLACDGIWDVMSNEELCEFVKSRLEVSD
T_cruzi_PP2C     -AFIVIACDGVWDVLSNEECCDLVKKGLKETD
L_major_PP2C     -EFIVIGCDGIWDVLSNEECCDLVKTLIQNSD
Tb_PP2C          -AFVVVACDGVWDVLSNDDCCQLIHQSFKDTE
S_pombe_PP2C     -EFLILACDGIWDCKSSQQVVEFVRRGIVARQ
consensus        -s----gCdgyWd--s----c----tg-----
                       **
                       *  **
```

Figure 3.10: ClustalW alignment of the conserved motif that is found in ESAG9 proteins and Protein Phosphatase 2C (PP2C). The top 12 sequences are the ESAG9s in *T. b. brucei* and *T. equiperdum*. The bottom six sequences are PP2Cs in *Homo sapiens, Mus musculus*, *Trypanosoma cruzi*, *Leishmania major*, *T. b. brucei*, and *Schizosaccharomyces pombe*. The bottom line shows the consensus sequence with residues that are present in all sequences indicated as a capital letter, resides that are present in most sequences indicated as a lower case letter, and dashes indicating where there is no conservation. The residues marked with red asterisks are involved in binding metal ions in *H. sapiens* PPM1A (Das et al., 1996). The residues marked with black asterisks are required for enzymic activity of PP2C in *M. musculus* (Kusuda et al., 1998).

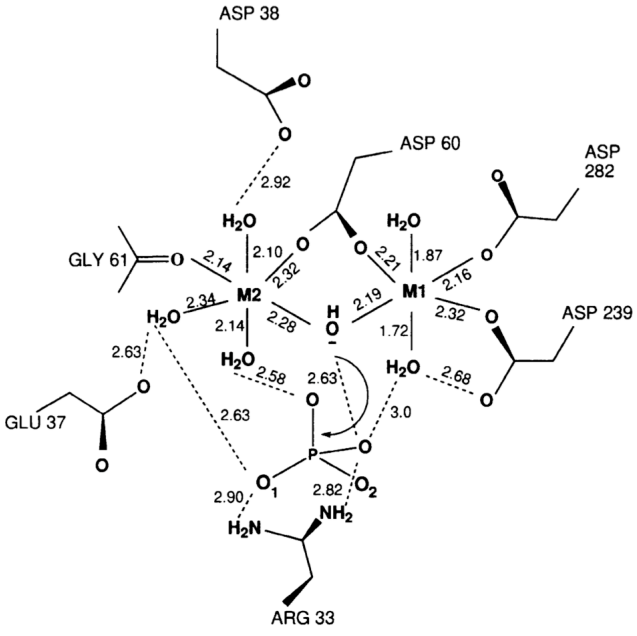


Figure 3.11: The metal-binding site in *Homo sapiens* PPM1A protein. The two Manganese ions, M1 and M2, are shown interacting with the surrounding amino acids. The ASP 239 residue is found in the short conserved motif which is also present in ESAG9 protein sequences. This diagram is from Das *et al.,* 1996.

3.2.5 BLAST and PSI-BLAST searches using ESAG9 gene family sequences

As well as searching the genome databases for specific protein motifs, it is possible to search databases for sequences with similarity to the whole protein, or proteins, of interest. The Basic Local Alignment Search Tool, or BLAST, is a tool for searching databases of nucleotide or protein sequences for similarity to the query sequence (Altschul *et al.,* 1990). The presence of similar sequences or genes in the genomes of other organisms can therefore be ascertained. Position-Specific Iterated BLAST, or PSI-BLAST (Altschul *et al.,* 1997) is a more advanced search tool for protein sequences whereby alignments generated by BLAST are used to search the genome database, such that weak similarities between sequences can be identified that would be missed by a normal BLAST search.

Using the 12 ESAG9 protein sequences from *T. b. brucei* and *T. b. rhodesiense* as a query, the genomes of a large number of organisms were searched using either BLAST or PSI-BLAST. PSI-BLAST is accessible via the NCBI (National Centre for Biotechnology Information) server and this incorporates many genome databases, including mammals, plants, fungi and a number of protozoa including *Leishmania spp*., *Giardia intestinalis*, *Eimeria tenella*, *Trichomonas vagnialis*, and *Trypanosoma brucei* (strain 927). However this database does not include the fully annotated genome of the trypanosomatids *T. cruzi*, *T. vivax*, *T. annulata*, *T. congolense*, *T. b. gambiense*, or the Lister 427 strain of *T. b. brucei*. These genomes were therefore searched using the normal BLAST search, which is accessible via GeneDB. The predicted proteins of the salivary glands and midgut of the tsetse fly vector *Glossina mortisans* were also searched via GeneDB.

The PSI-BLAST searches resulted in only a limited number of hits that were better than the default E-value threshold set by the programme. The E-value represents the number of times you would expect to get a hit with the same score by chance in a database of a given size. These hits were unanimously either other ESAG9 genes or pseudogenes and occasionally hypothetical proteins of unknown function, all from the *T. b. brucei* genome. There were no hits above the threshold from proteins found

in any other organisms. Of those hits which occurred which were worse than the E-value threshold, no particular patterns emerged. There were only two instances of an ESAG9 protein hitting a particular type of protein from another organism more than once. ESAG9-K69 and Tb09.160.5430 hit two zinc metalloproteases (accessions: ZP_02711752 and ZP_01818470.1 respectively) in *Streptococccus pneumoniae*; and Tb09.160.5430 and Tb09.v1.0330 hit multidrug resistance proteins in *Plasmodium knowlesi* (accession CAQ41661.1) and *P. vivax* (accession XP_001615913.1) respectively. The two zinc metalloproteases are fully conserved at the N-terminus and are overall 50% similar to each other, and the two multidrug resistance proteins are 78% similar to each other. To verify that these hits were meaningful, reciprocal blasts of the *T. b. brucei* 927 genome were carried out with these four protein sequences as queries. Out of the four reciprocal BLAST searches, only the *S. pneumoniae* zinc metalloprotease ZP_01818470.1 hit any ESAG9 sequences in *T. b. brucei.* It did not hit Tb09.160.5430, which is the sequence that originally selected it, but instead hit ESAG9-EQ, with an E value of 0.22.

PSI-BLAST searches were also carried out with the ESAG9 sequences from *T. b. gambiense* and *T. equiperdum*. Interestingly, the ESAG9u gene in *T. equiperdum* selected two secreted bacterial proteins: a putative secreted hemolysin-type calcium-binding bacteriocin in *Roseobacter sp.* (accession ZP_01058523), and a secreted protein in *Streptomycins sp*. (accession EDX26763). However again the E-values for these hits were above the threshold and so the hits may be spurious. The proteins do not appear to be related to each other and it was not possible to do a reciprocal BLAST of the *T. equiperdum* genome as this species has not been sequenced.

More useful data regarding the potential function of ESAG9 were gained from doing BLAST searches of the other trypanosomatid genomes. Here, a threshold E-value of less than 1×10^{-4} was used to designate a database hit as being significant. This value was chosen because the PSI-BLAST programme designated hits with an E-value in the region of 1×10^{-4} and lower to be significant. Out of the 12 ESAG9 genes used to query the *Trypanosoma cruzi* database, seven of them resulted in one or more hits to *T. cruzi* proteins with an E-value better than the threshold. These were unanimously either mucin-associated surface proteins (MASPS), mucins, or mucin-like

glycoproteins. The results of this BLAST search are summarised in Table 3.2, where either all the hits for a given ESAG9 protein sequence are shown, or the top five if there were more than five.

To verify these data, reciprocal BLASTS were carried out using the highest scoring mucin or MASP hit from each ESAG9 protein sequence query to BLAST the *T. brucei* strain 927 genome. In eight out of eleven reciprocal BLASTS, at least one ESAG9 sequence was pulled out, though often with a low E value. An ESAG9 (ESAG9-EQ) was the highest scoring hit for only two of these BLASTS.

BLAST searches of trypanosomatids also resulted in hits to *T. vivax* proteins that were better than the threshold E-value: ESAG9-EQ (Tb927.5.4620) and Tb427.BES122.10 both hit a 347 amino-acid *T. vivax* protein (accession tviv651a01.q1k_1) with E values of 1.1×10^{-8} and 2.1×10^{-7} respectively. However this was a hypothetical protein with no useful annotation assigned to it, and the degree of similarity to the ESAG9s was low. No other kinetoplastids appeared to have any related sequences, nor did the tsetse fly vector *G. morsitans*.

ESAG9 gene name and/or accession	Number of hits	Accession of all hits or top five hits	Description	E-value
Tb09.v1.0330	1	Tc00.1047053504081.350	Mucin-associated surface protein (MASP), putative	5×10^{-5}
Tb09.160.5400	9	Tc00.1047053510693.130	MASP, putative	1.5×10^{-6}
		Tc00.1047053508743.10	Mucin-like glycoprotein, putative	5.7×10^{-6}
		Tc00.1047053508741.440	MASP, putative	5.7×10^{-6}
		Tc00.1047053506245.200	MASP, putative	1.3×10^{-5}
		Tc00.1047053506501.220	MASP, putative	2.2×10^{-5}
Tb11.1000	1	Tc00.1047053504039.130	MASP, putative	9.7×10^{-5}
ESAG9-EQ Tb927.5.4620	17	Tc00.1047053508365.200	MASP, putative	1.2×10^{-7}
		Tc00.1047053508295.40	MASP, putative	7.7×10^{-7}
		Tc00.1047053509527.80	MASP, putative	1.7×10^{-6}
		Tc00.1047053507957.320	MASP, putative	4.2×10^{-6}
		Tc00.1047053508761.90	MASP, putative	4.7×10^{-6}
Tb927.1.5220	3	Tc00.1047053503859.40	MASP, putative	1.4×10^{-6}
		Tc00.1047053510377.430	MASP, putative	1.8×10^{-5}
		Tc00.1047053510025.160	MASP, putative	5.2×10^{-5}
ESAG9-K69	1	Tc00.1047053506769.120	MASP, putative	8.6×10^{-5}
Tb427.BES122.10	4	Tc00.1047053508501.140	Mucin, TcMUCII, putative	1×10^{-5}
		Tc00.1047053511255.80	MASP, putative	3.7×10^{-5}
		Tc00.1047053507957.320	MASP, putative	7.8×10^{-5}
		Tc00.1047053508495.50	Mucin, TcMUCII, putative	9.4×10^{-5}

Table 3.2: The results of BLAST searches of the *T. cruzi* genome using ESAG9 protein sequences. Hits are only shown that were better than the E-value cut off of 1×10^{-4}. If there were more than five hits with an E-value better than the threshold, then the accession numbers of the top five hits only are listed.

3.2.6 ESAG9 gene family sequences show some similarity to mucin-associated surface proteins

The outcome of the BLAST searches of the *Trypanosoma cruzi* genome, as described in section 3.2.5, were the most compelling and least ambiguous results of the database searches carried out. Mucin-associated surface proteins (MASPs) were the most frequent hits. *T. cruzi* has an extensive and fairly well-characterised array of GPI-anchored and glycosylated surface proteins, including mucin-like glycoproteins (TcMUC), the small mucin-like gene family (TcSMUG) and MASPs, and these represent an impressive 6% of all *T. cruzi* genes (Buscaglia *et al.,* 2006). Different sub-families of these proteins cover the cell surface in both the insect and mammalian stages of the life cycle and are proposed to have different roles in these stages. Trypomastigotes are the extracellular stage of development in the mammal host and their cell surface is covered with TcMUC proteins which, when sialylated, are involved in protecting the parasites from antibody-mediated lysis (Pereira-Chioccola *et al.,* 2000) The family of MASP genes is extensive and contains over 1300 genes (El-Sayed *et al.,* 2005); members of this family have been shown to be membrane associated and glycosylated by N-linked glycosylation (Atwood *et al.,* 2006) but the specific function of MASPs has not yet been elucidated. The insect stages of the *T. cruzi* life cycle also express a variety of mucin or mucin-like molecules on their surface, including TcSMUGs and SAPs. The SAP (serine, alanine, and proline rich protein) family of proteins, which bears some resemblance to MASPs and mucins, may play a role in the invasion of mammalian cells by metacyclics, and are also shed by metacyclics (Baida *et al.,* 2006).

The similarity of ESAG9s to these GPI-anchored, N-glycosylated surface proteins could provide an insight into the potential function of ESAG9s. To explore the homology between ESAG9s and these *T. cruzi* surface proteins, a ClustalW alignment was carried out with ESAG9-EQ (which hit the most MASPs in the *T. cruzi* genome), and the top three MASPs that it hit (see Figure 3.12). The ESAG9-EQ sequence was only between 14% and 19% similar to the MASP sequences, and was shorter at only 296 amino acids long, compared to the MASPs that are around 427 residues long. Nonetheless, there were some conserved residues shared by the

ESAG9-EQ and MASP sequences. The conserved resides are clustered in specific regions of homology rather than scattered through the length of the alignment. The MASP sequences do not contain the Cx_3Wx_8C motif that is conserved amongst the ESAG9s.

```
CLUSTAL 2.0.8 multiple sequence alignment


Tc00.1047053508365.200      MAMRMTGRVLLVCALCVLWCVFSSVSADARDDCEDSGQAVVGSVGQPSSP 50
Tc00.1047053509527.80       MAM-MTGRVLLVFVLCVLWCVFSSVSADAGDDCEDSGQAVVGSVGQTSSP 49
Tc00.1047053508295.40       MAM-MTGRVLLVCALCVLWCVFSSVSADARDDFEDSGQAVVGSVGQPSSP 49
Tb927.5.4620                ----MHRLATVPFLTLILLVVCGGSKASVQISSASPG------------- 33
                                *   . :     :*  * .. .*..  .  ..*

Tc00.1047053508365.200      VGGHGDNGVVTDGSSLSSGSHLPGSAPTGGKPPLPGVGVGVGLNESKDAL 100
Tc00.1047053509527.80       VGGHGDNGVVTDGSSLSSGSHLPGSAPTGGKPPLPGVGVGVGLNESKDAL 99
Tc00.1047053508295.40       VAGHGDNGVVTDGSSLSSGSHLPGSAPTGGKPPLPGVGEGVGLNESKDAL 99
Tb927.5.4620                ----CDGYWTYTGVKKTCWTQNMGTTKRSGVSPNP--------------- 64
                                 *.  .  * . :. ::  *::  .* .* *

Tc00.1047053508365.200      TLAQEGREVVQDQHELVPQVSHHTDAGTGGKSDPNAPEQSPNAKTELENG 150
Tc00.1047053509527.80       TLAKEGREVVQDQHELLPHVSHHTDAGTGGKSDPNAPEQSPNAKTELEDG 149
Tc00.1047053508295.40       TLAQEGREVVQDQHELVPQVSHHTDAGTGGKSDPNAPEQSPNAKTELEDG 149
Tb927.5.4620                -----------------PSVSRH--------------------RSQQPSS 77
                                             * **:*                    :::  ..

Tc00.1047053508365.200      GGVCGDDAAAAEGEVKTTVNEGQLGGSSTSDSHSPAADRGSREQEGTVEQ 200
Tc00.1047053509527.80       GGVCGDDAAAAEGEVKTTVNEGQLGGSSTSDSHPPAADRGSREQEGTVEQ 199
Tc00.1047053508295.40       GGVCGDDAAAAEGEVKTTVNEGQLGGSSTSDSHPPAADRGSREQEGTVEQ 199
Tb927.5.4620                GAGVGDQQQETQGSVSSSSTTS----RGTNQLGSDGEDTFPKEDVKQVYG 123
                            *.  **:   ::*.*.:: . .     .*.:  . . *  .:*:   *

Tc00.1047053508365.200      AASTAALTPVVGRETPPKGDTKEDSPEDKAATGAGITQGIPAVSQQQTHS 250
Tc00.1047053509527.80       AASTAALTPVVGRETPPKGDTEEDSPEDNAATGAGITQDIPAVSQQQTHS 249
Tc00.1047053508295.40       AASTAALTPVVGRETPPKGDTKEDSPEDKAATGAGNTQDIPAVSQQQTHS 249
Tb927.5.4620                SAGTLSMNPIAQPSGSPREPTDATPVVDRKEEGQVDQRGLESGSAKQQN- 172
                            :*.* ::.*:.  . .*:  *.  .  *.   *    :.: : * :* :

Tc00.1047053508365.200      PSTSTTGNGLTSTLGKGRAAEDISSNNERSGKALLQEGAQHETVAGSQSQ 300
Tc00.1047053509527.80       SSTPTTGNGLTSTLGKGRAAEEISSNNERSGKALLQEGAQHETVAGSQSQ 299
Tc00.1047053508295.40       SSTPTAGNGLTSTLGKGRAAEDISSNNERSGKALLQEGAEHETVAGSQSQ 299
Tb927.5.4620                ---TTNGNGEEENGGKGRGSGDHATGHDRGG------------------- 200
                               .* ***  .. ****.: : ::.::*.*

Tc00.1047053508365.200      TIPAATARNTLGTTLPGDSDNSTTITTAVRSDTGTEGTPTSNHPNRQSIE 350
Tc00.1047053509527.80       TIPAATARNTLGTTLPGDSDNSTTITTAVRSDTGTEGTPTSNHPNRKSIE 349
Tc00.1047053508295.40       TIPAATARNTLGTTVSGDSDNSTTITTAVRSDTGTEGTPTSNHPNRQSIE 349
Tb927.5.4620                -----MAQGVVSMSGQGSAQRVETVAEQPVKNKEDD------------IP 233
                                  *:..:. :  *.::.  *::    .:.  :            *

Tc00.1047053508365.200      GATSPGMNSDGEAASAKKYDTVSQSAGSSTAPTTNNKTRDTASHGNSDGS 400
Tc00.1047053509527.80       GATSPGMNFDGEAASTKKYDTVSQSAGSSTAPTTNNKTRDTASHGDRDSS 399
Tc00.1047053508295.40       GATSPGMNSDGEAASTKKYDTVSQSAGSSTAPTTNNKTRDTASHGNSDSS 399
Tb927.5.4620                GENVKGPGSSHVVSTPERRNTLETVTDNGDQ----QKSGESNAQGENSGS 279
                            * .  * . .  .::.:: :*:.  :...      :*: :: ::*: ..*

Tc00.1047053508365.200      TAVSHTTSPLLLLVVVTCAAAAAVVAA--- 427
Tc00.1047053509527.80       NAVSHTTSPLLLLLLVACAAAAAVVAA--- 426
Tc00.1047053508295.40       NAVSHTTSPLLLLVVVACAAAPVAVVAGPA 429
Tb927.5.4620                CKGRAVILPVILILMYS------------- 296
                                 .  *::*::: :
```

Figure 3.12: ClustalW alignment of ESAG9-EQ with three *T. cruzi* Mucin-Associated Surface Protein (MASP) sequences. Asterisks indicate a residue that is conserved in all sequences; ":" indicates conserved substitutions, and "." indicates semi-conservative substitutions.

3.2.7 Search for potential functional motifs in ESAG9 gene family members using online databases

There are a number of online databases that contain information about protein families and domains and these can be searched with query protein sequences. These are a useful additional resource to BLAST searches. Whereas a BLAST search will compare the query sequence to individual sequences in a genome, these domain databases contain information about protein families that is generated from large data sets, and can therefore represent a more powerful tool.

The databases queried with ESAG9 protein sequences were: Pfam, Prosite, SMART, TMHMM v2.0, and SignalP v3.0. The web addresses at which these tools can be found are listed in Table 3.1. Pfam is a database of protein families created from alignments using large numbers of proteins, and from hidden Markov models (Sonnhammer *et al.,* 1997), which currently contains information on 9318 protein families (Finn *et al.,* 2008). Prosite is a database of 1319 patterns and 745 profiles which are used to identify protein families (Hulo *et al.,* 2008). SMART (Letunic *et al.,* 2006) is constructed in a similar way to Pfam but is more successful at finding regulatory motifs. TMHMM v.20 predicts transmembrane helices (Sonnhammer *et al.,* 1998). Finally SignalP v3.0 (Bendtsen *et al.,* 2004) predicts the cleavage sites for potential signal peptides.

These databases were queried with the protein sequences of the eleven ESAG9s so far identified in *T. b. brucei* (nine in strain TREU 927/4, one in Lister 427 and ESAG9-K69 in EATRO 2340). The results of these queries are shown in Table 3.3. The most striking outcome of these queries was that no ESAG9 sequence was found to have a significant hit to any protein family, which is suggestive of ESAG9 proteins performing a novel and previously undescribed function. However the query of the Pfam database with ESAG9-K69 generated an insignificant hit (meaning that the E-value was higher than the cut-off) to a mucin-like glycoprotein family. Although the hit was not significant, this was supportive of the outcome of the

BLAST searches (section 3.2.5) which also suggested similarities between ESAG9s and the mucin glycoprotein family.

TMHMM v2.0 did not predict ESAG9s to have any transmembrane domains. The SignalP v3.0 and SMART hidden Markov models predicted that all the ESAG9s, excepting ESAG9-K69 in the case of the SMART model, have a signal peptide which is cleaved at between 18 and 35 residues from the N terminus. A signal peptide would target the transport of a protein to a specific organelle or trafficking pathway.

Accession number and/or gene name	Results of database queries					
	Pfam	Prosite	SMART	InterPro	TMHMM v2.0	SignalP v3.0
Tb927.1.5080	No significant hits	No hits	Signal peptide position 1-18	No hits	No transmembrane domains	Signal peptide position 1-18
Tb09.v1.0330	No hits	No hits	Signal peptide position 1-21	No hits	No transmembrane domains	Signal peptide position 1-21
Tb927.5.120	No significant hits	No hits	Signal peptide position 1-19	No hits	No transmembrane domains	Signal peptide position 1-19
Tb927.7.170 ESAG9-K9	No hits	No hits	Signal peptide position 1-21	No hits	No transmembrane domains	Signal peptide position 1-21
Tb09.160.5430	No significant hits	No hits	Signal peptide position 1-23	No hits	No transmembrane domains	Signal peptide position 1-23
Tb09.160.5400	No significant hits	No hits	Signal peptide position 1-22	No hits	No transmembrane domains	Signal peptide position 1-22
Tb11.1000	No hits	No hits	Signal peptide position 1-35	No hits	No transmembrane domains	Signal peptide position 1-35
Tb927.5.4620 ESAG9-EQ	No significant hits	No hits	Signal peptide position 1-23	No hits	No transmembrane domains	Signal peptide position 1-23
Tb927.1.5220	No hits	No hits	Signal peptide position 1-23	No hits	No transmembrane domains	Signal peptide position 1-23
ESAG9-K69	No significant hits, but one insignificant hit to a mucin-like glycoprotein family	No hits	Signal peptide position 1-23	No hits	No transmembrane domains	Signal peptide position 1-23
Tb427.BES122.10	No hits	No hits	Signal peptide position 1-24	No hits	No transmembrane domains	Signal peptide position 1-19

Table 3.3: Predictions of functional motifs in ESAG9 proteins using online databases. The eleven ESAG9 protein sequences in *T. b. brucei* were used to query six different databases that contain information about protein families, functional domains, and protein features. The web addresses at which these databases and programmes can be found are listed in Table 1.

3.2.8 Search for potential post-translational modification sites in ESAG9 gene family members using online prediction tools

In addition to the resources available to search for functional motifs in a protein sequence, it is also possible to predict whether post-translational modifications of a protein are likely to occur using online prediction tools. The occurrence of post-translational modification can assist in the prediction of the sub-cellular location of a protein, in particular whether or not it is located to the cell surface membrane. The ESAG9 protein sequences were searched for potential sites for GPI-anchor addition and N-glycosylation. Both of these post-translational modifications occur in the *T. cruzi* mucin-like and MASP protein families, as discussed in section 1.2.6.

3.2.8.1 Some ESAG9 proteins may be GPI-anchored

There are a number of tools to predict the addition of a GPI-anchor to a peptide. Big-PI uses an algorithm based on a set of known GPI-anchor sites from the SWISS-PROT database (Eisenhaber *et al.,* 1999). GPI-SOM uses a Kohonen self-organising map to predict GPI-anchor sites (unpublished). DGPI was the programme used in the annotation proteins in the *T. b. brucei* genome database; it is no longer possible to access the DGPI server, though the ESAG9 genes that are annotated in GeneDB had already been subjected to this analysis.

The results of the predictions are shown in Table 3.4. The outcome varies depending upon which prediction tool was used. GPI-SOM predicted that all eleven *T. b. brucei* ESAG9 proteins would have a GPI anchor. DGPI predicted that seven of the proteins would have an anchor, and Big-PI predicted that only three would (Tb09.v1.0330, Tb927.5.4620, Tb427.BES122.10). Moreover, DGPI and Big-PI did not agree on the predictions, so it was not just a case of Big-PI being more stringent. Despite the variations in the outcomes of these programmes, the fact that some of the ESAG9 proteins were predicted to have an anchor by at least two out of the three programmes suggested that it is probable that at least some ESAG9s are post-translationally modified in this way. The most likely candidates are Tb09.v1.0330 and Tb927.5.4620 (ESAG9-EQ) as they were predicted to have an anchor by all three programmes.

Gene accesion number / name	GPI Prediction tool					
	D-GPI		Big-PI		GPI-SOM	
	GPI anchor predicted?	Predicted cleavage site	GPI anchor predicted?	Predicted cleavage site	GPI anchor predicted?	Predicted cleavage site
Tb927.5.120	yes	C-31	no	n/a	yes	C-32
Tb09.160.5430	yes	C-20	no	n/a	yes	C-27
Tb11.1000	no	n/a	no	n/a	yes	C-20
Tb09160.5400	no	n/a	no	n/a	yes	C-31
Tb927.1.5220	yes	C-19	no	n/a	yes	C-20
Tb09.v1.0330	yes	C-19	yes	C-19	yes	C-20
Tb927.7.170 - K9	yes	C-19	no	n/a	yes	C-20
Tb927.1.5080	yes	C-19	no	n/a	yes	C-20
Tb927.5.4620 - EQ	yes	C-20	yes	C-18	yes	C-21
ESAG9-K69	*	*	no	n/a	yes	C-9
Tb427.BES122.10	no	n/a	yes	C-18	yes	C-21

Table 3.4: Predictions of GPI-anchor addition in ESAG9 proteins. Two online prediction tools, Big-PI and GPI-SOM, were used to predict which ESAG9 proteins are likely to have a GPI anchor. The gene annotation at GeneDB used D-GPI to predict addition of GPI anchors and these results are also listed.

* D-GPI was not used to predict GPI-anchor addition to K69 because it was not possible to access the server

3.2.8.2 Some ESAG9 proteins may be N-glycosylated

The sites in a protein sequence which are likely to be modified by the addition of glycans to asparagine residues can be predicted using the online prediction tool NetNGlyc (http://www.cbs.dtu.dk/services/NetNGlyc/). NetNGlyc uses artificial neural networks based on N-glycosylation in human proteins; it is therefore not optimised for protozoa, but there are no tools currently available that are designed specifically for use on protozoal sequences. Therefore, the eleven *T. b. brucei* ESAG9 protein sequences were subjected to N-glycosylation prediction, and the results are shown in Table 3.5 and in a schematic diagram in Figure 3.13. Four of the eleven sequences did not have any predicted glycosylation sites (Tb927.1.5080, Tb09.160.5340, Tb927.1.5220, Tb427.BES122.10); three have one potential glycosylation site (Tb927.5.120, Tb11.1000, Tb927.5.4620 [ESAG9-EQ]); three have two potential glycosylation sites (Tb09.v1.0330, Tb09.160.5400, ESAG9-K69); and ESAG9-K9 has three potential glycosylation sites. It is therefore likely that some ESAG9 proteins will be post-translationally modified by N-glycosylation, whereas others will not.

Gene accession number / name	N-glycosylation predicted to occur?	Sites at which N-glycosylation might occur	Potential of N-glycosylation occuring at these sites
Tb927.1.5080	no	n/a	n/a
Tb09.v1.0330	yes - at two potential sites	199	0.737
		235	0.547
Tb927.5.120	yes - at one potential site	260	0.611
Tb927.7.170 - K9	yes - at three potential sites	127	0.564
		225	0.603
		234	0.653
Tb09.160.5340	no	n/a	n/a
Tb09.160.5400	yes - at two potential sites	104	0.638
		229	0.601
Tb11.1000	yes - at one potential site	249	0.623
Tb927.5.6430 - EQ	yes - at one potential site	172	0.557
Tb927.1.5220	no	n/a	n/a
ESAG9-K69	yes - at two potential sites	232	0.643
		244	0.317
Tb427.BES122.10	no	n/a	n/a

Table 3.5: Prediction of N-glycosylation of ESAG9 proteins. The online tool NetNGlyc was used to predict likely sites of N-glycosylation of ESAG9 proteins. The sites at which N-glycosylation could occur (the amino acid position where 1 is the N-terminus of the protein), and the potential of this occurring, are listed.

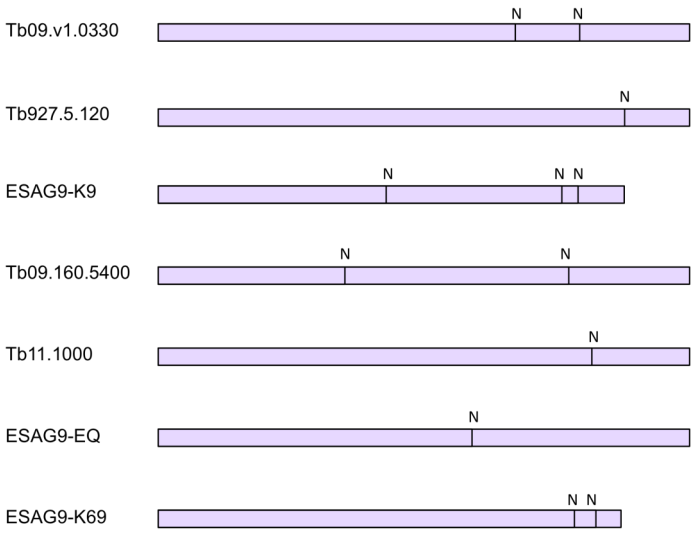


Figure 3.13: Schematic diagram showing the positions of predicted N-glycosylation sites in ESAG9 protein sequences. The online tool NetNGlyc was used to predict likely sites of N-glycosylation of ESAG9 proteins. The sites at which N-glycosylation could occur are indicated by an 'N'. The diagram is drawn to scale.

3.2.9 The 3' Untranslated Regions of ESAG9 genes

The 3' untranslated regions (UTRs) of ESAG9 genes may be involved in regulation of gene expression. The trypanosomatids are unusual in that they control gene expression almost uniquely via post-transcriptional mechanisms. Some genes, including nuclear-encoded mitochondrial components that are up-regulated in procyclic form cells, have been shown to contain regulatory motifs within the 3'UTRs that are involved in stage-specific gene expression (Mayho *et al.,* 2006). Due to the unusual mRNA and protein expression profiles of ESAG9 genes, it is likely that there are novel signals in their 3'UTRs involved in controlling gene expression.

The first step in this analysis was to identify the UTRs of the ESAG9 genes. Benz et al. (Benz *et al.,* 2005) have written an algorithm that predicts all the likely trans splice sites in the *T. b. brucei* strain 927 genome. They also investigated all sequenced cDNAs from *T. b. brucei* and found that polyadenylation at the 3' end of an mRNA transcript tends to occur between 80 and 140 nucleotides upstream of the trans splice site for the next gene. The full sequences of the relevant *T. b. brucei* chromosomes were downloaded from GeneDB. Artemis software was then used to view the annotated chromosomes, and the trans splice site predictions, in parallel. In the case of the Tb427.BES122.10, which is the final gene prior to the 70 base pair repeat region at the end of the chromosome, the 3'UTR sequence was taken to be from the 3' end of the coding region to the beginning of the 70 base pair repeats downstream. The ten 3'UTRs of *T. b. brucei* ESAG9s were predicted to have lengths varying from 510 to 8,159 base pairs long. The 3'UTR of ESAG9-K69 was not included in this analysis because it has not been sequenced.

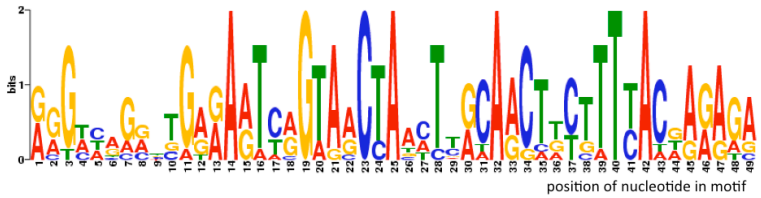
The 3'UTRs were aligned using the programme ClustalW, as described in section 3.2.3. The purpose of this analysis was to explore how similar the 3'UTRs are, and to see if there were any obviously conserved motifs. The alignment is shown in Appendix E (the final 4,000 base pairs of the 3'UTR of gene Tb927.5.120 were left out of the alignment because this predicted UTR was so much longer than the others). The alignment shows that there is not a high degree of conservation between

these sequences. There are 24 fully conserved residues, and these are marked with an asterisk. Some pairs or groups of sequences did however show a higher degree of similarity: Tb427.BES122.10 and Tb927.5.4620 (ESAG9-EQ) are 77% similar. Tb09.v1.0330 and Tb09.160.5400 are 77% similar to each other, and are 62% and 54% similar to Tb927.7.170 (ESAG9-K9) respectively. Tb11.1000 and Tb927.1.5080 are 61% similar.

No obvious highly conserved motifs were found by aligning the 3'UTRs, so a more powerful method was used to find motifs common between the sequences. A tool called Multiple Em for Motif Elicitation, or MEME (Bailey & Elkan, 1995) was used. This programme is not based on a database of known motifs, but instead searches for any motifs which are found more than once in a sequence, or set of sequences. Included in the output is a E-value which represents the chance of finding this motif purely by chance in a set of sequences of the same length. Two motifs present in all ESAG9 3'UTRs were identified by MEME and these are shown in Figure 3.14. The motifs, termed A and B, had E values of 5.1×10^{-37} and 3.5×10^{-29}. These motifs could be involved in post-transcriptional gene regulation.

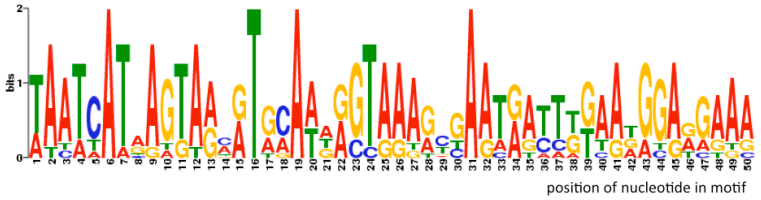
FOLLOWING PAGE: Figure 3.14: The two conserved motifs in ESAG9 3'UTRs identified by MEME. The online tool MEME, or Multiple Em for Motif Elicitation, was used to identify motifs common to all ESAG9 3'UTRs. Two motifs (A and B) were found in all ten 3'UTRs of the ESAG9 genes in *T. b. brucei* (ESAG9-K69 was excluded as this UTR has not been sequenced). The charts represent the consensus sequence with the height of each letter proportional to how many of the sequences that nucleotide is found in. For example the A at position 14 is the full height of the y-axis as it is found in that position in every single sequence. The sequence of the motif in each 3'UTR is shown, along with what strand the motif is found on, the start point of the motif, and the P-value. The P-value represents the probability of finding this motif by chance in a random set of sequences of the same size. In this case all the P-values were extremely low meaning that the occurrence of these motifs is highly unlikely to have occurred by chance.

Motif A E-value = 5.1×10^{-37}



NAME	STRAND	START	P-VALUE	SITES
Tb927.1.5080	+	68	5.78e-21	CGTGACACAG GGGCTGGACGGAGAGTCAGTAACTAAATCACAGCTTCGTTTACGAGAGA AACACGTGTG
Tb09.v1.0330	+	68	1.20e-20	CATAACACAA GGGACAGATTGAGAATCAGTGGCCAAATGGCAACTTCTTTTACAAGAGA CACACTTCCG
Tb427.BES122.10	+	68	1.35e-18	CATGATGCGG AGGAAGAGTTGAAAGTTGGAAGCTATCTTGCAACTACTTTCACTAGGGA TCATGTGGAT
Tb927.1.170	+	68	1.62e-17	TGTAACGCAT GGGTCAGAATGGAAATTGGTACCTAACTCGAAACTTTCTTTACAAGAAA AGCAGATGTG
Tb927.5.4620	+	68	2.13e-17	CATGATGAGG AGGAAGAGTTGAAAGTTGGAAACTAGCTTGCAACTGCTTTCAATGAAGA TCATGTGGAT
Tb927.5.120	+	51	9.10e-17	AGTATGACAC GGGTTCGAATGGGAAACAGTAACCAATTAACAACTTCTTTTACGAGAAC AAAATTTTTG
Tb09.160.5400	+	67	9.87e-17	GCATAGGTCG AAGTAAGGCTGTGAATCAGTAACTAACTTGCAAGTGTTTTCATTAAGGG ATATGTGCAT
Tb11.1000	+	69	2.16e-16	TATAACACAA GAGCTACCGCGGAAGTCGGTAGCTAAATCATAACCACTTTTACGAAAGA ATCACTTGTG
Tb927.1.5220	+	70	1.17e-15	CAATGTGCAT ACTTCAGCGCGAGAATCAGTAACTATTCTACAGCTTCGTTCACGGGATA AACACTTATG
Tb09.160.5430	+	67	1.39e-14	GTCAAGTACA AAGACTGGTGAAAAATACGTAACTACCTTGCAGCAGTTATTACAAAAGG GCTTGTTTAA

Motif B E-value = 3.5×10^{-29}



NAME	STRAND	START	P-VALUE	SITES
Tb927.5.4620	+	173	8.70e-22	AACACTAATG TAATCATAAGTAGCGTACATGAGTAAAACCAGTAACTGGAATGGAGGAAA AGCAGAATAA
Tb427.BES122.10	+	173	1.85e-21	AGCTCCAATG TAATCATAAGTAGAGTACATGAGTAAAACCAGTAACTGGAATGGAGGAAA AGCAGAATAA
Tb927.1.5220	+	192	1.48e-18	AGCTCCAGCA TAATCATAAGGAAAATGAAAAAGTAGAGCGAACGATTTTTGTGGAAGAAG ATAGGTTAAT
Tb927.5.120	+	173	2.04e-18	GACTCTGCCC TAATCATGAGTAAAATGAAAAGGTAAAACGAAAAGACAGAAGGGAGGGAA CGTAACTTAA
Tb927.1.5080	+	190	3.10e-18	AACCCTAATG TAATCATGAGTAACGTGCAAGAGTGGAGTGAATGTCCTGAAAGGGAGGGA ACCAAGTTAA
Tb11.1000	+	190	2.47e-17	AATTCCGATG AAATCATAAGTAAGATGGAATGGTGAAGTGAATGATTTTCAAGTATGAAA AATAAAAAAA
Tb09.v1.0330	+	190	1.67e-15	AGCATCAACA TATTTATCGGTAATGTGCAAAACTAAGGAGAATAGTTGGAATGGAAAAAA TCAAGTTAAA
Tb927.1.170	+	184	1.12e-14	GATGATAACC ATAATATGAGTAGCATTCAAAGGCAAAGTTAATGATTTTAGGGGGTAAAG AATAAATAAA
Tb09.160.5430	+	173	1.36e-14	ACTTTAATAT AAAAAATAATTAACATACATTGGTAATTCGAATAATTGTAAGACAGGAAC TGTAACTCAA
Tb09.160.5400	+	571	8.60e-12	AAGGGACGAA TACTCAATAAGTGAGTGAATTGCTAAAGGTAAAGATATGTAAAGAACTTA AGTCAATTGG

The potential 2-dimensional folding of the 3' UTRs was also investigated. A programme called s-fold uses algorithms which incorporate a Boltzmann distribution of possible RNA secondary structures to predict the most likely structure, called the ensemble centroid. The 3'UTRs assigned using the predictions by Benz *et al*. were too long to use in this programme so only the first 300 nucleotides of each 3' UTR was subjected to analysis. Ensemble centroid structures were generated using the ten abridged 3'UTRs of *T. b. brucei* ESAG9s (ESAG9-K69 was excluded as the 3'UTR has not been sequenced). The outputs from the s-fold programme were visually examined and no conserved folds were seen (data not shown). Either the secondary structure of the ESAG9 3'UTRs was not conserved between genes, or the programme failed to elucidate the correct folding structure. It is important to note that the s-fold algorithm takes into account factors such as temperature and ionic conditions that may not be physiological for trypanosomes.

To investigate whether other genes in the *T. b. brucei* genome had similar 3'UTRs to any of the ESAG9 genes, BLAST searches were carried out using the ESAG9 3'UTR sequences to search the contigs (rather than just the open reading frames). These searches resulted in a number of hits. When investigated in more detail all these hits turned out to be either intergenic regions far from any annotated open reading frames, the 3' UTRs of other ESAG9 genes or pseudogenes, or the 3' UTRs of two genes which have been annotated as similar to ESAG9. These genes were Tb09.142.0380 (354 base pairs long) and Tb927.5.150 (816 base pairs long) and they were annotated as being 46% and 43% similar to the ESAG9 Tb927.1.5220. The 3'UTRs of ESAG9s did not select any 3' UTRS of genes known to be developmentally regulated, such as procyclin. Hence it is likely that any signals in the 3'UTRs of ESAG9s that regulate gene expression are novel and previously undescribed.

3.2.10 The 5' UTRs of ESAG9 genes

The 5'UTRs of the ten ESAG9 genes in *T. b. brucei* were ascertained using the predictions by Benz *et al*. The 5'UTRs were taken to be the region from the

beginning of the predicted trans splice site to the start codon for each given gene. The 5'UTRs ranged in length from 22 to 223 nucleotides. The 5'UTRs were aligned using ClustalW, as previously described (see Figure 3.15). This revealed that there was very little conservation. The two ESAG9s that were most conserved in their 3'UTRs, ESAG9-EQ and Tb927.BES122.10, were only 13% similar to each other in their 5'UTRs. The greatest degree of conservation in the 5'UTRs was between Tb927.1.5220 and Tb927.1.5080 which were 82% similar to each other. The overall lack of conservation in the 5'UTRs suggested that there are no universal mechanisms for gene regulation present in the 5'UTRs of ESAG9 genes so this was not investigated further.

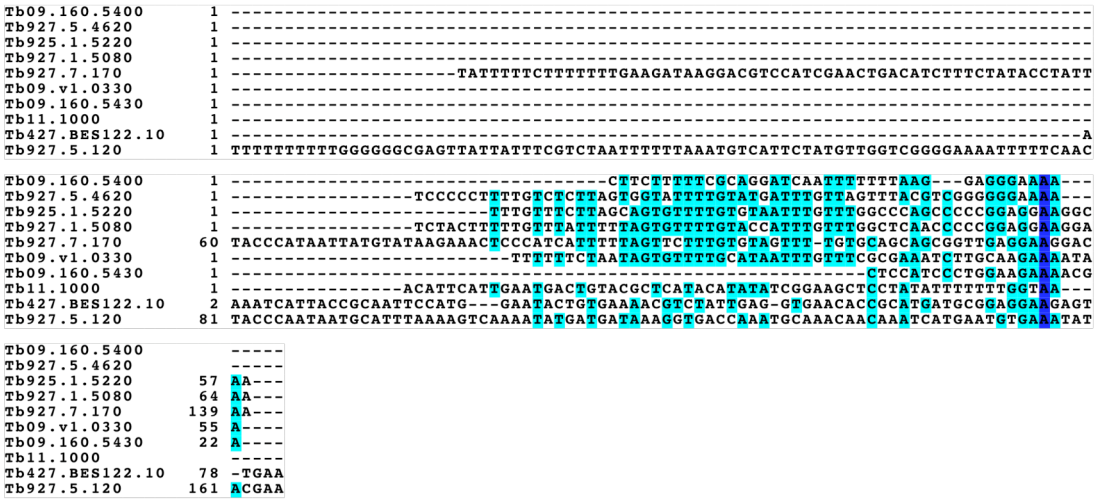


Figure 3.15: ClustalW alignment of the 5'UTRs of *T. b. brucei* ESAG9 genes. The 5'UTRs were predicted using the algorithm written by Benz *et al*. and aligned using ClustalW. There is one full conserved nucleotide, coloured in blue. The nucleotides coloured in cyan are found in that position in 50% or more of the sequences.

3.3 Summary

Prior to the commencement of this PhD project, two ESAG9 genes (ESAG9-K9 and ESAG9-K69) were identified as bloodstream stumpy form-enriched transcripts. No gene has been previously described to have the expression profile of ESAG9-K9 and ESAG9-K69. Moreover, Western blot analysis using an anti-peptide antibody against one of these, ESAG9-K9, revealed that this protein is expressed during differentiation from stumpy to procyclic forms *in vitro*.

The ESAG9 gene family so far identified in trypanosomes includes: nine genes in *T. b. brucei* strain TREU 927/4; one gene in both *T. b. brucei* strain Lister 427 and *T. b. gambiense* (though note that the sequencing of these genomes is not complete); one cDNA clone in *T. b. gambiense* which looks like an ESAG9; one gene in *T. b. brucei* strain EATRO 2340 (called ESAG9-K69) which is not present in the genome reference strain; and two genes in *T. equiperdum*.

Bioinformatic analyses were carried out using ESAG9 sequences. The purpose of these *in silico* experiments was to find any potential clues in the sequence data about characteristics and functions of ESAG9 proteins. These analyses revealed that the ESAG9 gene family, though most often not found within VSG ESs, are always found near VSG genes or other ESAGs. The family is not highly conserved, though retains a short motif similar to one found in PP2C. ESAG9 sequences bear some similarity to *T. cruzi* MASPS and mucin-like glycoproteins, and it is likely that at least some ESAG9 proteins would be post-translationally modified. The 3'UTRs of these genes are not highly conserved but contain two motifs which could be relevant in terms of post-transcriptional control of gene expression.

These bioinformatic analyses were used to inform decisions about what experiments should be carried out to ascertain the function of ESAG9 proteins. Of particular interest were the potential for post-translational modification of ESAG9 proteins, and also the potential functional similarity to mucin-like glycoproteins. These possibilities were explored with wet-bench experiments as described in later chapters. Based on the expression profiles of ESAG9 genes and the attributes

described above, we postulated that ESAG9 proteins could be involved in host-parasite interactions or play a role in the early establishment of an infection in the tsetse fly mid gut.

Chapter 4 Generation and subsequent analysis of transgenic cell lines which ectopically express one or more ESAG9 protein

4.1 Introduction

Two ESAG9 transcripts (ESAG9-K9 and ESAG9-K69) were shown to be upregulated in bloodstream stumpy form parasites (see Chapter 3 section 3.1.2). An anti-peptide antibody was raised against one ESAG9 protein, ESAG9-K9, and this revealed that expression of this protein peaked between six and nine hours into *in vitro* differentiation to the procyclic life form. By 30 hours into differentiation there was no protein detected (see Chapter 3 section 3.1.5).

To further explore the characteristics of ESAG9 proteins it was necessary to create transgenic cell lines in order to manipulate their expression and analyse the consequences. African trypanosomes are amenable to stable transfection with linearised plasmids, and these can be used as vectors for, for example, ectopic over-expression of a gene, or gene knockdown by RNA interference (RNAi) (Ngo *et al.,* 1998).

RNAi was not an option for this project. There are nine ESAG9 genes so far identified in the *T. b. brucei* genome reference strain and these genes and their 3' UTRs are quite divergent, as shown by the bioinformatic analyses carried out in Chapter 3. Hence, RNAi knockdown of one gene may well not have shown a phenotype due to potential redundancy of function between family members, and a knockdown of multiple genes simultaneously with one RNAi construct would have required a region very highly conserved between genes.

Ectopic over expression of ESAG9 proteins provided a potentially powerful tool however. ESAG9 genes are known not to be constitutively expressed by monomorphic slender bloodstream forms or by procyclic forms, though it is worth noting that the expression profile of all ESAG9 genes has not yet been explored.

Therefore transgenic cell lines were engineered to express an ESAG9 protein (or proteins) in each of these life cycle stages, with the assumption that no other ESAG9 protein was being expressed (however it is worth noting that not all of the expression profiles of ESAG9 genes in *T. b. brucei* have been characterised). The system used also had the advantage of protein expression being inducible by the addition of the drug tetracycline to the culture media of transfected trypanosomes (Biebinger et al., 1997). A construct for ectopic overexpression of proteins called pHD451 was utilised. This construct contains a procyclin promoter to drive gene expression, downstream of which is a TetR operator and then a site for insertion of the required open reading frame (ORF) (see Figure 4.3) (Wirtz and Clayton, 1995; Biebinger et al., 1997). The TetR (Tet Repressor) is an element of the bacterial tetracycline resistance cassette (Wirtz and Clayton, 1995). In the absence of tetracycline, the TetR protein binds to the TetR operator and this interferes with transcription from the procyclin promoter by blocking the progression of the polymerase enzyme. In the presence of tetracycline, the tetracycline binds to the TetR and therefore prevents it from having this blocking effect, allowing transcription to continue. A stable *T. b. brucei* cell line (named 427-449) which expesses the Tet repressor protein (TetR) had been previously generated. Therefore pHD451 constructs containing the required ESAG9 ORF were linearised with NotI restriction enzyme and transfected into this cell line where they integrated into the intergenic spacer of the ribosomal RNA locus. Another tool that was very useful to this project was the epitope-tagging of proteins using an epitope tag developed by the laboratory of Professor Keith Gull (Bastin *et al.,* 1996). The Ty-epitope tag is a ten amino-acid sequence (EVHTNQDPLD) found in a *Saccharomyces cerevisiae* virus-like particle called Ty1. This Ty tag is recognised by an antibody called BB2 (Brookman *et al.,* 1995),which means that Ty-tagged proteins can be visualised on western blots, and in fixed cells by immunofluorescence.

Cell lines were generated that expressed Ty-tagged versions of ESAG9-K9, ESAG9-K69, and ESAG9-EQ. Cell lines were also generated which simultaneously expressed pairs of ESAG9 genes: either ESAG9-K9 and ESAG9-K69 together, or ESAG9-EQ and ESAG9-K69 together.

These various cell lines were then used to determine the subcellular location (or shedding from the cell) of the expressed proteins, the occurrence of post-translational modifications, and in functional studies that will be discussed in later chapters. An anti-peptide antibody was also raised against ESAG9-K69 to assist in determining the expression profile and localisation of this protein in 'wild-type' pleiomorphic stumpy cells.

4.2 Generation of stable transgenic cell lines which ectopically and inducibly express ESAG9 proteins

4.2.1 Sub-cloning of Ty-tagged ESAG9s into a plasmid vector for ectopic expression in trypanosomes

Prior to the commencement of this PhD project, bloodstream form cell lines were generated which ectopically and inducibly over-expressed either un-tagged ESAG9-K9 protein (this cell line will subsequently be referred to as 'bsf K9(no tag)'), or ESAG9-K9 protein with an internal Ty tag (which will subsequently be referred to as 'bsf K9') (K.M., unpublished data). It was also necessary to clone tagged versions of ESAG9-EQ and ESAG9-K69 for ectopic expression. The tag was required as no antibodies were available against either of these proteins, although an anti-peptide antibody against ESAG9-K69 was subsequently generated (see section 4.4 of this chapter). For both ESAG9-EQ and ESAG9-K69 the tag was inserted internally in the polypeptide sequence, in a region of low homology between ESAG9 proteins, and replaced existing amino acids. This was necessary because it was predicted that either an N-terminal or a C-terminal epitope tag would disrupt the trafficking signals of the encoded protein (a signal peptide and GPI-anchor addition site respectively). The location of the tag in ESAG9-K9, ESAG9-EQ and ESAG9-K69 sequences is indicated by a box in Figure 4.1. It was hoped that by choosing a region of low homology between the ESAG9 proteins, the tag would not unduly disrupt the structure of the proteins, though this was not confirmed in *in silico* or *in vitro* experiments. The tag was not located near any of the N-glycosylation sites discussed in Chapter 3.

Figure 4.1: Alignment showing the location of the inserted Ty-tag in ESAG9-EQ, ESAG9-K69 and ESAG9-K9 protein sequences. The ten amino acids that were replaced with the Ty-tag are indicated by a box. This region does not include any of the predicted N-glycosylation sites discussed in chapter 3.

The tag was introduced into the nucleotide sequences by three rounds of PCR, as shown in a schematic diagram in Figure 4.2. A template was generated by a PCR reaction with 5' and 3' oligonucleotide primers for the ESAG9 gene of interest (primers 1 and 2 in Figure 4.2). Next, two PCR reactions were carried out with oligonucleotide primers which flanked the region to be replaced. These contained the sequence of the Ty tag in the 3' oligonucleotide primer of the upstream amplicon (primer 3 in Figure 4.2), and the 5' primer of the downstream amplicon (primer 4 in Figure 4.2). Finally, these two PCR products were used as the template for a final round of PCR resulting in a full-length PCR product incorporating the Ty tag. Sites for digestion by restriction enzymes were incorporated into the 5' and 3' primers

(primers 1 and 2 in Figure 4.2) to allow sub-cloning of the PCR products. A polymerase with proof-reading capability was used to minimise the probability of insertions, deletions, or substitutions being introduced into the PCR products. The Expand High Fidelity PCR System (Roche) incorporates both Taq DNA polymerase (which is standardly used for PCR) and Tgo DNA polymerase which has a 3'-5' proof-reading activity. Once generated, the resulting PCR products were sequenced to ensure that the sequence was correct and that the inserted Ty-tag was in frame.

The PCR products were then cloned into a pHD-451 vector backbone (Biebinger *et al.,* 1997). This vector incorporates a procyclin promoter to drive ectopic expression of the gene of interest, a sub-cloning region containing restriction enzyme digestion sites, an aldolase 3'UTR, and a drug resistance cassette as indicated in Figure 4.3. Either a hygromycin (used in cloning the ESAG9-EQ gene) or puromycin (used in cloning the ESAG9-K69 gene) drug resistance cassette enabled the selection of successful transfectants. Two different drug resistant cassettes were used so that cell lines could be generated containing two distinct constructs.

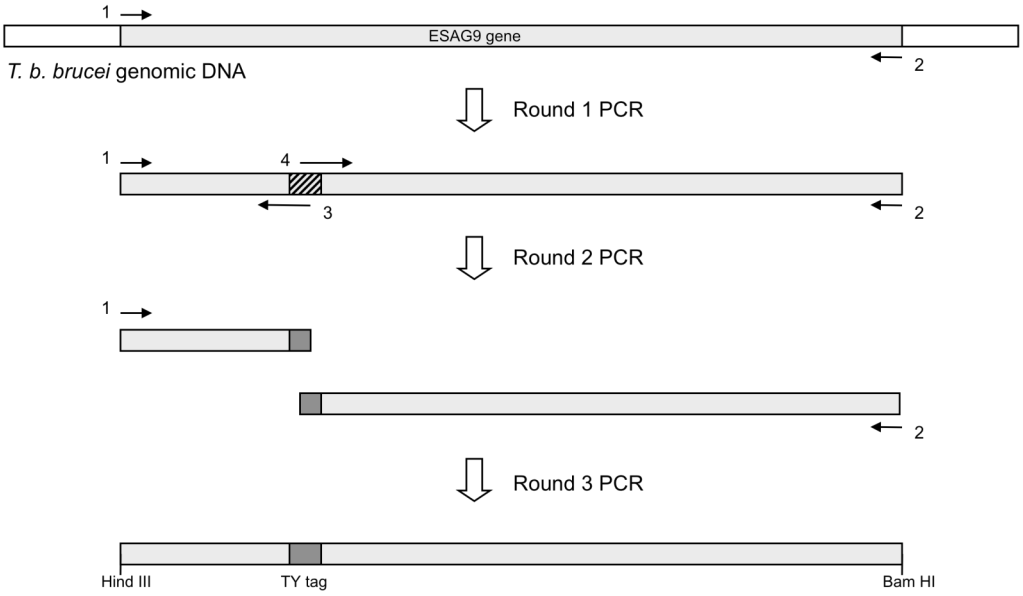


Figure 4.2: Insertion of the Ty-tag sequence into ESAG9 genes by three rounds of PCR. An initial template was generated, using primers 1 and 2, which was then used to generate two fragments that incorporated the Ty-tag sequence in the 3' primer of the upstream fragment and the 5' primer of the downstream fragment. These two fragments were then used as a template to generate a full length PCR product containing the Ty-tag.

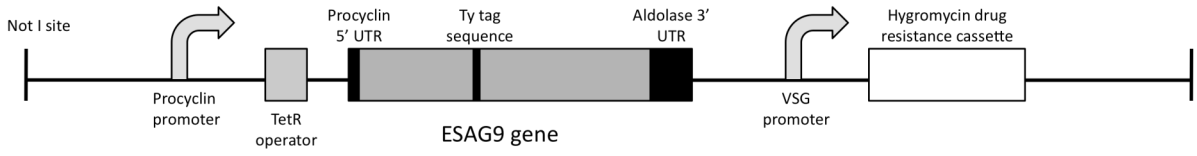


Figure 4.3: Schematic diagram showing the a Ty-tagged ESAG9 gene incorporated into a linearised pHD 451 plamid vector. The cloned gene is inserted into a multi-cloning site downstream of the procyclin promoter, and preceded by the EP procyclin 5'UTR and followed by the aldolase 3' UTR. Either a hygromycin or a puromycin drug resistance cassette is driven by a VSG promoter, and this allowed drug selection of successful transfectants. The construct is linearised with the Not I restriction enzyme prior to transfection.

4.2.2 Generation and protein expression analysis of procyclic form cell lines that ectopically express ESAG9-K9 or ESAG9-EQ

Procyclic form cell lines were generated which stably expressed either Ty-tagged ESAG9-K9 protein, or Ty-tagged ESAG9-EQ protein. Procyclic form *T. b. brucei* 427-449 cells were transfected with 10μg of linearised plasmid DNA (linearised with Not I enzyme) and successful transfectants selected with 30μg/ml Hygromycin, as described in Chapter 2, section 2.1.5. These cell lines will subsequently be referred to as 'pcf K9' and 'pcf EQ'.

Inducible ectopic expression of ESAG9-K9 protein in the stable procyclic form transfectants was confirmed by Western blotting, shown in Figure 4.4. Pcf K9 cells were induced for ESAG9-K9 protein expression by addition of 2μg/ml tetracycline, and protein samples made after 48 hours of induction. Western blotting with the BB2 antibody, which recognises the Ty tag, revealed an inducible band of approximately 40 kDa. This was 13 kDa bigger than the expected size of ESAG9-K9 (27.1 kDa) according to the GeneDB online database. This difference in size could be explained by post-translational modification of the protein. The pcf K9 + (i.e. induced with tetracycline) cells grew at the same rate as the pcf K9 - (i.e. uninduced) cells, as shown in Figure 4.4. This suggested that ectopic expression of the ESAG9-K9 protein was not having a detectable negative effect on the growth of the cells.

Inducible ectopic expression of ESAG9-EQ was also successfully achieved in pcf EQ cells. Cells were induced by the addition of 2μg/ml tetracycline, and protein samples made after 48 hours of induction. Western blotting with the BB2 antibody revealed that a band of approximately 34kDa was present in both the induced and uninduced samples, though the signal was much stronger in the induced sample, as

shown in Figure 4.5. In this cell line the tet repressor protein was not completely silencing transcription of the ectopic ESAG9-EQ gene in the absence of tetracycline and so the expression was leaky. The band migrated at approximately 3kDa higher than the 30.8kDa predicted by the GeneDB *Trypanosoma brucei* database. The induced and uninduced populations grew at the same rate, as shown in Figure 4.5.

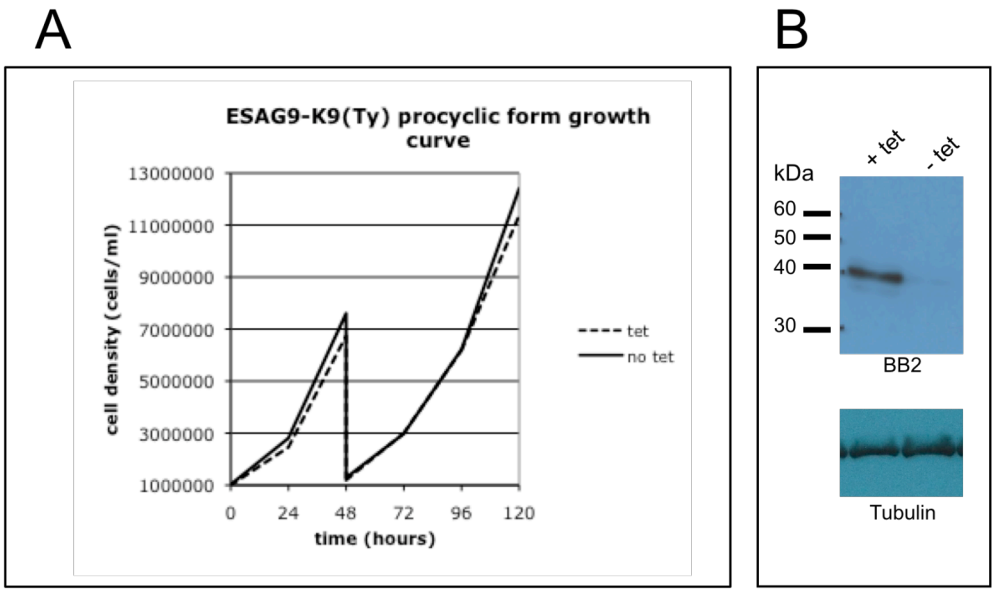


Figure 4.4: Growth and Western blot analysis of transgenic procyclic forms expressing Ty-tagged ESAG9-K9. The growth curve (panel A) shows uninduced cells (solid line) and cells induced for ectopic expression of ESAG9-K9 by addition of 2μg/ml tetracycline (dashed line) growing over a 120 hour time-course. The cells were counted every 24 or 48 hours and were passaged to a concentration of 1×10^6 cells/ml after 48 hours. Panel B shows a Western blot using the BB2 antibody which recognises the Ty-tag (top) and an antibody detecting α tubulin (bottom) which indicated equal loading across lanes. The '+' indicates the induced (with tetracycline) lane, and the '-' the uninduced (no tetracycline) lane. Size markers are in kilo Daltons (kDa). A clearly induced band of approximately 40 kDa was observed.

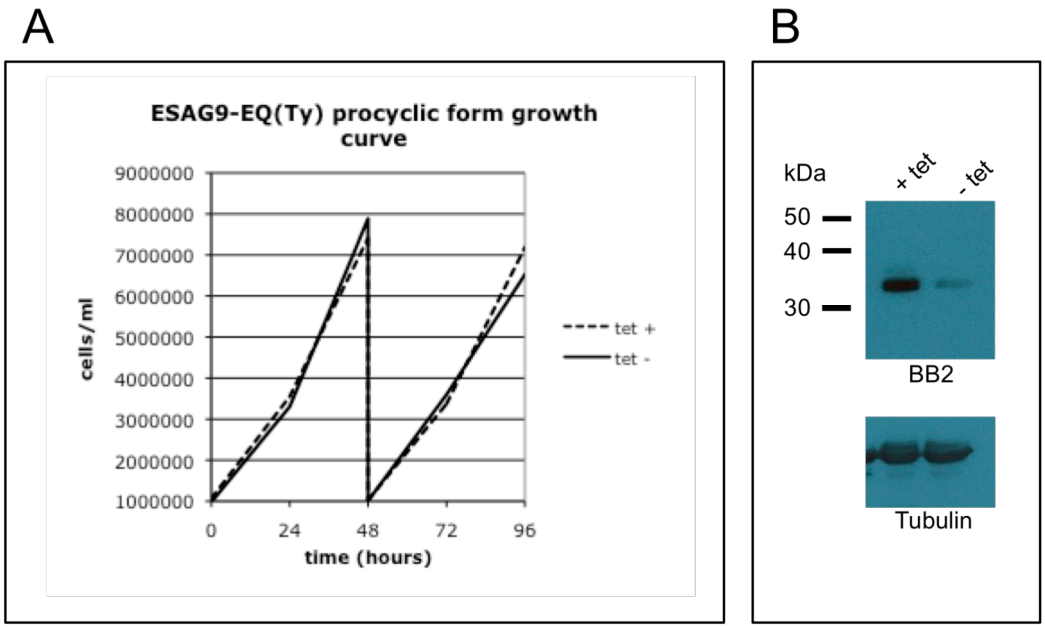


Figure 4.5: Growth and Western blot analysis of transgenic procyclic forms expressing Ty-tagged ESAG9-EQ. The growth curve (panel A) shows uninduced cells (solid line) and cells induced for ectopic expression of ESAG9-EQ by addition of 2μg/ml tetracycline (dashed line) growing over a 96 hour time-course. The cells were counted every 24 hours and were passaged to a concentration of 1×10^6 cells/ml after 48 hours. Panel B shows a Western blot using the BB2 antibody which recognises the Ty-tag (top), and an antibody detecting α tubulin (bottom), which indicated equal loading across lanes. The '+' indicates the induced (with tetracycline) lane, and the '-' the uninduced (no tetracycline) lane. Size markers are in kilo Daltons (kDa). A band of approximately 34 kDa was observed in both lanes but is stronger in the '+' lane.

4.2.3 Generation and Western analysis of bloodstream form cell lines that ectopically express ESAG9-K9, ESAG9-K69 or ESAG9-EQ

T. b. brucei bloodstream form cell lines that inducibly express either Ty-tagged (referred to as 'bsf K9') or untagged (referred to as 'bsf K9(no tag)') ESAG9-K9 protein had previously been generated (K. M., unpublished data). Inducible expression of the untagged or tagged protein was confirmed by induction with tetracycline and Western blotting, shown in Figures 4.6 and 4.7. Growth curves indicated that expression of ESAG9-K9 protein expression did not cause a growth phenotype in either cell line.

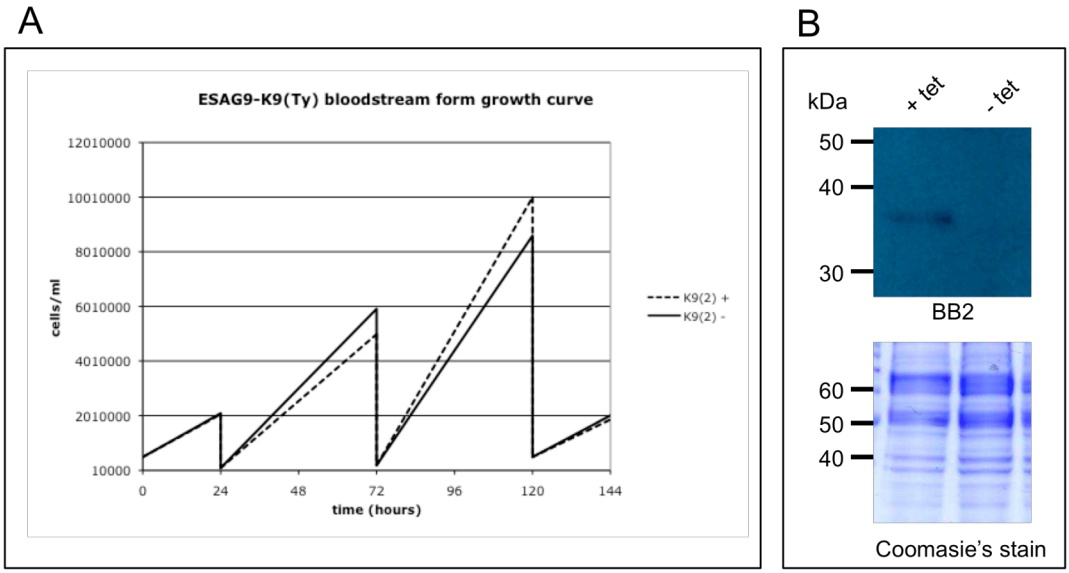


Figure 4.6: Growth and Western blot analysis of transgenic bloodstream forms expressing Ty-tagged ESAG9-K9 protein. The growth curve (panel A) shows uninduced cells (solid line) and cells induced for ectopic expression of ESAG9-K9 by addition of 2μg/ml tetracycline (dashed line) growing over a 144 hour time-course. The cells were counted every 24 or 48 hours and were passaged to a concentration of 1×10^5 cells/ml after 24, 72 and 120 hours. Panel B shows a Western blot using the BB2 antibody which recognises the Ty-tag. The '+' indicates the induced (with tetracycline) lane, and the '-' the uninduced (no tetracycline) lane. Size markers are in kilo Daltons (kDa). A band of approximately 37 kDa was observed in the '+' lane. The Coomasie's stained image of the gel (bottom) illustrates equivalent loading in each lane.

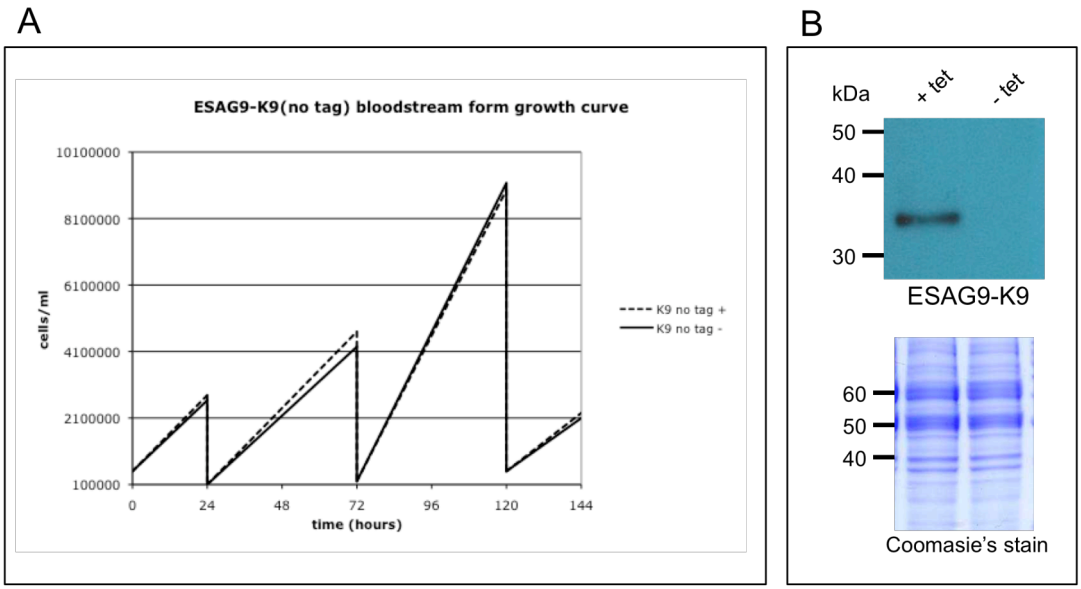


Figure 4.7: Growth and Western blot analysis of transgenic bloodstream forms expressing un-tagged ESAG9-K9 protein. The growth curve (panel A) shows uninduced cells (solid line) and cells induced for ectopic expression of ESAG9-K9 by addition of 2μg/ml tetracycline (dashed line) growing over a 144 hour time-course. The cells were counted every 24 or 48 hours and were passaged to a concentration of 1×10^5 cells/ml after 24, 72 and 120 hours. Panel B shows a Western blot using the ESAG9-K9 anti-peptide antibody. The '+' indicates the induced (with tetracycline) lane, and the '-' the uninduced (no tetracycline) lane. Size markers are in kilo Daltons (kDa). A band of approximately 34 kDa is seen in the '+' lane. The Coomasie's stained image of the gel (bottom) demonstrates equivalent loading across lanes.

T. b. brucei bloodstream form cell lines were also generated that inducibly expressed a Ty-tagged version of ESAG9-EQ or ESAG9-K69. *T. b. brucei* strain 427-449 cells were transfected with 10μg of linearised plasmid DNA. Successful transfectants were selected with 5μg/ml hygromycin, in the case of ESAG9-EQ, or with 0.5μg/ml puromycin, in the case of ESAG9-K69, as described in the Materials and Methods. Inducible ectopic expression of each ESAG9 protein was confirmed by Western

blotting with the BB2 antibody, as shown in Figure 4.8 (for ESAG9-EQ) and Figure 4.9 (for ESAG9-K69). In both these cases, induction of protein expression was not shown to have any affect on the growth rate of the cells, also shown in Figures 4.8 and 4.9. In the bsf EQ cell line, the ESAG9-EQ protein ran at the same size in the procyclic form and the bloodstream form transgenic cells (34 kDa). In the bsf K69 cell line, ESAG9-K69 ran at approximately 50kDa, which is considerably larger than the mass of 26.6kDa predicted using the online prediction tool PeptideMass. (www.expasy.org/tools/peptide-mass). This suggested that the protein was being extensively post-translationally modified. Also the quantity of ESAG9-K69 protein produced by the cells was much lower than for ESAG9-EQ. Thus, a sample from 6×10^6 cells was required to achieve a clear signal on Western blots whereas only 2×10^6 cells were used for all other Westerns described. Moreover, the ESAG9-K69 Western was visualised using the LI-COR® system. This detects proteins on Western blots using infrared fluorescence, and is more sensitive than the HRP (horse radish peroxidase) system used for the bsf EQ Western blot. This indicated that the levels of cell-associated ESAG9-K69 were less than for the other ectopically-expressed ESAG9 proteins assayed.

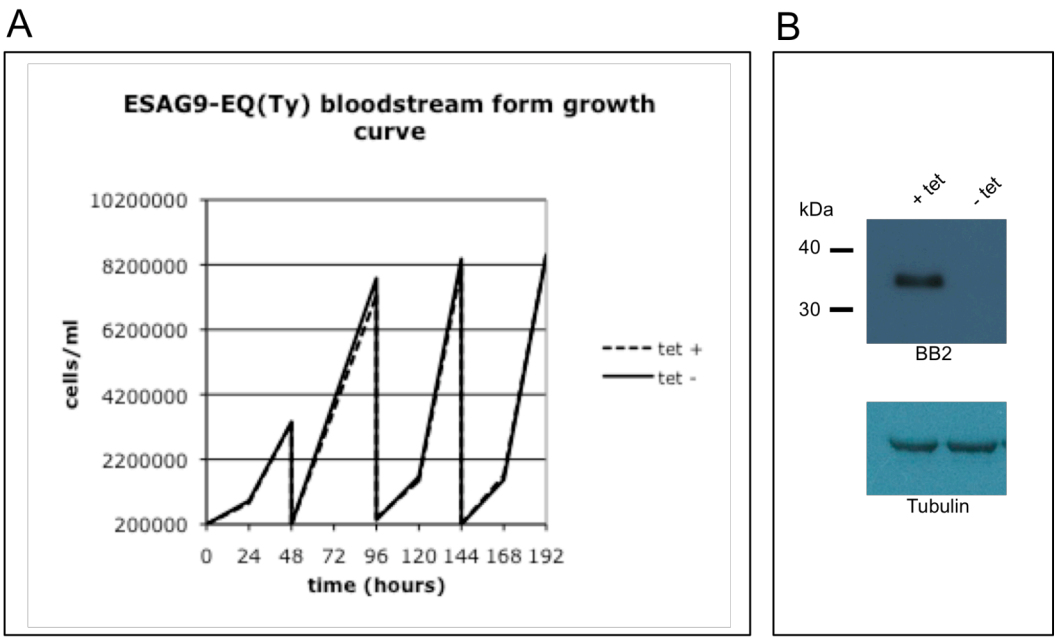


Figure 4.8: Growth and Western blot analysis of transgenic bloodstream forms expressing Ty-tagged ESAG9-EQ. The growth curve (panel A) shows uninduced cells (solid line) and cells induced for ectopic expression of ESAG9-EQ by addition of 2μg/ml tetracycline (dashed line) growing over a 192 hour time-course. The cells were counted every 24 or 48 hours and were passaged to a concentration of 1×10^5 cells/ml after 48, 96 and 144 hours. Panel B shows a Western blot using the BB2 antibody which recognises the Ty-tag (top) and an α tubulin antibody (bottom) which indicated equal loading across lanes. The '+' indicates the induced (with tetracycline) lane, and the '-' the uninduced lane. Size markers are in kilo Daltons (kDa). A band of approximately 34 kDa was observed in the '+' lane.

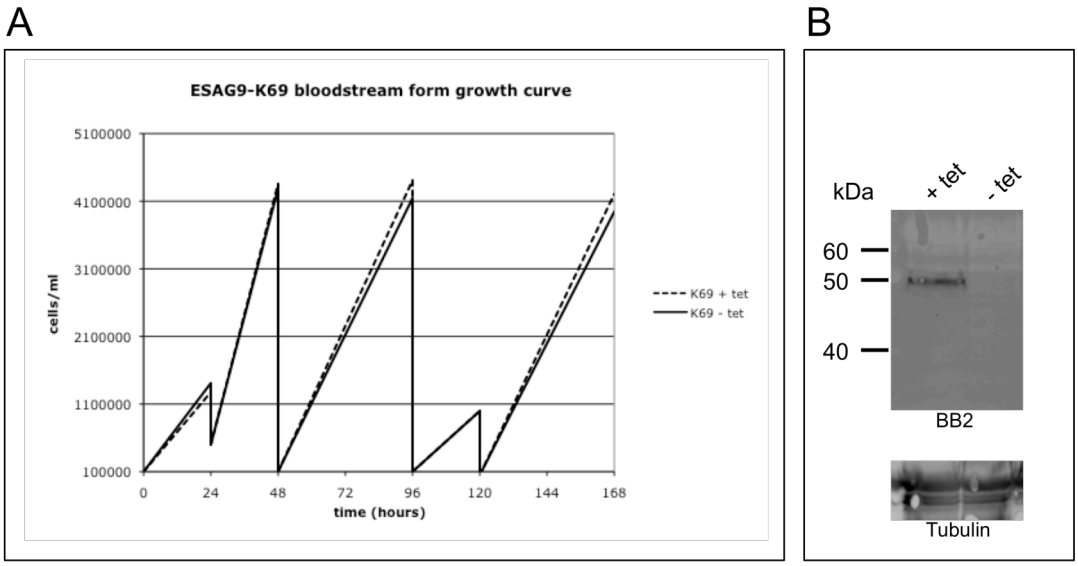


Figure 4.9: Growth and Western blot analysis of transgenic bloodstream forms expressing Ty-tagged ESAG9-K69. The growth curve (panel A) shows uninduced cells (solid line) and cells induced for ectopic expression of ESAG9-K69 by addition of 2μg/ml tetracycline (dashed line) growing over a 168 hour time-course. The cells were counted every 24 or 48 hours and were passaged to a concentration of 5×10^5 cells/ml after 24 hours, and to a concentration of 1×10^5 cells/ml after 48, 96 and 120 hours. Panel B shows a Western blot using the BB2 antibody which recognises the Ty-tag (top) and an α-tubulin antibody (bottom) which indicated equal loading across lanes. The '+' indicates the induced (with tetracycline) lane, and the '-' the uninduced lane. Size markers are in kilo Daltons (kDa). A band of approximately 50 kDa was observed in the '+' lane.

4.2.4 Generation and Western analysis of bloodstream form cell lines that express two different ESAG9 proteins concurrently

Pleiomorphic cells were shown to express more than one ESAG9 mRNA concurrently by Northern blotting, as shown in Chapter 3, Figure 3.2 (K. M., unpublished data). It is possible that ESAG9 proteins interact with one another in some way, in which case the ectopic expression of one ESAG9 protein alone might not generate a phenotype. There is a precedent for this in the ESAG gene family, in that ESAGs 6 and 7 combine to form the transferrin receptor complex (Steverding *et al*., 1994).

Bloodstream form cell lines that expressed two ESAG9 proteins concurrently were therefore generated. The combination of ESAG9-K9 and ESAG9-K69 was considered most interesting because these two genes are co-expressed in the *T. b. brucei* EATRO 2340 strain, and both mRNAs are expressed at a high level (see Chapter 3, Figure 3.2). The dual expression of these genes in monomorphic cells was made possible by the fact that the previously generated bsf K9(no tag) cell line was resistant to hygromycin, and the Ty-tagged ESAG9-K69 gene was cloned into a vector containing a puromycin drug resistance cassette. It was also possible to distinguish between the two proteins when co-expressed because the un-tagged ESAG9-K9 was recognised by the ESAG9-K9 anti-peptide antibody, whereas the Ty-tagged K69 was recognised by the BB2 antibody. There was no cross-reaction between the K9 anti-peptide antibody and the K69 protein (data not shown). A bloodstream form cell line that co-expressed ESAG9-EQ and ESAG9-K69 was also generated in the same manner, although it was not possible to distinguish between these two proteins as they were both Ty-tagged, except on the basis of size.

Thus, bsf K9(no tag) and bsf EQ cell lines were transfected with 10μg of linearised vector DNA containing the tagged ESAG9-K69 gene. Successful transfectants were selected with 0.5μg/ml puromycin, as described previously.

Expression of ESAG9-K69 protein concurrently with ESAG9-K9 was confirmed by Western blotting, shown in Figure 4.10 (this cell line will subsequently be referred to as bsf K9:K69). Stable transfectants were induced to express both proteins by addition of 2μg/ml tetracycline, and protein samples made after 48 hours of induction. Two Western blots were carried out, one with the ESAG9-K9 anti-peptide antibody, and the other with the BB2 antibody (detecting ESAG9-K69). The inducible expression of ESAG9-K9 was confirmed, and two bands were seen of 38 and 42 kDa, which differed from the bsf K9 cell line, where one band of 37 kDa was seen. The signal from the BB2 antibody, which detects the ESAG9-K69 protein, was extremely weak but a very faint band of 40 kDa was seen in both lanes and was stronger in the induced lane. The weakness of the signal could indicate that there is less expression of this ectopic gene occurring for some reason, or that less protein was cell associated. The protein could be degraded by the lysosomal machinery in the cell, or it could be secreted out of the cell. Interestingly, the detected ESAG9-K69 protein was running at a smaller size in this cell line, 40 kDa, compared to 50 kDa in the bsf K69 cell line which only expressed ESAG9-K69.

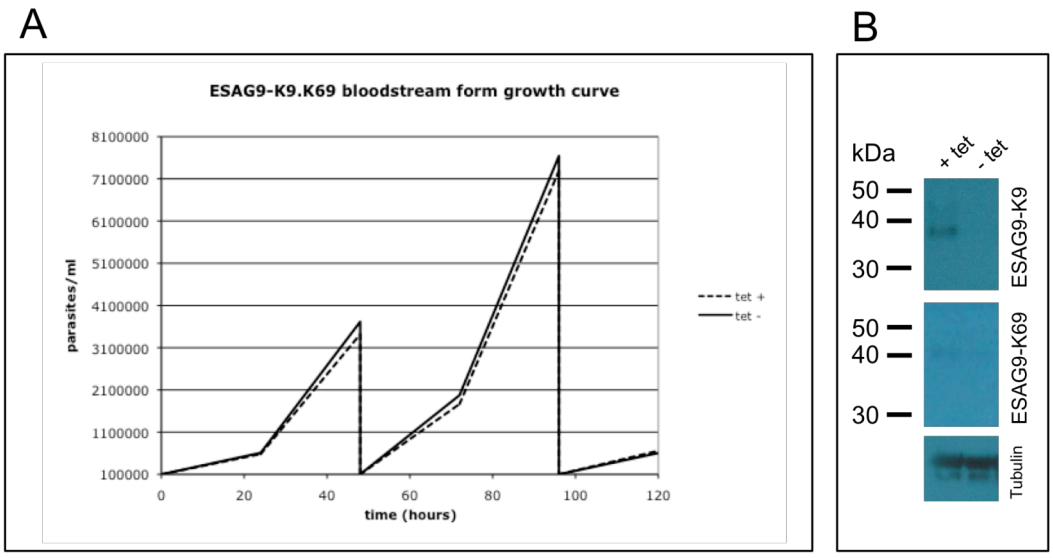


Figure 4.10: Growth and Western blot analysis of K9:K69 transgenic bloodstream forms which express un-tagged ESAG9-K9 and Ty-tagged ESAG9-K69 concurrently. The growth curve (panel A) shows uninduced cells (solid line) and cells induced for ectopic expression of the ESAG9 proteins by addition of 2μg/ml tetracycline (dashed line) growing over a 120 hour time-course. The cells were counted every 24 or 48 hours and were passaged to a concentration of 1×10^5 cells/ml after 48 and 96 hours. Panel B shows Western blots using the ESAG9-K9 anti-peptide antibody (top), the BB2 antibody which recognises the Ty-tagged ESAG9-K69 protein (middle) and an α tubulin antibody (bottom) which indicated equal loading across lanes. The '+' indicates the induced lane, and the '-' the uninduced lane. Size markers are in kilo Daltons (kDa). A faint band of approximately 43 kDa and a stronger band of approximately 38 kDa were both observed in the '+' lane of the ESAG9-K9 Western blot. Very faint bands of 40 kDa were observed in the BB2 Western blot (middle), indicating that a low level of expression of ESAG9-K69 protein was achieved, and that protein expression was leaky.

Expression of ESAG9-K69 and ESAG9-EQ concurrently in the bsf EQ:K69 cell line was also confirmed by Western blotting. Stable transfectants were induced as described above, and the resultant protein samples analysed by Western blotting. Two bands were seen, as shown in Figure 4.11. Although it was not possible to definitively ascribe one protein to each band, it is likely that ESAG9-EQ constitutes the lower band of 35kDa, which is the size the ESAG9-EQ protein runs at in the single-expressing cell line (see Figure 4.8). ESAG9-K69 is likely to constitute the upper band, as it runs at the same size as ESAG9-K69 in the bsf K9:K69 cell line (see Figure 4.10). The double expression of ESAG9 proteins in either bsf K9:K69 or bsf EQ:K69 cell lines did not result in a growth phenotype.

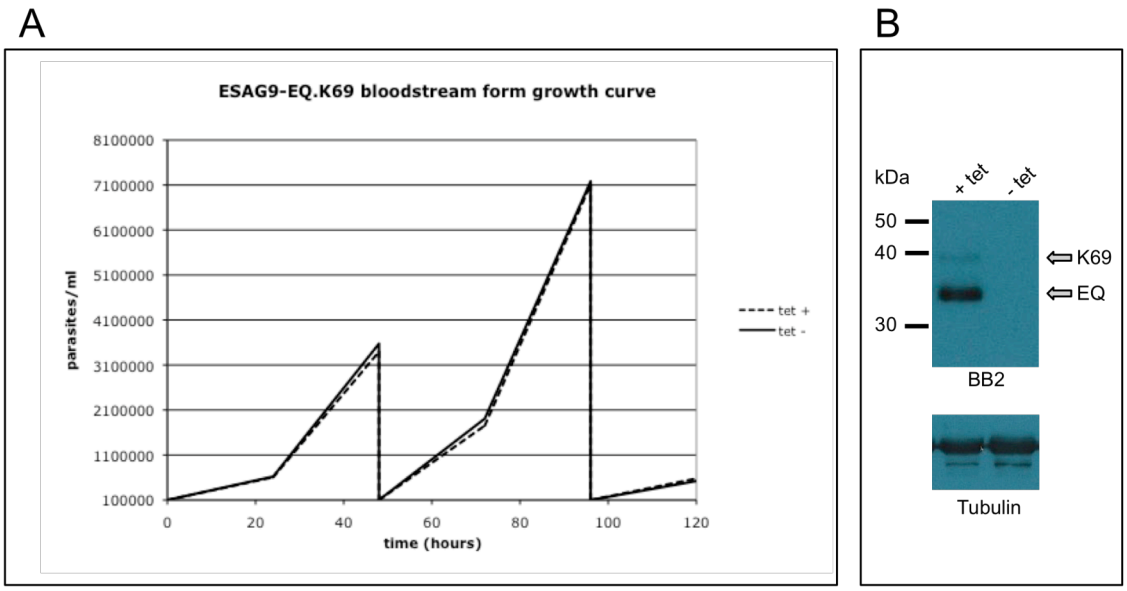


Figure 4.11: Growth and Western blot analysis of EQ:K69 transgenic bloodstream forms which express Ty-tagged ESAG9-EQ and Ty-tagged ESAG9-K69 concurrently. The growth curve (panel A) shows uninduced cells (solid line) and cells induced for ectopic expression of the ESAG9 proteins by addition of 2μg/ml tetracycline (dashed line) growing over a 120 hour time-course. The cells were counted every 24 or 48 hours and were passaged to a concentration of 1×10^5 cells/ml after 48 and 96 hours. Panel B shows a Western blot using the BB2 antibody which recognises the Ty-tag (top) and an α tubulin antibody (bottom) which indicated equal loading across lanes. The '+' indicates the induced (with tetracycline) lane, and the '-' the uninduced lane. Size markers are in kilo Daltons (kDa). A faint band of approximately 40 kDa and a stronger band of approximately 34 kDa were both observed in the '+' lane. It was not possible to distinguish definitively between the ESAG9-EQ and the ESAG9-K69 proteins as they were modified with the same Ty-tag, but it is likely that ESAG9-K69 constitutes the upper band and ESAG9-EQ constitutes the lower band.

4.3 Immunoprecipitation analysis of bloodstream form conditioned media

The level of ESAG9-K69 protein that was detected in the cell extracts from the bsf K69 and bsf K9:K69 cell lines was low. Therefore the possibility that some of the protein was being secreted or shed by the cells into the medium was explored by performing immunoprecipitations with trypanosome-conditioned media. No naturally secreted proteins have previously been characterised in bloodstream form *T. brucei* apart from a molecule called TLTF that was reported to induce CD8+ T cells to produce interferon gamma (Olsson *et al.,* 1991). However the gene was subsequently cloned and the protein characterized as being in the flagellum (Hill *et al.,* 2000). The effect on CD8+ T cells was not seen when other labs attempted to repeat the experiments so the work is considered highly suspect (personal communication, Professor John M. Mansfield, University of Wisconsin-Madison).

4.3.1 Immunoprecipitation of Ty-tagged protein using the BB2 antibody

The ability to successfully immunoprecipitate (IP) Ty-tagged proteins using the BB2 antibody was initially verified using 1ml of bsf K9 cell lysate derived from 1×10^8 cells. A schematic diagram summarising how the IP was carried out is shown in Figure 4.12. As a control, blocking peptide was generated that prevents the BB2 antibody from binding to the Ty tag. This blocking peptide was of the same sequence as the epitope recognised by the BB2 antibody, allowing it to block detection of Ty-tagged proteins. Induced bsf K9 + cell lysate was incubated with protein G beads for 45 minutes prior to the IP and the beads subsequently removed. Protein G beads should only bind antibody but the blocking step was required to remove any proteins from the cell lysate that bind non-specifically to the protein-G beads. The antibody was also incubated for 45 minutes prior to the IP with and without the blocking peptide. This was to ensure that any positive result seen was due to a Ty-tagged

protein binding to the BB2 antibody and to rule out non-specific binding. The blocked and unblocked antibody was then added to the cell lysate and incubated at 4ºC for one hour to allow the Ty-tagged ESAG9-K9 protein to bind to the BB2 antibody. Protein-G beads were then added to the mixture and again incubated for one hour at 4ºC to allow the bound BB2 antibody to bind to the protein-G beads. The antibody and the Ty-tagged protein were then eluted from the G-beads by incubation at 100ºC for ten minutes in Laemmli buffer.

The results of this test assay are shown in Figure 4.13. A faint band of the correct size for ESAG9-K9 was detected in the IP lane (left lane) and this band disappeared when the blocking peptide was used to prevent antibody binding (right lane). This indicated that the IP reaction conditions were effective.

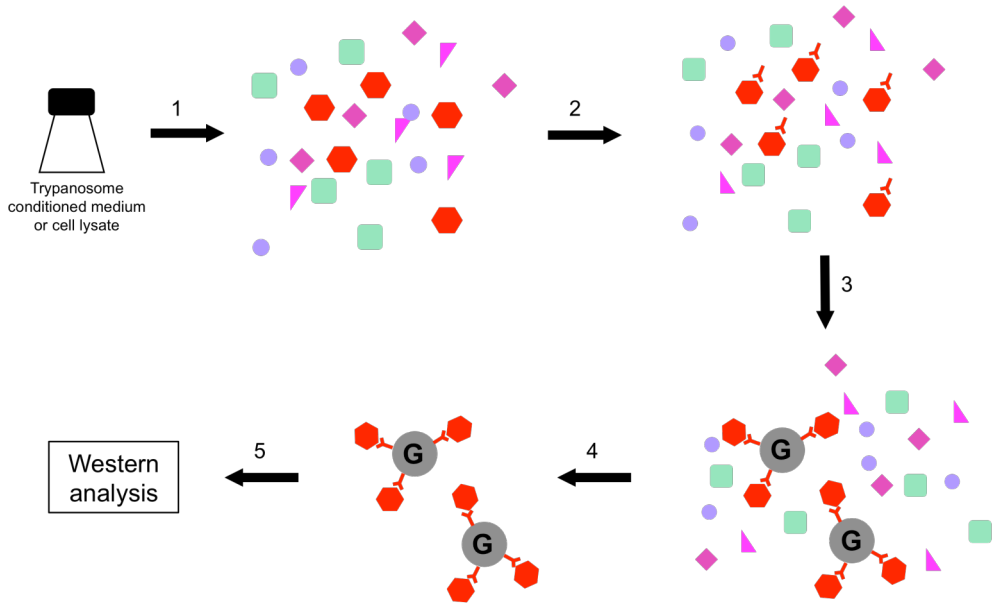


Figure 4.12: Schematic diagram showing the selection of Ty-tagged protein from cell lysate or conditioned media by immunoprecipitation with BB2 antibody. The steps are as follows, in simple terms: 1. cell lysate is made from a cell pellet, or the conditioned medium is reduced in volume, 2. BB2 is incubated with the mixture and this recognises and binds to Ty-tagged proteins (in red), 3. Protein-G beads are incubated with the mixture and the BB2 antibody binds to the beads, 4. The beads are isolated by centrifugation and wash steps, 5. The proteins are eluted from the beads and analysed by Western blotting.

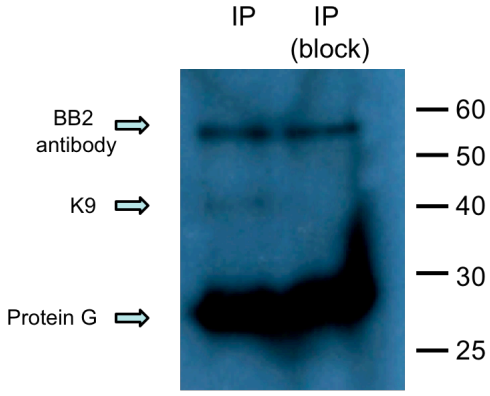


Figure 4.13: Test immunoprecipitation with Ty-tagged ESAG9-K9 protein. Bsf K9 + cells were harvested and the Ty-tagged protein pulled down by immunoprecipitation (IP) with BB2 antibody as described in the Materials and Methods. The left hand lane is the IP without blocking peptide and the right hand lane is with blocking peptide. The top band is the BB2 antibody, the faint middle band (which is only present in the absence of blocking peptide) is the Ty-tagged ESAG9-K9 protein, and the large band at the bottom is the Protein G. Size markers are in kilo Daltons.

4.3.2 ESAG9-K69 protein is not secreted/shed by bsf K69 cells

IPs were carried out using bsf K69 + cell lysate and conditioned medium. The cell lysate IP was carried out as described in Section 4.2.2. For the media IP, 180mls of conditioned media was reduced in volume to approximately 4mls by centrifugation through Vivaspin columns with a 10 kDa cut-off (VWR cat. no. 512-3783). The IP was then carried out as for the cell lysate, excepting that the volumes of antibody and G-beads used were scaled up to adjust for the larger volume.

No signal for K69 was seen from the conditioned media but a strong signal was seen from the cell lysate (see Figure 4.14). This indicated that in the bsf K69 cell line, all the K69 protein remained cell associated.

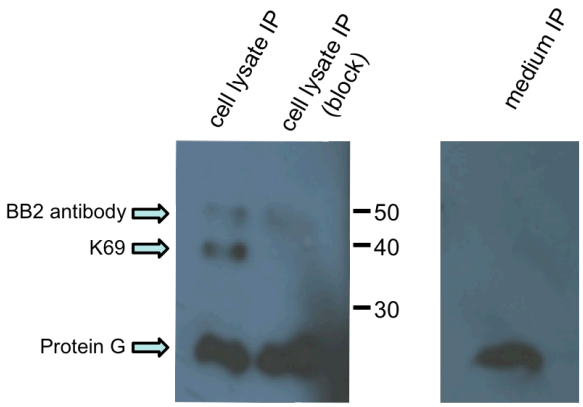


Figure 4.14: IP with bsf K69 cells and conditioned medium. ESAG9-K69 protein was immunoprecipitated from the cell lysate in the absence of blocking peptide (far left lane, middle band). However, no signal was seen from bsf K69 + conditioned medium (right panel) indicating that this protein was not being secreted out of the cells. Size markers are in kilo Daltons.

4.3.3 ESAG9-K69 protein is secreted/shed by bsf K9:K69 cells

An IP with cell lysate and conditioned medium from bsf K9:K69 + cells was also carried out as described in section 4.3.2. In this assay, the IP would only pull down the ESAG9-K69 protein since the ESAG9-K9 protein did not have a Ty tag. Significantly, there was no signal for K69 from the cell lysate IP, but there was a strong signal from the conditioned medium (see Figure 4.15). A blocking peptide control was not used in this particular assay, but the experiment was repeated twice with a blocking peptide control and the same result of a strong band from the conditioned medium was seen on both occasions when the blocking peptide was absent (data not shown).

To ensure that the signal seen in the medium was not a result of cells lysing prior to the IP and releasing their contents into the medium, a control IP was also carried out with an antibody to TbZFP3. TbZFP3 is a small zinc-finger protein involved in stage-specific expression in the trypanosome life cycle (Paterou *et al.,* 2006) that always remains cell-associated. Cell lysate and conditioned medium IPs were carried out with bsf K9:K69 + cells and the anti-TbZFP3 antibody. A signal for TbZFP3 was

only seen from the cell lysate IP and not from the medium (see figure 4.16) confirming that this constitutively expressed protein was remaining cell associated. Hence ESAG9-K69 seemed to be released to the medium specifically from the bsf K9:K69 cell line.

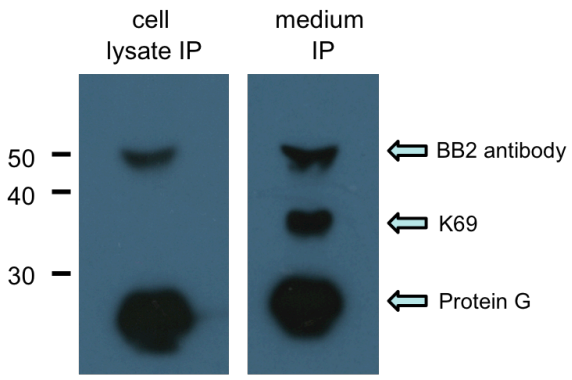


Figure 4.15: Cell lysate and conditioned media IPs using K9:K69 bsf cells. The ESAG9-K69 protein was immunoprecipitated using BB2 antibody. A strong signal was seen from the conditioned medium IP (right hand panel, middle band). No signal was seen for ESAG9-K69 from the cell lysate IP (left hand panel). In this experiment no blocking peptide control was included. Size markers are in kilo Daltons.

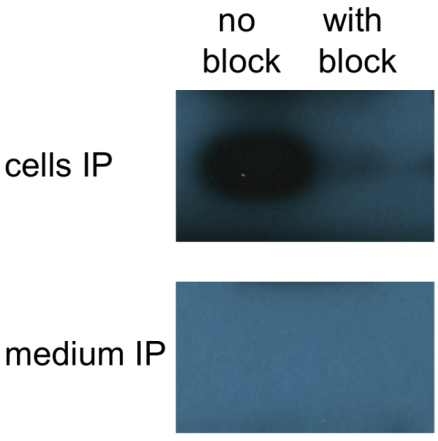


Figure 4.16: Cell lysate and media IPs with ZFP3 antibody. Bsf K9:K69 + cells were subjected to immunoprecipitation with the α TbZFP3 antibody which recognises constitutively expressed and cell-associated TbZFP3 protein. As expected, a signal for TbZFP3 was seen in the cell lysate IP in the absence of blocking peptide (top panel) and no signal was seen from the conditioned medium IP (bottom panel).

4.3.4 ESAG9-EQ protein is not shed by bsf EQ cells

Immunoprecipitations using the BB2 antibody were also carried out with bsf EQ + conditioned medium as previously described. In this assay, no signal was seen from the medium for Ty-tagged ESAG9-EQ protein, indicating that this protein remained cell associated and was not shed from the cell (data not shown).

4.3.5 Ty-tagged protein is shed by bsf EQ:K69 cells

To explore whether Ty-tagged protein was being shed from bsf EQ:K69 + cells, IPs were carried out with BB2 on the conditioned medium from this cell line. In this cell line, both the ESAG9-EQ and ESAG9-K69 proteins were tagged with the Ty-tag and so it was not possible to distinguish between them; either or both could be pulled down by immunoprecipitation. Figure 4.17 shows that a band of approximately 36 kDa was seen in the medium IP lane indicating that at least one of the proteins was being shed from the trypanosomes. In the cell lysate there were three bands visible: a double band at 36 kDa and a single band at 42 kDa. The band seen in the IP lane is likely to correspond to the double band seen in the cell lysate as they are running at the same size. Given the size of this band, it is more likely to correspond to ESAG9-EQ protein because ESAG9-EQ migrated at approximately 35 kDa in the bsf EQ cell line (see Figure 4.8).

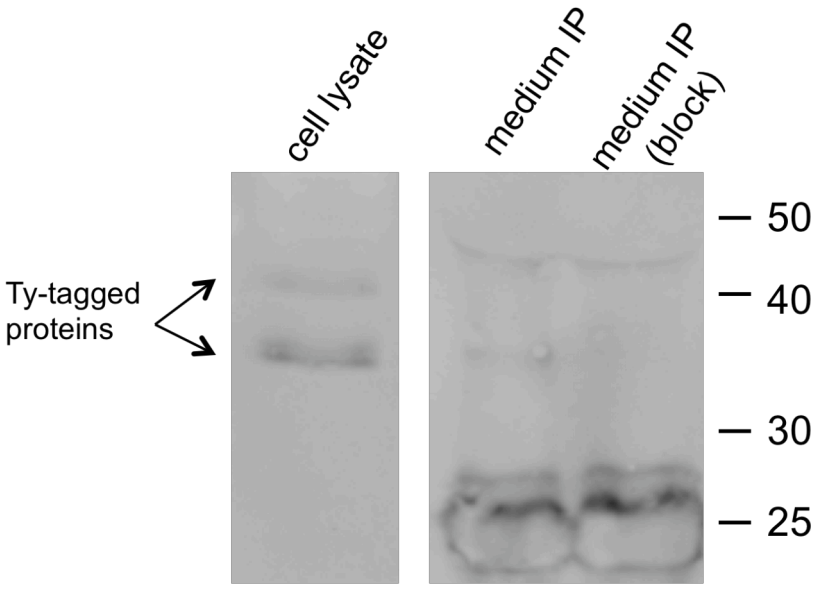


Figure 4.17: Conditioned medium IP with bsf EQ:K69 conditioned media. The cell lysate is shown as a confirmation that there was protein expression. Three bands were seen in the cell lysate: a double band at about 36 kDa and a single band at about 40 kDa (left lane). It was not possible to distinguish between the ESAG9-K69 and ESAG9-EQ proteins as they are both Ty-tagged and recognised by BB2 antibody. In the conditioned medium IP a band is seen at 36 kDa which could be either ESAG9-EQ or ESAG9-K69. This band was not present when the blocking peptide was used (far right lane). Size markers are in kilo Daltons.

4.4 Analysis of shedding of ESAG9-K69 protein by stumpy parasites

ESAG9-K69 protein was shown to be shed by transgenic bsf K9:K69 and bsf EQ:K69 cells. It was vital therefore to address the question of whether this phenomenon was seen purely in transgenic cells, or whether ESAG9-K69 protein was also shed by 'wild-type' pleiomorphic parasites. To this end, an anti-peptide antibody was raised against ESAG9-K69 to enable detection of this protein in wild-type cells.

4.4.1 Generation and verification of ESAG9-K69 anti-peptide antibody

Two peptides were chosen against which two antibodies would be raised in rabbits. The peptides chosen are shown in Figure 4.18, and are compared with the position of the peptides used to raise the ESAG9-K9 antibody. The sequences were searched against the *T. b. brucei* 927 genome to confirm that they were not homologous to other proteins in the genome. The antibodies were generated by Eurogentec.

```
ESAG9-K9        MLKVAVTALLTLRSCAGPAESRGTSLKVG-----TGGVGVTVSTYSGCSESWSPDGRNMV 55
ESAG9-K69       MLSFRTALFLSLILLVVCGESKASVEVTANAPRCSGYWDLSSGNYHCSPQPTGVVGPNGR 60
                **.. .: :*:*   .  .**:.:   ..     :*  .:: ..*  ..:. .  * *

ESAG9-K9        CGSPSTGPSKTVRRSNSVTSTPQKIQGQPVSGSIAGAKRQEMQTAAPSSGGAPHLTSVTL 115
ESAG9-K69       SPPPTPVSPTPVSRPARVASARTQVHEAPKDAVSSSTTESGTSRPPPDSDETSPTVSIQP 120
                . .*:. ....* *.  *:*:  :::  * ..  :.:...  . ..*.*. :.  .*:

ESAG9-K9        DNPREGKTQTKNTSDAVIRGPTERLTDLSSRGEPAHAPVGNGKATERVKQVHDGEQRDLE 175
ESAG9-K69       NDGVNG----HESKDQLVQPPDNSRVSESATS-PVADPDGQ-EVEQTVQELGTTEQRNAK 174
                ::  :*    :::.* ::: * :  .. *: . *.  * *: :. : *:::   ***: :

ESAG9-K9        GRGCKDQSPARSEPASGLMRDDGMDVSRPPARRTVSGAEQESTRELMTRNLSEQESAENS 235
ESAG9-K69       G-TIRESGKGEKVIAGDKAQGPNSGSVGQPLWSVVTQSGQSYTVGVAAVGVDVQESTENG 233
                *   ::.. ...  *..  :. . .    *   .*: : *. *  : : .:. ***:**.

ESAG9-K9        TQEKKSAANKGHAVMISAALTLISF 260
ESAG9-K69       TQGVRSVAKGNHSAMLSAALGLMSF 258
                **  :*.*: .*:.*:**** *:**
```

Figure 4.18: Alignment of ESAG9-K9 and ESAG9-K69 peptide sequences showing the peptides used to raise antibodies. The two peptides used to raise antibodies against ESAG9-K9 and ESAG9-K69 are highlighted in blue and red respectively.

The two anti-peptide antibodies, A and B, were tested by Western analysis at a range of concentrations against cell lysate from bsf K69 cells, induced and uninduced for ESAG9-K69 protein expression (data not shown). Antibody B was found to give a stronger signal, and further testing confirmed that a concentration of 1:250 gave a clean signal. Figure 4.19 shows a Western blot using the ESAG9-K69 anti-peptide antibody B against bsf K69 cell lysate. The same membrane was also probed with BB2 antibody that recognised the Ty-tag on the ESAG9-K69 protein. The LI-COR Odyssey system was used which allowed the detection of two antibodies simultaneously because they were raised in different animals and so could be visualised with two secondary antibodies with fluorochromes which were excited at different wavelengths.

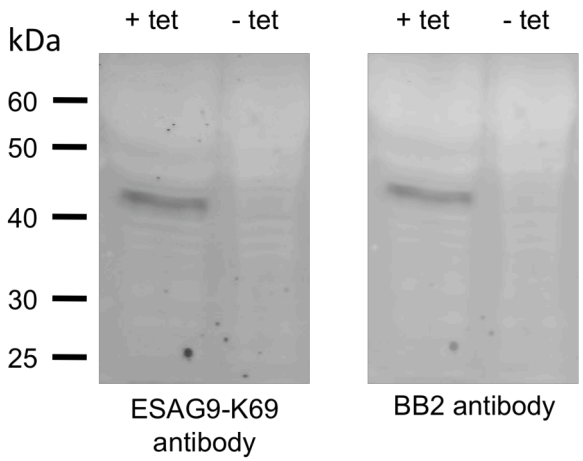


Figure 4.19: Confirmation that the ESAG9-K69 antibody recognises ectopically expressed ESAG9-K69 protein. Bloodstream form K69 cells were induced with tetracycline for ectopic expression of ESAG9-K69 protein (+tet, left hand lanes), and analysed by Western blotting along with uninduced cells (- tet, right hand lanes) using the α-ESAG9-K69 antipeptide antibody (left) and BB2 antibody that recognises the Ty-tag (right). This confirmed that there was an inducible signal with both antibodies. Size markers are in kilo Daltons.

4.4.2 Stumpy conditioned medium contains ESAG9-K69 protein

A mouse was inoculated with slender form pleiomorphic *T. b. brucei* EATRO 2340 parasites and the infection allowed to progress until the population was predominantly stumpy in morphology (5 days). The parasites were harvested from the mouse (a total of 1×10^9 cells), re-suspended in serum-free HMI-9 medium at a concentration of 3×10^6 cells per ml, and incubated at 37°C for 4 hours. The length of incubation was a balance between allowing sufficient time for any proteins secreted by the stumpy cells to accumulate in the medium, and the fact that the stumpy cells would not survive indefinitely in serum-free medium. The stumpy cells were harvested by centrifugation and cell lysate made at a concentration of 2×10^5 cells per μl in Laemmli buffer. The stumpy-conditioned medium was filtered to remove any remaining trypanosomes, and reduced in volume by centrifugation through Vivaspin columns. The volume was reduced by a factor of 1,332 to 250μl. The concentrated stumpy-conditioned medium was then boiled in 50% v/v laemmli buffer. The cell lysate and stumpy-conditioned medium samples were analysed by Western blotting using the ESAG9-K69 anti-peptide antibody. A control antibody, Hsp-70, was also used, as this protein should always remain cell associated and should not be shed into the medium.

Figure 4.20 shows the Western analyses using ESAG9-K69 and Hsp-70 antibodies. There was a signal from both antibodies from the stumpy-conditioned medium and the stumpy cell lysate. Several bands were detected by the ESAG9-K69 antibody. At 50 kDa, there was a single band in the cell lysate lane, and a double band in the conditioned medium lane. At 25 kDa there was a very faint single band in the cell lysate lane, and a faint double band in the conditioned medium lane. The predicted size of ESAG9-K69 is 26.6 kDa. The lower band is therefore of approximately the correct size for ESAG9-K69, and the higher band could be the protein having been extensively post-translationaly modified. The higher band migrates at approximately twice the mass of the lower band so it could represent a dimer of the protein, however it is likely that such an interaction would have been disrupted by the

denaturing conditions of the SDS-PAGE gel. Alternatively, the protein may be running aberrantly and therefore the size marker would not be equivalent to the true size of the protein. It is also worth noting that in the transgenic bsf K69 cell line, the Ty-tagged ESAG9-K69 protein runs at 50kDa (refer to Figure 4.9).

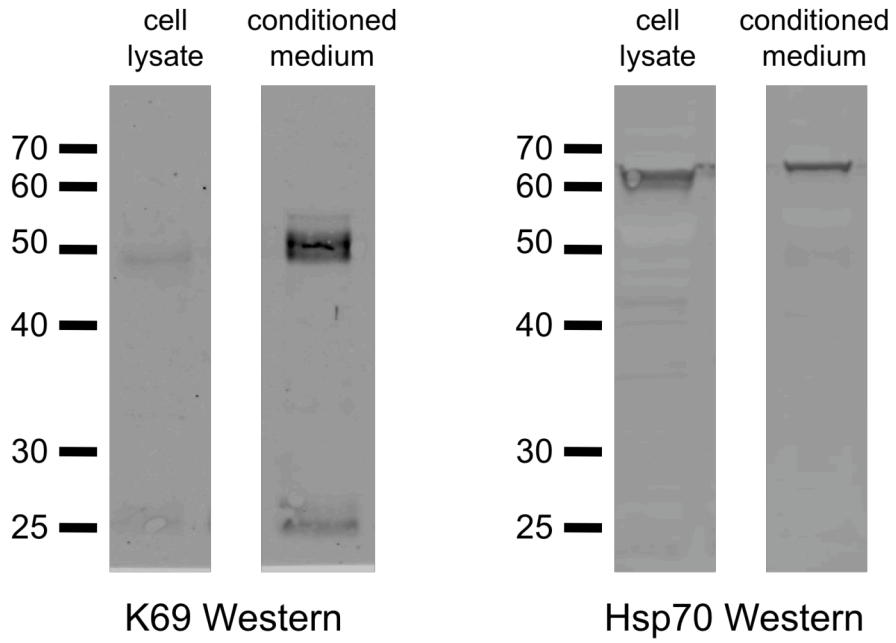


Figure 4.20: Western analysis of cell lysate and conditioned medium from *T. b. brucei* EATRO 2340 stumpy cells. Stumpy cells were harvested from a mouse and incubated at 37°C in serum-free medium for 4 hours. The cells were harvested by centrifugation and boiled in Laemmli buffer. The stumpy-conditioned medium was reduced in volume by centrifugation through Vivaspin columns. The samples were then analysed by Western blotting with ESAG9-K69 antibody (left hand lanes) and Hsp70 antibody (right hand lanes). There is a signal for both antibodies from both the cell lysate and conditioned medium lanes; the signal for ESAG9-K69 is stronger in the conditioned medium lane and the signal for Hsp70 is stronger in the cell lysate lane. The lanes were not equally loaded, with approximately 13.3 times more cell equivalents loaded in the conditioned medium lanes. The cell lysate and conditioned medium lanes are from the same blot scanned in at the same time, but irrelevant lanes were cropped for clarity. Size markers are in kilo Daltons.

The signal for the ESAG9-K69 protein was much stronger in the medium, and the signal for Hsp-70 was stronger in the cell lysate. However an equal number of cell equivalents were not loaded in each lane; 2.4×10^6 cell equivalents were loaded in the cell lysate lane, whereas 3.2×10^7 cell equivalents were loaded in the stumpy-conditioned medium lane. To quantify the relative signal from the lanes, the intensity of each band was measured using a LI-COR Odyssey 2.1 infrared imager and software. The measured intensity from the conditioned medium lanes was adjusted to take account of the fact that it was overloaded in comparison to the cell lysate lanes, and the percentage of signal from each antibody in the cell lysate versus the conditioned medium was then calculated. These values are shown in Table 4.1. The percentage of ESAG9-K69 remaining cell-associated in this experiment was 60.6%, and 39.4% of the signal for the protein was in the stumpy-conditioned medium (the upper band of 50kDa was used to calculate these values). Contrastingly, 95.6% of the signal for Hsp-70 protein remained cell-associated and 4.4% was in the stumpy-conditioned medium. It is not surprising that there was some Hsp-70 protein in the medium because some of the stumpy cells may have lysed due to the stress of incubation in the serum-free medium, or other manipulations such as the centrifugation step.

The fact that the percentage of ESAG9-K69 protein in the stumpy-conditioned medium was much higher than for Hsp-70 was strongly suggestive of ESAG9-K69 protein being shed actively by stumpy cells.

Antibody	Intensity of signal from conditioned media band	Intensity of media band corrected for cell equivalents	Intensity of signal from cell lysate band	**% total signal from cell lysate**	**% total signal from conditioned media**
K69	8.18	0.61	0.94	**60.6**	**39.4**
Hsp-70	9.96	0.75	16.22	**95.6**	**4.4**

Table 4.1: The quantification of the percentage of the signal from ESAG9-K69 and Hsp-70 antibodies that was cell-associated or from the conditioned medium. The intensities of the relevant bands were measured using LI-COR infrared imaging software, and the bands from the medium were then adjusted for cell equivalents. The percentage of the total signal that was from the cell lysate or the conditioned medium was then calculated.

4.4.3 The expression profile of ESAG9-K69 protein during differentiation

It was also interesting to use the newly-generated ESAG9-K69 anti-peptide antibody to explore the expression of the protein during *in vitro* differentiation from stumpy to procyclic cells. A Western blot was carried out using the same samples that were used to explore ESAG9-K9 protein expression during differentiation, as described in Chapter 3 section 3.1.5. Samples were taken at 0, 3, 6, 9, 12, 15 and 30 hours into differentiation and the Western blot is shown in Figure 4.21. The observed protein expression was weak and seemingly fairly consistent across the lanes, although comparison with the tubulin loading control showed that the lanes from the later time points were relatively overloaded. This indicated that, in fact, there was less ESAG9-K69 protein present at 9, 12, 15 and 30 hours than there was at 0, 3 and 6 hours into differentiation. The ESAG9-K69 protein migrated at approximately 48 kDa, and there was no band seen at the lower size of 25 kDa, unlike in Figure 4.20. The protein expression profile contrasts with that of ESAG9-K9 during differentiation, whereby ESAG9-K9 expression peaks between 6 and 9 hours into differentiation and by 30 hours its expression was longer detectable.

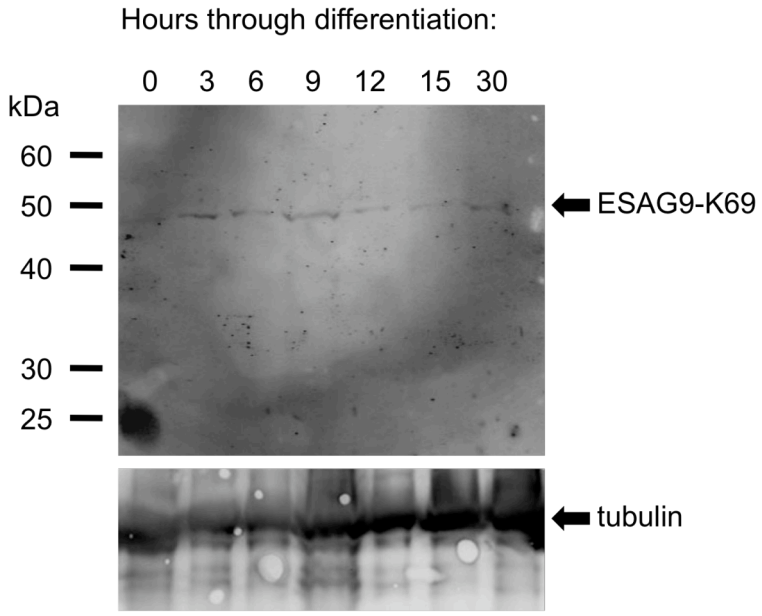


Figure 4.21: Expression profile of ESAG9-K69 protein during *in vitro* differentiation of *T. b. brucei* EATRO 2340 stumpy cells. Stumpy cells were harvested from a mouse and triggered to differentiate by the addition of cis-aconitate (6mM) and a drop in temperature to 27°C. Protein samples were prepared at 0, 3, 6, 9, 12, 15 and 30 hours into differentiation to procyclic forms. The protein expression profile of ESAG9-K69 protein is shown (top) along with a loading control of anti-tubulin antibody (bottom). The loading control indicates that loading was not equal across all the lanes, with the 9, 12, 15 and 30 hour time points being overloaded in comparision to the earlier time points. The size markers are in kilo Daltons.

4.5 Immunofluorescence analysis of the sub-cellular location of ESAG9 proteins in transgenic cell lines

4.5.1 ESAG9-K9 is detectable in the endoplasmic reticulum

In order to determine the precise sub-cellular location of ESAG9-K9 protein in transgenic cell lines, immunofluorescence staining of paraformaldehyde-fixed procyclic form cells was carried out. Pcf K9 cells were induced by addition of 2μg/ml tetracycline and after 48 hours cells were fixed in 2% paraformaldehyde and adhered to glass slides. The fixed cells were permeabilised with 0.5% Triton X-100 and blocked with 1% BSA to prevent non-specific binding of antibody. The BB2 antibody was then applied for one hour in a humid chamber, followed by a FITC-conjugated anti-mouse secondary antibody to allow visualisation. To ensure that any signal seen was due to the presence of ESAG9-K9 protein and not the result of non-specific binding, two controls were also used. Pcf K9- (i.e. uninduced) cells were stained with primary and secondary antibody as described. Also, wild-type (WT) cells were exposed to the secondary antibody alone. These controls were included in all subsequent immunofluorescence staining experiments described later in this chapter, but will not be shown.

Figure 4.22 shows that the the immunofluorescent signal was seen as a distinct web-like pattern throughout the cell. The pattern of staining was compatible with the ESAG9-K9 protein being located to the endoplasmic reticulum. To further explore this possible localisation, bsf K9 + cells were fixed in paraformaldehyde as described previously. The cells were then co-labelled with BB2 and an antibody against the endoplasmic reticulum marker BiP (Bangs *et al.,* 1993). BiP antibody was raised in rabbits and so could be visualised with an anti-rabbit Alexa Fluor® 633 secondary antibody, which is visible through a different fluorescent channel to FITC.

The co-labelled cells were then subjected to analysis using a confocal microscope. A series of pictures of a chosen cell were taken through the Z-axis in both green (for

BB2) and the red (for the BiP endoplasmic reticulum marker) channels. These pictures were then combined using Zeiss LSM Image Browser software to create a 'Z-stack', which is a 3-dimensional approximation of a cell. The Z-stacks were converted to movies, available on the Supplementary CD, and these can be viewed using programmes such as Windows Medium Player or Quicktime (available to download free from http://www.apple.com/quicktime/download/).

Movie 1 shows the localisation, in 3D, of the ESAG9-K9 protein in bsf K9 + cells. Movie 2 shows the localisation of the endoplasmic reticulum marker BiP. Movie 3 is the combined data from Movies 1 and 2 and any yellow staining indicates a co-localisation of the two proteins. These data show that K9 and BiP are co-localised to the endoplasmic reticulum. Two-dimensional images of the same cell are also shown in Figure 4.23.

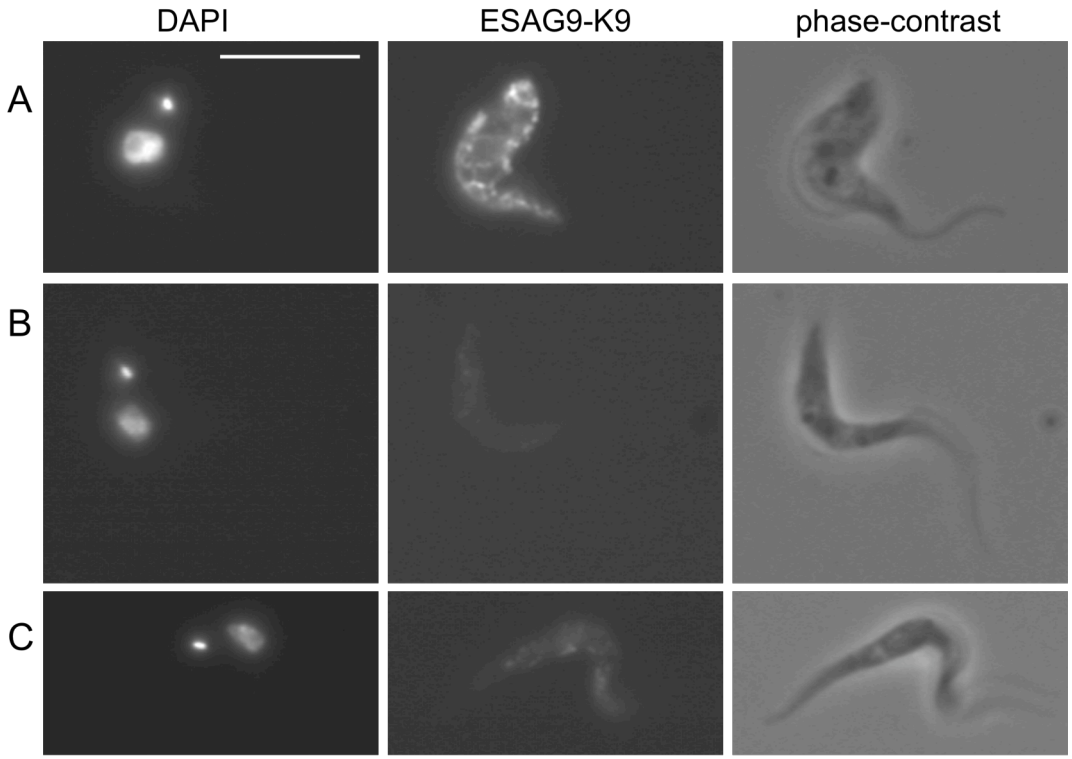


Figure 4.22: Location of Ty-tagged ESAG9-K9 protein in procyclic form cells. Procyclic form cells were fixed and stained with both BB2 primary antibody and an anti-mouse FITC secondary antibody, or with just the FITC secondary antibody. Row A is pcf K9 + cells stained with both antibodies, row B is WT cells stained with both antibodies, and row C is WT cells stained with the FITC secondary antibody alone. The signal from the FITC channel (middle column) is much stronger in the K9 + cells and indicates that the ESAG9-K9 protein is found in a web-like pattern through the cell. The left column shows DAPI staining of the cell nucleus and kinetoplast, and the right column shows phase-contrast images of the cells. The scale bar is 15 microns.

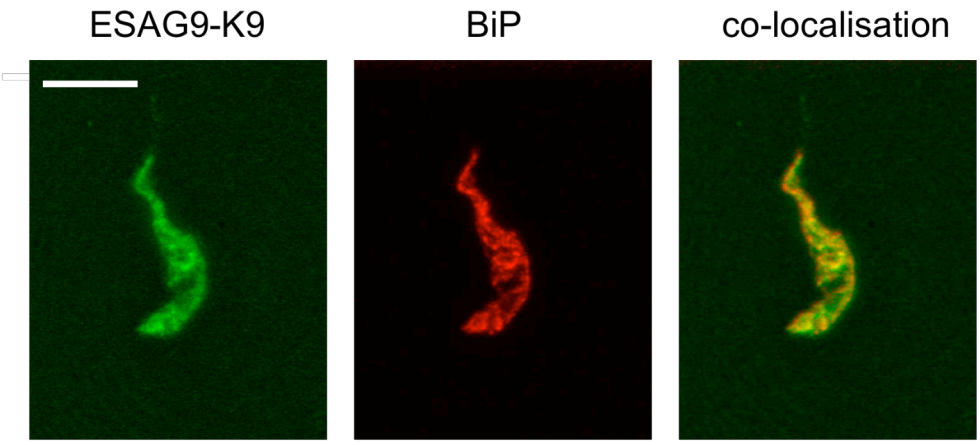


Figure 4.23: Co-localisation of Ty-tagged ESAG9-K9 protein and the endoplasmic reticulum marker BiP in bloodstream form cells. Bsf K9 + cells were co-stained with BB2, which recognises Ty-tagged ESAG9-K9 (green channel, left), and αBiP which is an endoplasmic reticulum marker (red channel, middle). Cells were visualized using a Zeiss confocal microscope and images analysed using Zeiss LSM Image Browser software. The right panel is an overlay of the red and green channels and co-localisation of the two signals is indicated in yellow. Scale bar represents 15 microns.

4.5.2 ESAG9-EQ is more diffuse in the cell than ESAG9-K9

Pcf EQ + and bsf EQ + cells were fixed and stained with BB2 antibody as described in Section 1.3.2. The procyclic-form staining is shown in Figure 4.24. The signal was more diffuse through the cell than for ESAG9-K9, with a gap where the nucleus is located. It was not completely evenly distributed throughout the cell however and was slightly flocculant. This pattern of staining was not indicative of ESAG9-EQ localising to a particular organelle.

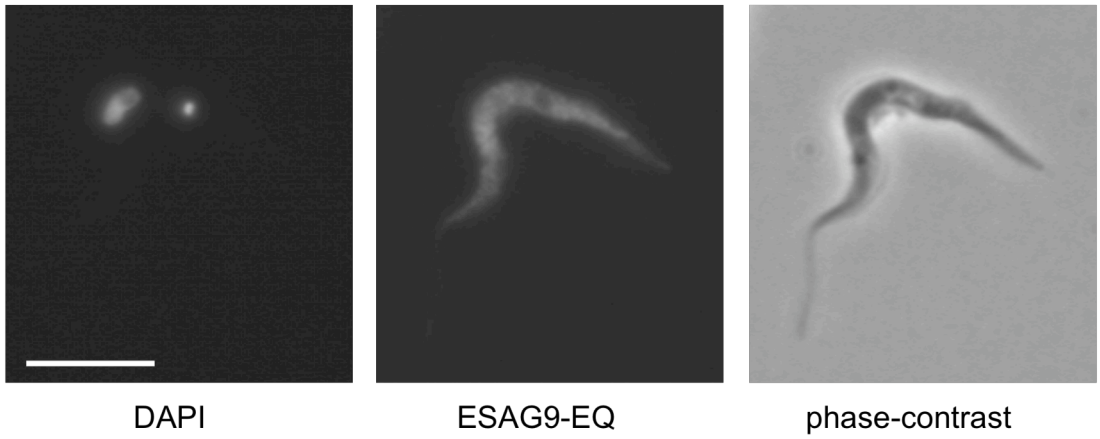


Figure 4.24: Localisation of Ty-tagged ESAG9-EQ protein in procyclic form cells. Pcf EQ + cells were stained with BB2 antibody, which recognises the Ty-tagged ESAG-EQ protein (middle panel). The left panel shows the nucleus and kinetoplast stained with DAPI and the right panel the phase-contrast image of the cell. The scale bar represents 15 microns.

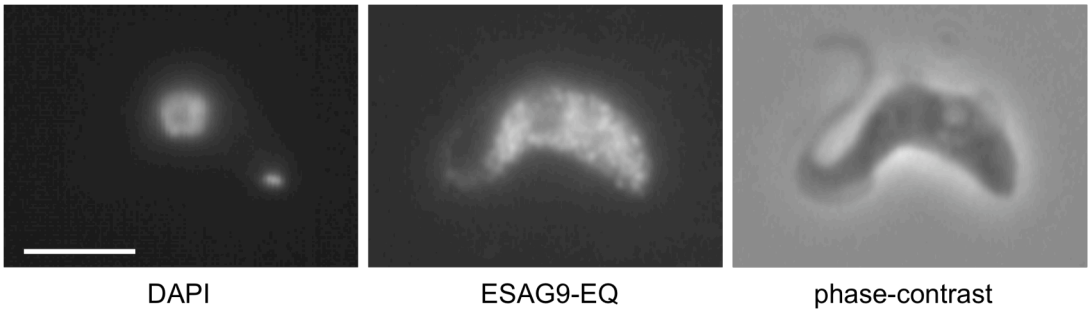


Figure 4.25: Localisation of Ty-tagged ESAG9-EQ protein in bloostream form cells. Bsf EQ + cells were stained with BB2 antibody, which recognises the Ty-tagged ESAG9-EQ protein (middle panel). The left panel shows the nucleus and kinetoplast stained with DAPI and the right panel the phase-contrast image of the cell. The scale bar represents 15 microns.

The localisation of ESAG9-EQ in bsf EQ + cells is shown in Figure 4.25. The pattern of staining was similar to that of procyclic cells, with a diffuse signal through the cell and a gap where the nucleus was located. ESAG9-EQ protein may be in the cytoplasm in both bsf and pcf EQ + cells. However it is also possible that this signal is from the cell surface. The gap in signal where the nucleus is could be due to the DAPI, which is used to visualise DNA, quenching the signal from the FITC-conjugated secondary antibody. This has been reported in other experiments (K.M., personal comment). To definitively determine whether ESAG9-EQ protein was internal or cell-surface located, confocal microscopy or electron microscopy would have been required, but this was not carried out as part of this project.

4.5.3 ESAG9-K69 is sometimes seen in the flagellar pocket or surrounding the cell

The location of Ty-tagged ESAG9-K69 protein in bsf K69, bsf EQ:K69 and bsf K9:K69 cells was also investigated. Bsf K69 +, bsf EQ:K69 + and bsf K9:K69 + cells were fixed and stained with BB2 antibody as described previously. In the bsf K69 + cells no fluorescent signal above the background was visible (data not shown). This may be because the expression of ESAG9-K69 protein in this cell line was quite low, as discussed in section 4.2.3. The bsf K9:K69 cell line ectopically expresses both un-tagged ESAG9-K9 protein, and Ty-tagged ESAG9-K69 that is recognised by the BB2 antibody. In this cell line, a signal from BB2 was seen from the flagellar pocket in some cells, and an example of such a staining is shown in Figure 4.26. A ‘dotty’ pattern of staining was also seen on the slide surrounding the fixed cells (see Figure 4.27). This was indicative of some of the ESAG9-K69 protein being released by the cells, consistent with the immunoprecipitation data shown in section 4.3.3. A signal from the flagellar pocket or from around the cells on the slide was not seen when bsf K9+ cells were immunofluorescently stained with the ESAG9-K9 anti-peptide antibody (data not shown). However, when stumpy *T. b. brucei* EATRO

2340 cells were stained with the ESAG9-K9 anti-peptide antibody a signal was sometimes seen from the flagellar pocket (refer to Chapter 3 Figure 3.5).

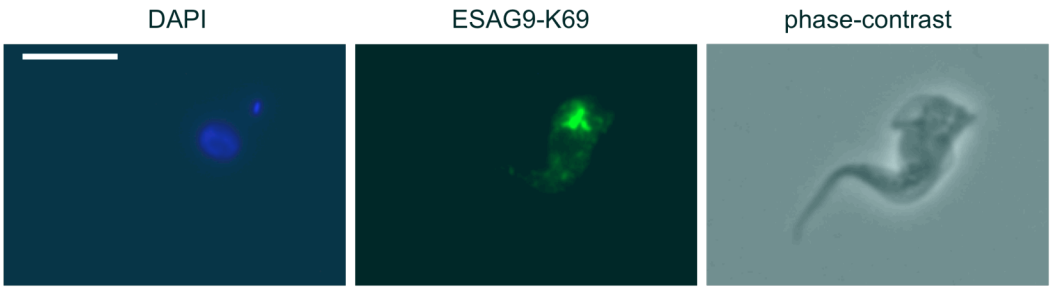


Figure 4.26: The location of ESAG9-K69 protein in a bloodstream form cell. Bsf K9:K69 + cells were fixed and stained with BB2 antibody which recognises the Ty-tagged ESAG9-K69 protein (middle panel) but does not recognise un-tagged K9 protein. A faint signal is seen from the cell body and a stronger signal from the flagellar pocket. The left panel shows the nuclei and kinetoplast stained with DAPI, and the right panel is a phase-contrast image of the cell. The scale bar represents 15 microns.

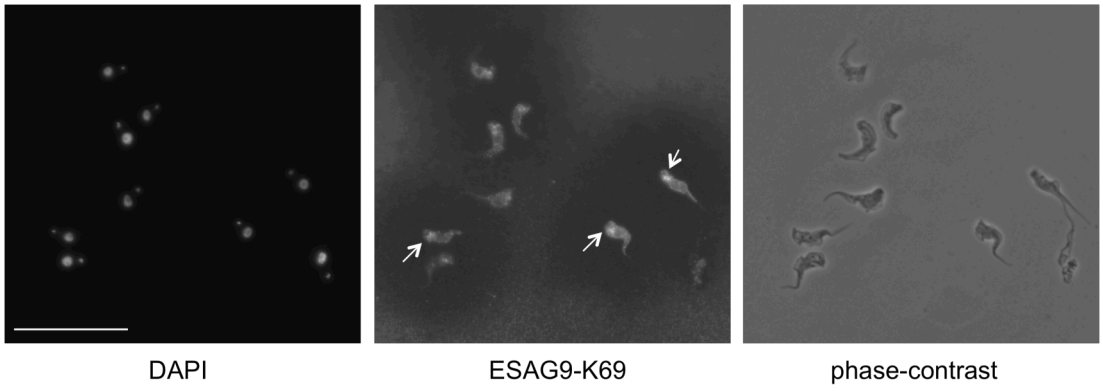


Figure 4.27: The location of K69 protein in bloodstream form K9:K69 cells. Bsf K9:K69 + cells were fixed and stained with BB2 antibody which recognises the Ty-tagged ESAG9-K69 protein (middle panel) but does not recognise un-tagged K9 protein. A faint signal is seen at the cell body and in some cells a signal is also seen at the flagellar pocket (as indicated by white arrows). There was also BB2 antibody adhered to the slide, as indicated by the dotty staining surrounding the cells. This was an indication that ESAG9-K69 protein was possibly being secreted and supports the immunoprecipitation analysis. The left panel shows the nuclei and kinetoplasts of the cells stained with DAPI, and the right panel is a phase-contrast image of the cells. The scale bar represents 50 microns.

The ESAG9-K69 antibody was also utilised in immunofluorescent assays to explore the localisation of the protein in the stumpy cells. No signal was seen for the protein, which could have been because the protein was being shed, or it alternatively it could have been because the ESAG9-K69 anti-peptide antibody was not effective for immunofluorescence. This was not explored any further due to time constraints.

In conclusion, from the immunofluorescence assays discussed in this section, the distribution of different ESAG9 proteins seems to differ in the cell. ESAG9-K9 was associated with the endoplasmic reticulum, ESAG9-EQ was cytosolic or at the cell surface, and ESAG9-K69 was released from the cell, at least when expressed concurrently with another ESAG9 (ESAG9-K9 or ESAG9-EQ).

4.6 Analysis of post-translational modification of ESAG9 proteins

4.6.1 N-glycosylation of ESAG9 proteins

As discussed in Chapter 3, Section 3.2.6, the ESAG9 gene family appears to bear some resemblance to the mucin and MASP families of proteins in the related parasite *Trypanosoma cruzi*. These proteins are extensively post-translationally modified by addition of N-glycan chains to the peptides. Bioinformatic analysis using an online prediction tool (NetNGlyc, see Chapter 3 section 3.2.8.2) revealed potential sites for N-glycosylation in many of the ESAG9 protein sequences (a schematic diagram of the locations of potential N-glycosylation sites in the peptide sequences is shown in Chapter 3 Figure 3.13).

Therefore the question of whether the ESAG9 proteins expressed by transgenic cells were N-glycosylated was addressed using a simple assay. The enzyme Peptide N-Glycosidase F (PNGaseF) cleaves N-glycans from proteins and is available from New England Biosciences as a kit. To investigate this potential modification, bsf K9, bsf K69 and bsf EQ cells were induced by addition of 2μg/ml tetracycline. After 48 hours the parasites were harvested and the cell pellet resuspended in a Glycoprotein Denaturing Buffer. Samples were boiled for ten minutes then split into two aliquots

and incubated for one hour at 37°C either with or without the addition of PNGaseF enzyme. The samples were then analysed by Western blotting with either the ESAG9-K9 anti-peptide antibody or the BB2 antibody, as required.

The ESAG9-K9 protein ran at a lower size after incubation with PNGaseF enzyme, as shown in Figure 4.28. The decrease in size was approximately 4 kDa. Also, the ESAG9-K69 protein ran approximately 2kDa lower after treatment with PNGase F. However the ESAG9-EQ protein did not show a band-shift after enzyme treatment. This indicates that whilst the ESAG9-K9 and ESAG9-K69 proteins were N-glycosylated, ESAG9-EQ was not.

Bioinformatic analyses had revealed 3 potential sites for N-glycosylation in the ESAG9-K9 protein sequence, 2 sites for N-glycosylation in ESAG9-K69 and one potential site in ESAG9-EQ. When VSG221 receives one glycan chain, it increased in size from 53 kDa to 54.5 kDa and when it receives two glycan chains, it increased in size from 53 kDa to 57 kDa (Jones *et al.,* 2005). Another ESAG protein, ESAG7, has three potential sites of N-glycosylation (Schell *et al.,* 1991b), and treatment with PNGase F resulted in a band shift of approximately 5 kDa (Salmon *et al.,* 1994). Although approximate, the size of the band-shifts of ESAG9 proteins therefore suggests that ESAG9-K9 may be glycosylated at two of its 3 potential glycosylation sites and ESAG9-K69 may be glycosylated at one of its two potential glycosylation sites.

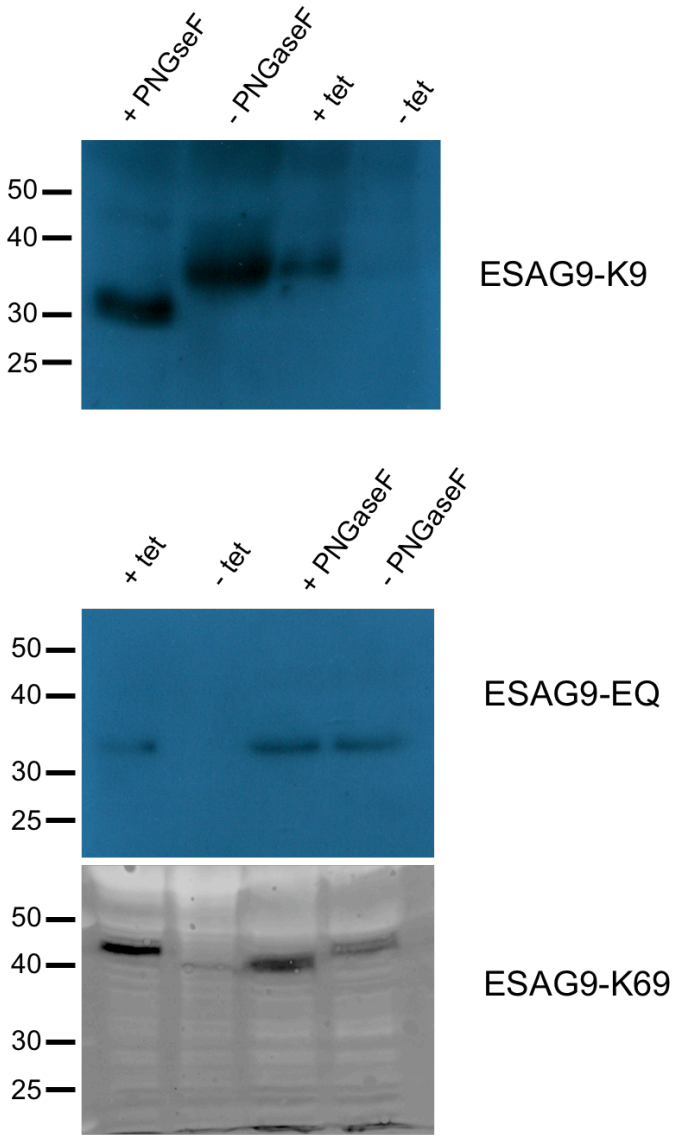


Figure 4.28: N-glycosylation of ESAG9 proteins. Bloodstream form cells expressing either ESAG9-EQ, -K9, or -K69 protein were subjected to enzymatic digestion with PNGaseF which cleaves N-glycans. '+tet' indicates cell lysate from cells induced for protein expression; '-tet' indicates uninduced cells; '+PNGaseF' cell lysate was incubated with the enzyme for one hour at 37°C; and '-PNGaseF' was incubated in the enzyme buffers for one hour at 37°C without the addition of the enzyme. Western blotting with BB2 antibody indicated that all cell lines showed inducible expression of the Ty-tagged protein. ESAG9-EQ protein did not change size as a result of PNGaseF treatment, whereas ESAG9-K9 protein changed size by approximately 4 kDa, and ESAG9-K69 by 2 kDa, indicating that both these proteins were N-glycosylated. Size markers are in kilo Daltons.

4.6.2 GPI-anchor addition to ESAG9 proteins

Bioinformatic analyses revealed the potential for addition of GPI anchors to ESAG9 proteins (see Chapter 3, section 3.2.8.1). Three different online tools were used to predict GPI anchor addition and all the ESAG9 proteins were predicted by at least one of them to have a GPI anchor. The most likely candidates were Tb09.v1.0330 and ESAG9-EQ as these were predicted to have an anchor by all three programmes.

The usual function of a GPI anchor would be to anchor a protein to the plasma membrane of the cell, as is the case for *T. brucei* surface proteins such as VSG (Ferguson *et al.,* 1985a). Immunofluorescent staining of transgenic cells expressing ESAG9-K9, ESAG9-K69 or both proteins concurrently did not reveal any staining of the cell surface. Work by the Bangs lab has shown that there is an alternative fate for reporter proteins with only one GPI anchor. Specifically, procyclin was ectopically expressed as a transgenic protein in bloodstream form cells, and this resulted in the addition of a single bloodstream form GPI anchor to the protein and its subsequent release into the medium (Schwartz *et al.,* 2005). The location of ESAG9-EQ in transgenic bsf EQ cells was not definitively characterised, but it is possible that the protein was at the cell surface.

Bloodstream form cells ectopically expressing either ESAG9-K9 or ESAG9-EQ were subjected to a hypotonic lysis assay to explore whether they were GPI anchored. This assay utilises the endogenously expressed GPI-PLC enzyme to cleave any GPI anchors in lysed cells, resulting in proteins that were GPI anchored being released into the supernatant, and those which were not anchored remaining in the cell pellet after centrifugation. This assay was not carried out with bloodstream form cells expressing ESAG9-K69 because the level of protein that was cell-associated in this cell line was insufficient. Transgenic bloodstream form cells were induced for expression of either ESAG9-K9 or ESAG9-EQ by addition of 2μg/ml tetracycline and the cells harvested by centrifugation after 48 hours and washed in ice-cold VPBS (Voorheis-modified PBS, for recipe see Materials and Methods). The cell pellets

were then resuspended in 100μl of pre-warmed hypotonic lysis buffer and incubated at 37°C to allow the GPI-PLC enzyme to act. Cells expressing ESAG9-K9 were incubated for 0, 5 and 10 minutes at 37°C as a range-finding experiment. Cells expressing ESAG9-EQ were only incubated for 5 minutes as this turned out to be the optimal length of time for incubation.

The samples were analysed by Western blotting and probed with either the ESAG9-K9 anti-peptide antibody, or the BB2 antibody in the case of ESAG9-EQ. The results are shown in Figure 4.29. In the case of ESAG9-K9 (panel A) there was a band of the correct size for the protein in only the pellet fraction (P) for the three incubation times used. The protein was not released into the supernatant, which suggested that the protein was not GPI anchored. In the case of ESAG9-EQ (panel C) there was a strong signal for the protein from the supernatant (S) as well as the pellet fractions, which suggests that this ESAG9 protein was GPI anchored and was being cleaved by the action of endogenous GPI-PLC enzyme. VSG was used as a positive control in both of these experiments. In both cases there was a signal from the VSG-221 antibody (which detects the VSG expressed by these cell lines) from the supernatant as well as the pellet, indicating that GPI-PLC activity was intact in both of these experiments.

The experiment was also repeated to see whether the ESAG9-K9 protein was GPI-anchored in WT stumpy-form *T. b. brucei* strain EATRO 2340 cells (Figure 4.29 panel B). Again the ESAG9-K9 localised to the pellet fraction after five minutes incubation at 37°C and there was no indication of cleavage by GPI-PLC. Unfortunately it was not possible to run a positive control for VSG in this instance because it was not known which VSG gene the *T. b. brucei* EATRO cells were expressing.

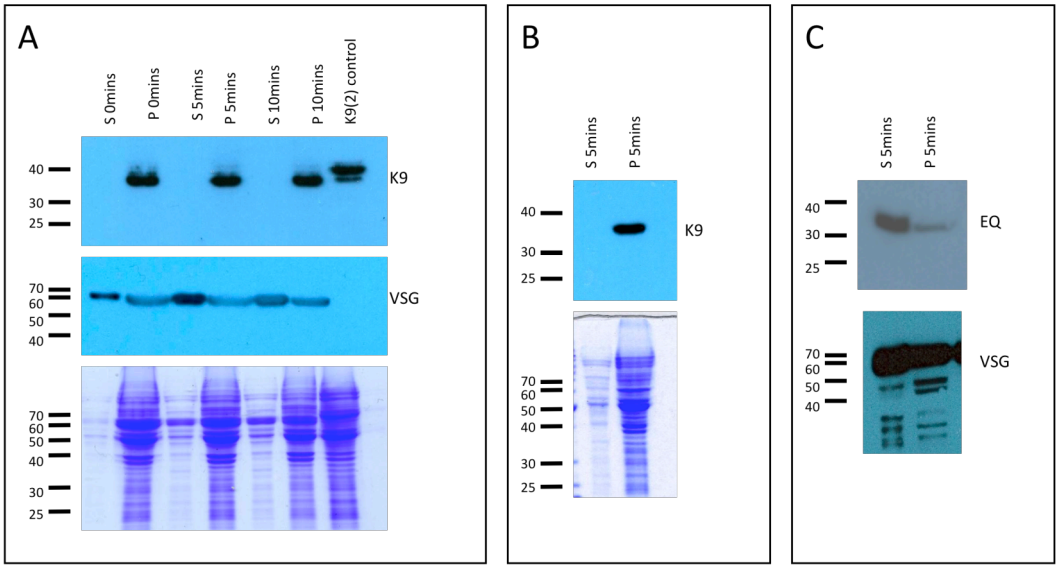


Figure 4.29: Hypotonic lysis experiments with transgenic bsf K9 and bsf EQ cells and pleiomorphic *T. b. brucei* EATRO 2340 cells. Cells were lysed with hypotonic lysis buffer and incubated at 37°C for 0, 5 and 10 minutes in the case of bsf K9 cells (panel A), or just for 5 minutes for the pleiomorphic *T. b. brucei* EATRO 2340 cells (panel B) or bsf EQ cells (panel C). Panel A shows a Western blot with the ESAG9-K9 antipeptide antibody (top), a Western blot with VSG-221 antibody (middle) and a coomassie stained gel (bottom). Panel B shows a Western blot with the ESAG9-K9 antibody (top) and a coomassie stained gel, but it was not possible to do a Western with an antibody against VSG as it was not known which VSG the *T. b. brucei* EATRO 2340 cells were expressing. Panel C shows a Western blot with BB2 antibody (top), and a Western blot with VSG-221 antibody (bottom). In this instance, there was proteolytic degradation of the VSG occurring, which explains the presence of multiple bands. Size markers indicated are in kilo Daltons.

To explore whether ESAG9-K69 protein was GPI anchored, two different techniques were employed. The first technique involved radioactively labelling cells with tritiated myristate. The triated myristate is incorporated into any newly synthesised GPI anchors, and hence GPI-anchored proteins can be detected as they become radioactively labelled.

Two cell types were used for this assay: bsf K9:K69 and bsf EQ. The bsf EQ cells were included as a positive control because the ESAG9-EQ protein is thought to be GPI anchored (see Figure 4.29). Cells were induced by addition of 2μg/ml tetracycline, harvested after 48 hours, and equilibrated in fatty-acid free medium (modified RPMI medium, see Buxbaum *et al*., 1994) for 30 minutes. A radioactive 'pulse' of 80 microcuries of triated myristate was then added and the cells incubated at 37°C for three hours. The cells were harvested by centrifugation, and both the cells and the conditioned medium subjected to immunoprecipitation to pull down the Ty-tagged protein (ESAG9-EQ protein from bsf EQ cells or medium, and ESAG9-K69 protein from bsf K9:K69 cells or medium), as previously described. The IPs were resolved on an SDS-PAGE gel and the gel dried and exposed to an autoradiography film for 5 weeks in a -80°C freezer.

Unfortunately no signals were seen from the cell or medium IPs from either the bsf EQ or the bsf K9:K69 samples (data not shown). This negative result had two potential explanations. Either the quantity of radioactive myristate used was insufficient to get a signal, or neither the ESAG9-EQ or ESAG9-K69 proteins were GPI-anchored. It was not possible to distinguish between these outcomes.

It was not known by what precise mechanism the ESAG9-K69 protein exits the cell in bsf K9:K69 cells. If the ESAG9-K69 protein was GPI-anchored then one possible pathway involved trafficking of the protein to the cell surface, followed by cleavage of the GPI anchor by endogenous GPI-PLC enzyme, which would release the protein from the cell. Cleavage of a bloodstream form GPI anchor from a protein reveals a domain called the Cross Reacting Domain (CRD), which is recognised by the anti-

CRD antibody (Cardoso de Almeida & Turner, 1983). Therefore Western blot analysis was used to determine whether the ESAG9-K69 protein, which was pulled out of conditioned medium by IP, cross-reacted with the anti-CRD antibody.

The Western analysis is shown in Figure 4.30. No signal was seen in the ESAG9-K69 IP lane with the anti-CRD antibody. Therefore the ESAG9-K69 protein was not having a GPI anchor cleaved when it exited the cell. This did not rule out the possibility, however, that the ESAG9-K69 protein had a GPI anchor that was not cleaved.

To confirm that the anti-CRD antibody was working, it was also used in Western analysis of the bsf EQ GPI hypotonic lysis experiment previously discussed in section 4.5.2. Multiple bands were seen in the supernatant (S) fraction, also shown in Figure 4.30, with the very strong signal likely to be representing cleavage of VSG protein. This confirmed that the anti-CRD antibody was recognising proteins from which the GPI anchor had been cleaved.

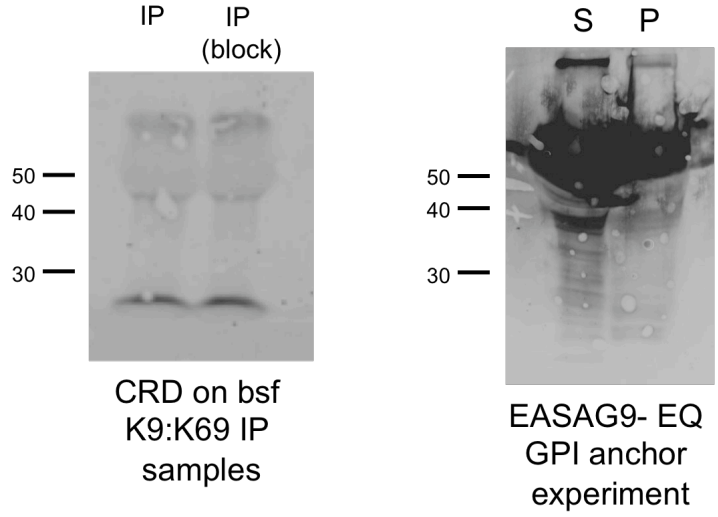


Figure 4.30: Determining presence of proteins from which a GPI anchor has been cleaved using the α-CRD antibody. The α-CRD antibody was cross-reacted against a bsf K9:K69 conditioned media IP (left hand panel). However no signal was seen from the IP in the absence of blocking peptide, indicating that the IP had not specifically selected any protein with a cross-reacting domain. To confirm that the α-CRD worked, it was also reacted against a Western blot of the bsf EQ hypotonic lysis experiment (refer to Figure 4.29, panel C), shown in the right-hand panel. There were multiple bands in the supernatant fraction (S) which confirmed that the antibody worked. Size markers indicated are in kilo Daltons.

4.7 Summary

In this chapter, transgenic cell lines and anti-peptide antibodies were used in combination with various techniques to attempt to resolve the sub-cellular localisation and the post-translational modifications of three ESAG9 proteins. The findings are summarised in Table 4.2.

ESAG9-K9 protein was found to be located to the endoplasmic reticulum, and is N-glycosylated. It was not shed from bloodstream form cells, nor did it appear to have a GPI anchor. The endoplasmic reticulum is the site of addition of the N-glycans (Shental-Bechor & Levy, 2008), so it is possible that once the protein is glycosylated, the transgenic cell lines then for some reason fail to properly traffic the ectopic ESAG9-K9 protein to its final destination.

ESAG9-EQ protein is located either in the cytoplasm or at the cell surface. An immunoprecipitation (IP) using BB2 antibody and conditioned medium from bsf EQ:K69 cells resulted in a signal for a Ty-tagged protein. This could have been either the ESAG9-EQ protein or the ESAG9-K69 protein although the size that the protein migrated at was indicative of it being ESAG9-EQ. ESAG9-EQ protein is GPI-anchored, as shown by a hypotonic lysis experiment. It is not N-glycosylated.

ESAG9-K69 protein was shed by both transgenic bsf K9:K69 cells and by *T. b. brucei* EATRO 2340 stumpy cells (and possibly by bsf EQ:K69 cells, see previous paragraph). Immunofluorescent staining of bsf K9:K69 cells revealed that some cells had a signal from the flagellar pocket, and also that there was a dotty pattern of staining on the slides which could be secreted protein. ESAG9-K69 protein showed a small band-shift upon treatment with PNGase F enzyme, indicating that it is N-glycosylated, and probably only at one out of two possible sites predicted bioinformatically.

It was not possible to determine whether or not ESAG9-K69 protein was GPI-anchored, for the reasons described in section 4.6.2.

The different sizes that ESAG9-K69 migrated at in SDS-PAGE gels was somewhat perplexing. In the bsf K69 cell line, which ectopically expresses ESAG9-K69 protein alone, the protein ran at 50 kDa (see Figure 4.9). When the ESAG9-K69 protein was expressed concurrently with either ESAG9-K9 or ESAG9-EQ, the ESAG9-K69 protein ran at approximately 40 kDa (see Figures 4.10 and 4.11). An anti-peptide antibody was raised against ESAG9-K69 and reacted against stumpy cell lysate and stumpy-conditioned medium (see Figure 4.20). In the cell lysate, single bands were detected at approximately 25 kDa and 50 kDa. In the stumpy-conditioned medium, double bands were seen at 25 kDa and 50 kDa. The different sizes of the ESAG9-K69 protein could be the result of dimerisation or post-translational modification. However N-glycosylation is probably not solely responsible for the different sizes of the proteins; treatment of ectopically-expressed ESAG9-K69 protein with PNGase F only resulted in a small change in size of the protein.

ESAG9 protein	Expressed in bsf transgenic	Expressed in pcf transgenic	N-glycosylated	GPI-anchored	Localisation
ESAG9-K9	✓ Both individually and with ESAG9-K69	✓	✓	✖	In the endoplasmic reticulum
ESAG9-EQ	✓ Both individually and with ESAG9-K69	✓	✖	✓	Either in the cytoplasm or surface-located. May be secreted when expressed concurrently with ESAG9-K69
ESAG9-K69	✓ Both individually and with either ESAG9-K9 or ESAG9-EQ. Weak expression.	✖	✓	**?**	Secreted when expressed concurrently with ESAG9-K9

Table 4.2: Summary of findings regarding the characterisation of three ESAG9 proteins. The properties of three ESAG9 proteins, ESAG9-K9, ESAG9-EQ and ESAG9-K69, were investigated by generating transgenic cell lines that ectopically expressed one or more of these proteins. These cell lines were then used in assays to determine the occurrence of post-translational modifications and the localisation of the proteins.

Chapter 5 : Functional analysis of the role of ESAG9 proteins in the early colonisation of the tsetse fly midgut: *in vitro* and *in vivo* experiments

5.1 Introduction

Bloodstream stumpy forms are taken up in the tsetse fly bloodmeal, and upon entry into the fly mid-gut, receive signals that cause them to differentiate to the next life cycle stage, the procyclic form. Procyclic forms do not have the protective VSG coat, which is lost during the first 4 to 6 hours by pleiomorphic differentiating parasites (Roditi *et al.,* 1989; Ziegelbauer *et al.,* 1990), and therefore are susceptible to any mammalian immune factors that remain in the digesting blood meal.

Due to the unusual expression profile of ESAG9-K9 protein, whereby the protein is expressed in the early stages of differentiation from stumpy forms to procyclic forms *in vitro* (see Chapter 3 Figure 3.5), we hypothesised that the protein could have a function in the early colonisation of the tsetse fly midgut. For example, it may protect the parasites in some way from complement present in the blood meal or the fly innate immune system.

There is not, to date, any published material regarding the activity of the mammalian immune response in the tsetse fly mid-gut. However, in mosquitoes, the alternative pathway of complement present in rats' blood remains active in the mid-gut for up to six hours post-feed (Margos *et al.,* 2001). The tsetse fly innate immune system also responds to trypanosome infection, involving the secretion of, for example, antimicrobial peptides (Hu & Aksoy, 2006).

To test the potential functions of ESAG9, a number of approaches were used. In particular, an alamarBlue assay was used which allowed parasite growth or survival to be monitored, such that the ability of ESAG9 proteins to protect the parasites

against lysis by mammalian sera could be assessed. Procyclic form parasites were used in these experiments, which are usually lysed by the complement in mammalian serum. AlamarBlue is a redox reagent that undergoes a change in redox potential as the nutrients in the parasite culture medium are depleted. The percent reduction of the alamarBlue can be calculated from the absorption at 540 and 595nm, allowing quantitative data to be obtained from a microplate reader. The difference in redox potential can also be observed by the naked eye, as seen in Figure 5.1, which shows an assay in which procyclic cells have been incubated in medium containing alamarBlue for 1 hour and 24 hours at 27°C.

As a second approach, *in vivo* experiments with tsetse flies were carried out whereupon flies were fed blood containing procyclic or bloodstream form transgenic trypanosomes, induced or not induced for ESAG9 protein expression. In these experiments, both microscopy and quantitative PCR were used to attempt to quantify parasite numbers, but with limited success, as will be discussed.

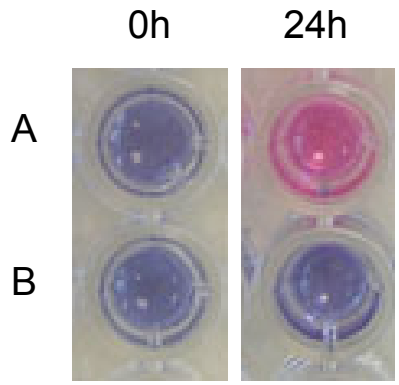


Figure 5.1: Procyclic form parasites growing in AlamarBlue over a period of 24 hours. Parasites were grown in a 96-well plate and the plate photographed at one and 24 hours. Row A shows procyclic form parasites grown in normal media with 10% AlamarBlue. The colour change from blue to pink indicates that the parasites have grown during a 24 hour incubation at 27°C. Row B shows media and AlamarBlue with no parasites present and there is no colour change after 24 hours.

5.2 ESAG9-K9 protein expression by procyclic form cells in vitro does not protect them from lysis by complement in sera from a range of different mammals

Transgenic procyclic parasites, engineered to express ESAG9-K9 protein under tetracycline induction (pcf K9 cells), were exposed to serial dilutions of various mammalian sera. Guinea Pig Serum (GPS) was used due to its known high concentration of active complement. Horse and Bovine sera were also utilised. It has been hypothesised by Bitter *et al*. (1998) and Pays *et al*. (2001) that a function of ESAGs may be to enhance the ability of African trypanosomes to survive in a range of mammalian hosts. It was possible that the expression of different ESAG9 proteins could protect parasites in the fly gut from complement from different mammals.

The mechanism of parasite killing by the serum can be determined by incubating procyclic form parasites in serum in the presence and absence of EDTA (data not shown). This chelates divalent cations, thereby inhibiting the alternative pathway of complement, which requires magnesium ions (Mg^{2+}) for its activity. Confirming this, microscopic observation revealed that parasite lysis by guinea pig serum was inhibited by the presence of 0.1M EDTA, indicating that the alternative pathway of complement was responsible for lysis rather than any other factors of the immune system. The classical pathway of complement, part of the adaptive immune response, would also be inhibited by the presence of EDTA, but this could not be responsible for lysis since the animals used for the sera were never exposed to African trypanosomes (the bovine and horse sera were sourced from New Zealand). Importantly, the sera used in these assays were never freeze-thawed more than once to ensure that the complement retained activity.

Further controls for theses assays involved exposing the parasites to guinea pig, bovine, and horse serum that had been heat-inactived by incubation at 57°C for 30 minutes, or to heat-inactivated Foetal Bovine Serum sourced from Gibco (HI-FCS).

Figure 5.2 shows an example of a 96-well plate used for this type of assay with procyclic form parasites growing in serial dilutions of GPS. Figure 5.3 represents graphically the reduction of alamarBlue by pcf K9+ (induced) and pcf K9- (uninduced) parasites, when grown for 24 hours in GPS. These data show that there

was no difference in growth between the induced and uninduced parasites, as assayed by the percentage reduction of alamarBlue, in either GPS, heat-inactivated GPS, or HI-FCS. The parasites grew most effectively in FCS, probably because they are culture-adapted to grow in this serum.

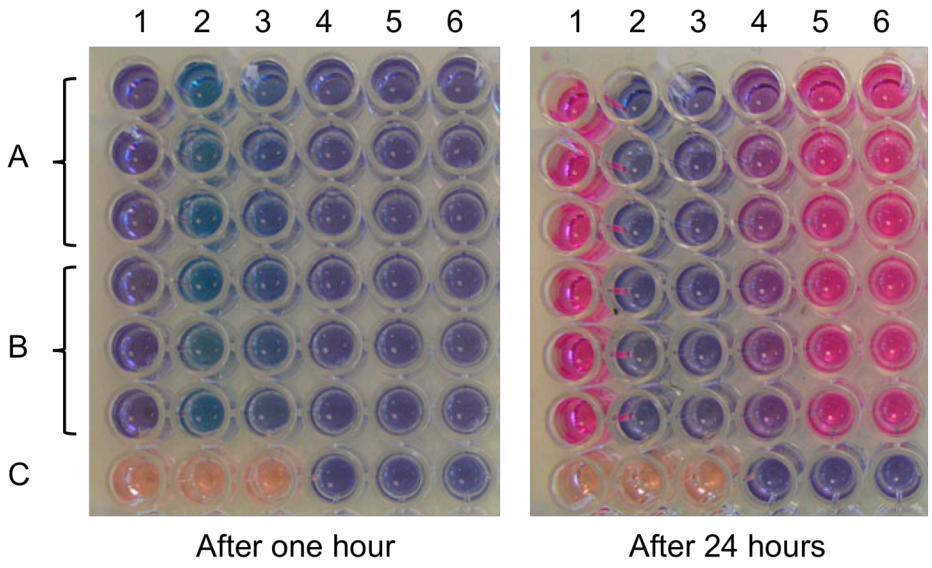


Figure 5.2: An alamarBlue assay with procyclic form parasites growing in serial dilutions of Guinea Pig Serum. Column 1 is a positive control with no guinea pig serum (GPS) present, and columns 2 to 6 are decreasing concentrations of GPS. Column 2 is 20% GPS and columns 3 to 6 are serial 1:2 dilutions. Rows A are procyclic forms induced for expression of ESAG9-K9; rows B are the uninduced controls, and row C is blank wells without (columns 1-3) and with (columns 4-6) alamarBlue. The pink wells are those in which the trypanosomes have grown the most after 24 hours incubation at 27°C.

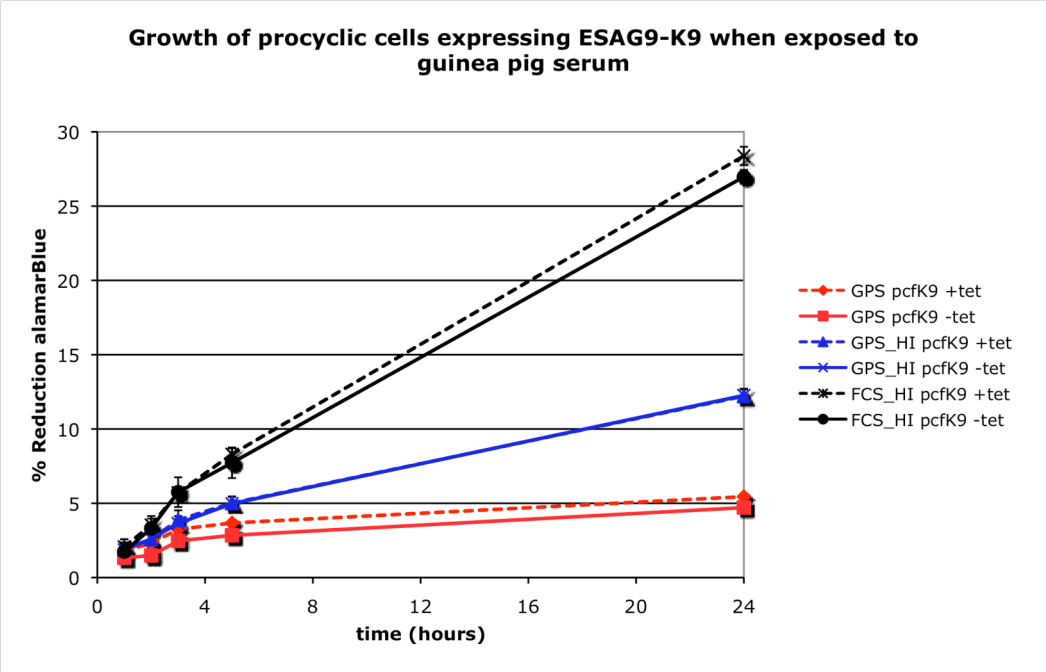


Figure 5.3: Growth of procyclic cells expressing ESAG9-K9 in guinea pig serum. Procyclic cells induced (+ tet; dashed line) and uninduced (-tet; solid line) for ESAG9-K9 protein expression were grown for 24 hours in either 10% active guinea pig serum (GPS; red); 10% heat inactivated GPS (GPS_HI; blue) or 10% heat-inactivated FCS (FCS_HI; black). Each point represents the mean of three wells and the error bars are the standard deviations from the means.

Figure 5.4 shows similar results for the growth of parasites in active bovine serum, heat-inactivated bovine serum, and heat-inactivated foetal calf serum. Parasite growth was not fully inhibited by active bovine serum. This could be because the bovine serum contained a low quantity or activity of complement. Another point to note is that the pcf K9+ grew less than the pcf K9- parasites when exposed to the active serum. However this is unlikely to be an important effect as it was not reproduced in other assays, and variability between wells was observed (as represented by the error bars which are the standard deviations from the mean).

Figure 5.5 shows the growth of pcf K9+ and pcf K9- parasites in active and heat-inactivated horse serum, and in heat-inactivated foetal calf serum. Interestingly, in this case, the parasites grew faster in the heat-inactivated horse serum than in the heat-inactivated foetal calf serum during the first five hours of the assay, as indicated by a greater reduction of alamarBlue over time. Again, the pcf K9+ parasites did not survive better than the pcf K9- parasites.

Hence, in these assays the ectopic expression of ESAG9-K9 protein did not favour survival or growth of procyclic form parasites in active guinea pig, horse, or bovine serum.

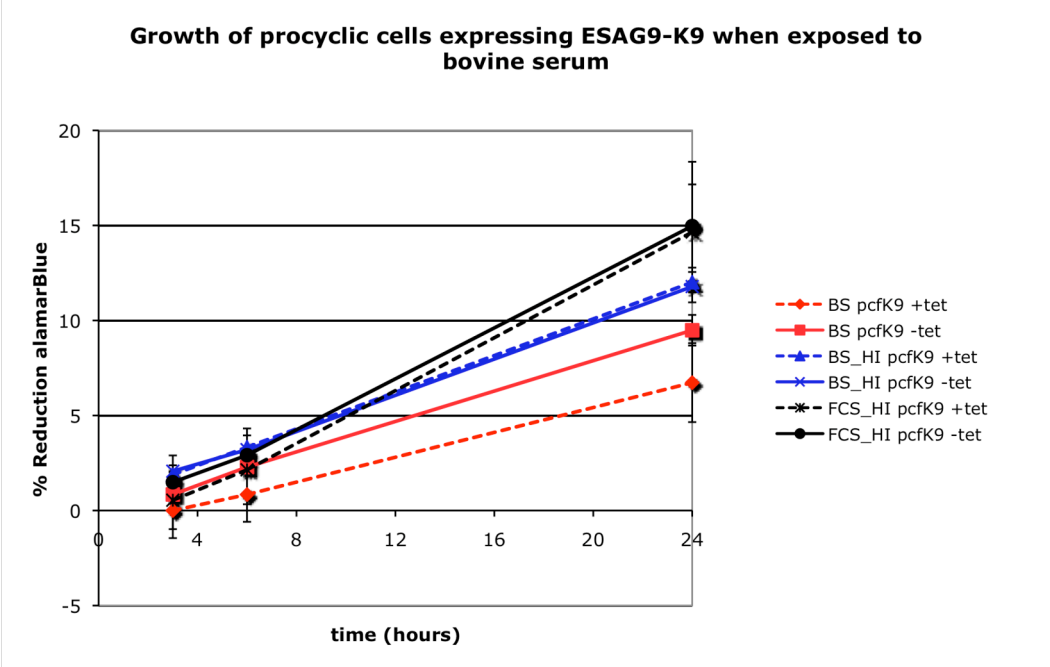


Figure 5.4: Growth of procyclic cells expressing ESAG9-K9 in bovine serum. Procyclic cells induced (+ tet; dashed line) and uninduced (-tet; solid line) for ESAG9-K9 protein expression were grown for 24 hours in either 10% active bovine serum (BS; red); 10% heat inactivated bovine serum (BS_HI; blue) or 10% heat-inactivated FCS (FCS_HI; black). Each point represents the mean of three wells and the error bars are the standard deviations from the means.

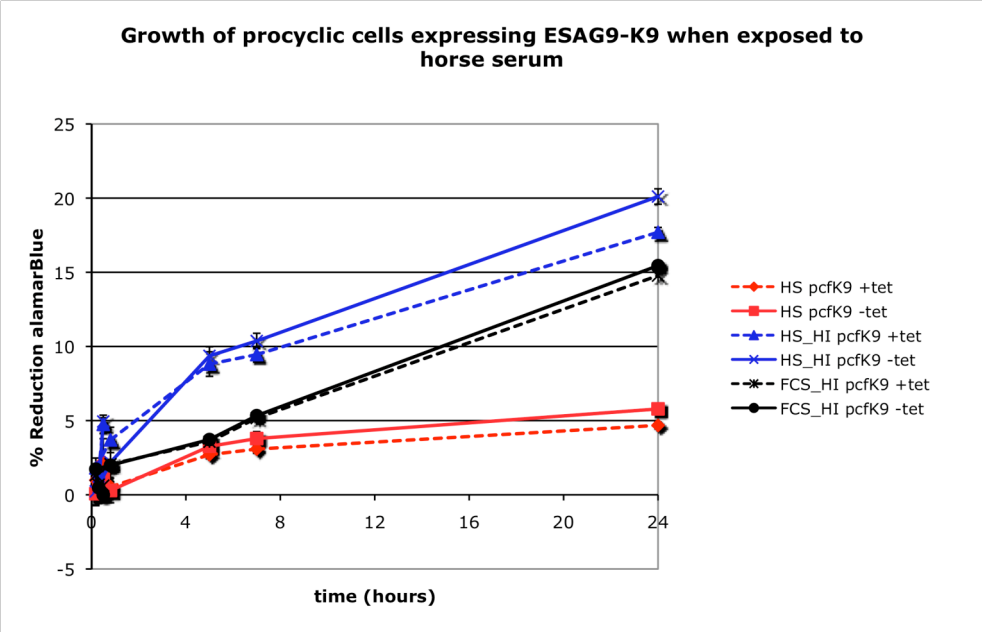


Figure 5.5: Growth of procyclic cells expressing ESAG9-K9 in horse serum. Procyclic cells induced (+ tet; dashed line) and uninduced (-tet; solid line) for ESAG9-K9 protein expression were grown for 24 hours in either 10% active horse serum (HS; red); 10% heat in-activated horse serum (HS_HI; blue) or 10% heat-inactivated FCS (FCS_HI; black). Each point represents the mean of three wells and the error bars are the standard deviations from the means.

5.3 ESAG9-EQ protein expression by procyclic form cells does not protect them from lysis by complement in guinea pig serum

ESAG9-EQ expression has also been shown to be regulated in a stage-specific manner. Like ESAG9-K9, its mRNA was up-regulated in stumpy form parasites, as shown by Northern blotting (see Chapter 3, Figure 3.2). This version of ESAG9 is named EQ because it is the most similar ESAG9 in the *T. b. brucei* genome database to the ESAG9 gene that was originally identified in *Trypanosoma equiperdum* (Florent *et al*., 1991).

Since there was no antibody available against ESAG9-EQ, it was not possible to confirm that the protein is expressed in a stage-specific manner. However, it was assumed that this was likely to be the case. In order to test whether this protein protects cells from lysis by the complement present in guinea pig serum, pcf EQ

parasites were exposed to GPS and their growth monitored by the alamarBlue colorimetric assay as detailed in Section 5.2.

As was the case for ESAG9-K9, it was found that the ectopic expression of ESAG9-EQ protein by procyclic form cells did not increase their survival in GPS (Figure 5.6). The relative growth of the uninduced and induced pcf EQ parasites was the same in the active serum, the heat-inactivated GPS, and in the heat-inactivated FCS. This demonstrated that the expression of ESAG9-EQ protein did not confer complement resistance on procyclic form parasites.

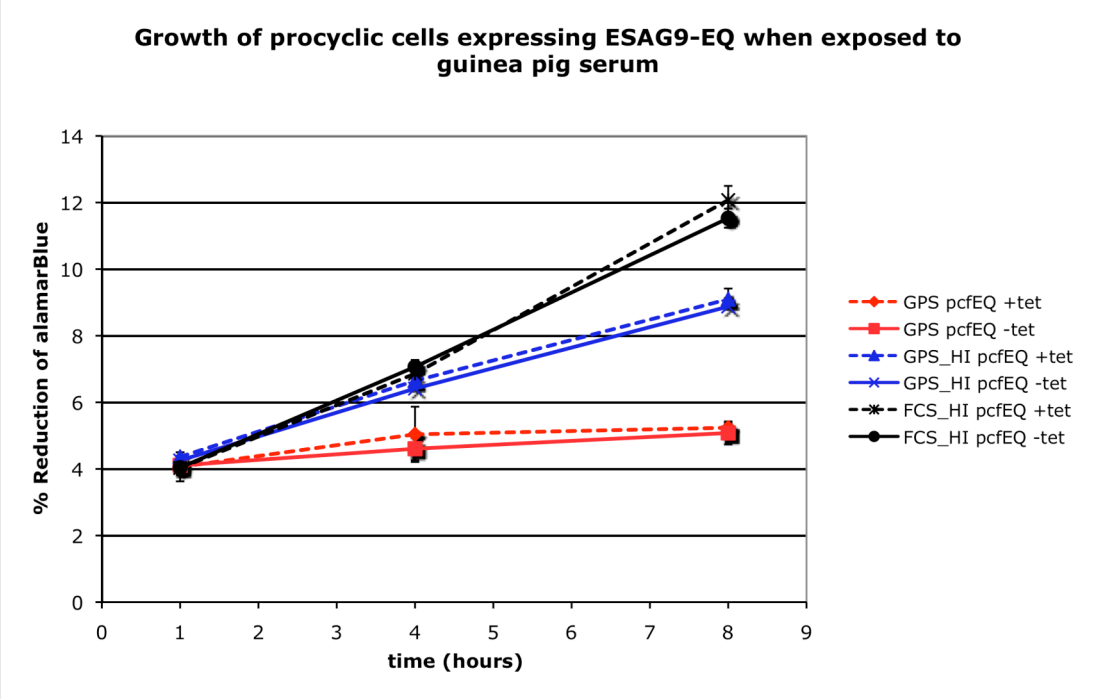


Figure 5.6: Growth in guinea pig serum of procyclic cells expressing ESAG9-EQ. Procyclic cells induced (+ tet; dashed line) and uninduced (-tet; solid line) for ESAG9-EQ protein expression were grown for 8 hours in either 10% active guinea pig serum (GPS; red), 10% heat inactivated guinea pig serum (GPS_HI; blue), or 10% heat-inactivated FCS (FCS_HI; black). Each point represents the mean of three wells and the error bars are the standard deviations from the mean.

5.4 ESAG9-K69 protein secreted from bloodstream form cells does not protect procyclic cells from lysis by complement in guinea pig serum

The transgenic bloodstream form cell line (called bsf K9:K69) that expresses both ESAG9-K9 and ESAG9-K69 proteins simultaneously has been shown to secrete ESAG9-K69 protein into the culture medium (see Chapter 4 Figure 4.15). This led us to hypothesise that, in an *in vivo* situation, pleiomorphic stumpy form cells could be secreting ESAG9-K69 whilst they are differentiating to procyclic forms in the tsetse fly mid-gut. A potential function of the secreted protein could be to inhibit the action of, for example, blood meal complement. This seems a more likely scenario than a protein which is intracellular having an inhibitory effect, since complement, unlike trypanosome lytic factor (see Introduction section 1.5.4.2), operates at the cell surface.

Given that bloodstream forms are not, due to their VSG coat, susceptible to lysis via the membrane attack complex, we used procyclic form cells to assay the potential function of the secreted ESAG9-K69 protein. Bloodstream form K9:K69 cells, induced or uninduced for the simultaneous expression of ESAG9-K9 and ESAG9-K69 protein, were grown in HMI-9 medium for 48 hours and the conditioned medium was then isolated. Procyclic form 427-449 cells (which do not express any ESAG9 proteins) were then exposed to: active GPS, heat-inactivated GPS, or heat-inactivated FCS, in the presence of the conditioned medium from induced or uninduced K9:K69 bsf cells. Haemin at a concentration of 0.1% was also included in the assay because procyclic form cells tend to clump in the absence of this supplement.

Figure 5.7 shows that the growth of procyclic cells was inhibited by the presence of 10% GPS whether they were incubated with conditioned medium from induced or uninduced bsf K9:K69 cells, as assayed by the percentage reduction of alamarBlue. In the control assays using either 10% heat-inactivated GPS or 10% heat-inactivated

FCS, the procyclic cells grew to the same extent whether they were incubated with conditioned medium from induced or uninduced cells.

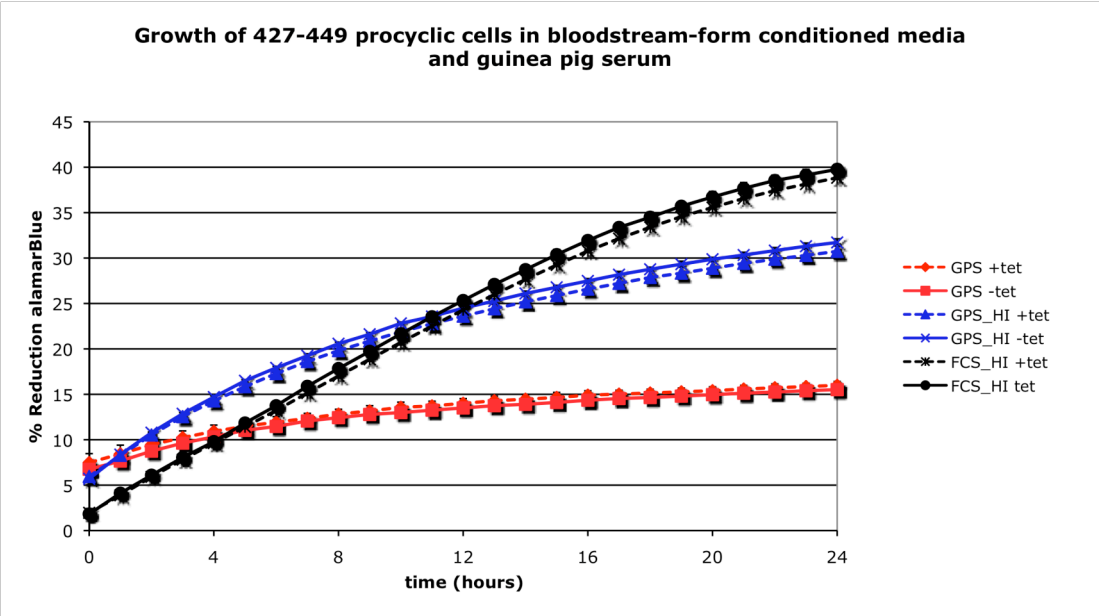


Figure 5.7: Growth of procyclic cells in guinea pig serum and conditioned media from induced or uninduced bsf K9:K69 cells. Procyclic cells (427-449 cells) were grown for 24 hours in 10% v/v of either active guinea pig serum (GPS; red); heat in-activated guinea pig serum (GPS_HI; in blue) or heat-inactivated FCS (in black). Growth in conditioned media isolated from bsf K9:K69 cells induced (+tet; dashed line) or uninduced (-tet; solid line) for expression ESAG9-K69 and ESAG9-K9 was assayed. Each point represents the mean of three wells and the error bars are the standard deviations from the means.

In case the presence of conditioned medium would only have an effect at lower concentrations of GPS, a range of serial dilutions of GPS were assayed, represented in Figure 5.8. At all of the concentrations of GPS the procyclic form cells grew to a similar degree whether they were in conditioned medium from induced or uninduced bloodstream form parasites. In these assays, reduction in alamarBlue was followed over a 24-hour time course using a Biotek® plate reader and integrated Gen5™ software.

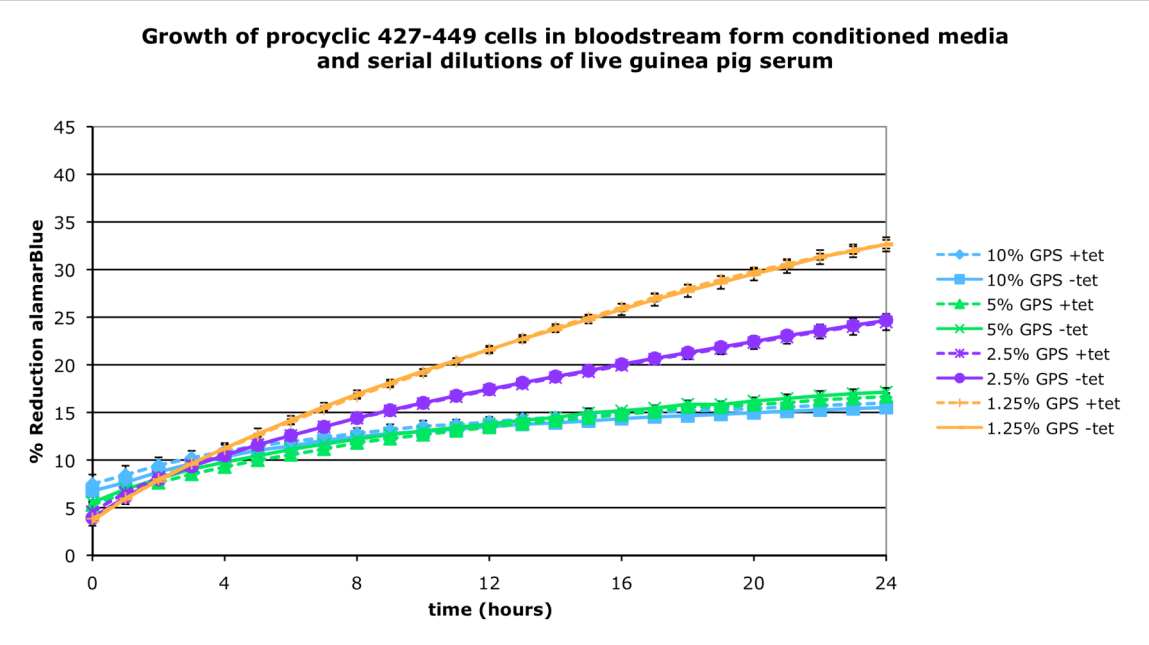


Figure 5.8: Growth of procyclic cells in bloodsteam form conditioned media and serial dilutions of guinea pig serum. Procyclic cells were grown for 24 hours in serial dilutions of active guinea pig serum, the highest being 10% v/v (pale blue), and the lowest being 1.25% v/v (orange). Growth in conditioned media isolated from bsf K9:K69 cells induced (+tet; dashed line) or uninduced (-tet; solid line) for expression of ESAG9-K69 and ESAG9-K9 was assayed. Each point represents the mean of three wells and the error bars are the standard deviations from the means.

5.5 Analysis of the tsetse infection efficiency of bloodstream form parasites expressing ESAG9-K9 protein

The assays described in sections 5.2 to 5.4 show that ESAG9 proteins do not protect procyclic cells from lysis by complement under the *in vitro* conditions used.

However there are other factors that are responsible for parasite death in the tsetse fly midgut. These include antimicrobial peptides secreted in response to invasion (Hu and Aksoy 2006) and lectins (Chandra *et al.,* 2004), as discussed in the Chapter 1 section 1.5.5. Rather than design *in vitro* assays to test all these factors, it was possible to look directly at the *in vivo* situation by feeding tsetse flies on horses' blood which is supplemented with cultured trypanosomes. Tetracycline can be included in this blood meal to induce the expression of ESAG9 protein in transgenic cells whilst they are in the fly mid-gut. This technique of induction of expression has been verified by Peacock *et al*. (Peacock *et al.,* 2005). The experiment in summarised in a flow chart, depicted in Figure 5.9.

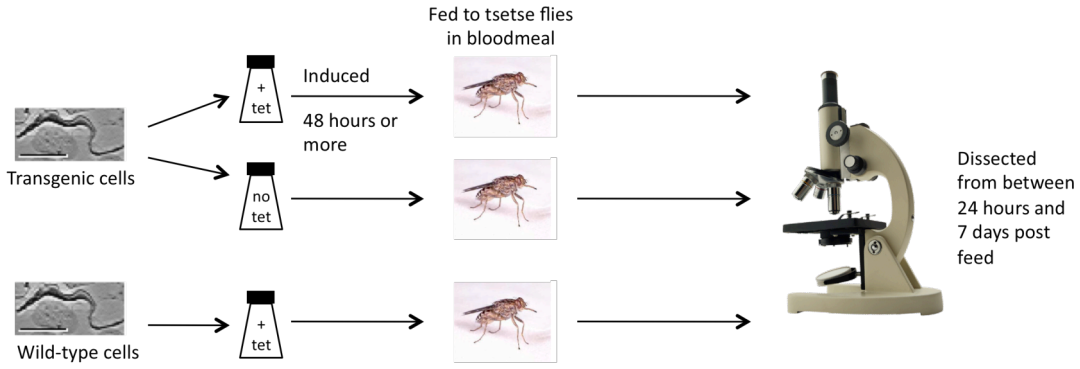


Figure 5.9: Schematic diagram of experimental infections of tsetse flies with cultured trypanosomes. Bloodstream or procyclic form transgenic (either bsf K9 or bsf EQ cell lines) or wild-type cells were cultured in the presence or absence of tetracycline, then added to the tsetse fly bloodmeal. Thereafter, the tsetse flies were dissected between 24 hours and 7 days post-blood meal and the mid-gut examined microscopically for the presence of trypanosomes.

Tsetse flies were fed (i) horse blood containing wild-type (427-449) bloodstream form cells with tetracycline, (ii) transgenic parasites induced with tetracycline to ectopically express one version of ESAG9, or (iii) uninduced transgenic parasites with no tetracycline. The addition of the tetracycline to the blood feed with wild-type cells was an important control because the tetracycline could potentially affect the population of symbiotic micro-organisms in the tsetse fly mid-gut. These in turn have been shown to be involved in the refractoriness of tsetse flies to trypanosome infection (Dale & Welburn, 2001; Moloo *et al.,* 1998). However, the concentration of tetracycline that we used, 25μg/ml, has been shown to have no effect on the longevity of tsetse flies (Peacock *et al*., 2005).

Dissections were carried out after either 24 hours or 72 hours. These time points were chosen in order to balance, on the one hand, the fact that in differentiating stumpy forms the ESAG9-K9 protein is only expressed for a few hours (see Chapter 3 Figure 3.5) and so is likely to function early on in the colonisation of the fly mid-gut and, on the other hand, the need to allow the parasites sufficient time to establish an infection so that their numbers could be quantified. The standard method used for quantifying tsetse fly infections by other investigators involves the scoring of the mid-gut infection as being 'heavy', 'intermediumte', 'weak', or 'negative' (for example:(Liniger *et al.,* 2004; Vassella *et al.,* 2009)). This seems subjective and so we decided to attempt to assess the infections in a more quantitative fashion by parasite counting.

In general, the level of infection that could be achieved in the tsetse fly mid-guts was very low. The infection level was estimated by removing the fly gut and dicing it in a small volume of PBS, and then observing a drop of the material under a microscope. After counting how many parasites were visible, the number of parasites in the whole volume of the mid-gut was estimated, although often either none or only one parasite was visible.

Figure 5.10 shows the detectable infection level in the flies 24 hours after feeding with horses' blood containing pcf K9 or wild-type (427-449) parasites. The number of flies with a detectable infection was low (the highest being 4 positives out of a total of 18 flies in the bsf K9+ group). Lower infection levels were achieved in the

flies fed with wild-type cells and those fed with bsf K9 – cells (uninduced for ESAG9-K9 expression).

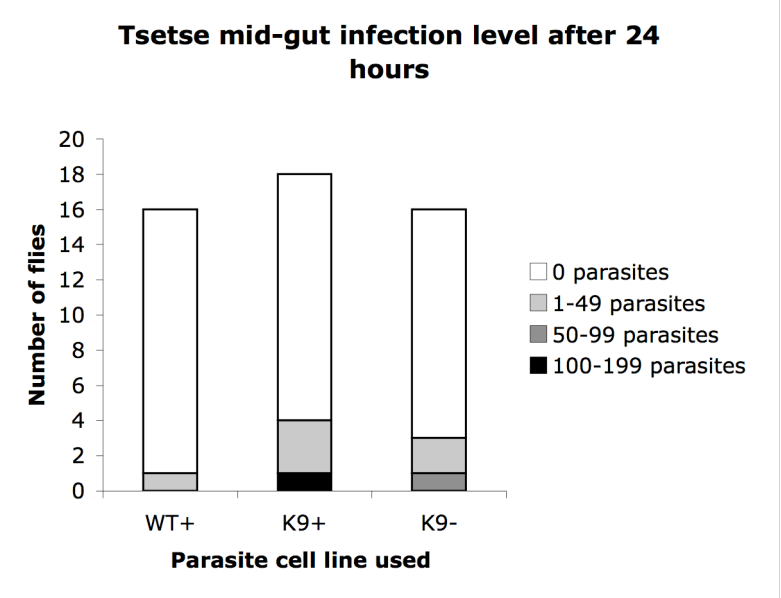


Figure 5.10: Experimental infection of tsetse flies with transgenic or wild-type parasites over 24 hours. Either wild-type bloodstream form (bsf) cells with tetracycline (WT+), bsf cells induced for ectopic expression of ESAG9-K9 protein (K9+), or uninduced cells (K9-) were fed to flies. These flies were then dissected and the mid-guts excised after 24 hours. The number of parasites present in the midgut was counted and the counts grouped into: zero parasites (white), 1-49 parasites per fly (pale grey), 50-99 parasites per gut (dark grey), or 100-199 parasites per gut (black).

The experiment was also carried out over a 72 hour period with the same cell lines. Figure 5.11 shows the results of an experiment where between 13 and 17 flies were used in each group. Figure 5.12 shows the results of a lager scale experiment where between 27 and 30 flies were used in each group. Again the detectable infection level was low, although there was a larger number of flies with a positive infection than after 24 hours. In both these experiments (Figure 5.11, Figure 5.12), the bsf K9+ achieved the most intensive infection level in individual flies. However, in the larger scale experiment (Figure 5.12) it was the bsf K9- parasites that showed the highest proportion of flies with a detectable infection.

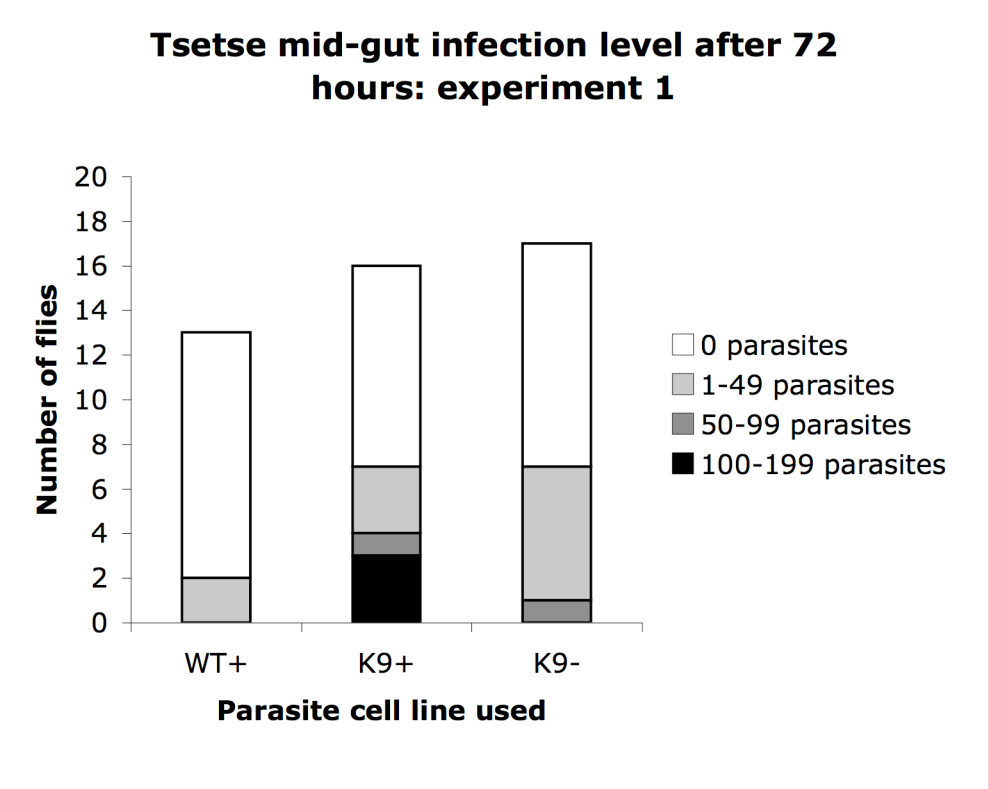


Figure 5.11: Experimental infection of tsetse flies with transgenic or wild-type parasites over 72 hours. Either wild-type bloodstream form (bsf) cells with tetracycline (WT+), bsf cells induced for ectopic expression of ESAG9-K9 protein (K9+), or uninduced cells (K9-) were fed to flies. These flies were then dissected and the mid-guts excised after 72 hours. The number of parasites present in the midgut was counted and the counts grouped into: zero parasites (white), 1-49 parasites per fly gut (pale grey), 50-99 parasites per gut (dark grey), or 100-199 parasites per gut (black).

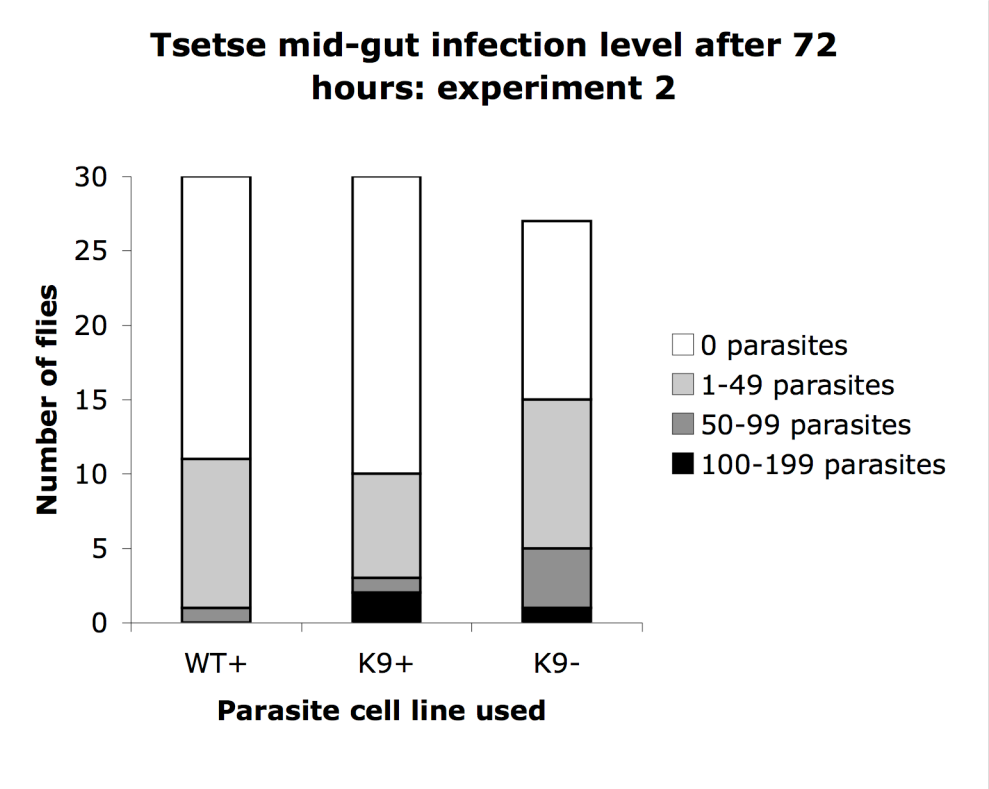


Figure 5.12: Experimental infection of tsetse flies with transgenic or wild-type parasites over 72 hours; repeat experiment with larger number of flies. Either wild-type bloodstream form (bsf) cells with tetracycline (WT+), bsf cells induced for ectopic expression of ESAG9-K9 protein (K9+), or uninduced cells (K9-) were fed to flies. These flies were then dissected and the mid-guts excised after 72 hours. The number of parasites present in the midgut was counted and the counts grouped into: zero parasites (white), 1-49 parasites per fly gut (pale grey), 50-99 parasites per gut (dark grey), or 100-199 parasites per gut (black).

5.6 Analysis of the tsetse infection efficiency of procyclic form parasites expressing ESAG9-EQ protein

As discussed in section 5.3, ESAG9-EQ is another ESAG9 gene that is up-regulated in stumpy form parasites. ESAG9-EQ protein could therefore possibly have a role in the establishment of the parasite infection in the early colonisation of the tsetse fly midgut, though it is important to note that an antibody has not been raised against ESAG9-EQ so the protein expression profile has not been investigated in wild-type cells. The microscopic counting of bloodstream-form infections in tsetse flies had proved challenging so it was decided in these experiments to use pcf EQ parasites that are inducible for ectopic expression of ESAG9-EQ protein. Procyclic form parasites are more adapted to the midgut environment and so were expected to survive better and provide higher levels of infection for microscopic counting. Although using procyclic forms was unrealistic because these parasites would not normaly express ESAG9s, procyclic forms are still susceptible to the immune factors in the mammalian blood feed and, possibly, the antimicrobial factors produced by the tsetse fly.

Teneral tsetse flies (those which are newly hatched and have not yet received a blood meal) were fed with blood containing (i) procyclic form parasites induced for ectopic expression of ESAG9-EQ, (ii) the same cell line but uninduced, or (iii) wild-type parasites with tetracycline. The fly mid-guts were dissected and the numbers of parasites present were counted by microscopy 48 hours and seven days post-feed. The results of the first assay in which the flies were dissected after 48 hours are shown in Figure 5.13. Figure 5.13, panel A shows the percentage the tsetse flies in each group that showed a detectable level of parasite infection by microscopy. Figure 5.13, panel B, shows the infection level in those flies that were positive for an infection. The graphical representation used is a box plot. The two boxes are the interquartile ranges, which contain twenty five percent of the values either side of the mediumn value. The bars represent the full spread of the data, with outliers represented by asterisks. These results indicated, unexpectedly, that those parasites

that were most effective at colonising the mid-gut of the flies were transgenic parasites that were not induced for ectopic ESAG9-EQ protein expression.

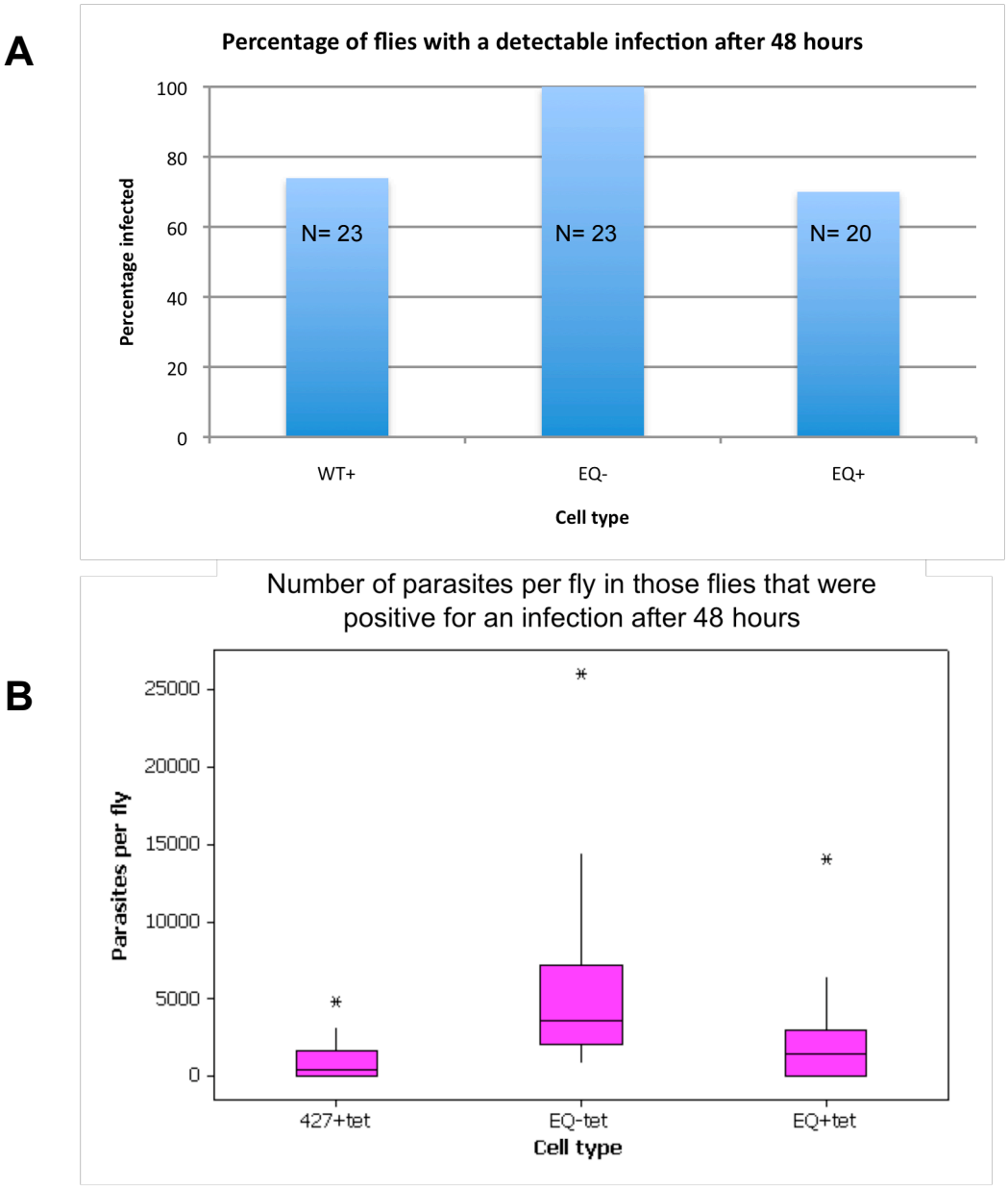


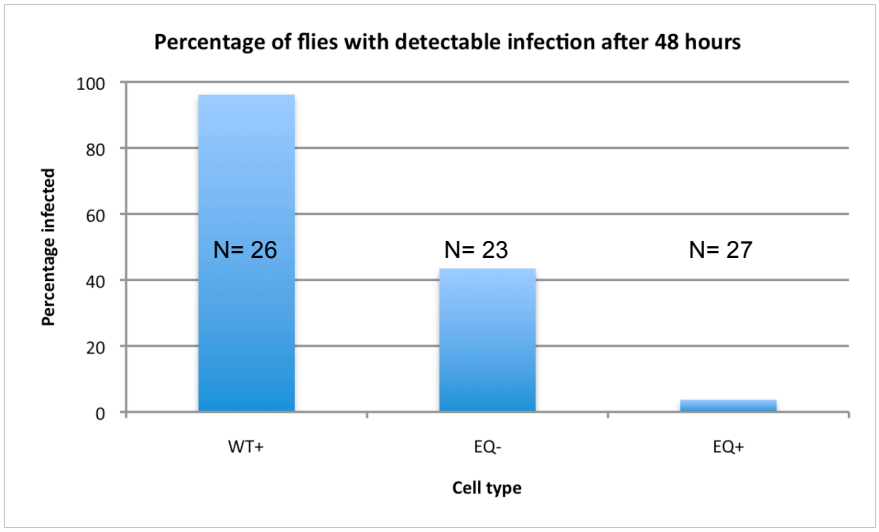
Figure 5.13: Experimental infections with procyclic form wild-type and transgenic parasites. Flies were fed with blood containing either wild-type procyclic form parasites with tetracycline (WT or 427 +); transgenic procyclic form parasites induced for ectopic expression of ESAG9-EQ protein (EQ+); or uninduced parasites (EQ-). The midguts were dissected after 48 hours and the trypanosome infection level counted via microscopy. Panel A shows the percentage of infected flies in each group; the number of flies used for each group is indicated on the bars. Panel B shows the number of parasites per fly in those flies which were infected; the results are represented as a Box Plot where the boxes represent the two interquartile ranges either side of the median value, and the asterisks represent anomalous outliers.

However, when this assay was repeated a quite different result was obtained. In this second experiment the parasite numbers were so low that they were scored only as 'positive' or 'negative', without quantification of the level of infection. The percentage of flies in each group with a positive infection are shown in Figure 5.14, panel A. These results suggested that the expression of ESAG9-EQ protein was in fact having a detrimental affect on the parasites' ability to colonise the tsetse midgut. Indeed, the wild-type cells achieved the highest number of infected tsetse flies whereas transgenic cells induced for expression of ESAG9-EQ protein achieved the lowest. There was obviously a great deal of variability in this assay so it was not possible to draw any firm conclusions from these data.

The infection levels after seven days were also assayed in a separate experiment, as shown in Figure 5.14, panel B. Here, there was a 2-fold enhancement of the number of flies infected when ESAG9-EQ protein was being expressed, compared to the wild-type or uninduced infections. This could indicate that ESAG9-EQ protein was having a beneficial effect on the parasites ability to survive and grow in the tsetse fly mid-gut over the longer term. However this assay was not repeated, and if it had been, the results may have been different due to the overall variability of the assay. Moreover the role of ESAG9 genes is unlikely to involve the longer-term survival of parasites in the fly because the ESAG9 mRNA is not normally expressed by procyclic forms.

It is hard to interpret the results from sections 5.5 and 5.6 with any great confidence. The method of detecting trypanosomes, by visualising down a microscope, is problematic. For example, the dissections and counting were performed concurrently by more than one person due to the time involved, and so different investigators may have had differing abilities to detect low level infections. Also, as fly guts were kept on ice for perhaps several hours whilst the counting was completed, it is possible that some death of the parasites would have occurred during this period, introducing another variable. So, although the results of these three experiments detailed in section 5.5 appear to show that the expression of ESAG9-K9 protein in the fly gut increased the ability of the parasites to colonise the gut, extreme caution should be used when attributing this as a function of ESAG9-K9 protein.

A



B

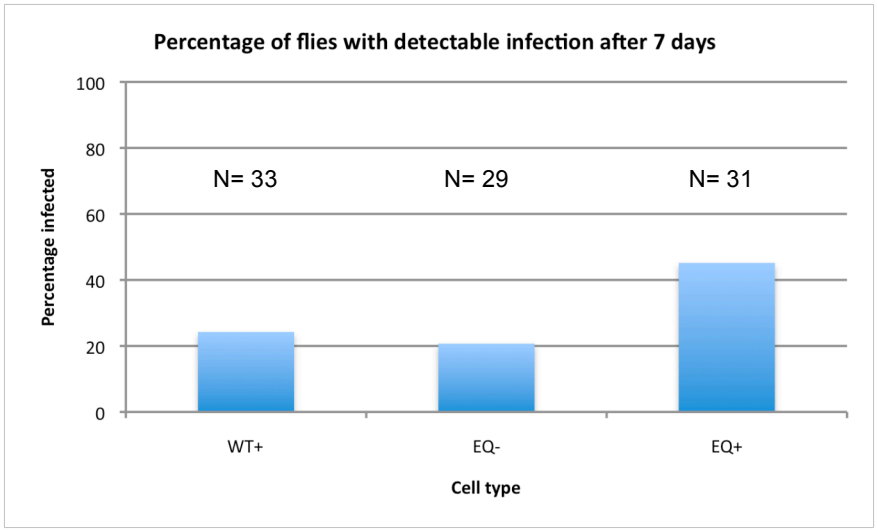


Figure 5.14: Percentage of flies infected after 48 hours and seven days post-feed. In two separate experiments, flies were fed with either wild-type procyclic form cells with tetracycline (WT); transgenic procyclic form cells induced for ectopic expression of ESAG9-EQ protein (EQ+); or uninduced cells (EQ-). In the first experiment, the mid-guts were excised after 48 hours (panel A) and the number of infected flies were scored, and in the second the mid-guts were excised after seven days (panel B). The number of flies used in each group for each experiment is indicated.

5.7 Quantitative Real-Time PCR as a method to quantify low-level trypanosome infections in tsetse fly midguts

In a bid to find a way to reliably quantify the infection level of trypanosomes in tsetse fly mid-guts, we decided to test quantitative PCR.

Quantitative PCR (qPCR) involves a real-time PCR reaction where a gene is amplified and a fluorescent probe incorporated into the amplicon (in the assays to be described TaqMan® probes were used). The fluorescence is measured during the reaction and, by comparing to known standards, the number of genomes present in the sample can be extrapolated. This method is increasingly being used in parasitology (Zarlenga & Higgins, 2001), and has been used to accurately quantify the infection level of malaria parasites in mosquitoes (Bell & Ranford-Cartwright, 2004).

To date, there is no published protocol for use of qPCR to quantify midgut infections in tsetse flies. Quantitative PCR from isolated RNA has been used to analyse relative procyclin mRNA expression in tsetse flies (Urwyler *et al.,* 2005), but in this case the salivary glands only were isolated so the technique is not comparable. Also, PCR has been used to confirm the presence of trypanosomes in tsetse fly field samples (Masiga *et al.,* 1992; Morlais *et al.,* 1998), but not in a quantitative manner.

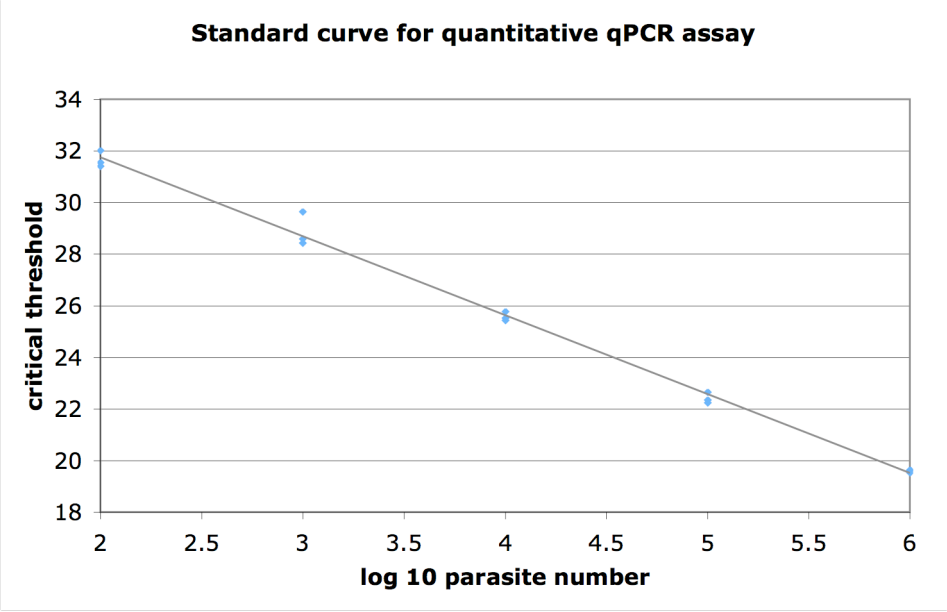
As a first step to develop a protocol for qPCR, different techniques for extracting the DNA were tested. The method used by Bell and Ranford-Cartwright (2004) involved extracting the mosquito mid-gut into a buffer and then doing a phenol chloroform extraction of the DNA. For this study, two buffers were compared; TNES-Urea (Asahida *et al.,* 1996) and CTAB (Stewart & Via, 1993). These have been used for DNA extraction from fish and plant samples, respectively. Flies were infected with 427-449 procyclic form trypanosomes and the mid-guts dissected after 8 days; the first ten mid-guts were also examined microscopically for the presence of a trypanosome infection but no trypanosomes were seen. A comparison of the number of parasite genomes detectable in eleven tsetse fly mid-guts in each buffer by qPCR revealed that the CTAB buffer was more effective for this purpose (the absolute quantification is shown in Table 5.1). Five out of 11 flies were positive for trypanosome infection when the CTAB buffer was used, whereas zero were positive

in the TNES-Urea buffer. Figure 5.15 shows an example of a standard curve for a qPCR assay, which demonstrates that the standards used were reliable.

Buffer used for DNA extractions	Parasite counts from individual fly guts
Midgut stored in TNES-Urea buffer prior to DNA extraction	0
	0
	0
	0
	0
	0
	0
	0
	0
	0
	0
Midgut stored in CTAB buffer prior to DNA extraction	0
	0
	0
	0
	0
	0
	276
	316
	1120
	1721
	3051

LEFT. Table 5.1: Test qPCR assays with mid-gut DNA from flies infected with procyclic 427-449 cells. The fly gut DNA was placed in either TNES-Urea or CTAB buffer and then phenol-chloroform extracted prior to quantifying the infection by real-time quantitative PCR.

BELOW. Figure 5.15: Example of a standard curve for the qPCR assay. The standard used was DNA extracted from a known number of cultured trypanosomes. Serial dilutions of the standards were carried out (x-axis) and the critical threshold measured (y-axis). The critical threshold is the PCR cycle at which the fluorescence from a given sample increases above the background fluorescence. The points should lie as close to a straight line as possible.



To further test the reliability of this assay, a number of midguts were excised from uninfected flies, and then a known number of trypanosome cells added to the midgut mixture prior to carrying out the phenol-chloroform DNA extraction. Also, a number of different treatments of the fly midguts were tested to inform how best to carry out the assay. The results of these tests are shown in Table 5.2. The absolute quantification of the number of trypanosome genomes by qPCR is compared with the number of cells added to the midgut. It is clear from this data that the CTAB-phenol-chloroform extraction is not a satisfactory technique; the quantification by qPCR is not very reproducible and does not tally accurately with the number of cells added to the fly gut.

To improve this assay a kit specifically designed for extracting DNA from tissues was evaluated. The Puregene Genomic DNA Purification Kit (Tissue Kit D-7000A from Gentra Systems) is a system based on using salt rather than organic solvents for DNA extraction (Buffone & Darlington, 1985) and can be used on a variety of tissues. As before, a number of midguts were excised from uninfected flies, a known number of trypanosome cells added, the DNA extracted according to manufacturer's instructions, and the infections quantified using qPCR. The results of two assays, without and with an overnight Proteinase K digestion step, are shown in Tables 5.3. The Proteinase K digestion resulted in an increase in the number of genomes detected by qPCR, presumably because there were less contaminants.

Fly	Midgut treatment	Genomes by qPCR	No of parasites added to midgut
1	Midgut resuspended in PBS	1461	1×10^3
		233	
2		4456	1×10^3
3		84233	1×10^6
4		248614	1×10^6
5	Midgut resuspended in CTAB and DNA extraction carried out straight away	501	1×10^3
		572	
6		2052	1×10^3
7		39967	1×10^6
8		9418	1×10^6
9	Midgut resuspended in CTAB and frozen overnight prior to DNA extraction	128	1×10^3
		184	
10		669	1×10^3
11		32604	1×10^6
12		4303	1×10^6
13	Midgut resuspended in CTAB and stored at room temperature overnight prior to DNA extraction	1193	1×10^3
		637	
14		4155	1×10^3
15		22755	1×10^6
16		130353	1×10^6

Table 5.2: Results of quantitative PCR tests on DNA samples extracted using a range of treatments. Midguts were dissected from the flies and either 1×10^3 or 1×10^6 cultured procyclic form trypanosomes added, prior to the midguts receiving a range of treatments (see column 2). The counts derived by qPCR are shown in column 3 and compared to the number of parasites that should have been present, shown in column 4.

A

Number of genomes per fly by qPCR	Number of parasites added to fly gut
145	1×10^4
187	1×10^4
743	1×10^4
1095	1×10^4
1496	1×10^6
10723	1×10^6
19855	1×10^6
289564	1×10^6

B

Number of genomes per fly by qPCR	Number of parasites added to fly gut
Undetected	1×10^4
Undetected	
3345	1×10^4
2364	
72762	1×10^6
69643	
245871	1×10^6
221236	
375269	1×10^6
308951	

Tables 5.3: Results of quantitative PCR tests with DNA samples extracted using a Gentra DNA Purification Kit. Tstse fly midguts were excised and a known number of procyclic form cells were added. The DNA was extracted with the Gentra kit and either not treated with Proteinase K (table A); or treated overnight with Proteinase K (table B). The number of parasites counted by qPCR (left hand columns) is compared with the number of cells added to the gut prior to carrying out the DNA extractions (right hand columns).

Although the results achieved using this assay were not as good as those derived when qPCR has been used to quantify infections in mosquitoes, it was still an improvement on microscope counting. Moreover, it allowed the assay of a much larger number of flies at once, the flies could be processed quickly, and it removed the human error of microscope counting. We therefore concluded that it was valid to attempt this assay on a larger scale experiment.

5.8 Analysis of the effect of ESAG9-K9 protein expression on tsetse fly infection level using qPCR

The ability of transgenic bloodstream form cells, ectopically expressing ESAG9-K9 protein, to infect tsetse flies was compared with that of uninduced transgenic cells and wild-type cells, as in section 5.5. In this experiment the infections were quanitified using qPCR rather than microscopy. Another variable was also introduced into this experiment. One of our hypotheses for the function of ESAG9 protein was that it may protect differentiating cells from the complement in the tsetse fly blood meal. To test this hypothesis we compared the outcome of the infection when the serum in the horses' blood used to feed the flies (which contains the complement factors) was either present or absent. Thus in the presence of serum ESAG9 protein might promote survival whereas in its absence no clear effect of ESAG9 expression would be seen.

Three different parasite cell types were used: transgenic bloodstream form parasites induced to express ESAG9-K9 protein; the same cells but uninduced; and wild type cells with tetracycline. Two different types of blood were used: whole blood; or blood which had been centrifuged, the serum removed, and the volume made up with SDM-79 medium. A total of around 240 flies were used in the six treatment groups and, of these, 178 flies took a blood meal and so could be used for quantification. The flies were dissected after 72 hours, their midguts excised, and the DNA extracted according to the Puregene Genomic DNA Purification Kit instructions.

The results of this experiment are shown in Figure 5.16. The graphical representation used is a box plot, as in Figure 5.13. The data did not support the hypothesis that the role of ESAG9-K9 is to protect cells from complement present in serum. The level of infection was similar in the flies fed with wild-type parasites and those fed with parasites induced to express ESAG9-K9. In these groups there was no difference in the level of parasite numbers whether whole blood, or blood without serum, was used. This suggests that the presence of serum was not causing parasite death.

Further, the highest parasite numbers achieved were in those flies fed with whole blood containing transgenic cells which were not induced for ESAG9-K9 expression.

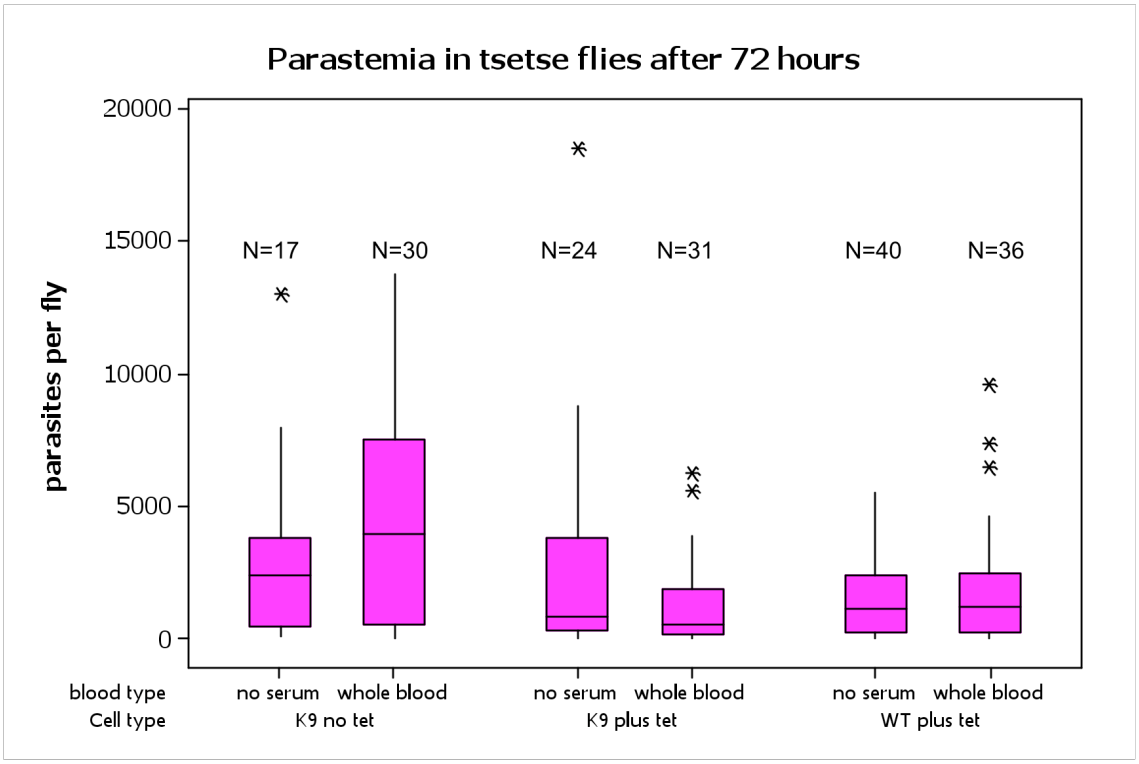


Figure 5.16: Results of tsetse fly infections with bloodstream form transgenic and wild-type cells as quantified by qPCR. Flies were fed with either wild-type bloodstream form cells with tetracycline (WT plus tet); bsf K9 cells induced for expression of ESAG9-K9 protein (K9 plus tet); or uninduced cells (K9 no tet). The horse blood used to feed the flies was either used complete (whole blood); or it was centrifuged, the serum removed, and the volume made up to the original with additional SDM-79 media (no serum). The boxes represent the interquartile ranges, containing 25% of values either side of the median value. The bars represent the spread of the data and the asterixes represent outlying values. The number of flies used for each treatment group are indicated.

In order to test the reproducibility of the assay, internal repeats were carried out on eighteen of the samples whereby they were run three times (in different 96-well plates) instead of once in the qPCR machine. These counts, shown in Table 5.4, indicate how unreliable this assay was; there was a large variation in the derived counts between the three plates. This may indicate that the DNA was not cleanly isolated from the mid-gut material, potentially leaving proteins in the sample, which could inhibit the PCR reaction.

The apparent unreliability of this assay means that, although there is no apparent difference between the groups, we cannot assert that ESAG9-K9 protein does not play role in the survival of the parasites in the tsetse fly mid-gut. It is likely that anything but a very large difference between the infection levels would be missed using this assay.

Tsetse fly sample number	qPCR first attempt	qPCR second attempt	qPCR third attempt
1	76	810.2	210.5
10	185.6	462.7	172
20	169	2783.5	201.3
30	630.6	1114.2	283.6
40	61.5	101.9	31.3
50	156.9	321.1	50.1
60	370.1	594.7	143.8
70	1982.8	2892.5	729.2
80	1478.1	2447.5	473.3
90	7523.1	12163	3285.7
100	229.9	6.3	6.55
110	7532.1	10926.1	4507.9
120	919.5	1383.7	417.8
130	3200.9	3910.2	1384.1
140	1463.5	2200.6	623
150	1675.4	3010.7	900.7
160	5249.2	5174.2	1266.3
170	897	1243.3	228.2

Table 5.4: Repeats of qPCR assay using samples from procyclic form infections. Eighteen samples from the tsetse fly infections shown in Figure 5.16 were run three times rather than once in three separate qPCR plates. The counts from the three attempts are shown in colums 2-4.

5.9 Summary

In this chapter an attempt was made to address the hypothesis that ESAG9 proteins have a function whilst parasites are differentiating in the tsetse midgut. This hypothesis was posed due to the unusual mRNA expression profiles of ESAG9-K9, ESAG9-EQ and ESAG9-K69 whereby expression was up-regulated in bloodstream form stumpy parasites (see Chapter 3 Figure 3.2). ESAG9-K9 and ESAG9-K69 proteins were expressed during *in vitro* differentiation from stumpy to procyclic form (refer to Chapter 3 Figure 3.5 and Chapter 4 Figure 4.21).

The growth assays with procyclic form cells detailed in sections 5.2, 5.3 and 5.4 indicate that the expression of the three ESAG9 proteins investigated do not give the parasites a survival advantage in mammalian serum. It can therefore be concluded that, in the *in vitro* conditions used, ESAG9 proteins are not interacting with the complement pathway to prevent parasite killing by the alternative pathway of complement. A potential flaw of the *in vitro* growth assays was that in reality the ESAG9 proteins are expressed by differentiating stumpy cells, and not by fully differentiated procyclic forms. The procyclic cells used are culture adapted and may have lost some of the phenotypic characteristics of their forebears. A more realistic model would be to use parasites that are undergoing differentiation, although the asynchronous differentiation of transgenic monomorphic parasites would make these experiments very difficult to interpret.

The tsetse fly assays proved problematic in terms of accurate quantification of the infections, and in experimental design. In deciding whether to use procyclic form or bloodstream form parasites for these assays, there were advantages and disadvantages of both cell types. Bloodstream form parasites should in theory provide a more 'realistic' model in that tsetse flies would normally ingest bloodstream forms. The caveat, however, is that the culture-adapted monomorphic parasites never produce stumpy forms, and so may not be adapted for the mid-gut environment. Also their ability to infect the flies was very low. Using procyclic form parasites avoids these problems, but if our hypothesis is that ESAG9 protein is important in differentiating parasites, it may have not have a role to play in parasites

which are fully adapted for this environment. Using pleiomorphic parasites would not provide a solution to these problems. Pleiomorphic stumpy bloodstream form cells have been shown to express multiple ESAG9 genes simultaneously (see Chapter 3 Figure 3.2). Therefore creating transgenic cell lines which over-express one version of ESAG9 is unlikely to enhance infections over wild type levels. Moreover, RNAi could not be used to ablate multiple copies of ESAG9 in pleiomorphic cells because the genes are too divergent to be targeted with one RNAi vector.

In general, the infections resulted in very low levels of parasites in the flies. The level of infection was difficult to quantify either microscopically, or by qPCR. A further issue was that anecdotal evidence from other investigators in the field (Alvaro Acosta-Serrano, personal comment) has indicated that the tsetse fly populations routinely used in these types of assays in labs across Europe have become infected with another protozoa, *Gregarine spp*., which has resulted in low and unpredictable levels of trypanosomes in experimental infections.

We have not been able to either confirm or rule out a function of ESAG9 proteins in the tsetse fly. However given that two of the proteins, ESAG9-K9 and ESAG9-EQ, are not secreted, we have to question how they might function. ESAG9-K69 protein is secreted by transgenic monomorphic parasites and by 'wild-type' stumpy form parasites (refer to Chapter 4 sections 4.3.3 and 4.4.2). Therefore, further functional assays focused on that stage of the life cycle were carried out, utilising the mouse model of trypanosome infection, and these will be detailed in Chapter 6.

Chapter 6 *In vivo* and *ex vivo* analysis using the murine model of trypanosome infection

6.1 Introduction

Bloodstream form K9:K69 parasites were shown to secrete ESAG9-K69 protein out of the cell (see Chapter 4 Figure 4.15). Similarly, pleiomorphic *T. b. brucei* EATRO 2340 stumpy form parasites were also shown to secrete ESAG9-K69 protein (Chapter 4 Figure 4.20). We therefore postulated that ESAG9-K69 protein could be involved in host-parasite interactions in the host. It has been shown that other ESAG proteins have functions in host-parasite interactions. For example the expression of slightly different ESAGs 6 and 7 (which form the transferrin receptor) are proposed to enhance parasite survival in serum from different mammals (Bitter et al., 1998) though this is still somewhat controversial.

The ESAG9-K69 protein could potentially interact with the mammalian immune system in some way. Other organisms, for example nematode worms, have been shown to secrete molecules that mimic human cytokines (Maizels *et al.,* 2004). Furthermore, the protozoal parasites *Plasmodium spp.* and *Leishmania major* both have homologues to the cytokine Macrophage Inhibitory Factor (MIF)(Augustijn *et al.,* 2007; Ivens *et al.,* 2005). The presence of a MIF-like protein in the secretome of *Leishmania donovani* has been verified, though a biological function has not yet been verified for this molecule (Silverman *et al.,* 2008). BLAST searches using the ESAG9 protein sequences did not reveal any similarity to any known immunomodulatory molecule. This does not rule out the possibility of ESAG9-K69 protein having a previously undescribed function however.

6.2 In vivo analysis of growth rates of bsf K9:K69 cells in mice

To determine whether the secretion of ESAG9-K69 protein by bsf parasites had any effect on parasitaemia in the mammal host, mice were infected with bsf K9:K69 parasites. This double-expressing cell line was chosen as it had been shown

previously that ESAG9-K69 protein is secreted by these cells. MF1 mice were inoculated with parasites by intra-peritoneal injection. The induction of ectopic gene expression in the parasites was achieved by adding 200μg/ml doxycycline to the drinking water of the mice, with the control mice being inoculated with the same parasites, but without the provision of doxycycline. Doxycycline is taken up by the parasites whilst in the bloodstream and is sufficiently similar to tetracycline to function in the same way and so induce gene expression from constructs silenced by the Tet repressor. The flavour of the doxycycline was disguised by addition of 5% sucrose to the drinking water, whereas the '-dox' control mice received water with 5% sucrose only.

The results of five independent experiments will be presented. In the first two experiments, a starting inoculation of 1×10^4 parasites per mouse was used. This had been shown to be an appropriate initial dose from previous experiments carried out in the Matthews lab. Figure 6.1 shows the outcome of the first experiment. Sixteen mice were used, split into two groups of eight mice. One group was injected with 1×10^4 K9:K69 parasites, and the other with 1×10^4 WT parasites. Half the mice in each group were provided with water with 200μg/ml doxycycline and 5% sucrose, and half were provided with water containing only 5% sucrose. The parasitaemia in each mouse was then recorded over a period of five to seven days by counting the number of parasites visible in tail smears using the rapid matching method of Herbert and Lumsden (Herbert & Lumsden, 1976). This method involves making a smear on a slide using a drop of blood from a tail snip. The number of parasites visible in a field of view at a specified magnification is then compared to a key, which indicates the likely cell density this represents in cells per ml of blood. The mice were culled or found dead usually at either day five or day six post-infection. This is typical of infections with monomorphic parasites, which are very virulent and kill a mouse host more rapidly than pleiomorphic parasites.

Figure 6.1 panel A shows the parasitaemia in individual mice infected with bsf K9:K69 parasites (mice A-D were provided with doxycycline; mice E-H were not). The data are split into two graphs for clarity. All of the -dox mice survived until day 6, whereas three out of four of the +dox mice were culled due to sickness or died

naturally on day 5. Panel B shows the mean parasitaemias in the -dox and + dox mice, with the mean only being calculated as long as all the mice in each group were still alive. This graph indicates that the mean growth of +dox parasites was greater than the -dox parasites from day 3 to day 4, but that the mean growth then decreased from day 4 to day 5. Panel C shows the mean growth of the WT controls, and a similar effect was seen in that the +dox parasites grew slightly faster from day 3 to day 4.

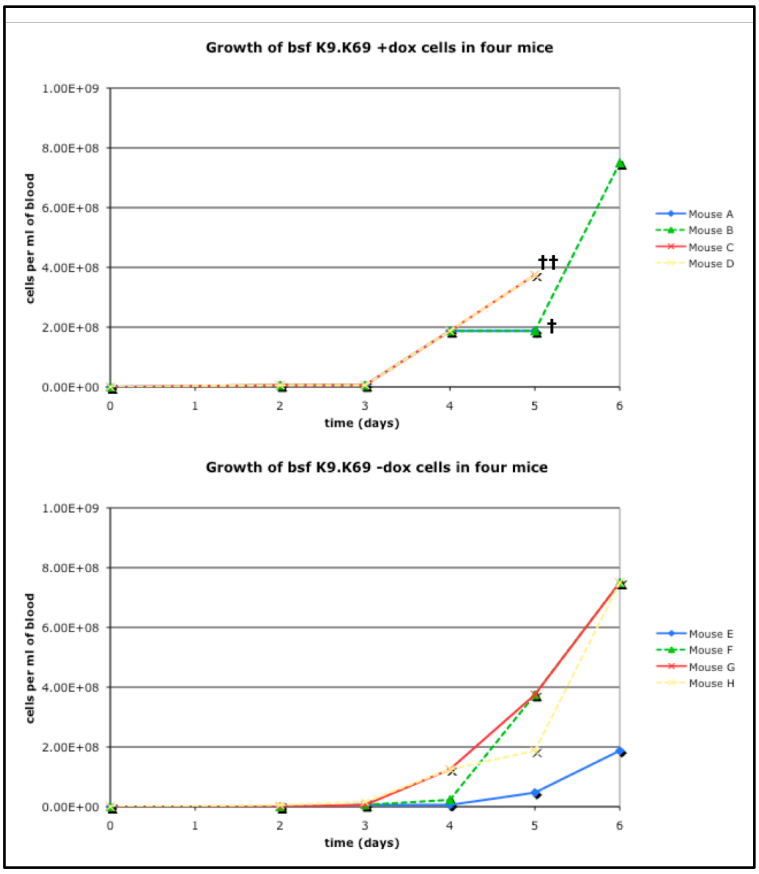
The growth rates from day 3 to day 4 and from day 4 to day 5 of the bsf K9:K69 parasites in the +dox and -dox mice were compared statistically using an Analysis of Variance (ANOVA). An ANOVA tests whether the variance within samples is the same as the variance between samples, and can therefore be used to determine whether the means of two samples are significantly different. An ANOVA assumes that data are normally distributed, and in all cases this was verified prior to carrying out the ANOVA. A cut-off p-value of 0.05 was used to determine whether the outcome was significant. For more details on the p-value and the statistical package used, refer to Chapter 2 section 2.12.

There was a significant difference between growth rate of +dox and -dox parasites from day 3 to day 4 ($F_{1,6} = 8.19$, $p = 0.029$). The difference in growth rate from day 4 to day 5 was not significant however ($F_{1,6} = 4.89$, $p = 0.069$).

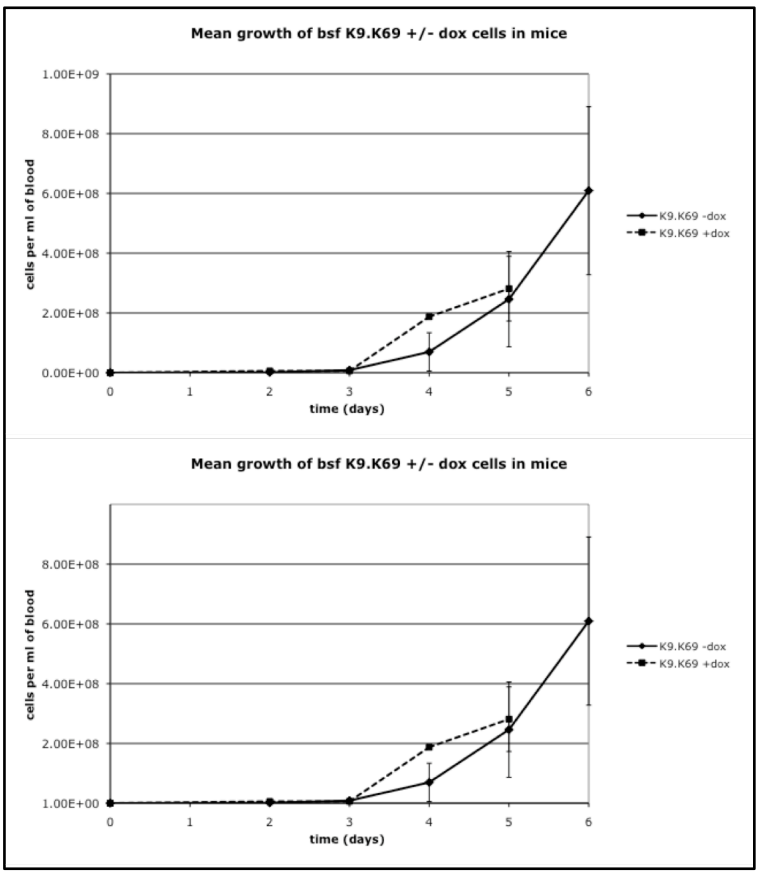
The longevity of the mice was not analysed using statistical methods as it was not possible to quantify this accurately. For example, some of the mice were found dead in their cages, and without knowing at exactly what time the mice had died, an accurate lifespan could not be assigned to the mice. Moreover, others were culled when exhibiting distress, and so the natural time of death for these mice would not be known either. However the trend indicated that +dox mice, in which the bsf K9:K69 parasites were induced to express ESAG9-K9 and ESAG9-K69, tended to live less long than the -dox mice.

Figure 6.1 (following page): Growth of bsf K9:K69 and WT parasites in mice. Eight mice were inoculated with 1×10^4 bsf K9:K69 parasites, and eight mice with 1×10^4 WT parasites. Half the mice in each group were provided with drinking water with 200µg/ml doxycycline and 5% sucrose (+dox mice), and half provided with water containing 5% sucrose only (-dox). Parasites per ml of blood were estimated daily from day 2 by counting the number of parasites visible in a tail smear using the rapid matching method of Herbert and Lumsden (1976). Panel A shows the parasitaemias in individual mice infected with K9:K69 parasites. Panel B shows the mean parasitaemias from mice infected with K9:K69 parasites (top) or WT parasites (bottom). Error bars are standard deviations from the mean.

A


Growth of bsf K9.K69 +dox cells in four mice
cells per ml of blood
1.00E+09
8.00E+08
6.00E+08
4.00E+08
2.00E+08
0.00E+00
0
1
2
3
4
5
6
time (days)
††
†
Mouse A
Mouse B
Mouse C
Mouse D
Growth of bsf K9.K69 -dox cells in four mice
cells per ml of blood
time (days)
Mouse E
Mouse F
Mouse G
Mouse H

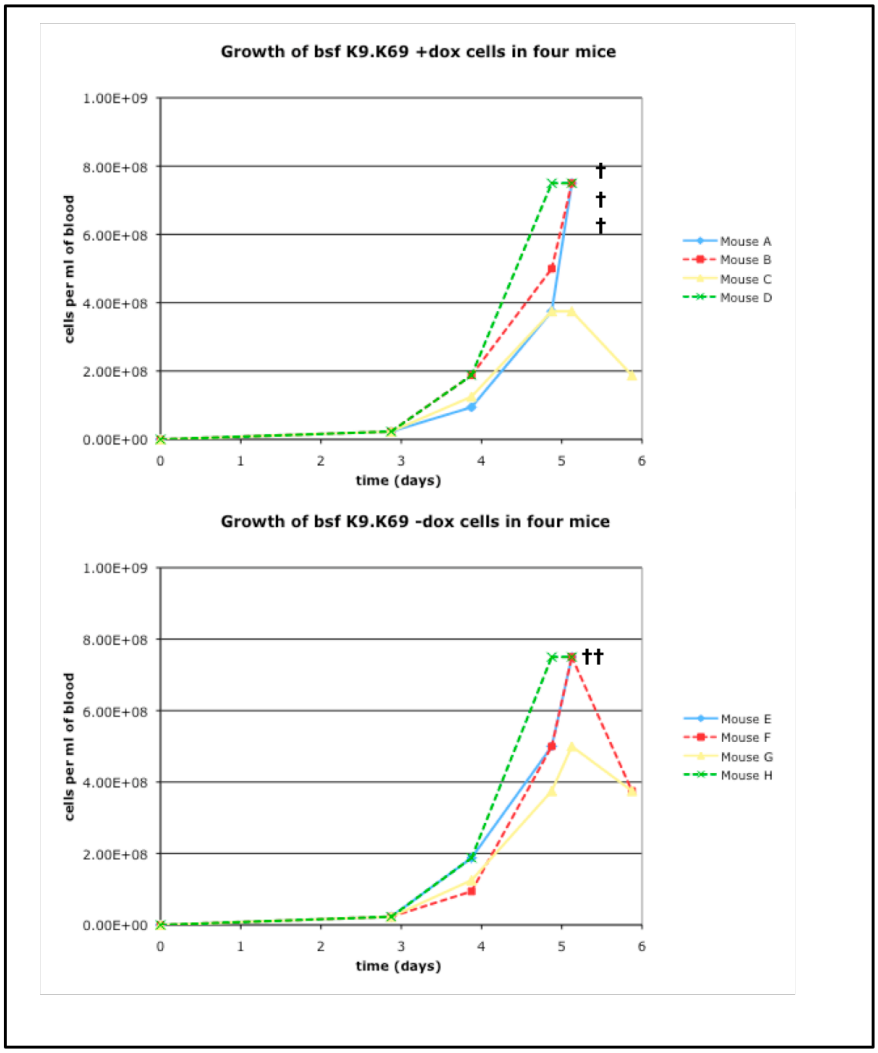
B


Mean growth of bsf K9.K69 +/- dox cells in mice
cells per ml of blood
1.00E+09
8.00E+08
6.00E+08
4.00E+08
2.00E+08
0.00E+00
0
1
2
3
4
5
6
time (days)
K9.K69 -dox
K9.K69 +dox
Mean growth of bsf K9.K69 +/- dox cells in mice
cells per ml of blood
1.00E+00
time (days)
K9.K69 -dox
K9.K69 +dox

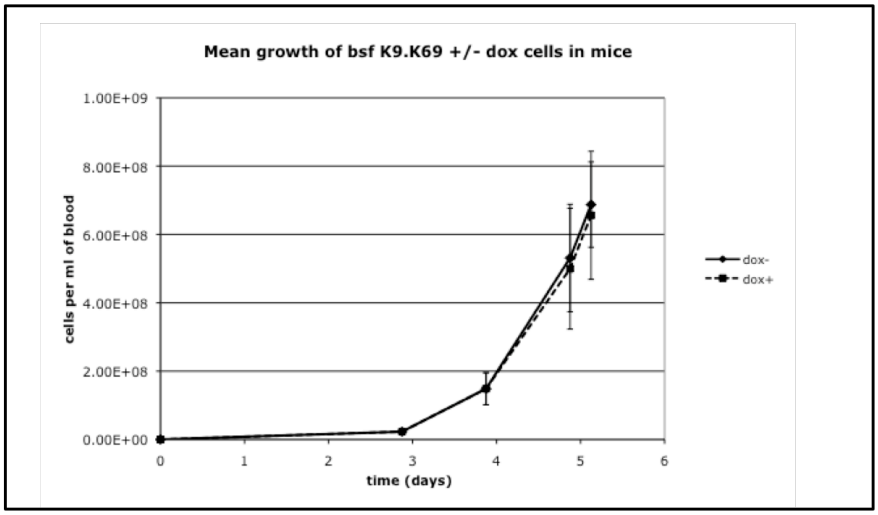
To determine if this trend was reproducible, the experiment was repeated in exactly the same way, excepting that the WT controls were excluded. The results of this experiment are shown in Figure 6.2. Panel A shows the parasitaemia in four +dox mice and four -dox mice, again represented on separate graphs for clarity. In this experiment, three +dox mice and two -dox mice died or were culled on day 5. The mean parasitaemias are shown in panel B. This shows that there was clearly no difference in the mean parasitaemia of the +dox and -dox mice up until day 5. It can be concluded from these data that in this experiment there was no difference between parasitaemia in +dox and -dox mice.

Figure 6.2 (following page): Growth of bsf K9:K69 parasites in mice. Eight mice were inoculated with 1×10^4 bsf K9:K69 parasites. Half the mice were provided with drinking water with 200µg/ml doxycycline and 5% sucrose (+dox mice), and half provided with water containing 5% sucrose only (-dox). Parasites per ml of blood were estimated daily from day 2 by counting the number of parasites visible in a tail smear by the rapid matching method of Herbert and Lumsden (1976). Panel A shows the parasitaemias in individual mice infected with K9:K69 parasites. Panel B shows the mean parasitaemias from mice infected with K9:K69 parasites. Error bars are standard deviations from the mean.

A


Growth of bsf K9.K69 +dox cells in four mice
cells per ml of blood
1.00E+09
8.00E+08
6.00E+08
4.00E+08
2.00E+08
0.00E+00
0
1
2
3
4
5
6
time (days)
†
†
†
Mouse A
Mouse B
Mouse C
Mouse D
Growth of bsf K9.K69 -dox cells in four mice
cells per ml of blood
1.00E+09
8.00E+08
6.00E+08
4.00E+08
2.00E+08
0.00E+00
0
1
2
3
4
5
6
time (days)
††
Mouse E
Mouse F
Mouse G
Mouse H

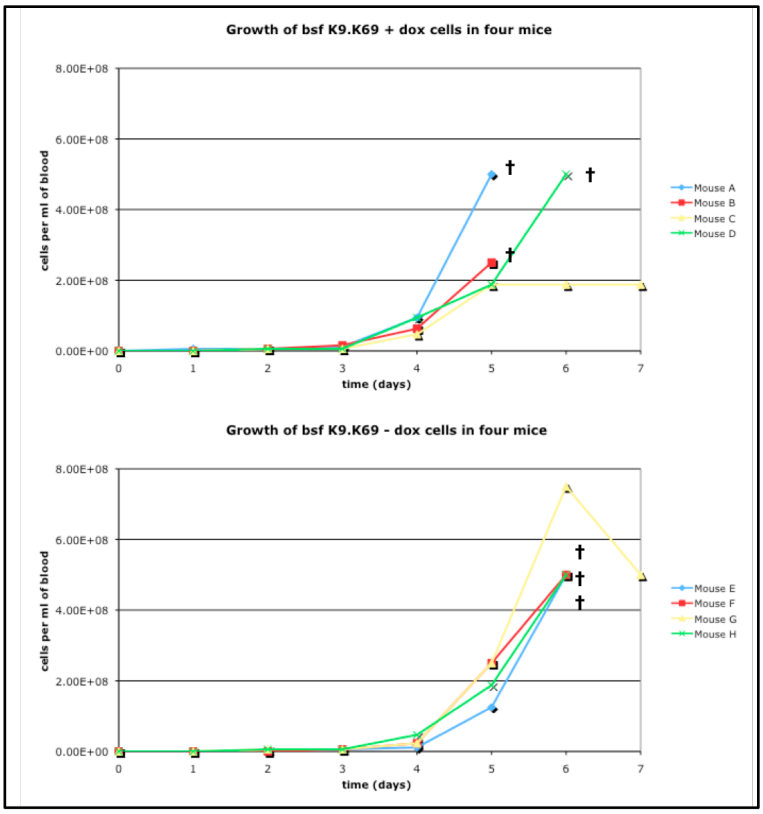
B


Mean growth of bsf K9.K69 +/- dox cells in mice
cells per ml of blood
1.00E+09
8.00E+08
6.00E+08
4.00E+08
2.00E+08
0.00E+00
0
1
2
3
4
5
6
time (days)
dox-
dox+

With a starting inoculation of 1×10^4 parasites, the mice were found dead or were being culled due to sickness on day 5 or day 6. A lower inoculation was therefore also used to determine if it was possible to achieve a longer-term infection that might be more informative. Previous experiments with other monomorphic cell lines had shown that an inoculation of 1×10^3 parasites resulted in the mice surviving the first wave of parasitaemia, and not being killed by the infection until the second wave (K.M., unpublished data). Therefore three independent experiments were set up in exactly the same way as described for Figures 6.1 and 6.2, except that a lower starting dose of 1×10^3 parasites was used.

The first of the lower dose experiments is shown in Figure 6.3. Eight mice were inoculated with 1×10^3 bsf K9:K69 parasites and the parasitaemia in each mouse is shown in panel A. In the +dox group, two mice died or were culled on day 5, one on day 6, and one survived until day 7. In the -dox group, three mice survived until day 6, and one until day 7. The mice did not survive as long as they were expected to, and in no case did the mice control the infection, allowing development of a second wave of parasitaemia. This particular cell line could for some reason be more virulent than ones previously tested. Panel B shows the mean parasitaemia whilst all the mice were alive, and demonstrates that the +dox parasites were, on average, increasing in number more rapidly from days 3 – 4 and days 4 – 5. The growth rates in +dox and –dox mice were analysed with ANOVAs. The growth rates from days 3 – 4 were not found to be significantly different ($F_{1,6} = 3.23$, $p = 0.123$), but there was a significant difference in growth rate from days 4 – 5 between the +dox and -dox mice ($F_{1,6} = 6.89$, $p = 0.039$). The fact that the parasites are increasing in number more rapidly could be due to a faster rate of cell division, or it could be due to host factors.

A



B

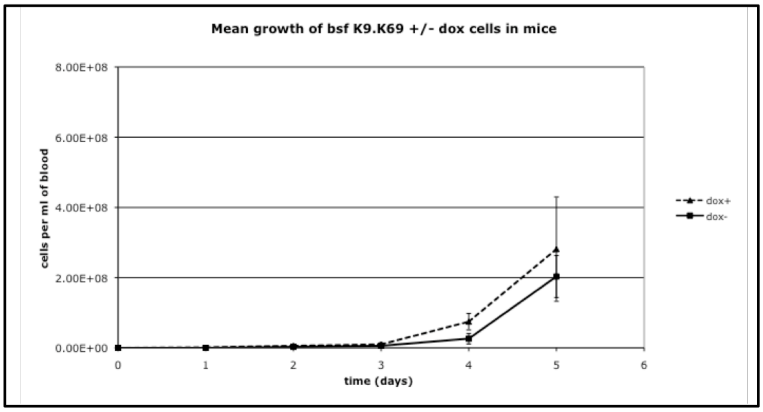
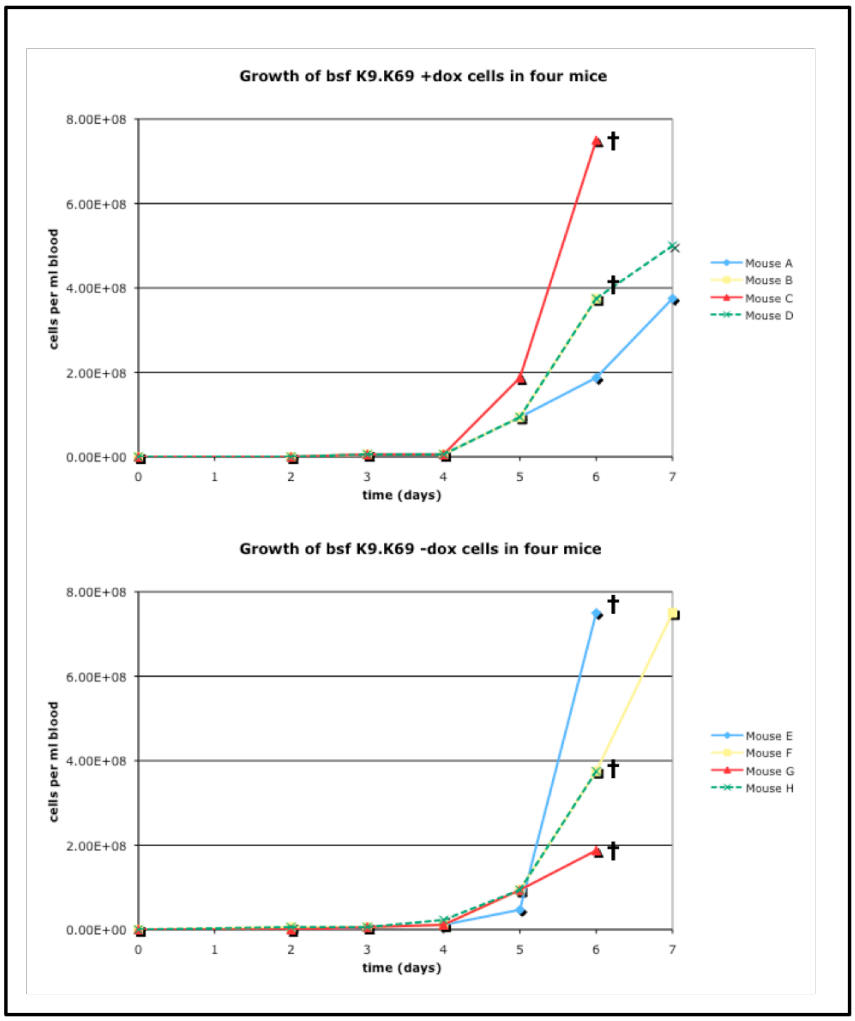


Figure 6.3: Growth of bsf K9:K69 parasites in mice. Eight mice were inoculated with 1×10^3 bsf K9:K69 parasites. Half the mice were provided with drinking water with 200µg/ml doxycycline and 5% sucrose (+dox mice), and half provided with water containing 5% sucrose only (-dox). Parasites per ml of blood were estimated daily from day 2 by counting the number of parasites visible in a tail smear using the rapid matching method of Herbert and Lumsden (1976). Panel A shows the parasitaemias in individual mice infected with K9:K69 parasites. Panel B shows the mean parasitaemias from mice infected with K9:K69 parasites. Error bars are standard deviations from the mean.

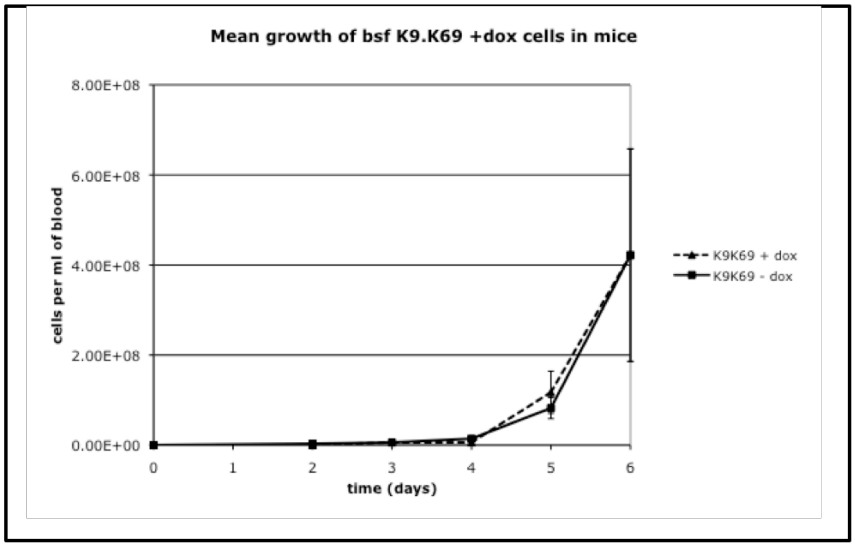
To verify (or otherwise) the results from this experiment, the experiment was repeated in exactly the same way and these results are shown in Figure 6.4. However, the results of this experiment differed from those represented in Figure 6.3. In the +dox group, two mice survived until day 6 and two until day 7. In the -dox group, three mice survived until day 6 and one until day 7. Panel B shows the mean parasitaemias. The difference in growth rate from days 3-4 and 4-5 were again analysed using ANOVAs. In this case, the difference in growth rate between +dox and –dox groups was significant on both days 3-4 ($F_{1,6} = 22.6$, $p = 0.003$) and days 4-5 ($F_{1,6} = 19.69$, $p = 0.004$).

Figure 6.4 (following page): Growth of bsf K9:K69 parasites in mice. Eight mice were inoculated with 1×10^3 bsf K9:K69 parasites. Half the mice were fed drinking water with 200µg/ml doxycycline and 5% sucrose (+dox mice), and half fed with 5% sucrose only (-dox). Parasites per ml of blood were estimated daily from day 2 by counting the number of parasites visible in a tail smear. Panel A shows the parasitaemias in individual mice infected with K9:K69 parasites. Panel B shows the mean parasitaemias from mice infected with K9:K69 parasites. Error bars are standard deviations from the mean.

A


Growth of bsf K9.K69 +dox cells in four mice
8.00E+08
6.00E+08
4.00E+08
2.00E+08
0.00E+00
cells per ml blood
0
1
2
3
4
5
6
7
time (days)
Mouse A
Mouse B
Mouse C
Mouse D
†
†
Growth of bsf K9.K69 -dox cells in four mice
8.00E+08
6.00E+08
4.00E+08
2.00E+08
0.00E+00
cells per ml blood
0
1
2
3
4
5
6
7
time (days)
Mouse E
Mouse F
Mouse G
Mouse H
†
†
†

B


Mean growth of bsf K9.K69 +dox cells in mice
8.00E+08
6.00E+08
4.00E+08
2.00E+08
0.00E+00
cells per ml of blood
0
1
2
3
4
5
6
time (days)
K9K69 + dox
K9K69 - dox

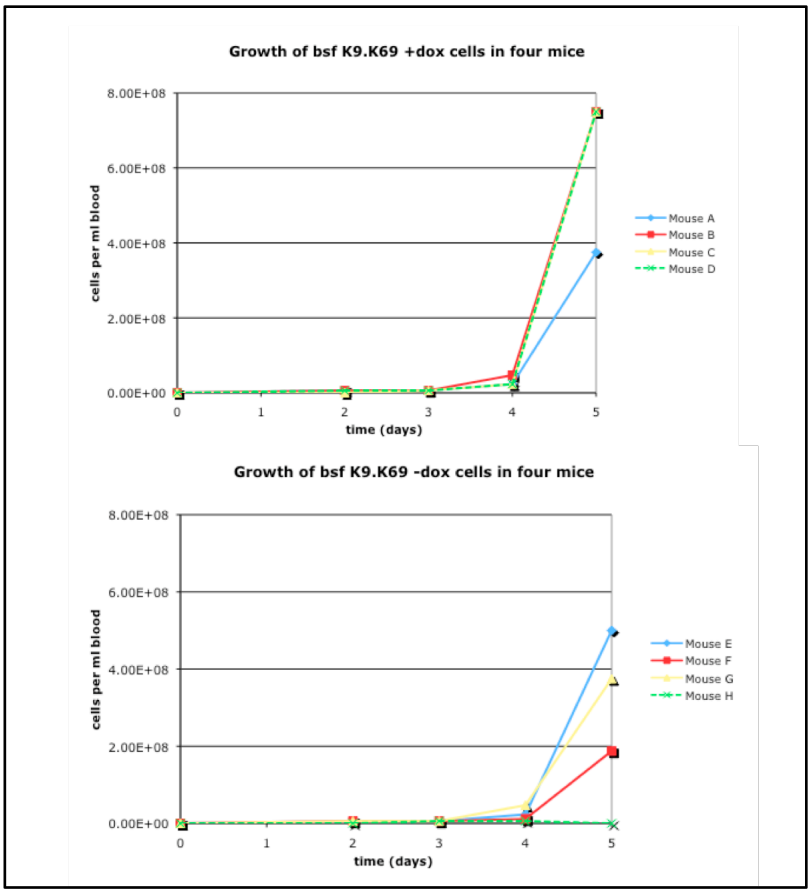
Because of the inconsistencies between the experiments, the experiment was again repeated but this time WT control inoculations were included. This was to ensure that the addition of doxycycline did not have an affect on parasitaemia, independent of cell line, when a low initial starting dose was used. Figure 6.5 panel A shows the parasitaemias in four mice inoculated with 1×10^3 bsf K9:K69 parasites. All mice were culled on day 5, as they were exhibiting discomfort. Panel B shows the mean parasitaemias in the bsf K9:K69 infected mice, and there was a clear difference in mean parasite density on day 5 between the +dox and -dox groups. However when the growth rates between days 3 - 4 and days 4 - 5 were analysed using an ANOVA no significant difference between +dox and -dox was observed in either case [($F_{1,6}$ = 1.14, p = 0.327) and ($F_{1,6}$ = 2.5, p = 0.165) respectively].

Panel B also shows the mean parasitaemias in the WT infections, and this shows that the addition of doxycycline had no effect on the number of parasites.

A summary of the five mouse infection experiments with bsf K9:K69 and WT parasites is shown in Table 6.1. Overall there was no reliably reproducible effect of ESAG9-K9 and ESAG9-K69 expression by bsf K9:K69 parasites on parasiteamia in mice; however in some experiments the expression of these proteins did result in an increase in the virulence of the infection.

Figure 6.5 (following page): Growth of bsf K9:K69 and WT parasites in mice. Eight mice were inoculated with 1×10^3 bsf K9:K69 parasites, and eight mice with 1×10^3 WT parasites. Half the mice in each group were fed drinking water with 200µg/ml doxycycline and 5% sucrose (+dox mice), and half fed with 5% sucrose only (-dox). Parasites per ml of blood were estimated daily from day 2 by counting the number of parasites visible in a tail smear. Panel A shows the parasitaemias in individual mice infected with K9:K69 parasites. Panel B shows the mean parasitaemias from mice infected with K9:K69 parasites (top) or WT parasites (bottom). Error bars are standard deviations from the mean.

A


Growth of bsf K9.K69 +dox cells in four mice
cells per ml blood
8.00E+08
6.00E+08
4.00E+08
2.00E+08
0.00E+00
0
1
2
3
4
5
time (days)
Mouse A
Mouse B
Mouse C
Mouse D
Growth of bsf K9.K69 -dox cells in four mice
cells per ml blood
8.00E+08
6.00E+08
4.00E+08
2.00E+08
0.00E+00
0
1
2
3
4
5
time (days)
Mouse E
Mouse F
Mouse G
Mouse H

B

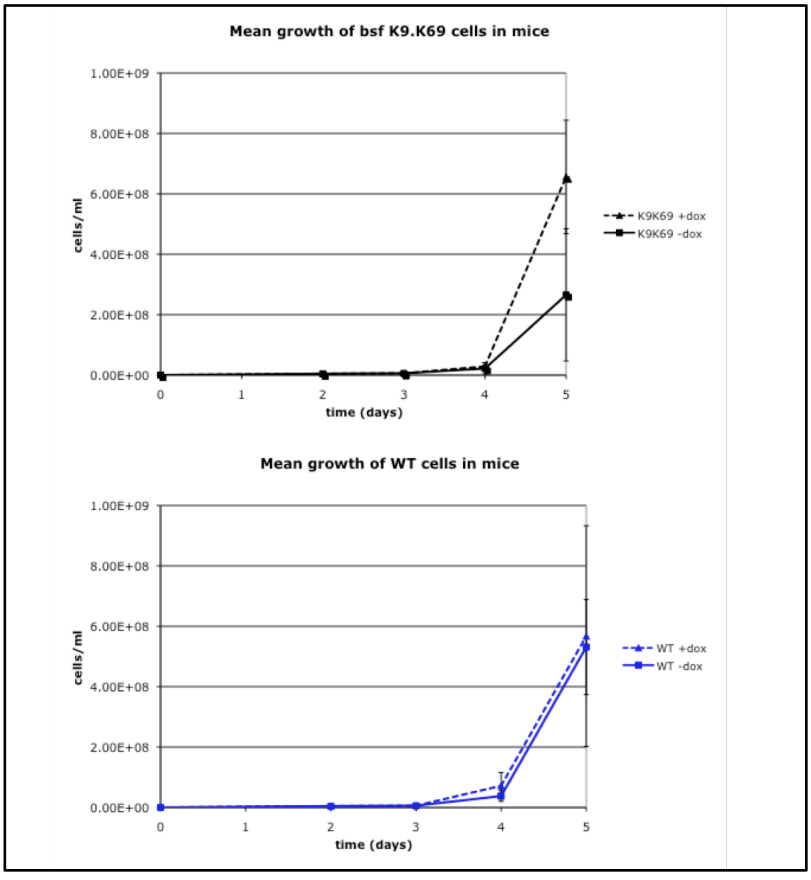

Mean growth of bsf K9.K69 cells in mice
cells/ml
1.00E+09
8.00E+08
6.00E+08
4.00E+08
2.00E+08
0.00E+00
0
1
2
3
4
5
time (days)
K9K69 +dox
K9K69 -dox
Mean growth of WT cells in mice
cells/ml
1.00E+09
8.00E+08
6.00E+08
4.00E+08
2.00E+08
0.00E+00
0
1
2
3
4
5
time (days)
WT +dox
WT -dox

Figure	Number of mice	Cell types used	Inoculation	Outcome - longevity of mice	Outcome - growth rate of parasites in mice
6.1	16	Bsf K9.K69 and WT	1×10^4 parasites per mouse	The +dox mice survived less long than the -dox mice.	The growth rate was higher in +dox mice than -dox mice and there was a statistically significant difference in growth rate from day 3-4 using an ANOVA
6.2	8	Bsf K9.K69 only	1×10^4 parasites per mouse	No convincing difference between +dox and -dox mice.	No difference in growth rate between the +dox and -dox groups.
6.3	8	Bsf K9.K69 only	1×10^3 parasites per mouse	The +dox mice survived slightly less long than the -dox mice.	The growth rate was higher in +dox mice than -dox and there was a statistically significant difference in growth rate from day 4-5 using an ANOVA
6.4	8	Bsf K9.K69 only	1×10^3 parasites per mouse	No convincing difference between +dox and -dox mice.	There were statistically significant differences in growth rates on both days 3-4 and days 4-5 using ANOVAs.
6.5	16	Bsf K9.K69 and WT	1×10^3 parasites per mouse	All the mice were culled on day 5 due to obvious discomfort	The parasitaemia in +dox mice was higher on day 5 but there were no statistically significant differences in growth rates.

Table 6.1: Summary of mouse infection experiments with bsf K9:K69 parasites. The figures in which the full data are shown are indicated in the far left column. The number of mice used, the cell types used, and the starting inoculations are shown in columns 2 to 5. The two far right columns summarise the longevity and growth rate outcomes of the experiments.

6.3 *In vivo analysis of growth rates of bsf EQ cells in mice*

To explore whether the ectopic expression of ESAG9-EQ protein altered parasitaemia in mice, bsf EQ cells were injected into parasites as described previously. The ectopic expression of ESAG9-EQ protein by transgenic parasites differs from that of ESAG9-K69 protein in that the ESAG9-EQ protein remains cell-associated and is not secreted.

Eight mice were inoculated with 1×10^4 parasites and four of those mice were provided with doxycycline in their drinking water to induce gene expression by the parasites. The parasitaemia in each mouse was then recorded over a period of five days by scoring the number of parasites visible in tail smears using the Herbert and Lumsden rapid matching method (Herbert and Lumsden, 1976).

The results are shown in Figure 6.6. Panel A shows the parasitaemias in individual mice and Panel B shows the mean parasitaemia for the first five days when all the mice were still alive. The +dox mice had a lower mean parasitaemia overall, and one of these mice survived to day 6 whereas all the -dox mice died or were culled on day five. This differs from the results when mice were infected with bsf K9:K69 parasites – in these experiments the trend was for the +dox mice to experience a more virulent infection. The parasitaemias were analysed using an ANOVA as described previously. There was not found to be any statistically significant difference between growth rates in +dox and -dox groups on day 3-4 ($F_{1,6} = 0.59$, $p = 0.473$) or on day 4-5 ($F_{1,6} = 0.43$, $p = 0.537$).

It was not possible to draw firm conclusions form the data in sections 6.2 and 6.3 of this chapter. Effects on virulence were seen in some experiments, but these were subtle and experimentally variable. The results were not considered sufficiently convincing to conclude that ESAG9 was having an effect on virulence.

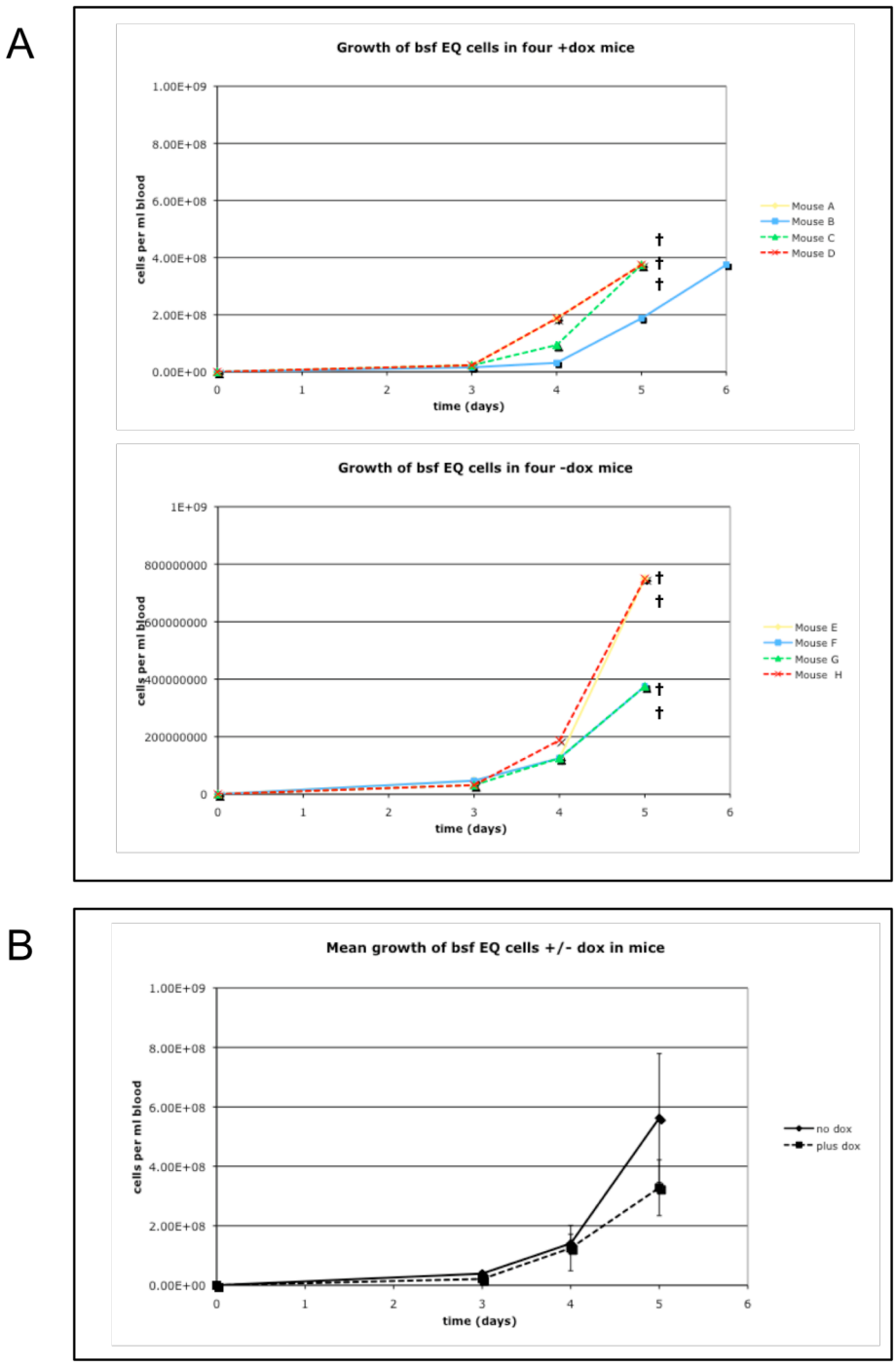


Figure 6.6: Growth of bsf EQ parasites in mice. Eight mice were inoculated with 1×10^4 bsf EQ parasites. Half the mice were provided with drinking water with 200µg/ml doxycycline and 5% sucrose (+dox mice), and half provided with water containing 5% sucrose only (-dox). Parasites per ml of blood were estimated daily from day 2 by counting the number of parasites visible in a tail smear according to the rapid matching method of Herbert and Lumsden (1976). Panel A shows the parasitaemias in individual mice infected with K9:K69 parasites. Panel B shows the mean parasitaemias from mice infected with K9:K69 parasites. Error bars are standard deviations from the mean.

6.4 Ex vivo FACS analysis of murine spleen cells in mice infected with bsf K9:K69 cells

6.4.1 Introduction

When mice were infected with bsf K9:K69 parasites and provided with drinking water containing doxycycline, in some experiments this resulted in the parasite infection being more virulent and in earlier death of the mice (see Section 6.2). However this was not the case in every experiment, and it was not possible to statistically analyse the longevity of the mice for the reasons previously described. To determine whether the ectopic expression of ESAG9 proteins by the bsf K9:K69 parasites was having an effect on the immune responses of the mice, *ex vivo* fluorescent-activated cell sorting (*ex vivo* FACS) of the spleen cells was utilised. The spleen contains large numbers of both erythrocytes, and more importantly, white blood cells. The technique is called '*ex vivo*' because the mice are infected with parasites and the infection allowed to progress, and then the mice are culled and the spleens removed for analysis of the white blood cell populations. FACS analysis combines flow cytometry and fluorescent staining of the cells and can provide a lot of information about the cell populations. Flow cytometry uses the forward scatter (FSC) and side scatter (SSC) of visible light to separate cell populations on the basis of cell size and granularity. Cells can also be stained with primary antibodies to certain cell markers, which are conjugated to fluorochromes to allow detection.

For this experiment, three groups of four mice were used. Two groups of the mice were inoculated with 1×10^4 bsf K9:K69 parsites by intra-peritoneal injection. One of these groups was provided with drinking water with 200μg/ml doxycycline and 5% sucrose ('+dox' mice), whilst the other group was provided with drinking water with 5% sucrose alone ('-dox' mice). The final group were the naïve mice that were inoculated with medium alone via intra-peritoneal injection (these mice were also provided with drinking water supplemented with 200μg/ml doxycycline and 5% sucrose). This was to control for any stimulatory effect the injection itself might have on the murine immune system. For a schematic diagram of the infections, see Figure

6.7. Four days after infection, the mice were culled and the spleens removed. The parasitaemias in the mice were counted prior to culling and these data are shown in Table 6.2. The spleens were the placed in 2mls RPMI medium (Gibco) with 5% FCS. The spleens were placed between two pieces of mesh and mashed with forceps to release the spleen cells into the medium. Following this, the cells were stained with antibodies for FACS analysis.

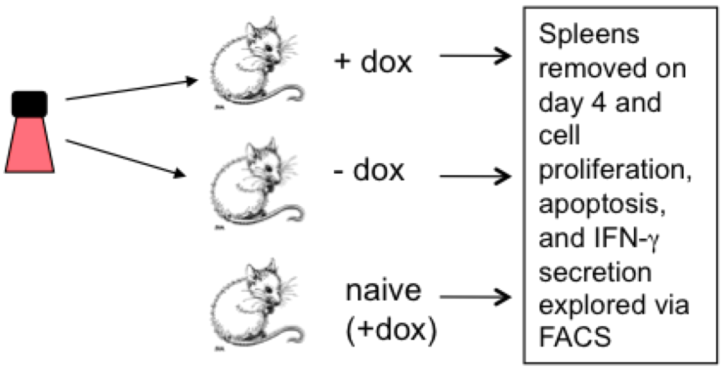


Figure 6.7: Schematic representation of *ex vivo* FACS experiment. Three different groups were used and each group contained four mice. Two groups of mice were each inoculated with 1×10^4 parasites; one group was fed water with 5% sucrose and 200μg/ml doxycycline, and the other group was fed water with only 5% sucrose. The naïve group (ie not inoculated with parasites) was also fed water with 5% sucrose and 200μg/ml doxycycline. After four days the spleens were harvested and the spleen cell populations analysed by FACS.

Treatment	Mouse	Parasitaemia day 4 (cells per ml of blood)
+ DOX	R1	4.7×10^7
	R2	2.3×10^7
	R3	2.3×10^7
	R4	5.95×10^6
- DOX	B1	1.145×10^7
	B2	2.3×10^7
	B3	1.145×10^7
	B4	5.95×10^6

Table 6.2: Parasitaemias in mice after four days. A smear of blood from a tail snip was taken from each mouse and the parasitaemia scored using the rapid matching method of Herbert and Lumsden (1976).

The spleen cells were stained with a number of different markers. The secretion of IFN-γ by T cells has been linked with susceptibility to trypanosmiasis in mice (Shi *et al.,* 2006), and so secretion of this cytokine by T cells was explored using the appropriate markers. Trypanosome infection of mice has been reported to result in the inhibition of proliferation of lymphocytes (Beschin *et al.,* 1998; Sternberg & Mabbott, 1996). Hence the proportion of different classes of lymphocytes (CD4+ T cells, CD8+ T cells, and B cells) was also looked at, though the proliferation of lymphocytes was not explored. It had been suggested this inhibition might come about by apoptosis of lymphocytes (Sternberg and Mabbott, 1996); also the apoptosis of B cells as a result of trypanosome infection has been reported (Radwanska *et al.,* 2008), and so the apoptosis of lymphocytes was also investigated using the appropriate antibodies. Finally, the proliferation of the different classes of granulocytes was investigated. Macrophages are granulocytes, and the activation of, and secretion of nitric oxide by, macrophages is an important

early event in murine trypanosomiasis (Baetselier *et al.,* 2001). A list of the antibodies and suppliers can be found in Appendix D.

6.4.2 Secretion of interferon gamma by T cells

To determine whether T cells were secreting the cytokine interferon gamma (IFN-γ), they were first incubated in a protein transport inhibitor named GolgiStop™ (BD Pharmingen) for five hours so that the cytokine remained cell-associated. The spleen cells were then co-labelled with an antibody for IFN-γ, and an antibody to either CD4+ or CD8+ T cells to determine cell type. Initially, the sub set of white blood cells which were lymphocytes was determined from the FSC/SSC plot and a gate was set to exclude other cell types. Figure 6.8 shows examples of the raw data for one mouse from each group, and also the isotype control and single-staining controls.

The percentages of each T cell type that were producing IFN-γ were then calculated, and are represented in Individual Value Plots, and these data are shown in Figure 6.9. Panel A shows the percentage of CD4+ T cells that were secreting IFN-γ. The +dox mice had the highest mean percentage of CD4+ T cells positive for IFN-γ. The –dox mice had an intermediumte mean percentage, and the naïve mice the lowest mean percentage. To determine whether these differences were statistically significant, ANOVAs were carried out. The -dox mice had significantly more CD4+ T cells secreting IFN-γ than the naïve mice ($F_{1,6} = 8.8$, $p = 0.025$). The +dox mice also had significantly more CD4+ T cells secreting IFN-γ than the naïve mice ($F_{1,6} = 14.78$, $p = 0.009$). However the difference between the number of CD4+ T cells secreting IFN-γ in +dox mice and -dox mice was not significant ($F_{1,6} = 1.3$, $p = 0.298$). It is worth noting that overall the percentages of cells which were positive for IFN-γ were very low.

Figure 6.9 panel B shows the percentage of CD8+ T cells that were secreting IFN-γ. Again the percentages of CD8+ T cells which were secreting IFN-γ were very low, and although the +dox and -dox mice had higher mean percentages than the naïve mice, the difference was not great so it was not considered worthwhile testing whether the difference was statistically significant.

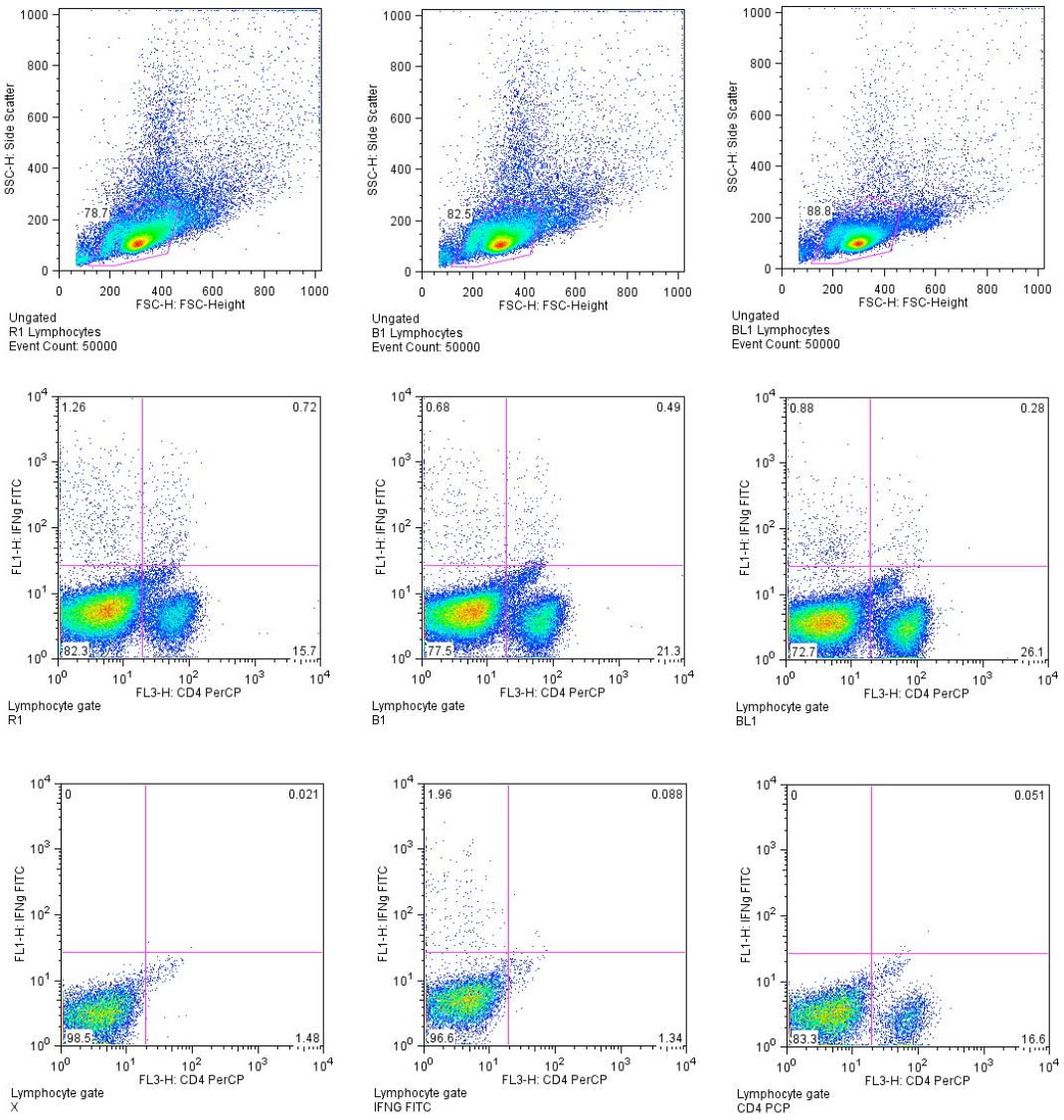


Figure 6.8: An example of raw FACS data and controls. An example of FACS plots for one mouse from each group is shown. Mouse R1 = +dox, B1 = -dox, and BL1 = naïve. The top row shows the FSC/SSC plots for each mouse, and the region that was judged to be lymphocytes is shown surrounded by a pink line; this was used to gate the lymphocyte analysis. The middle row shows the lymphocyte-gated plots for staining of an antibody against CD4+ T cells versus an antibody against IFN-γ. All the cells in the top right portion of the graph are positive for both markers. FITC and PerCP refer to the fluorochrome markers that are conjugated to the antibodies. The bottom row shows the controls: the left graph is the isotype control, the middle is a single staining control with α IFN-γ and the right hand graph is a single staining control with α CD4.

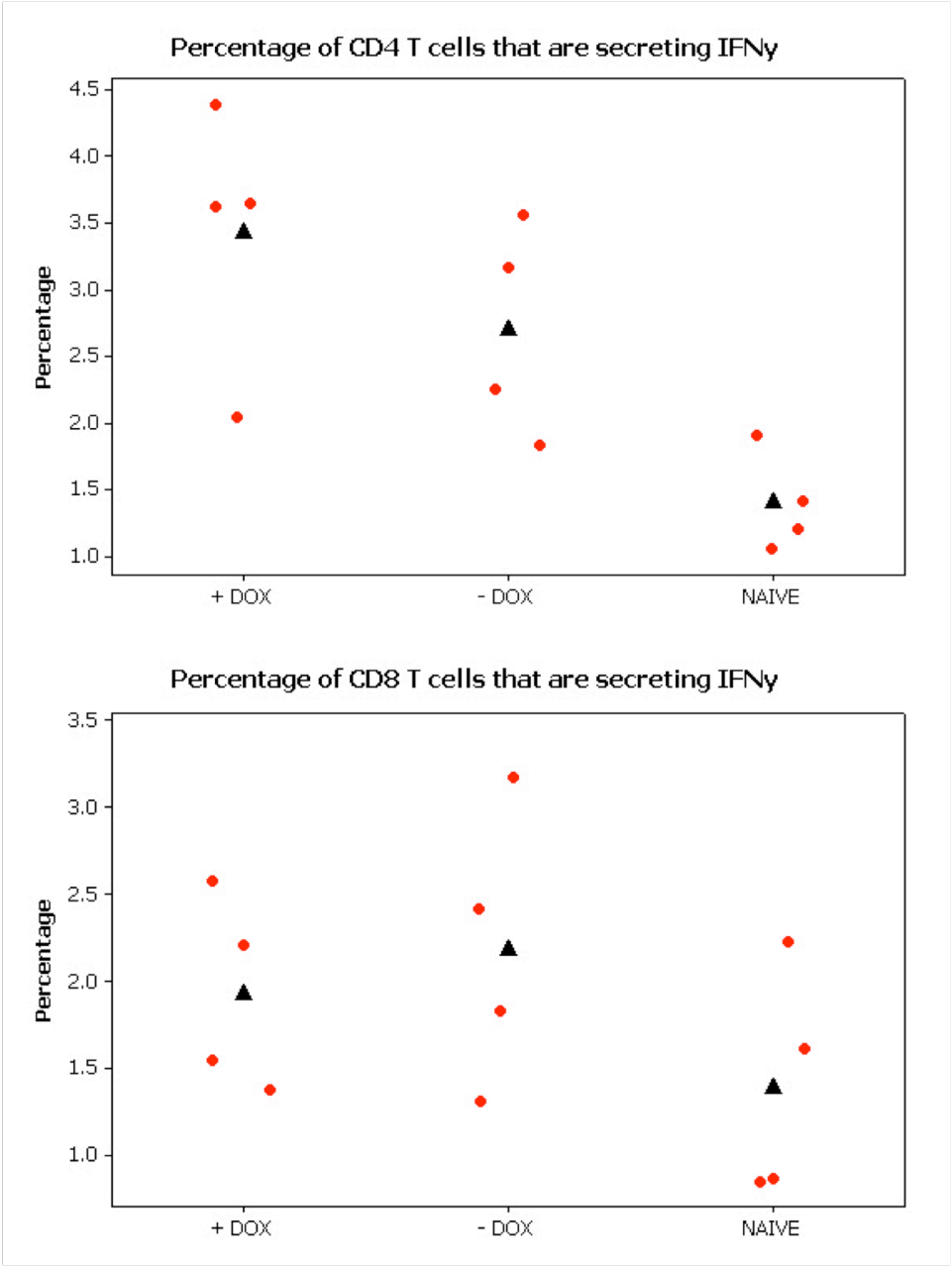


Figure 6.9: Percentage of CD4 and CD8 T cells that were secreting interferon gamma. Eight mice were inoculated with bsf K9:K69 parasites. Half of these were provided with water containing 200μg/ml doxycycline and 5% sucrose (+ DOX) and half were provided with water containing 5% sucrose alone (-DOX). Four mice were inoculated with media alone (NAÏVE) and fed 200μg/ml doxycycline and 5% sucrose. Each red circle represents the percentage of that cell type in an individual mouse, and each black triangle represents the mean value for that group. Panel A is an Individual Value Plot for the percentage CD4+ T cells that were secreting IFN-γ, and panel B is a plot for the percentage of CD8+ T cells that were secreting IFN-γ.

6.4.3 Proportion of different classes of lymphocytes in the spleen

To determine the proportions of different classes of lymphocytes, CD4+ T cells, CD8+ T cells, and B cells were identified by staining with α-CD4, α-CD8a and α-B220 antibodies respectively. For each mouse, the percentage of total lymphocytes which were CD4+ T cells, CD8+ T cells or B cells was then calculated. These data are shown in Figure 6.10.

Panel A shows the percentage of lymphocytes that were CD4+ T cells. The value for each mouse is represented by a circle, and the mean for each group is represented by a triangle. This plot shows that there were differences between the percentages of CD4+ T cells in each group. The naïve mice had the highest mean percentage of CD4+ T cells, the -dox mice an intermediumte mean percentage of CD4+ T cells, and the +dox mice the lowest mean percentage. To determine whether these differences were statistically significant, ANOVAs were carried out as previously described. The +dox mice had significantly less CD4+ T cells than the -dox mice ($F_{1,6} = 8.73$, $p = 0.025$). However the -dox mice did not have significantly less CD4+ T cells than the naïve mice ($F_{1,6} = 5.21$, $p = 0.063$). This was strongly suggestive of the expression of the ectopic ESAG9 proteins by the bsf K9:K69 cells having an effect on the CD4+ T cell population in +dox mice.

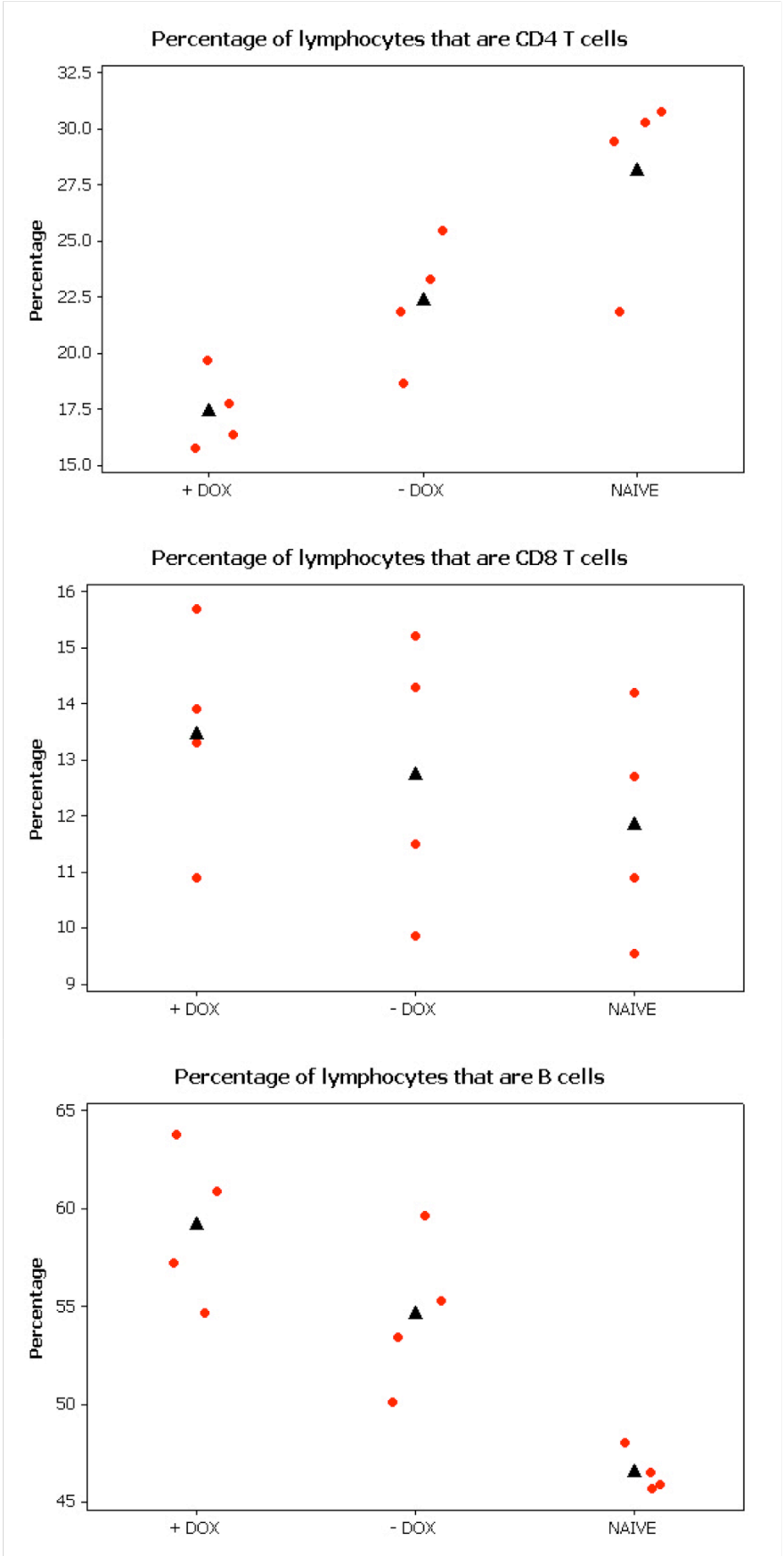
Figure 6.10 Panel B shows the percentage of lymphocytes that were CD8+ T cells. Although there was an increase in the mean percentage when comparing the -dox mice with the naïve mice, and an increase when comparing the +dox mice with the -dox mice, the increase was very small and so was not considered important.

Figure 6.10 Panel C shows the percentage of lymphocytes that were B cells. Both the -dox and +dox mice showed an increase in B cells when compared to naïve mice, and +dox mice have slightly more B cells than -dox mice. To determine whether these differences were statistically significant, ANOVAs were carried out as previously described. The -dox mice had significantly more B cells than the naïve mice ($F_{1,6} = 15.33$, $p = 0.008$). The +dox mice also, not surprisingly, had significantly more B cells than the naïve mice ($F_{1,6} = 36.7$, $p = 0.001$). However the +dox mice did not have significantly more B cells than the -dox mice ($F_{1,6} = 2.55$, $p = 0.162$). This

indicates that although the inoculation with parasites resulted in a higher number of B cells compared to the naïve mice, the presence of doxycycline and induction of gene expression by the bsf K9:K69 parasites was not having a significant effect on the B cell population.

Figure 6.10 (following page): Percentage of spleen lymphocytes that were CD4 T cells, CD8 T cells or B cells. Eight mice were inoculated with bsf K9:K69 parasites. Half of these were provided with water containing 200μg/ml doxycycline and 5% sucrose (+ DOX) and half were provided with water containing 5% sucrose alone (-DOX). Four mice were inoculated with media alone (NAÏVE) and provided with water containing 200μg/ml doxycycline and 5% sucrose. Each red circle represents the percentage of that cell type in an individual mouse, and each black triangle represents the mean value for that group. Panel A is an Individual Value Plot for the percentage of lymphocytes that were CD4+ T cells, panel B is a plot for the percentage of lymphocytes that were CD8+ T cells, and panel C is a plot for the percentage of lymphocytes that were B cells.

A


Percentage of lymphocytes that are CD4 T cells
Percentage
32.5
30.0
27.5
25.0
22.5
20.0
17.5
15.0
+ DOX
- DOX
NAIVE
Percentage of lymphocytes that are CD8 T cells
Percentage
16
15
14
13
12
11
10
9
+ DOX
- DOX
NAIVE
Percentage of lymphocytes that are B cells
Percentage
65
60
55
50
45
+ DOX
- DOX
NAIVE

B

C

6.4.4 Lymphocyte populations undergoing apoptosis

To determine what proportion of CD4+ T cells, CD8+ T cells and B cells were undergoing apoptosis, the spleen cells were stained with the appropriate cell type marker (as described in Section 6.5.2) and an apoptosis marker called Annexin V.

Annexin V is a phospholipid-binding protein (Kaplan *et al.,* 1988). The exposure of phosphatidylserine residues on cell surface membranes is an early event in apoptosis (Naito *et al.,* 1997) and the Annexin V marker recognises these exposed residues (Koopman *et al.,* 1994; van Engeland *et al.,* 1998).

The results of this FACS analysis are shown in Figure 6.11. Panel A shows the percentage of CD4+ T cells that were undergoing apoptosis in +dox, -dox and naïve mice. There was no difference between the mean percentage of CD4+ T cells undergoing apoptosis between the +dox and -dox mice. In the naïve mice there were less CD4+ T cells undergoing apoptosis. The differences in percentage of CD4+ T cells undergoing apoptosis between +dox and naïve mice, and between –dox and naïve mice, were analysed using ANOVAs and found to be statistically significant [($F_{1,6} = 8.02$, $p = 0.03$) and ($F_{1,6} = 13.29$, $p = 0.011$) respectively]. This suggests that trypanosome infection in some way triggers the apoptosis of CD4+ T cells, independent of whether or not the expression of ectopic genes in the bsf K9:K69 parasites was induced.

There were no large differences between the percentages of CD8+ T cells undergoing apoptosis in +dox, -dox or naïve mice, as shown in Figure 6.11 Panel B. Panel C shows that in B cells the mean percentage of cells undergoing apoptosis in naïve and -dox mice was very similar. The mean was lower however in +dox mice. The percentages of B cells undergoing apoptosis in +dox and -dox mice were analysed using an ANOVA. The +dox mice were not found to have significantly less B cells undergoing apoptosis than the -dox mice ($F_{1,6} = 3.07$, $p = 0.13$). However the +dox mice were found to have significantly less B cells undergoing apoptosis than the naïve mice ($F_{1,6} = 7.32$, $p = 0.035$).

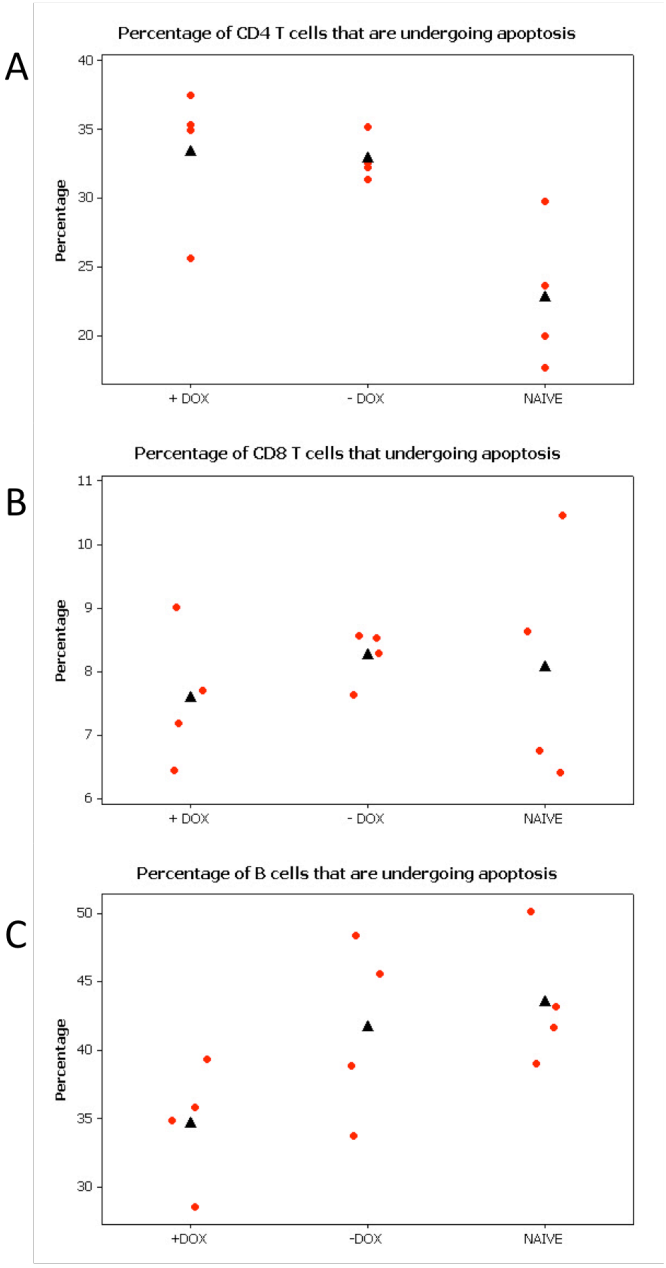


Figure 6.11: Percentage of spleen lymphocytes that were undergoing apoptosis. Eight mice were inoculated with bsf K9:K69 parasites. Half of these were provided with water containing 200μg/ml doxycycline and 5% sucrose (+ DOX) and half were provided with water containing 5% sucrose alone (-DOX). Four mice were inoculated with media alone (NAÏVE) and provided with water containing 200μg/ml doxycycline and 5% sucrose. Each red circle represents the percentage of that cell type in an individual mouse, and each black triangle represents the mean value for that group. Panel A is an Individual Value Plot for the percentage CD4+ T cells that were undergoing apoptosis, panel B is a plot for the percentage of CD8+ T cells that were undergoing apoptosis, and panel C is a plot for the percentage of B cells that were undergoing apoptosis.

6.4.5 Proportion of different classes of granulocytes

Granulocytes can be differentiated from the rest of the spleen cell population by their granularity. This is determined by the side scatter of visible light by the FACS machine. Different sub-classes of granulocytes can then be identified by staining with cell-type specific markers. Macrophages were identified using F4/80, which is a general macrophage marker (Schaller *et al.,* 2002). Eosinophils and neutrophils were identified by staining with Siglec-F (Angata *et al.,* 2001) and Gr-1 (Hestdal *et al.,* 1991) respectively.

The results are shown in Figure 6.12. Panel A shows the percentage of all granulocytes which were macrophages, Panel B the percentage which were eosinophils and Panel C the percentage which were neutrophils. It is clear from these data that, although in both +dox and -dox mice there is an up-regulation of all three types of granulocytes when compared to naïve mice, there is no difference between the percentages in the +dox and -dox mice. This indicates that a trypanosome infection resulted in increased numbers of macrophages, neutrophils and eosinophils, but that induction of ectopic gene expression in the bsf K9:K69 parasites did not have any effect on this. The differences between the percentages of macrophages, neutrophils, and eosinophils in –dox and naïve mice were tested using an ANOVA and in all cases the difference was found to be statistically significant [($F_{1,6} = 6.91$, $p = 0.039$), ($F_{1,6} = 66.98$, $p < 0.0005$) and ($F_{1,6} = 13.3$, $p = 0.011$) respectively].

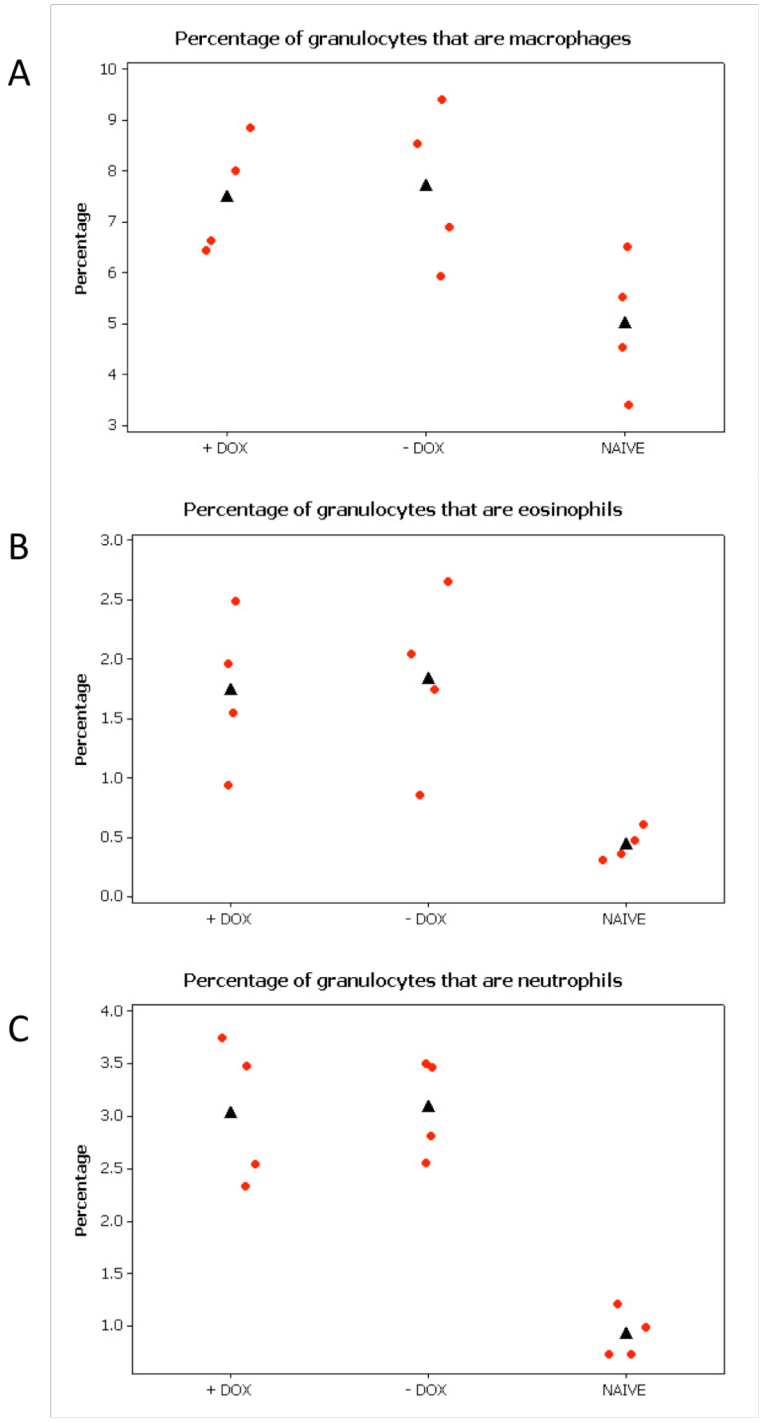


Figure 6.12: Percentage granulocytes that were macrophages, eosinophils, or neutrophils. Eight mice were inoculated with bsf K9:K69 parasites. Half of these were provided with water containing 200μg/ml doxycycline and 5% sucrose (+ DOX) and half were provided with water containing 5% sucrose alone (-DOX). Four mice were inoculated with media alone (NAÏVE) and fed 200μg/ml doxycycline and 5% sucrose. Each red circle represents the percentage of that cell type in an individual mouse, and each black triangle represents the mean value for that group. Panel A is an Individual Value Plot for the percentage granulocytes that were macrophages, Panel B is a plot for the percentage of granulocytes that were eosinophils, and Panel C is a plot for the percentage of granulocytes that were neutrophils.

6.5 Summary

The mice infection experiments discussed in Sections 6.2 and 6.3 gave varying results. The infection experiments with bsf K9:K69 parasites were carried out with two different starting inoculations and were repeated two or three times. The outcomes of these experiments are summarised in Table 6.1. It is hard to draw strong conclusions from data where there is such variability between individual experiments, though the +dox mice in some cases experienced an increased growth rate of parasites compared to the –dox, and also a shortened longevity.

The method used to count the parasite density in blood was not very precise. Tail smears were taken and the cells per ml of blood estimated from the number of parasites visible in a field of view under the microscope, according to the rapid matching method of Herbert and Lumsden (Herbert and Lumsden, 1976). There was therefore a possibility that human error occurred during counting. The experiment was blinded in terms of which mice were receiving the doxycycline. A more accurate way of quantifying the parasitaemia would have been to take blood samples for quantitative Real-Time PCR, however this was beyond the scope of this PhD project.

The longevity of the mice was hard to assess. To determine this accurately would have required the mice to be checked 24 hours a day, or for a heat sensitive camera to be set up which would help determine time of death.

The *ex vivo* FACS analysis gave some intriguing results. The down-regulation in the number of CD4+ T cells in the +dox mice (see Figure 6.10) is an interesting phenotype, and the data were statistically significant when analysed using an ANOVA. Other published data have shown an inhibition of T cell proliferation as a response to infection with *T. brucei* and this will be discussed in Chapter 7 section 7.9. There were also some smaller and not statistically significant differences between the +dox and –dox groups in terms of the number of B cells, the number of CD4 and T cells undergoing apoptosis, and the number of CD4 cells secreting IFN-γ.

Although these results are fascinating, it is important to repeat these experiments. Other controls also need to be included. As well as naïve mice, and mice inoculated with bsf K9:K69 parasites, it would be very useful to have mice inoculated with bsf WT parasites, both with and without doxycycline. This would control for the fact that the ectopic gene expression by bsf K9:K69 parasites may be leaky, and therefore some ESAG9-K69 protein could be secreted by bsf K9:K69 parasites in the absence of doxycycline. It would also determine whether the presence of doxycycline had any effect when mice were inoculated with parasites, independent of cell line.

Chapter 7 Discussion

African trypanosomes cause a lethal disease of humans, and a debilitating disease in cattle, and are therefore of huge socio-economic importance.

The African trypanosome has a complicated life cycle, whereby it is passed between mammalian hosts by the tsetse fly vector. The stumpy cells are of special interest as they are the agents of transmission between the mammal and the insect vector. The specific attributes of stumpy cells, which allow them to survive in both a mammalian and tsetse fly host, have not been extensively characterised as of yet. Prior to the commencement of this PhD project, members of the ESAG9 gene family had been identified as stumpy-specific messenger RNAs, which is a novel expression profile in *Trypanosoma brucei*.

The aims of this project will be reiterated, followed by a summary of the findings. The overall conclusions that can be drawn from the findings will then be outlined, and these will be put in the context of other research in this field of biology. Potential future directions for this project will then be discussed.

7.1 Overview of aims

In brief, the aims of the PhD project were to:

- Determine the protein expression profile of ESAG9 proteins.
- Explore attributes of ESAG9 protein sequences using bioinformatics and use this information to predict post-translational modifications or functions of ESAG9s.
- Create transgenic cell lines with which to determine sub-cellular localisation and post-translational modification of ESAG9 proteins.
- Use these transgenic cell lines in functional assays.

7.2 Stage-specific gene expression

A search for genes that are up-regulated in the stumpy stage of the trypanosome life cycle was carried out prior to the commencement of this PhD project. A family of genes called ESAG9s were found to be upregulated in the stumpy stage (see Chapter 3 sections 3.1.2-3.1.4). The raising of anti-peptide antibodies against two members of this gene family, ESAG9-K9 and ESAG9-K69 revealed that they had differing protein expression profiles. ESAG9-K9 protein was shown to be expressed at a low level by stumpy cells, and its expression peaked between 6 and 9 hours through differentiation to procyclic forms. By 30 hours into differentiation the ESAG9-K9 protein was no longer detectable (see Chapter 3 section 1.1.5, Figure 5). Such an expression profile had not been described before in *T. brucei*. ESAG9-K69 protein had a different protein expression profile whereby there was a low level of cell-associated protein in stumpy cells and differentiating cells. There is no antibody available against ESAG9-EQ so the protein expression in stumpy and differentiating cells could not be investigated.

ESAG9 genes were attributed to be ESAGs (expression site-associated genes) because when they were first identified in *T. equiperdum*, one ESAG9 gene was found to be within an expression site (ES) (Florent *et al*., 1991). None of the nine ESAG9 genes and ten pseudogenes so far identified in *T. b. brucei* 927 are in expression sites (Berriman et al 2005) but there was an absence of sequenced ESs in that analysis. However, one ESAG9 gene has been found in an ES in *T. b. brucei* strain Lister 427 (Hertz-Fowler *et al*., 2008), where a specific analysis of ESs has been carried out.

The genomic context of ESAG9 genes in *T. b. brucei* EATRO 2340 has not been explored during this project. This pleiomorphic strain of the parasite contains at least four ESAG9 genes (see Chapter 3, Figure 3.2), including the three genes that we have focused on: ESAG9-K9, ESAG9-K69 and ESAG9-EQ. It would be interesting to know whether or not these genes are in ESs in *T. b. brucei* EATRO 2340. This

could be deduced by investigating whether the transcription of these genes is sensitive to α-amanitin. The ESAGs and VSG in an ES are transcribed by pol I which is insensitive to α-amanitin (Gunzl *et al.,* 2003; Rudenko *et al.,* 1989) whereas genes outwith ESs are transcribed by pol II which is sensitive.

7.3 Bioinformatic analysis

The genomic context of the ESAG9 genes, conservation within the gene family, similarity to any other characterised genes, the presence of functional motifs or signals for post-translational modification, and the conservation of the UTRs were all explored using bioinformatic analyses.

The ESAG9 gene family is not highly conserved, with between 16% and 88% identity between the peptide sequences. There are no homologues in other organisms (excluding the very closely related *T. equiperdum*) as determined by BLAST searches. However there was a motif common to most of the ESAG9 peptide sequences that was also found in protein phosphatase 2C (PP2C) enzymes. The motif CX_3WX_8C is involved in the binding of divalent cations in PP2C enzymes (Das *et al.,* 1996; Kusuda *et al.,* 1998). It is unlikely to perform this function in ESAG9 proteins because they do not have the other amino acids involved in the binding site, but this conserved motif could be performing another function in ESAG9s.

Although there were no true homologues of ESAG9s in other organisms, the ESAG9 sequences did show some similarity to the mucin-associated surface protein (MASP) gene family of *T. cruzi* (see Chapter 3 sections 3.2.5 and 3.2.6). MASPs are membrane-associated, glycosylated proteins of unknown function in *T. cruzi*. These proteins do not contain the CX_3WX_8C motif.

ESAG9 protein sequences were queried for the presence of any characterised functional motifs, using a range of databases. The only useful information from these searches was that all ESAG9 sequences have a signal peptide at the N terminus.

ESAG9 proteins do not bear any significant resemblance to any characterised functional motifs.

Tools for predicting post-translational modifications of peptide sequences were used to explore whether ESAG9 proteins were likely to by modified by the addition of a GPI anchor and/or N-glycans (see Chapter 3 section 3.2.8). GPI-SOM predicted GPI anchor addition to all ESAG9 peptides. Seven out of the eleven ESAG9 sequences queried were predicted to be N-glycosylated at at least one position.

The un-translated regions (UTRs) of ESAG9 genes were explored for conserved regions, motifs, or 2D folds (see Chapter 3 section 3.2.9). The 3'UTRs were not found to be highly conserved, and contained no unanimous 2D fold structures. However MEME analysis did reveal that there were two conserved motifs common to the ESAG9 3'UTRs. There was no conservation between the 5' UTRs of ESAG9 genes.

To conclude, the bioinformatic analyses were not highly informative in terms of predicting function of the ESAG9 genes. This gene family has no homologues and the sequences contain no motifs of known function. However the similarity of ESAG9 protein sequences to MASPs in *T. cruzi* was interesting, especially given that the ESAG9s were predicted to be GPI-anchored and, in some cases, N-glycosylated, as are the MASP proteins (Atwood *et al.,* 2006). Unfortunately MASP proteins have not yet been assigned a function. The presence of two conserved motifs in the 3' UTRs could provide a mechanism by which the unusual expression profile of the ESAG9 genes is controlled; but experimental exploration of this possibility was beyond the realms of this project.

7.4 Sub-cellular localisation of ESAG9 proteins

Bloodstream form and procyclic form transgenic cell lines were engineered to ectopically express one or more of the following ESAG9 proteins: ESAG9-K9,

ESAG9-EQ and ESAG9-K69. These were then used to explore various attributes of ESAG9 proteins including their sub-cellular localisation.

7.4.1 Sub-cellular localisation of ESAG9-K9

ESAG9-K9 was located in the endoplasmic reticulum (ER) in both procyclic and bloodstream form transgenic cells. Co-staining of bsf K9 cells with the ER marker BiP and BB2 (which recognises the Ty-tag in the ectopic ESAG9-K9 protein) and confocal microscopy analysis revealed that these proteins co-localise in the ER (see Chapter 4 Figure 4.23). The localisation of ESAG9-K9 in differentiating stumpy *T. b. brucei* EATRO 2340 cells was explored using the ESAG9-K9 anti-peptide antibody. Although this antibody gave a certain degree of background in immunofluorescence, it was possible to compare slender cells (which do not express ESAG9-K9 protein) with differentiating stumpy cells and see that the protein was intracellular with a halo around the nucleus (see Chapter 3 Figure 3.6). This could represent an ER localisation. Some differentiating stumpy cells also had a signal from the ESAG9-K9 antibody at the flagellar pocket.

7.4.2 Localisation of ESAG9-EQ protein

ESAG9-EQ was seen as a diffuse signal with a gap coincident with the nucleus in both procyclic and bloodstream form transgenic cells. As discussed in Chapter 4, section 4.5.2, this staining could represent a cytoplasmic location for ESAG9-EQ, but it could also possibly be at the cell surface. This was not explored further as part of this project. However it would be very interesting to conclusively determine the localisation of ESAG9-EQ. This could be achieved by staining bsf EQ cells with BB2 antibody and analysing the staining pattern by confocal microscopy. This would give information about the localisation of the protein in 3D, which would make it easier to determine whether or not the protein is cell surface located. Another possible route would be to use a commercially-available cell fractionation kit which fractions lysed cells into cytosolic, membrane and organellar vesicle, nuclear, and cytoskeletal fractions, as utilised by Paterou et al. to localise TbZFP3 protein (Paterou *et al.,* 2006). The resultant fractions are then analysed by Western blotting.

If there was a signal for ESAG9-EQ in the membranous fraction, then it could be inferred that the protein was associated with the cell surface membrane, given that the immunofluorescent staining was not indicative of the protein being in the membrane of any sub-cellular organelle.

Procyclic cell lines have been engineered by the laboratory of Jeremy Mottram that are lacking a catalytic subunit required for the addition of GPI anchors to proteins (Lillico *et al.,* 2003). Using this mutant, ESAG9-EQ could be ectopically expressed and analysed by immunofluorescence to determine if its localisation changed when the protein was not modified by the addition of a GPI anchor. It would not be possible to do the same experiment in bloodstream forms because GPI anchor addition is essential for survival in that stage of the life cycle (Nagamune *et al.,* 2000).

7.4.3 Localisation of ESAG9-K69 protein

ESAG9-K69 protein, when expressed concurrently with ESAG9-K9 protein in bloodstream form transgenic cells, was secreted out of (or shed from) the cell into the medium (see Chapter 4 section 4.3.3), as assayed by immunoprecipitation of conditioned medium. When ESA9-K69 and ESAG9-EQ protein were expressed concurrently, a signal for Ty-tagged protein was seen from an IP of conditioned medium (section 4.3.5). This could be either protein as they were both Ty-tagged, however it was most likely to be ESAG9-EQ protein due to the size it migrated at. An ESAG9-K69 anti-peptide antibody was also used to explore whether *T. b. brucei* EATRO 2340 stumpy cells secreted ESAG9-K69. Western analysis using the LI-COR infrared imaging system revealed that stumpy cells secreted/shed 39.4% of the ESAG9-K69 protein. A cell-associated control, Hsp-70, was used and 4.4% of this protein was released into the medium. This indicates that the amount of ESAG-K69 protein that is in the medium was very unlikely to be due to cell death but will be an active process of secretion or shedding by the stumpy cells. The secretion of other proteins by *T. b. brucei* will be discussed in section 7.9.2.

7.5 Post-translational modification of ESAG9 proteins

7.5.1 N-glycosylation of ESAG9 proteins

N-glycosylation and GPI anchor addition to ESAG9 proteins was explored. ESAG9-K9 protein and ESAG9-K69 protein were both found to be N-glycosylated, as shown by their increased migration in an SDS-PAGE gel after treatment with PNGase F enzyme. ESAG9-EQ, in contrast, was not N-glycosylated. The altered migration profile of the ESAG9 proteins suggested that ESAG9-K9 was potentially glycosylated at two of its 3 potential glycosylation sites and ESAG9-K69 was glycosylated at one of its two potential glycosylation sites (see Chapter 4 section 4.6.1). This modification of ESAG9 proteins was predicted by the bioinformatic analyses carried out in Chapter 3. As mentioned above, ESAG9 protein sequences were found to have some similarities to the mucin-associated surface proteins (MASPs) and mucin-like glycoproteins of *T. cruzi*; these proteins are extensively glycosylated (Buscaglia *et al.,* 2006).

7.5.2 GPI-anchor addition to ESAG9 proteins

A hypotonic assay was used to explore whether the ESAG9 proteins investigated were modified by the addition of a GPI anchor. This assay indicated that ESAG9-EQ was indeed GPI anchored, whereas ESAG9-K9 was not. The fact that ESAG9-EQ was GPI anchored strongly suggests that the signal seen in the immunofluorescence assay with bloodstream form cells (see Chapter 4 section 4.5.2) was in fact a cell surface staining.

ESAG9-K9 had been predicted by bioinformatic analyses to be modified by the addition of a GPI anchor. It has been suggested that the addition of a GPI anchor acts as a signal on the protein that allows it to leave the endoplasmic reticulum (Muniz 2000). It could be hypothesised therefore that ESAG9-K9 localises to the

endoplasmic reticulum in transgenic cells (refer to section 1.4.1) due to a failure by the cells to successfully add a GPI anchor to an ectopically-expressed protein. However there are several lines of argument against this idea. Ectopically-expressed proteins have been found to be successfully GPI-anchored in other cell lines, for example Bangs et al. (Bangs *et al.,* 1997). When VSG is truncated so that it lacks the signal for the addition of a GPI anchor, it was found to remain in the ER in procyclic form cells (McDowell *et al.,* 1998), but importantly it was found to be quickly degraded in lysosomes by bloodstream form cells (Triggs & Bangs, 2003), whereas we saw a strong signal from the ER for ESAG9-K9 protein in the bsf K9 cell line. Moreover, in wild-type pleiomorphic stumpy cells there was no signal for ESAG9-K9 from the supernatant in the hypotonic lysis experiment (see Chapter 4 section 4.6.2). This indicates that endogenous ESAG9-K9 protein is not GPI anchored, and that the bioinformatic predictions were in this case incorrect.

It was not possible to determine whether or not ESAG9-K69 protein was GPI anchored (see Chapter 4 section 4.6.2), since there was insufficient ESAG9-K69 protein remaining cell-associated to perform a hypotonic lysis assay. An attempt at pulse-labelling cells with tritiated myristate did not give any signal; this may have been because the amount of tritiated myristate used was insufficient, or it could have been because the ESAG9-K69 protein was not GPI anchored. A Western blot of ESAG9-K69 protein selected from conditioned medium by immunoprecipitation did not give a signal with the anti-CRD (cross-reacting domain) antibody. This indicated that if ESAG9-K69 was GPI anchored, then this anchor was not being cleaved by GPI-PLC when the protein was shed from the cell. When ESAG6 was shed from the cell by hypotonic lysis a signal was seen with the anti-CRD antibody (Schell *et al.*, 1991). The addition of GPI anchors has been implicated in intracellular trafficking and secretion in trypanosomes (Schwartz *et al.,* 2005; Triggs & Bangs, 2003), and so would provide a pathway by which ESAG9-K69 is secreted. It would be very interesting to conclusively determine whether or not this protein is GPI anchored.

7.6 Analysis of role of ESAG9 proteins in the early colonisation of the tsetse fly midgut

Analysis of the expression profile of ESAG9-K9 protein revealed that it was expressed at a low level by stumpy cells, and its expression peaked 6 to 9 hours into differentiation, before being undetectable by 30 hours into differentiation to the procyclic form. This expression profile was indicative of ESAG9-K9 protein playing a possible role in the early colonisation of the tsetse fly midgut.

To explore this possibility, *in vitro* and *in vivo* assays were used. One potential role of the ESAG9 protein is in protecting the parasites from the mammalian complement active in the bloodmeal. To investigate this, procyclic form cells, engineered to inducibly express either ESAG9-K9 or ESAG-EQ, were incubated with serum (containing active complement) from various mammals and their growth assayed using a reagent called alamarBlue. It was found that the expression of either protein did not enhance survival of parasites under the *in vitro* conditions used.

In contrast to ESAG9-K9, ESAG9-K69 protein was secreted/shed by stumpy cells. Hence, the protein could also be secreted/shed by stumpy cells differentiating to procyclic forms, although this was not confirmed experimentally. Instead, wild-type procyclic form cells were incubated with bsf K9:K69 conditioned medium, from cells both induced and uninduced for ectopic protein expression, and guinea pig serum containing active complement. However, although the conditioned medium would contain the secreted ESAG9-K69, this did not enhance the survival of procyclic cells.

In vivo assays were also used to attempt to determine whether ESAG9 proteins play a role in the colonisation of the tsetse fly midgut. Tsetse flies were fed with horses' blood supplemented with cultured trypanosomes, as described in Chapter 5 section 5.4. The ability of bsf K9 and pcf EQ parasites to colonise the fly midgut was then assayed by counting the burden of infection by microscopy between 24 hours and seven days post-feed. However it proved extremely difficult to quantify the infection levels with any degree of accuracy due to the low levels of trypanosome infection and the difficulties in counting the infection levels accurately by microscopy.

Quantitative real-time PCR was also used to quantify the infection levels of trypanosomes in tsetse flies (see Chapter 5 section 5.6). This technique also proved problematic however; the qPCR reaction was not repeatable. Quantitative PCR has been used to successfully quantify levels of parasites in mosquitoes (Bell and Ranford-Cartwright 2004) and sandflies (Ranasinghe *et al.,* 2008) but these insects are much smaller than tsetse flies, which may make it easier to cleanly extract the DNA. Given the difficulties in quantifying the trypanosome infection levels in tsetse flies, it was not possible to draw any firm conclusions about whether the ESAG9 proteins play a role in the early colonisation of the tsetse fly midgut. The expression of ESAG9-K9 protein by bloodstream form cells did enhance their ability to infect tsetse flies to a small degree (Chapter 5 Figures 5.10, 5.11 and 5.12) but this would need to be investigated further and with a more rigorous assay before this could be assigned as a function of ESAG9-K9 protein.

7.7 Analysis of the effect of ESAG9 protein expression during murine trypanosomiasis

Given that the ESAG9-K69 protein was secreted/shed by both transgenic monomorphic cells and by wild-type pleiomorphic stumpy cells, it was possible that this protein played a role in parasite-host interactions. This was, therefore, explored by performing *in vivo* and *ex vivo* experiments with the murine model of trypanosomiasis.

7.7.1 Effect of ESAG9 protein expression on parasitaemia in mice

In a series of experiments, mice were inoculated with either 1×10^4 or 1×10^3 parasites, and their parasitaemias assessed by counting how many parasites were visible in blood smears.

In five separate experiments, mice were infected with bsf K9:K69 parasites (which had been shown to secrete/shed ESAG9-K69 protein). In the first two experiments, a

starting inoculation of 1×10^4 was used and in the final three experiments a starting inoculation of 1×10^3 parasites was used. The induction of ectopic gene expression was achieved in half of the mice by providing the mice with water containing doxycycline. The parasitaemias were then compared between the mice in which gene expression had been induced, and those in which gene expression had not been induced.

The experiments and outcomes are summarised in Chapter 6 Table 6.1. In the two experiments in which a starting inoculation of 1×10^4 parasites was used, the +dox experienced a higher growth rate of parasites than the -dox mice in one experiment, and this difference was statistically significant only between day 3 and day 4. In the other experiment there was no difference in growth rate of parasites between +dox and -dox mice.

Of the three experiments in which a starting inoculation of 1×10^3 parasites was used, in one experiment the +dox mice experienced a statistically significant higher growth rate than the -dox mice between both days 3 and 4, and days 4 and 5. In a second experiment, the growth rate in +dox mice was only significantly higher between days 4 and 5. In the final experiment there were no statistically significant differences between the growth rates of parasites in +dox and -dox mice.

Thus it can be summarised that the outcomes of these infections of mice with transgenic parasites were very variable. It was not therefore possible to draw firm conclusions about whether the expression of ESAG9-K9 and ESAG9-K69 proteins (and hence the secretion/shedding of ESAG9-K69 protein) was having an effect on virulence in the mouse model of trypanosomiasis. However the fact that sometimes an effect was seen was, in our opinion, sufficient justification to further explore a potential role for ESAG9 proteins in host-parasite interactions using *ex vivo* FACS analysis.

To determine with more certainty whether the expression of ESAG9-K9 and ESAG9-K69 has an effect on the virulence of the infection, quantitative real-time PCR could be utilised to determine the number of parasites in the bloodstream of the mammal. This would give more accurate data than the Herbert and Lumsden method of rapid matching (Herbert and Lumsden, 1976).

7.7.2 *Ex vivo* FACS analysis of spleen cell populations

We utilised *ex vivo* FACS to analyse the effect of ESAG9 protein expression on populations of spleen cells. The technique involved inoculating mice with trypanosomes, allowing the infection to develop *in vivo*, and then harvesting the spleens for FACS analysis. Twelve mice were used in this experiment. Eight mice were inoculated with 1×10^4 bsf K9:K69 parasites; half of these were administered drinking water containing doxycycline (to induce ectopic gene expression). Four mice were inoculated with medium alone, to control for any stimulatory effect on the murine immune system that might result from receiving an injection.

The proportions or apoptosis of, and the interferon gamma production by, different spleen cell populations were then analysed by FACS (see Chapter 6, sections 1.4.2 – 1.4.5). The aspects of the spleen cell populations that were assayed were: the proportions of CD4+ T cells, CD8+ T cells, B cells and granulocytes in the spleen; apoptosis of CD4+ T cells, CD8+ T cells and B cells; and secretion of interferon gamma by CD4+ and CD8+ T cells. The most interesting outcome of these experiments was that there was a statistically significant down-regulation of CD4+ T cells in the spleens of mice infected with bsf K9:K69 parasites induced with doxycycline, when compared to mice infected with bsf K9:K69 parasites that had not been induced (Chapter 6, section 6.4.2). No statistically significant differences were found between the +dox and -dox mice that were infected with bsf K9:K69 parasites in any of the other assays. In some of the assays, there was a statistically significant difference between trypanosome-infected (irrespective of induction of gene expression) and naïve mice, which was not surprising, and these differences are discussed in Chapter 6.

The mechanism by which the ectopic expression of ESAG9 and ESAGK69 proteins by the transgenic parasites might cause a down-regulation of CD4+ T cells is not known. Given that the ESAG9-K69 protein is secreted by the cells (see Chapter 4, Figure 4.15), whereas ESAG9-K9 remains internal, it is highly likely that the ESAG9-K69 protein is the causative agent of the effect seen.

The effect of an inhibition of T cell proliferation as a result of trypanosomiasis has been observed in *in vitro* and *ex vivo* assays a number of other investigators, and this will be discussed in section 7.9.4. Macrophages are classically activated during early-stage trypanosomiasis (Baetselier *et al.,* 2001), and will therefore be producing NO. The importance of macrophages was not explored as part of this study beyond whether there was a difference in proportions of macrophages in the spleen between the different treatment groups. It would be very interesting to explore whether ESAG9-K69 protein could be having an effect on macrophage activation and NO production given that other studies have highlighted a link between NO production and inhibition of lymphocyte proliferation during trypanosomiasis.

The activation of macrophages could be explored by various assays. The production of NO by cultured macrophages or macrophages harvested from the spleen or lymph nodes of trypanosome-infected mice could be measured using a Greiss assay (Green *et al.,* 1982). Also, quantitative reverse-transcriptase PCR could be used to determine the level of transcription of genes that are involved in NO production, such as NO synthase and arginase.

The proliferation of different classes of lymphocytes (as opposed to the proportion of different classes of lymphocytes in the spleen, expressed as a percentage of the total number of lymphocytes in the spleen, which is what was assayed in this project) could be determined by doing, for example, *ex vivo* or *in vitro* growth assays with tritiated thymidine.

A control that was not included in this experiment, but which would be useful for our understanding of this process of down-regulation of CD4+ T cells, would be to infect mice with transgenic parasites that secrete an alternative protein. This would help to determine whether it is the secretion of any protein that results in the effect seen, or whether the effect is in fact specific to the secretion of ESAG9-K69. However it might be challenging to engineer parasites to secrete an alternative protein. Stimulatory effects on the murine immune system by protozoal GPI-anchored proteins have been extensively reported (Magez *et al.,* 2002; Tachado *et al.,* 1997). It is not known whether ESAG-K69 is GPI-anchored.

There are also some other controls that would be beneficial that were not included in the original assay. Mice were inoculated with transgenic bsf K9:K69 parasites, or with medium for the naïve controls. The ectopic expression of ESAG9 proteins by the bsf K9:K69 cell line may have been leaky. Therefore it would also be very useful to include the control of mice inoculated with the 427-449 parental cell line (which does not express ESAG9 proteins), and provided with water containing doxycycline and sucrose or containing sucrose alone. This would also therefore control for the effect of administering doxycycline, independent of gene expression. Unfortunately these assays have not yet been attempted because our collaborator, Simmi Mahajan, is detained in India due to visa issues.

7.8 Overall conclusions

This PhD project has explored the expression profiles, post-translational modifications, and functions, of members of the ESAG9 family of proteins.

ESAGs are a diverse family of proteins found in the telomeric expression sites of VSG, and also in chromosome-internal positions. They have been implicated in host-parasite interactions (Pays *et al*., 2001). This has been shown conclusively for an ESAG gene called SRA (De Greef *et al*., 1989; De Greef *et al*., 1994; Xong *et al*., 1998). ESAGs 6 and 7, which form the transferrin receptor (Salmon *et al*. 1994; Steverding *et al*. 1994), have also been implicated in increasing the host range of African trypanosomes (Bitter *et al*., 1998), though evidence which is contrary to this theory has also been published (Salmon *et al*., 2005; Steverding *et al*., 2006).

In the context of the wider knowledge regarding the functions of some ESAG proteins, we explored the possibility of ESAG9 proteins having a function in host-parasite interactions. Our findings can be summarised as follows:

1. Initial bioinformatic analysis showed some similarity to *Trypanosoma cruzi* mucin-associated surface proteins (MASPs), and predicted the occurrence of post-translational modifications.
2. Messenger RNA expression profiles of various ESAG9 proteins had been characterised and the mRNAs shown to be up-regulated in stumpy forms; the protein-expression profiles of ESAG9-K9 and ESAG9-K69 showed that these proteins were expressed by stumpy and differentiating cells.
3. The three proteins investigated using transgenic cell lines, ESAG9-K9, ESAG9-K69, and ESAG9-EQ, varied in their sub-cellular location and post-translational modifications.
4. ESAG9-K69 was shown to be secreted/shed by bloodstream form transgenic and wild-type pleiomorphic cells.
5. The ability of ESAG9-K9 and ESAG9-EQ proteins to enhance the colonisation of the tsetse fly midgut by transgenic parasites was explored but these data were inconclusive.
6. The expression of ESAG9-K9 and ESAG9-K69 by transgenic bloodstream form parasites in the mouse host resulted in a significant decrease in the number of CD4+ T cells in the spleen when compared to controls. ESAG-K69 can therefore be tentatively assigned a role in suppression of the host immune response by the parasite, though it is important to note that this experiment has not yet been repeated.

7.9 The impact and relevance of these data

The characterisation of an ESAG9 protein as a potential immunomodulator that is secreted/shed by *Trypanosoma brucei brucei* is a novel and interesting one. The findings of this project can be compared and contrasted with what is known about other ESAGs, other stumpy cell specialisations, other trypanosomatid secreted proteins and immunomodulators, and also compared with what is known about the

MASP proteins of *T. cruzi*. This puts the data regarding ESAG9 in the wider context of what has been discovered so far in this field of trypanosomatid biology.

7.9.1 Comparison to other ESAG gene families

The attributes of the other ESAG gene families are summarised in Chapter 1, Table 1.1. Not all of the ESAG families have been characterised in much detail, indeed only ESAGs 6, 7 and SRA have had a function conclusively ascribed to them. Where characterised, however, the genes have been shown to be important in host-parasite interactions. ESAGs 6 and 7 form the transferrin receptor (Tf-R), which is responsible for binding and uptake via the flagellar pocket of host transferrin in bloodstream form parasites (Salmon *et al*., 1994; Steverding *et al*., 1994). Both these proteins are N-glycosylated and ESAG6 has a GPI anchor, as does ESAG9-EQ. However there is no sequence similarity between either of these ESAGs, which ruled out ESAG9 playing a similar role.

It was the discovery that the different Tf-Rs formed by ESAGs 6 and 7 from different VSG expression sites had varying abilities to bind transferrin from different mammal hosts (Bitter *et al*., 1998) that first bought about the hypothesis that a function of ESAGs is to increase the host range of *T. brucei* (Bitter *et al.,* 1998; Pays *et al*., 2001). Indeed, the serum resistance associated gene (SRA) is an ESAG, and has been shown to be responsible for the ability of *T. b. rhodesiense* to be human infective (whereas *T. b. brucei* is not human infective) (De Greef *et al*., 1994).

In this context, if the function of ESAG9s is, as discussed, to aid the survival of stumpy parasites by manipulating the mammal host immune system, it is possible that slightly different versions of ESAG9 could perform this role optimally in different mammal hosts. However, the mode of expression of ESAG9 genes is very different to that of ESAGs 6 and 7. ESAGs 6 and 7 could be described as canonical ESAG genes, in that they are present in all the ESs that have been sequenced so far

(Berriman *et al.,* 2002; Hertz-Fowler *et al.,* 2008). Members of the ESAG9 gene family are rarely found in expression sites however: there is one ESAG9 gene in an expression site in *T. equiperdum* (Florent *et al*., 1991), one in an expression site in *T. b. brucei* strain Lister 427 (Hertz- Fowler *et al*., 2008), and one ESAG9 pseudogene in a metacyclic expression site in *T. brucei* strain EATRO 795 (Graham *et al*., 1999) (though there may be other ESAG9 genes in expression sites that have not yet been sequenced). Therefore most expression sites do not in fact contain ESAG9. There does not seem to be selective expression of ESAG9 mRNAs by stumpy parasites; the *T. b. brucei* EATRO 2340 and AnTat1.1 strains express a different array of ESAG9 mRNAs, but this is due to the presence or absence of the genes in each strain (Chapter 3 section 3.1.4 and K.M. unpublished data) rather than selective expression. There could be differences between the relative abundance ESAG9 transcripts or proteins expressed within each strain but this has not been investigated.

It may be more pertinent to compare ESAG9 to ESAG5 than to the other ESAGs that have been characterised. There is not as yet any published experimental data regarding the function of ESAG5. However, bioinformatic analyses have predicted ESAG5 proteins to be N-glycosylated, and either secreted (due to the presence of a signal peptide), or membrane-bound (Barker *et al*., 2008). ESAG5 could therefore represent another family of glycosylated, secreted proteins, alongside ESAG9. ESAG5 and GRESAG5 (Gene Related to ESAG5) bear some resemblance to the bactericidal/permeability increasing (BPI) protein family, and it was hypothesised that ESAG5 may function as a receptor (Barker *et al*., 2008). There is no sequence similarity between ESAG5 and ESAG9 genes (data not shown). ESAG5 genes are more canonical ESAGs than ESAG9 in that they are found in 13 out of the 14 sequenced expression sites in *T. b. brucei* strain Lister 427 (Hertz-Fowler *et al*., 2008). ESAG5 genes are also found in *Trypanosoma vivax*, *T. congolense*, and there is on GRESAG5 gene in both *T. cruzi* and *Leishmania ssp*. Therefore they also appear to be more conserved in evolution than ESAG9 genes, which are apparently confined to *T. brucei ssp* and *T. equiperdum*.

It can be summarised from comparing ESAG9 with the other ESAGs that, although ESAG9 is not a canonical ESAG, the characterisation of ESAG9 as a potential immunomodulator is new and additional evidence that ESAG gene families are important in host-parasite interactions.

7.9.2 Other specialisations of stumpy form parasites

ESAG9 transcripts are enriched in stumpy forms. If ESAG9 proteins perform a function in immunomodulation, and therefore enhance parasite survival, it is surprising perhaps that there is not selective pressure for slender cells to express these proteins. However stumpy cells have other specialisations to help them avoid the mammalian immune response. Stumpy cells are more resistant to antibody-mediumted lysis, and can clear VSG-antibody complexes from their surface almost twice as fast as slender forms (Engstler *et al.,* 2007), as discussed in Chapter 1 section 1.2.2.3. Stumpy cells are cell cycle arrested and are not undergoing antigenic variation and any mechanisms that prolong their survival will increase the chance of transmission to the tsetse fly. Some of these mechanisms may be unique to stumpy cells because if they were expressed throughout the mammal stage of the life cycle they would make the disease more virulent. If this resulted in decreasing the longevity of the host it would be disadvantageous to the parasite.

7.9.3 Comparison to other secreted factors

ESAG7, which is a constituent of the transferrin receptor (Tf-R), has been identified in the conditioned medium of *T. b. brucei* bloodstream form cells by Western blotting (Salmon *et al*., 1994). However, it was not clear from this paper whether the conditioned medium was filtered prior to analysis to remove any remaining trypanosomes. Also, the Tf-R remains membrane associated in the flagellar pocket via ESAG6, which is GPI-anchored. The ESAG7 does not have a GPI anchor, so it is probably not, in fact, being actively secreted by bloodstream form cells, rather it is shed from the membrane if it comes loose from its heterodimeric partner, ESAG6.

Other factors secreted by *T. brucei* and other trypanosomatid parasites have been identified. A molecule called TLTF was identified as an immunomodulator secreted by *T. brucei* (Olsson *et al*., 1991) but this work has been called into question by other

researchers in the field (refer to Chapter 1 section 1.5.2.3). The conditioned medium of *T. brucei* parasites has been shown to have immunomodulatory properties (Holzmuller *et al.,* 2008) but the secretome of *T. brucei* has not, to date, been analysed in detail by proteomics so no other specific secreted molecules have been identified.

Trypanosoma cruzi parasites have been shown to shed SAP proteins, and these have been implicated in the process of host cell invasion by this parasite (Baida *et al*., 2006). Given that *T. brucei* parasites remain extracellular throughout their life cycle and do not need to invade host cells, it is unlikely that there are any functional similarities between ESAG9s and SAPs, although ESAG9s did select other members of the *T. cruzi* mucin-like glycoprotein superfamily in BLAST analysis, as described previously.

The secretome of the kinetoplastid parasite *Leishmania donovani* has been analysed in more detail. This study revealed that a large number of molecules are secreted by the promastigote stage of the parasites life cycle. Promastigotes are found in the midgut and proboscis of the insect vector of Leishmaniasis (the sand fly) and are the infective stage that are taken up into the bloodmeal and phagocytosed by macrophages. Amongst other molecules, an ortholog of the mammalian Macrophage Migration Inhibitory Factor (MIF) was secreted by these parasites (Silverman *et al*., 2008). In mammals, MIF is secreted primarily by activated macrophages (Calandra *et al.,* 1994) and is involved in T cell activation and in reducing the rate of macrophage apoptosis. There are in fact two MIF proteins in *Leishmania major*, LmjMIF1 and LmjMIF2, with LmjMIF1 only being found in amastigotes which is the intracellular stage (Richardson *et al.,* 2009). These proteins have some structural similarities to the mammalian MIF (Richardson *et al.,* 2009). No experimental evidence has been published to date to show a biological function for MIF in Leishmaniasis however this is no doubt an area of active investigation. There was no sequence similarity between MIF and the ESAG9 protein sequences (data not shown), so it is unlikely that ESAG9 proteins would be performing a similar function.

Although MIF and ESAG9-K69 are unlikely to perform the same function, the fact that *L. donovani* secretes an ortholog of a characterised immunomodulatory molecule, and that *T. brucei* conditioned medium has been shown to have immunomodulatory properties, does set a precedent for secretion of immunodulators by kinetoplastid parasites.

7.9.4 Inhibition of lymphocyte proliferation during trypanosomiasis

Published data indicate that an inhibition of lymphocyte proliferation occurs as a result of trypanosomiasis in mice. Macrophages were transferred from *T. brucei* infected mice to a new host and this resulted in an inhibition of proliferation of lymphocytes when stimulated with the mitogen Con-A (Mabbott *et al.,* 1995). The same experiment was carried out but using mice in which NO synthesis had been temporarily inhibited by administering L-NAME in the mouse drinking water, and this resulted in a partial recovery of the proliferation of lymphocytes (Mabbott *et al.*, 1995). In *in vitro* assays, spleen cells were incubated with soluble trypanosome lysate, or live *T. brucei* parasites, and an inhibition of the proliferation of lymphocytes was reported (Sternberg & Mabbott, 1996). Moreover, the inclusion of a nitric-oxide synthase inhibitor decreased the degree of inhibition. This experiment was not carried out with trypanosome-conditioned medium however, which would have been interesting. The results were suggestive of nitric oxide (NO) playing a role in this inhibition. Similar results were seen by Beschin *et al*., whereby an inhibition of T cell proliferation early on in murine trypanosomiasis was seen, and this process was partially NO-dependent (Beschin *et al*., 1998). Mutant mice have been generated by Wei et al. that are deficient in inducible Nitric Oxide Synthase (iNOS) (Wei *et al.,* 1995). The proliferation of CD4+ lymphocytes during trypanosomiasis was compared in iNOS-deficient mice and control mice and found to be inhibited in the control mice compared to the iNOS-deficient mice (Millar *et al.,* 1999).

Therefore there is a precedent in the literature for trypanosomiasis to cause an inhibition of CD4+ T cell proliferation early on in the course of the disease. This process seems to be partially dependent on the presence of NO. Secreted ESAG9-

K69 protein could be a candidate mediator of this process, and could be exerting an effect via the NO pathway. A number of experiments would have to be carried out to validate this hypothesis, including confirming that ESAG9-K69 results in an inhibition of proliferation of CD4+ T cells as well as a decrease in the percentage of CD4+ T cells.

7.9.5 Comparison to mucin and MASP proteins

Bioinformatic analysis revealed that ESAG9 proteins had some sequence similarity to the mucin-associated surface proteins (MASPs) of *T. cruzi*. Mucins were originally identified in vertebrate epithelial tissues. Genes encoding mucin-like glycoproteins, termed TcMUCs were then also discovered *Trypanosoma cruzi* (Di Noia *et al.,* 1995; Reyes *et al.,* 1994), which is the causative agent of Chagas disease. The mucin-like glycoprotein superfamily encompasses a large number of genes, which total 6% of the *T. cruzi* genome (Buscaglia *et al*., 2006). Within this are the mucin-associated surface proteins (MASPs), which are small GPI-anchored proteins with sequence similarity to TcMUCs (Atwood *et al.,* 2005; El-Sayed *et al.,* 2005), and the Serin-, Alanine-, and Proline-Rich (SAP) proteins which also have some sequence similarity to mucins (Baida *et al*., 2006). MASPs are glycosylated by N-linked glycosylation (Atwood *et al*., 2006), as are some ESAG9 proteins. MASPs are also assigned as being membrane-associated or extracellular (Atwood *et al*., 2006) and are expressed by trypomastigotes, which is the extracellular stage of the *T. cruzi* life cycle whilst in the mammal host. Therefore the similarities between MASPs and ESAG9s goes beyond the peptide sequences, but extends to their expression profile and post-translational modifications. When reciprocal BLAST analysis was carried out (refer to Chapter 3 section 3.2.5) it was in fact ESAG-EQ which was the highest scoring hit out of any of the ESAG sequences, which is surprising perhaps given that this particular ESAG9 does not appear to be modified by N-glycosylation.

Unfortunately, a biological function has not been proposed for MASPs, so it is not possible to contrast their function with the putative function of ESAG9s as

immunomodulators. However it will be interesting to see if future research in this area reveals any similarity of function between these gene families.

7.10 Future Directions

There are some questions that still remain from the work that was carried out during this PhD project. There are also avenues that would be very interesting to explore in future experiments.

The genomic context of ESAG9 genes has only been partially determined by sequencing efforts so far. The *T. b. brucei* genome was sequenced using strain 927 (Berriman *et al*., 2005) and the telomeric expression sites are not represented in this database. The telomeres of strain Lister 427 have subsequently been sequenced, and an ESAG9 was found in only one of the characterised bloodstream form expression sites (Hertz-Fowler *et al*., 2008). It would be very interesting to determine whether the three ESAG9 genes most extensively explored in this project (termed ESAG9-K9, ESAG9-K69 and ESAG9-EQ) are within expression sites in the pleiomorphic *T. b. brucei* EATRO 2340 strain in which they were first identified by Keith Matthews. RNA run-on analysis based on the sensitivity of gene expression to α-amanitin would answer this question. It is unlikely that all three genes are in expression sites however because they are expressed simultaneously, and therefore would have to be in the same expression site.

This project did not explore the control of ESAG9 gene expression. Bioinformatic analyses using a tool called MEME indicate that there are at least two motifs common to the 3' UTRs of ESAG9 genes; these could be involved in controlling gene expression. It would be possible to create transgenic cell lines in which certain ESAG9 gene(s) were lacking these motifs and use these to determine if this altered gene expression patterns.

Finally, and perhaps most importantly, the phenotype of immunomodulation resulting from ectopic expression of ESAG9 proteins could be explored in greater detail. Initially these experiments need to be repeated with further controls, as detailed in Section 7.7.2.

Bibliography

Abubakar LU, Bulimo WD, Mulaa FJ, Osir EO (2006) Molecular characterization of a tsetse fly midgut proteolytic lectin that mediumtes differentiation of African trypanosomes. *Insect Biochem Mol Biol* **36**(4)**:** 344-352

Acosta-Serrano A, Cole RN, Englund PT (2000) Killing of Trypanosoma brucei by concanavalin A: structural basis of resistance in glycosylation mutants. *J Mol Biol* **304**(4)**:** 633-644

Acosta-Serrano A, Vassella E, Liniger M, Kunz Renggli C, Brun R, Roditi I, Englund PT (2001) The surface coat of procyclic Trypanosoma brucei: programmed expression and proteolytic cleavage of procyclin in the tsetse fly. *Proc Natl Acad Sci U S A* **98**(4)**:** 1513-1518

Aksoy S, Gibson WC, Lehane MJ (2003) Interactions between tsetse and trypanosomes with implications for the control of trypanosomiasis. *Adv Parasitol* **53:** 1-83

Alarcon CM, Son HJ, Hall T, Donelson JE (1994) A monocistronic transcript for a trypanosome variant surface glycoprotein. *Mol Cell Biol* **14**(8)**:** 5579-5591

Alexandre S, Paindavoine P, Tebabi P, Pays A, Halleux S, Steinert M, Pays E (1990) Differential expression of a family of putative adenylate/guanylate cyclase genes in Trypanosoma brucei. *Mol Biochem Parasitol* **43**(2)**:** 279-288

Almeida IC, Camargo MM, Procopio DO, Silva LS, Mehlert A, Travassos LR, Gazzinelli RT, Ferguson MA (2000) Highly purified glycosylphosphatidylinositols from Trypanosoma cruzi are potent proinflammatory agents. *EMBO J* **19**(7)**:** 1476-1485

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* **215**(3)**:** 403-410

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**(17)**:** 3389-3402

Angata T, Hingroani R, Varki NM, Varki A (2001) Cloning and characterization of a novel mouse Siglec, mSiglec-F: differential evolution of the mouse and human (CD33) Siglec-3-related gene clusters. *J Biol Chem* **276**(48)**:** 45128-45136

Asahida H, Kobayashi T, Saitoh K, Nakayama I (1996) Tissue preservation and total DNA extraction from fish stored at ambient temperature using buffers containing high concentration of urea. *Fisheries Science* **62**(5)**:** 727-730

Atwood JA, 3rd, Minning T, Ludolf F, Nuccio A, Weatherly DB, Alvarez-Manilla G, Tarleton R, Orlando R (2006) Glycoproteomics of Trypanosoma cruzi trypomastigotes using subcellular fractionation, lectin affinity, and stable isotope labeling. *J Proteome Res* **5**(12)**:** 3376-3384

Atwood JA, 3rd, Weatherly DB, Minning TA, Bundy B, Cavola C, Opperdoes FR, Orlando R, Tarleton RL (2005) The Trypanosoma cruzi proteome. *Science* **309**(5733)**:** 473-476

Augustijn KD, Kleemann R, Thompson J, Kooistra T, Crawford CE, Reece SE, Pain A, Siebum AH, Janse CJ, Waters AP (2007) Functional characterization of the Plasmodium falciparum and P. berghei homologues of macrophage migration inhibitory factor. *Infect Immun* **75**(3)**:** 1116-1128

Baetselier PD, Namangala B, Noel W, Brys L, Pays E, Beschin A (2001) Alternative versus classical macrophage activation during experimental African trypanosomosis. *Int J Parasitol* **31**(5-6)**:** 575-587

Baida RC, Santos MR, Carmo MS, Yoshida N, Ferreira D, Ferreira AT, El Sayed NM, Andersson B, da Silveira JF (2006) Molecular characterization of serine-, alanine-, and proline-rich proteins of Trypanosoma cruzi and their possible role in host cell infection. *Infect Immun* **74**(3)**:** 1537-1546

Bailey TL, Elkan C (1995) The value of prior knowledge in discovering motifs with MEME. *Proc Int Conf Intell Syst Mol Biol* **3:** 21-29

Bangs JD, Ransom DM, McDowell MA, Brouch EM (1997) Expression of bloodstream variant surface glycoproteins in procyclic stage Trypanosoma brucei: role of GPI anchors in secretion. *EMBO J* **16**(14)**:** 4285-4294

Bangs JD, Uyetake L, Brickman MJ, Balber AE, Boothroyd JC (1993) Molecular cloning and cellular localization of a BiP homologue in Trypanosoma brucei. Divergent ER retention signals in a lower eukaryote. *J Cell Sci* **105 (Pt 4):** 1101-1113

Barker AR, Wickstead B, Gluenz E, Gull K (2008) Bioinformatic insights to the ESAG5 and GRESAG5 gene families in kinetoplastid parasites. *Mol Biochem Parasitol* **162**(2)**:** 112-122

Bastin P, Bagherzadeh Z, Matthews KR, Gull K (1996) A novel epitope tag system to study protein targeting and organelle biogenesis in Trypanosoma brucei. *Mol Biochem Parasitol* **77**(2)**:** 235-239

Becker M, Aitcheson N, Byles E, Wickstead B, Louis E, Rudenko G (2004) Isolation of the repertoire of VSG expression site containing telomeres of Trypanosoma brucei 427 using transformation-associated recombination in yeast. *Genome Res* **14**(11)**:** 2319-2329

Bell AS, Ranford-Cartwright LC (2004) A real-time PCR assay for quantifying Plasmodium falciparum infections in the mosquito vector. *Int J Parasitol* **34**(7)**:** 795-802

Bendelac A, Lantz O, Quimby ME, Yewdell JW, Bennink JR, Brutkiewicz RR (1995) CD1 recognition by mouse NK1+ T lymphocytes. *Science* **268**(5212)**:** 863-865

Bendtsen JD, Nielsen H, von Heijne G, Brunak S (2004) Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* **340**(4)**:** 783-795

Benz C, Nilsson D, Andersson B, Clayton C, Guilbride DL (2005) Messenger RNA processing sites in Trypanosoma brucei. *Mol Biochem Parasitol* **143**(2)**:** 125-134

Berriman M, Ghedin E, Hertz-Fowler C, Blandin G, Renauld H, Bartholomeu DC, Lennard NJ, Caler E, Hamlin NE, Haas B, Bohme U, Hannick L, Aslett MA, Shallom J, Marcello L, Hou L, Wickstead B, Alsmark UC, Arrowsmith C, Atkin RJ, Barron AJ, Bringaud F, Brooks K, Carrington M, Cherevach I, Chillingworth TJ, Churcher C, Clark LN, Corton CH, Cronin A, Davies RM, Doggett J, Djikeng A, Feldblyum T, Field MC, Fraser A, Goodhead I, Hance Z, Harper D, Harris BR, Hauser H, Hostetler J, Ivens A, Jagels K, Johnson D, Johnson J, Jones K, Kerhornou AX, Koo H, Larke N, Landfear S, Larkin C, Leech V, Line A, Lord A, Macleod A, Mooney PJ, Moule S, Martin DM, Morgan GW, Mungall K, Norbertczak H, Ormond D, Pai G, Peacock CS, Peterson J, Quail MA, Rabbinowitsch E, Rajandream MA, Reitter C, Salzberg SL, Sanders M, Schobel S, Sharp S, Simmonds M, Simpson AJ, Tallon L, Turner CM, Tait A, Tivey AR, Van Aken S, Walker D, Wanless D, Wang S, White B, White O, Whitehead S, Woodward J, Wortman J, Adams MD, Embley TM, Gull K, Ullu E, Barry JD, Fairlamb AH, Opperdoes F, Barrell BG, Donelson JE, Hall N, Fraser CM, Melville SE, El-Sayed NM (2005) The genome of the African trypanosome Trypanosoma brucei. *Science* **309**(5733)**:** 416-422

Berriman M, Hall N, Sheader K, Bringaud F, Tiwari B, Isobe T, Bowman S, Corton C, Clark L, Cross GA, Hoek M, Zanders T, Berberof M, Borst P, Rudenko G (2002) The architecture of variant surface glycoprotein gene expression sites in Trypanosoma brucei. *Mol Biochem Parasitol* **122**(2)**:** 131-140

Beschin A, Brys L, Magez S, Radwanska M, De Baetselier P (1998) Trypanosoma brucei infection elicits nitric oxide-dependent and nitric oxide-independent suppressive mechanisms. *J Leukoc Biol* **63**(4)**:** 429-439

Biebinger S, Wirtz LE, Lorenz P, Clayton C (1997) Vectors for inducible expression of toxic gene products in bloodstream and procyclic Trypanosoma brucei. *Mol Biochem Parasitol* **85**(1)**:** 99-112

Birnboim HC, Doly J (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res* **7**(6)**:** 1513-1523

Bitter W, Gerrits H, Kieft R, Borst P (1998) The role of transferrin-receptor variation in the host range of Trypanosoma brucei. *Nature* **391**(6666)**:** 499-502

Bogdan C (2001) Nitric oxide and the immune response. *Nat Immunol* **2**(10)**:** 907-916

Borden KL, Freemont PS (1996) The RING finger domain: a recent example of a sequence-structure family. *Curr Opin Struct Biol* **6**(3)**:** 395-401

Borst P, Bitter W, Blundell PA, Chaves I, Cross M, Gerrits H, van Leeuwen F, McCulloch R, Taylor M, Rudenko G (1998) Control of VSG gene expression sites in Trypanosoma brucei. *Mol Biochem Parasitol* **91**(1)**:** 67-76

Borst P, Ulbert S (2001) Control of VSG gene expression sites. *Mol Biochem Parasitol* **114**(1)**:** 17-27

Brookman JL, Stott AJ, Cheeseman PJ, Burns NR, Adams SE, Kingsman AJ, Gull K (1995) An immunological analysis of Ty1 virus-like particle structure. *Virology* **207**(1)**:** 59-67

Bruce D, A. E. Hamerton, H. R. Bateman, F. P. Mackie (1909) The Development of Trypanosoma gambiense in Glossina palpalis. *Proceedings of the Royal Society of London Series B, Containing Papers of a Biological Character* **81**(505)**:** 405-414

Bruce D, Nabarro G (1903) The Etiology of Sleeping Sickness. *British Medical Journal* **2:** 1343-1352

Brun R, Schonenberger M (1979) Cultivation and in vitro cloning of procyclic culture forms of Trypanosoma brucei in a semi-defined medium. *Acta Trop* **36:** 289-292

Brun R, Schonenberger M (1981) Stimulating effect of citrate and cis-Aconitate on the transformation of Trypanosoma brucei bloodstream forms to procyclic forms in vitro. *Z Parasitenkd* **66**(1)**:** 17-24

Buffone GJ, Darlington GJ (1985) Isolation of DNA from biological specimens without extraction with phenol. *Clin Chem* **31**(1)**:** 164-165

Bulow R, Griffiths G, Webster P, Stierhof YD, Opperdoes FR, Overath P (1989) Intracellular-Localization of the Glycosyl-Phosphatidylinositol-Specific Phospholipase-C of Trypanosoma-Brucei. *Journal of Cell Science* **93:** 233-240

Buscaglia CA, Campo VA, Frasch AC, Di Noia JM (2006) Trypanosoma cruzi surface mucins: host-dependent coat diversity. *Nat Rev Microbiol* **4**(3)**:** 229-236

Buxbaum LU, Raper J, Opperdoes FR, Englund PT (1994) Myristate exchange. A second glycosyl phosphatidylinositol myristoylation reaction in African trypanosomes. *J Biol Chem* **269**(48)**:** 30212-30220

Calandra T, Bernhagen J, Michell RA, Bucala R (1994) The macrophage is an important and previously unrecognized source of macrophage migration inhibitory factor. *J Exp Med* **179**(6)**:** 1895-1902

Caljon G, Van Den Abbeele J, Stijlemans B, Coosemans M, De Baetselier P, Magez S (2006) Tsetse fly saliva accelerates the onset of Trypanosoma brucei infection in a mouse model associated with a reduced host inflammatory response. *Infect Immun* **74**(11)**:** 6324-6330

Camargo MM, Almeida IC, Pereira ME, Ferguson MA, Travassos LR, Gazzinelli RT (1997a) Glycosylphosphatidylinositol-anchored mucin-like glycoproteins isolated from Trypanosoma cruzi trypomastigotes initiate the synthesis of proinflammatory cytokines by macrophages. *J Immunol* **158**(12)**:** 5890-5901

Camargo MM, Andrade AC, Almeida IC, Travassos LR, Gazzinelli RT (1997b) Glycoconjugates isolated from Trypanosoma cruzi but not from Leishmania species membranes trigger nitric oxide synthesis as well as microbicidal activity in IFN-gamma-primed macrophages. *J Immunol* **159**(12)**:** 6131-6139

Campillo N, Carrington M (2003) The origin of the serum resistance associated (SRA) gene and a model of the structure of the SRA polypeptide from Trypanosoma brucei rhodesiense. *Mol Biochem Parasitol* **127:** 79-84

Cardoso de Almeida ML, Turner MJ (1983) The membrane form of variant surface glycoproteins of Trypanosoma brucei *Nature* **302:** 349-352

Carruthers VB, Navarro M, Cross GA (1996) Targeted disruption of expression site-associated gene-1 in bloodstream-form Trypanosoma brucei. *Mol Biochem Parasitol* **81**(1)**:** 65-79

Chandra M, Liniger M, Tetley L, Roditi I, Barry JD (2004) TsetseEP, a gut protein from the tsetse Glossina morsitans, is related to a major surface glycoprotein of trypanosomes transmitted by the fly and to the products of a Drosophila gene family. *Insect Biochem Mol Biol* **34**(11)**:** 1163-1173

Coller SP, Mansfield JM, Paulnock DM (2003) Glycosylinositolphosphate soluble variant surface glycoprotein inhibits IFN-gamma-induced nitric oxide production via reduction in STAT1 phosphorylation in African trypanosomiasis. *J Immunol* **171**(3)**:** 1466-1472

Cross GA (1975) Identification, purification and properties of clone-specific glycoprotein antigens constituting the surface coat of Trypanosoma brucei. *Parasitology* **71**(3)**:** 393-417

Cross GA (1978) Antigenic variation in trypanosomes. *Proc R Soc Lond B Biol Sci* **202**(1146)**:** 55-72

Cross GA, Manning JC (1973) Cultivation of Trypanosoma brucei spp. in semi-defined and defined medium. *Parasitology* **67:** 315-331

Cully DF, Gibbs CP, Cross GA (1986) Identification of proteins encoded by variant surface glycoprotein expression site-associated genes in Trypanosoma brucei. *Mol Biochem Parasitol* **21**(2)**:** 189-197

Cully DF, Ip HS, Cross GA (1985a) Coordinate transcription of variant surface glycoprotein genes and an expression site associated gene family in Trypanosoma brucei. *Cell* **42**(1)**:** 173-182

Cully DF, Ip HS, Cross GaM (1985b) Coordinate Transcription of Variant Surface Glycoprotein Genes and an Expression Site Associated Gene Family in Trypanosoma Brucei. *Cell* **42**(1)**:** 173-182

Cunningham MP, Vickerman K (1962) Antigenic analysis in the Trypanosoma brucei gropu. using the agglutination reaction. *Trans R Soc Trop Med Hyg* **56**(1)**:** 48-59

Czichos J, Nonnengaesser C, Overath P (1986) Trypanosoma brucei: cis-aconitate and temperature reduction as triggers of synchronous transformation of bloodstream to procyclic trypomastigotes in vitro. *Exp Parasitol* **62**(2)**:** 283-291

Dale C, Welburn SC (2001) The endosymbionts of tsetse flies: manipulating host-parasite interactions. *Int J Parasitol* **31**(5-6)**:** 628-631

Das AK, Helps NR, Cohen PT, Barford D (1996) Crystal structure of the protein serine/threonine phosphatase 2C at 2.0 A resolution. *EMBO J* **15**(24)**:** 6798-6809

Daulouede S, Bouteille B, Moynet D, De Baetselier P, Courtois P, Lemesre JL, Buguet A, Cespuglio R, Vincendeau P (2001) Human macrophage tumor necrosis factor (TNF)-alpha production induced by Trypanosoma brucei gambiense and the role of TNF-alpha in parasite control. *J Infect Dis* **183**(6)**:** 988-991

De Greef C, Hamers R (1994) The serum resistance-associated (SRA) gene of Trypanosoma brucei rhodesiense encodes a variant surface glycoprotein-like protein. *Mol Biochem Parasitol* **68**(2)**:** 277-284

De Greef C, Imberechts H, Matthyssens G, Van Meirvenne N, Hamers R (1989) A gene expressed only in serum-resistant variants of Trypanosoma brucei rhodesiense. *Mol Biochem Parasitol* **36**(2)**:** 169-176

de Koning HP (2008) Ever-increasing complexities of diamidine and arsenical crossresistance in African trypanosomes. *Trends Parasitol* **24**(8)**:** 345-349

De Lange T, Borst P (1982) Genomic environment of the expression-linked extra copies of genes for surface antigens of Trypanosoma brucei resembles the end of a chromosome. *Nature* **299:** 451-453

Dean S, Marchetti R, Kirk K, Matthews K (2009) A surface transporter family conveys the differentiation signal in African trypanosomes. *In press*

Di Noia JM, Sanchez DO, Frasch AC (1995) The protozoan Trypanosoma cruzi has a family of genes resembling the mucin genes of mammalian cells. *J Biol Chem* **270**(41)**:** 24146-24149

Domenicali Pfister D, Burkard G, Morand S, Renggli CK, Roditi I, Vassella E (2006) A Mitogen-activated protein kinase controls differentiation of bloodstream forms of Trypanosoma brucei. *Eukaryot Cell* **5**(7)**:** 1126-1135

Eisenhaber B, Bork P, Eisenhaber F (1999) Prediction of potential GPI-modification sites in proprotein sequences. *J Mol Biol* **292**(3)**:** 741-758

El-Sayed NM, Myler PJ, Bartholomeu DC, Nilsson D, Aggarwal G, Tran AN, Ghedin E, Worthey EA, Delcher AL, Blandin G, Westenberger SJ, Caler E, Cerqueira GC, Branche C, Haas B, Anupama A, Arner E, Aslund L, Attipoe P, Bontempi E, Bringaud F, Burton P, Cadag E, Campbell DA, Carrington M, Crabtree J, Darban H, da Silveira JF, de Jong P, Edwards K, Englund PT, Fazelina G, Feldblyum T, Ferella M, Frasch AC, Gull K, Horn D, Hou L, Huang Y, Kindlund E, Klingbeil M, Kluge S, Koo H, Lacerda D, Levin MJ, Lorenzi H, Louie T, Machado CR, McCulloch R, McKenna A, Mizuno Y, Mottram JC, Nelson S, Ochaya S, Osoegawa K, Pai G, Parsons M, Pentony M, Pettersson U, Pop M, Ramirez JL, Rinta J, Robertson L, Salzberg SL, Sanchez DO, Seyler A, Sharma R, Shetty J, Simpson AJ, Sisk E, Tammi MT, Tarleton R, Teixeira S, Van Aken S, Vogt C, Ward PN, Wickstead B, Wortman J, White O, Fraser CM, Stuart KD, Andersson B (2005) The genome sequence of Trypanosoma cruzi, etiologic agent of Chagas disease. *Science* **309**(5733)**:** 409-415

Ellis DS, Maudlin I (1985) The behaviour of trypanosomes within the midguts of wild-caught tsetse from Zimbabwe. *Trans R Soc Trop Med Hyg* **79**(6)**:** 867-868

Engstler M, Boshart M (2004) Cold shock and regulation of surface protein trafficking convey sensitization to inducers of stage differentiation in Trypanosoma brucei. *Genes Dev* **18**(22)**:** 2798-2811

Engstler M, Pfohl T, Herminghaus S, Boshart M, Wiegertjes G, Heddergott N, Overath P (2007) Hydrodynamic flow-mediumted protein sorting on the cell surface of trypanosomes. *Cell* **131**(3)**:** 505-515

Fang J, McCutchan TF (2002) Thermoregulation in a parasite's life cycle. *Nature* **418**(6899)**:** 742

Fenn K, Matthews KR (2007) The cell biology of Trypanosoma brucei differentiation. *Curr Opin Microbiol* **10**(6)**:** 539-546

Ferguson MA (1999) The structure, biosynthesis and functions of glycosylphosphatidylinositol anchors, and the contributions of trypanosome research. *J Cell Sci* **112 (Pt 17):** 2799-2809

Ferguson MA, Haldar K, Cross GA (1985a) Trypanosoma brucei variant surface glycoprotein has a sn-1,2-dimyristyl glycerol membrane anchor at its COOH terminus. *J Biol Chem* **260**(8)**:** 4963-4968

Ferguson MA, Low MG, Cross GA (1985b) Glycosyl-sn-1,2-dimyristylphosphatidylinositol is covalently linked to Trypanosoma brucei variant surface glycoprotein. *J Biol Chem* **260**(27)**:** 14547-14555

Field MC, Allen CL, Dhir V, Goulding D, Hall BS, Morgan GW, Veazey P, Engstler M (2004) New approaches to the microscopic imaging of Trypanosoma brucei. *Microsc Microanal* **10**(5)**:** 621-636

Finn RD, Tate J, Mistry J, Coggill PC, Sammut SJ, Hotz HR, Ceric G, Forslund K, Eddy SR, Sonnhammer EL, Bateman A (2008) The Pfam protein families database. *Nucleic Acids Res* **36**(Database issue)**:** D281-288

Florent IC, Raibaud A, Eisen H (1991) A family of genes related to a new expression site-associated gene in Trypanosoma equiperdum. *Mol Cell Biol* **11**(4)**:** 2180-2188

Gerrits H, Mussmann R, Bitter W, Kieft R, Borst P (2002) The physiological significance of transferrin receptor variations in Trypanosoma brucei. *Mol Biochem Parasitol* **119**(2)**:** 237-247

Gibson WC (1986) Will the real Trypanosoma b. gambiense please stand up. *Parasitol Today* **2**(9)**:** 255-257

Gibson W, Peacock L, Ferris V, Williams K, Bailey M (2008) The use of yellow fluorescent hybrids to indicate mating in Trypanosoma brucei. *Parasit Vectors* **1**(1)**:** 4

Gibson W, Stevens J (1999) Genetic exchange in the trypanosomatidae. *Adv Parasitol* **43:** 1-46

Ginger ML, Blundell PA, Lewis AM, Browitt A, Gunzl A, Barry JD (2002) Ex vivo and in vitro identification of a consensus promoter for VSG genes expressed by metacyclic-stage trypanosomes in the tsetse fly. *Eukaryot Cell* **1**(6)**:** 1000-1009

Gottesdiener KM (1994) A new VSG expression site-associated gene (ESAG) in the promoter region of Trypanosoma brucei encodes a protein with 10 potential transmembrane domains. *Mol Biochem Parasitol* **63**(1)**:** 143-151

Graham SV, Barry JD (1991) Expression Site-Associated Genes Transcribed Independently of Variant Surface Glycoprotein Genes in Trypanosoma-Brucei. *Molecular and Biochemical Parasitology* **47**(1)**:** 31-41

Graham SV, Barry JD (1995) Transcriptional regulation of metacyclic variant surface glycoprotein gene expression during the life cycle of Trypanosoma brucei. *Mol Cell Biol* **15**(11)**:** 5945-5956

Graham SV, Terry S, Barry JD (1999) A structural and transcription pattern for variant surface glycoprotein gene expression sites used in metacyclic stage Trypanosoma brucei. *Mol Biochem Parasitol* **103**(2)**:** 141-154

Grandgenett PM, Otsu K, Wilson HR, Wilson ME, Donelson JE (2007) A function for a specific zinc metalloprotease of African trypanosomes. *PLoS Pathog* **3**(10)**:** 1432-1445

Green LC, Wagner DA, Glogowski J, Skipper PL, Wishnok JS, Tannenbaum SR (1982) Analysis of nitrate, nitrite, and [15N]nitrate in biological fluids. *Analytical Biochem* **126**(1)**:** 131-138

Greenwood BM, Whittle HC, Molyneux DH (1973) Immunosuppression in Gambian trypanosomiasis. *Trans R Soc Trop Med Hyg* **67**(6)**:** 846-850

Gruszynski AE, van Deursen FJ, Albareda MC, Best A, Chaudhary K, Cliffe LJ, del Rio L, Dunn JD, Ellis L, Evans KJ, Figueiredo JM, Malmquist NA, Omosun Y, Palenchar JB, Prickett S, Punkosdy GA, van Dooren G, Wang Q, Menon AK, Matthews KR, Bangs JD (2006) Regulation of surface coat exchange by differentiating African trypanosomes. *Mol Biochem Parasitol* **147**(2)**:** 211-223

Gull K (2002) The cell biology of parasitism in Trypanosoma brucei: insights and drug targets from genomic approaches? *Curr Pharm Des* **8**(4)**:** 241-256

Gunzl A, Bruderer T, Laufer G, Schimanski B, Tu LC, Chung HM, Lee PT, Lee MG (2003) RNA polymerase I transcribes procyclin genes and variant surface glycoprotein gene expression sites in Trypanosoma brucei. *Eukaryot Cell* **2**(3)**:** 542-551

Guther ML, Lee S, Tetley L, Acosta-Serrano A, Ferguson MA (2006) GPI-anchored proteins and free GPI glycolipids of procyclic form Trypanosoma brucei are nonessential for growth, are required for colonization of the tsetse fly, and are not the only components of the surface coat. *Mol Biol Cell* **17**(12)**:** 5265-5274

Hajduk SL, Moore DR, Vasudevacharya J, Siqueira H, Torri AF, Tytler EM, Esko JD (1989) Lysis of Trypanosoma brucei by a toxic subspecies of human high density lipoprotein. *J Biol Chem* **264**(9)**:** 5210-5217

Hao Z, Kasumba I, Aksoy S (2003) Proventriculus (cardia) plays a crucial role in immunity in tsetse fly (Diptera: Glossinidiae). *Insect Biochem Mol Biol* **33**(11)**:** 1155-1164

Hao Z, Kasumba I, Lehane MJ, Gibson WC, Kwon J, Aksoy S (2001) Tsetse immune responses and trypanosome transmission: implications for the development of tsetse-based strategies to reduce trypanosomiasis. *Proc Natl Acad Sci U S A* **98**(22)**:** 12648-12653

Hedengren M, Asling B, Dushay MS, Ando I, Ekengren S, Wihlborg M, Hultmark D (1999) Relish, a central factor in the control of humoral but not cellular immunity in Drosophila. *Mol Cell* **4**(5)**:** 827-837

Herbert WJ, Lumsden WHR (1976) Trypanosoma brucei: A Rapid "Matching" Method for Estimating the Host's Parsitemia. *Exp Parasitol* **40:** 427-431

Hertz-Fowler C, Figueiredo LM, Quail MA, 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 JE, Walker D, White B, Young R, Cross GA, Rudenko G, Barry JD, Louis EJ, Berriman M (2008) Telomeric expression sites are highly conserved in Trypanosoma brucei. *PLoS ONE* **3**(10)**:** e3527

Hestdal K, Ruscetti FW, Ihle JN, Jacobsen SE, Dubois CM, Kopp WC, Longo DL, Keller JR (1991) Characterisation and regulation of RB6-8C5 antigen expression on murine bone marrow cells. *J Immunol*

147: 22-28

Hill KL, Hutchings NR, Grandgenett PM, Donelson JE (2000) T lymphocyte-triggering factor of african trypanosomes is associated with the flagellar fraction of the cytoskeleton and represents a new family of proteins that are present in several divergent eukaryotes. *J Biol Chem* **275**(50)**:** 39369-39378

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* **75**(6)**:** 985-989

Hoare CA (1972) The Trypanosomes of Mammals. *Blackwell Publishing, Oxford*

Hoek M, Zanders T, Cross GA (2002) Trypanosoma brucei expression-site-associated-gene-8 protein interacts with a Pumilio family protein. *Mol Biochem Parasitol* **120:** 269-283

Holzmuller P, Biron DG, Courtois P, Koffi M, Bras-Goncalves R, Daulouede S, Solano P, Cuny G, Vincendeau P, Jamonneau V (2008) Virulence and pathogenicity patterns of Trypanosoma brucei gambiense field isolates in experimentally infected mouse: differences in host immune response modulation by secretome and proteomics. *Microbes Infect* **10**(1)**:** 79-86

Hu C, Aksoy S (2006) Innate immune responses regulate trypanosome parasite infection of the tsetse fly Glossina morsitans morsitans. *Mol Microbiol* **60**(5)**:** 1194-1204

Hu Y, Aksoy S (2005) An antimicrobial peptide with trypanocidal activity characterized from Glossina morsitans morsitans. *Insect Biochem Mol Biol* **35**(2)**:** 105-115

Hulo N, Bairoch A, Bulliard V, Cerutti L, Cuche BA, de Castro E, Lachaize C, Langendijk-Genevaux PS, Sigrist CJ (2008) The 20 years of PROSITE. *Nucleic Acids Res* **36**(Database issue)**:** D245-249

Hunt M, Brun R, Kohler P (1994) Studies on compounds promoting the in vitro transformation of Trypanosoma brucei from bloodstream to procyclic forms. *Parasitol Res* **80**(7)**:** 600-606

Hwa KY, Acosta-Serrano A, Khoo KH, Pearson T, Englund PT (1999) Protein glycosylation mutants of procyclic Trypanosoma brucei: defects in the asparagine-glycosylation pathway. *Glycobiology* **9**(2)**:** 181-190

Ikezawa H, Yamanegi M, Taguchi R, Miyashita T, Ohyabu T (1976) Studies on phosphatidylinositol phosphodiesterase (phospholipase C type) of Bacillus cereus I. Purification, properties and phosphatase-releasing activity. *Biochim Biophys Acta* **450**(2)**:** 154-164

Ivens AC, Peacock CS, Worthey EA, Murphy L, Aggarwal G, Berriman M, Sisk E, Rajandream MA, Adlem E, Aert R, Anupama A, Apostolou Z, Attipoe P, Bason N, Bauser C, Beck A, Beverley SM, Bianchettin G, Borzym K, Bothe G, Bruschi CV, Collins M, Cadag E, Ciarloni L, Clayton C, Coulson RM, Cronin A, Cruz AK, Davies RM, De Gaudenzi J, Dobson DE, Duesterhoeft A, Fazelina G, Fosker N, Frasch AC, Fraser A, Fuchs M, Gabel C, Goble A, Goffeau A, Harris D, Hertz-Fowler C, Hilbert H, Horn D, Huang Y, Klages S, Knights A, Kube M, Larke N, Litvin L, Lord A, Louie T, Marra M, Masuy D, Matthews K, Michaeli S, Mottram JC, Muller-Auer S, Munden H, Nelson S, Norbertczak H, Oliver K, O'Neil S, Pentony M, Pohl TM, Price C, Purnelle B, Quail MA, Rabbinowitsch E, Reinhardt R, Rieger M, Rinta J, Robben J, Robertson L, Ruiz JC, Rutter S, Saunders D, Schafer M, Schein J, Schwartz DC, Seeger K, Seyler A, Sharp S, Shin H, Sivam D, Squares R, Squares S, Tosato V, Vogt C, Volckaert G, Wambutt R, Warren T, Wedler H, Woodward J, Zhou S, Zimmermann W, Smith DF, Blackwell JM, Stuart KD, Barrell B, Myler PJ (2005) The genome of the kinetoplastid parasite, Leishmania major. *Science* **309**(5733)**:** 436-442

Jacobs SL, Lee ND (1964) Determination of Citric Acid in Serum and Urine Using Br82. *J Nucl Med* **5:** 297-301

Jenni L, Marti S, Schweizer J, Betschart B, Le Page RW, Wells JM, Tait A, Paindavoine P, Pays E, Steinert M (1986) Hybrid formation between African trypanosomes during cyclical transmission. *Nature* **322**(6075)**:** 173-175

Jenni L, Molyneux DH, Livesey JL, Galun R (1980) Feeding behaviour of tsetse flies infected with salivarian trypanosomes. *Nature* **283**(5745)**:** 383-385

Johnson PJ, Kooter JM, Borst P (1987) Inactivation of transcription by UV irradiation of T. brucei provides evidence for a multicistronic transcription unit including a VSG gene. *Cell* **51**(2)**:** 273-281

Jones DC, Mehlert A, Guther ML, Ferguson MA (2005) Deletion of the glucosidase II gene in Trypanosoma brucei reveals novel N-glycosylation mechanisms in the biosynthesis of variant surface glycoprotein. *J Biol Chem* **280**(43)**:** 35929-35942

Kaaya GP, Flyg C, Boman HG (1987) Induction of cecropin and attacin-like antibacterial factors in the haemolymph of Glossina morsitans morsitans. *Insect Biochem* **17**(2)**:** 309-315

Kaplan R, Jaye M, Burgess WH, Schlaepfer DD, Haigler HT (1988) Cloning and expression of cDNA for human endonexin II, a Ca2+ and phospholipid binding protein. *J Biol Chem* **263**(17)**:** 8037-8043

Koopman G, Reutelingsperger CP, Kuijten GA, Keehnen RM, Pals ST, van Oers MH (1994) Annexin V for flow cytometric detection of phosphatidylserine expression on B cells undergoing apoptosis. *Blood* **84**(5)**:** 1415-1420

Kusuda K, Kobayashi T, Ikeda S, Ohnishi M, Chida N, Yanagawa Y, Shineha R, Nishihira T, Satomi S, Hiraga A, Tamura S (1998) Mutational analysis of the domain structure of mouse protein phosphatase 2Cbeta. *Biochem J* **332 (Pt 1):** 243-250

Laemmli UK (1970) **Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4**. *Nature* **227:** 680-685

Lai DH, Hashimi H, Lun ZR, Ayala FJ, Lukes J (2008) Adaptations of Trypanosoma brucei to gradual loss of kinetoplast DNA: Trypanosoma equiperdum and Trypanosoma evansi are petite mutants of T. brucei. *Proc Natl Acad Sci U S A* **105**(6)**:** 1999-2004

Lanham SM, Godfrey DG (1970) Isolation of salivarian trypanosomes from man and other mammals using DEAE-cellulose. *Exp Parasitol* **28:** 521-534

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* **23**(21)**:** 2947-2948

Laxman S, Riechers A, Sadilek M, Schwede F, Beavo JA (2006) Hydrolysis products of cAMP analogs cause transformation of Trypanosoma brucei from slender to stumpy-like forms. *Proc Natl Acad Sci U S A* **103**(50)**:** 19194-19199

Lee MG, Van der Ploeg LH (1997) Transcription of protein-coding genes in trypanosomes by RNA polymerase I. *Annu Rev Microbiol* **51:** 463-489

Lehane MJ, Aksoy S, Gibson W, Kerhornou A, Berriman M, Hamilton J, Soares MB, Bonaldo MF, Lehane S, Hall N (2003) Adult midgut expressed sequence tags from the tsetse fly Glossina morsitans morsitans and expression analysis of putative immune response genes. *Genome Biol* **4**(10)**:** R63

Lehane MJ, Aksoy S, Levashina E (2004) Immune responses and parasite transmission in blood-feeding insects. *Trends Parasitol* **20**(9)**:** 433-439

Lehane MJ, Gibson W, Lehane SM (2008) Differential expression of fat body genes in Glossina morsitans morsitans following infection with Trypanosoma brucei brucei. *Int J Parasitol* **38**(1)**:** 93-101

Leppert BJ, Mansfield JM, Paulnock DM (2007) The soluble variant surface glycoprotein of African trypanosomes is recognized by a macrophage scavenger receptor and induces I kappa B alpha degradation independently of TRAF6-mediumted TLR signaling. *J Immunol* **179**(1)**:** 548-556

Letunic I, Copley RR, Pils B, Pinkert S, Schultz J, Bork P (2006) SMART 5: domains in the context of genomes and networks. *Nucleic Acids Res* **34**(Database issue)**:** D257-260

Lillico S, Field MC, Blundell P, Coombs GH, Mottram JC (2003) Essential roles for GPI-anchored proteins in African trypanosomes revealed using mutants deficient in GPI8. *Mol Biol Cell* **14**(3)**:** 1182-1194

Liniger M, Acosta-Serrano A, Van Den Abbeele J, Kunz Renggli C, Brun R, Englund PT, Roditi I (2003) Cleavage of trypanosome surface glycoproteins by alkaline trypsin-like enzyme(s) in the midgut of Glossina morsitans. *Int J Parasitol* **33**(12)**:** 1319-1328

Liniger M, Urwyler S, Studer E, Oberle M, Renggli CK, Roditi I (2004) Role of the N-terminal domains of EP and GPEET procyclins in membrane targeting and the establishment of midgut infections by Trypanosoma brucei. *Mol Biochem Parasitol* **137**(2)**:** 247-251

Lopez R, Demick KP, Mansfield JM, Paulnock DM (2008) Type I IFNs play a role in early resistance, but subsequent susceptibility, to the African trypanosomes. *J Immunol* **181**(7)**:** 4908-4917

Low MG, Finean JB (1977) Release of alkaline phosphatase from membranes by a phosphatidylinositol-specific phospholipase C. *Biochem J* **167**(1)**:** 281-284

Lucas R, Magez S, De Leys R, Fransen L, Scheerlinck JP, Rampelberg M, Sablon E, De Baetselier P (1994) Mapping the lectin-like activity of tumor necrosis factor. *Science* **263**(5148)**:** 814-817

Lundkvist GB, Kristensson K, Bentivoglio M (2004) Why trypanosomes cause sleeping sickness. *Physiology (Bethesda)* **19:** 198-206

Lythgoe KA, Morrison LJ, Read AF, Barry JD (2007) Parasite-intrinsic factors can explain ordered progression of trypanosome antigenic variation. *Proc Natl Acad Sci U S A* **104**(19)**:** 8095-8100

Mabbott NA, Sutherland IA, Sternberg J (1995) Suppressor macrophages in Trypanosoma brucei infection: nitric oxide is related to both suppressive activity and lifespan in vivo. *Parasite Immunol* **17:** 143-150

MacLeod A, Tweedie A, McLellan S, Hope M, Taylor S, Cooper A, Sweeney L, Turner CM, Tait A (2005) Allelic segregation and independent assortment in T. brucei crosses: proof that the genetic system is Mendelian and involves meiosis. *Mol Biochem Parasitol* **143**(1)**:** 12-19

Magez S, Geuskens M, Beschin A, del Favero H, Verschueren H, Lucas R, Pays E, de Baetselier P (1997) Specific uptake of tumor necrosis factor-alpha is involved in growth control of Trypanosoma brucei. *J Cell Biol* **137**(3)**:** 715-727

Magez S, Radwanska M, Beschin A, Sekikawa K, De Baetselier P (1999) Tumor necrosis factor alpha is a key mediumtor in the regulation of experimental Trypanosoma brucei infections. *Infect Immun* **67**(6)**:** 3128-3132

Magez S, Stijlemans B, Baral T, De Baetselier P (2002) VSG-GPI anchors of African trypanosomes: their role in macrophage activation and induction of infection-associated immunopathology. *Microbes Infect* **4**(9)**:** 999-1006

Magez S, Stijlemans B, Radwanska M, Pays E, Ferguson MA, De Baetselier P (1998) The glycosyl-inositol-phosphate and dimyristoylglycerol moieties of the glycosylphosphatidylinositol anchor of the trypanosome variant-specific surface glycoprotein are distinct macrophage-activating factors. *J Immunol* **160**(4)**:** 1949-1956

Maizels RM, Balic A, Gomez-Escobar N, Nair M, Taylor MD, Allen JE (2004) Helminth parasites - masters of regulation. *Immunol Rev* **201:** 89-116

Mancini PE, Patton CL (1981) Cyclic 3',5'-adenosine monophosphate levels during the developmental cycle of Trypanosoma brucei brucei in the rat. *Mol Biochem Parasitol* **3**(1)**:** 19-31

Margos G, Navarette S, Butcher G, Davies A, Willers C, Sinden RE, Lachmann PJ (2001) Interaction between host complement and mosquito-midgut-stage Plasmodium berghei. *Infect Immun* **69**(8)**:** 5064-5071

Masiga DK, Smyth AJ, Hayes P, Bromidge TJ, Gibson WC (1992) Sensitive detection of trypanosomes in tsetse flies by DNA amplification. *Int J Parasitol* **22**(7)**:** 909-918

Matthews KR, Gull K (1994) Evidence for an interplay between cell cycle progression and the initiation of differentiation between life cycle forms of African trypanosomes. *J Cell Biol* **125**(5)**:** 1147-1156

Mayho M, Fenn K, Craddy P, Crosthwaite S, Matthews K (2006) Post-transcriptional control of nuclear-encoded cytochrome oxidase subunits in Trypanosoma brucei: evidence for genome-wide conservation of life-cycle stage-specific regulatory elements. *Nucleic Acids Res* **34**(18)**:** 5312-5324

McCulloch R (2004) Antigenic variation in African trypanosomes: monitoring progress. *Trends Parasitol* **20**(3)**:** 117-121

McDowell MA, Ransom DM, Bangs JD (1998) Glycosylphosphatidylinositol-dependent secretory transport in Trypanosoma brucei. *Biochem J* **335 (Pt 3):** 681-689

McLintock LM, Turner CM, Vickerman K (1993) Comparison of the effects of immune killing mechanisms on Trypanosoma brucei parasites of slender and stumpy morphology. *Parasite Immunol* **15**(8)**:** 475-480

Mclintock LML, Turner CMR, Vickerman K (1990) A comparison of multiplication rates in primary and challenge infections of *Trypanosoma brucei* blood-stream forms. *Parasitology* **101:** 49-55

Melville SE, Gerrard CS, Blackwell JM (1999) Multiple causes of size variation in the diploid megabase chromosomes of African tyrpanosomes. *Chromosome Res* **7**(3)**:** 191-203

Millar AE, Sternberg J, McSharry C, Wei X, Liew F, Turner CM (1999) T-Cell Responses during Trypanosoma brucei Infections in Mice Deficient in Inducible Nitric Oxide Synthase. *Infect Immun* **67**(7)**:** 3334-3338

Moloo SK, Kabata JM, Waweru F, Gooding RH (1998) Selection of susceptible and refractory lines of Glossina morsitans centralis for Trypanosoma congolense infection and their susceptibility to different pathogenic Trypanosoma species. *Med Vet Entomol* **12**(4)**:** 391-398

Morlais I, Grebaut P, Bodo JM, Djoha S, Cuny G (1998a) Characterization of trypanosome infections by polymerase chain reaction (PCR) amplification in wild tsetse flies in Cameroon. *Parasitology* **116 (Pt 6):** 547-554

Morlais I, Grebaut P, Bodo JM, Djoha S, Cuny G, Herder S (1998) Detection and identification of trypanosomes by polymerase chain reaction in wild tsetse flies in Cameroon. *Acta Trop* **70**(1)**:** 109-117

Nagamune K, Nozaki T, Maeda Y, Ohishi K, Fukuma T, Hara T, Schwarz RT, Sutterlin C, Brun R, Riezman H, Kinoshita T (2000) Critical roles of glycosylphosphatidylinositol for Trypanosoma brucei. *Proc Natl Acad Sci U S A* **97**(19)**:** 10336-10341

Naito M, Nagashima K, Mashima T, Tsuruo T (1997) Phosphatidylserine externalization is a downstream event of interleukin-1 beta-converting enzyme family protease activation during apoptosis. *Blood* **89**(6)**:** 2060-2066

Navarro M, Gull K (2001) A pol I transcriptional body associated with VSG mono-allelic expression in Trypanosoma brucei. *Nature* **414**(6865)**:** 759-763

Ngo H, Tschudi C, Gull K, Ullu E (1998) Double-stranded RNA induces mRNA degradation in Trypanosoma brucei. *Proc Natl Acad Sci U S A* **95**(25)**:** 14687-14692

Oli MW, Cotlin LF, Shiflett AM, Hajduk SL (2006) Serum resistance-associated protein blocks lysosomal targeting of trypanosome lytic factor in Trypanosoma brucei. *Eukaryot Cell* **5**(1)**:** 132-139

Olsson T, Bakhiet M, Edlund C, Hojeberg B, Van der Meide PH, Kristensson K (1991) Bidirectional activating signals between Trypanosoma brucei and CD8+ T cells: a trypanosome-released factor triggers interferon-gamma production that stimulates parasite growth. *Eur J Immunol* **21**(10)**:** 2447-2454

Olsson T, Bakhiet M, Hojeberg B, Ljungdahl A, Edlund C, Andersson G, Ekre HP, Fung-Leung WP, Mak T, Wigzell H, et al. (1993) CD8 is critically involved in lymphocyte activation by a T. brucei brucei-released molecule. *Cell* **72**(5)**:** 715-727

Paindavoine P, Rolin S, Van Assel S, Geuskens M, Jauniaux J, Dinsart C, Huet G, Pays E (1992) A Gene from the Variant Surface Glycoprotein Expression Site Encodes One of Several Transmembrane Adenylate Cyclases Located on the Flagellum of Trypanosoma brucei. *Mol Cell Biol* **12**(3)**:** 1218

Park SH, Bendelac A (2000) CD1-restricted T-cell responses and microbial infection. *Nature* **406**(6797)**:** 788-792

Paterou A, Walrad P, Craddy P, Fenn K, Matthews K (2006) Identification and stage-specific association with the translational apparatus of TbZFP3, a CCCH protein that promotes trypanosome life-cycle development. *J Biol Chem* **281**(51)**:** 39002-39013

Pays E, Guyaux M, Aerts D, Van Meirvenne N, Steinert M (1985) Telomeric reciprocal recombination as a possible mechanism for antigenic variation in trypanosomes. *Nature* **316**(6028)**:** 562-564

Pays E, Lips S, Nolan D, Vanhamme L, Perez-Morga D (2001) The VSG expression sites of Trypanosoma brucei: multipurpose tools for the adaptation of the parasite to mammalian hosts. *Molecular and Biochemical Parasitology* **114**(1)**:** 1-16

Pays E, Tebabi P, Pays A, Coquelet H, Revelard P, Salmon D, Steinert M (1989) The genes and transcripts of an antigen gene expression site from T. brucei. *Cell* **57**(5)**:** 835-845

Pays E, Vanhollebeke B (2008) Mutual self-defence: the trypanolytic factor story. *Microbes Infect* **10**(9)**:** 985-989

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

Peacock L, Bailey M, Gibson W (2005) Tetracycline induction of gene expression in Trypanosoma brucei within the tsetse fly vector. *Mol Biochem Parasitol* **140**(2)**:** 247-249

Peacock L, Ferris V, Bailey M, Gibson W (2006) Multiple effects of the lectin-inhibitory sugars D-glucosamine and N-acetyl-glucosamine on tsetse-trypanosome interactions. *Parasitology* **132**(Pt 5)**:** 651-658

Pearson TW, Beecroft RP, Welburn SC, Ruepp S, Roditi I, Hwa KY, Englund PT, Wells CW, Murphy NB (2000) The major cell surface glycoprotein procyclin is a receptor for induction of a novel form of cell death in African trypanosomes in vitro. *Mol Biochem Parasitol* **111**(2)**:** 333-349

Pereira-Chioccola VL, Acosta-Serrano A, Correia de Almeida I, Ferguson MA, Souto-Padron T, Rodrigues MM, Travassos LR, Schenkman S (2000) Mucin-like molecules form a negatively charged coat that protects Trypanosoma cruzi trypomastigotes from killing by human anti-alpha-galactosyl antibodies. *J Cell Sci* **113 (Pt 7):** 1299-1307

Pfeffer K (2003) Biological functions of tumor necrosis factor cytokines and their receptors. *Cytokine Growth Factor Rev* **14**(3-4)**:** 185-191

Priest JW, Hajduk SL (1994) Developmental regulation of mitochondrial biogenesis in Trypanosoma brucei. *J Bioenerg Biomembr* **26**(2)**:** 179-191

Procopio DO, Almeida IC, Torrecilhas AC, Cardoso JE, Teyton L, Travassos LR, Bendelac A, Gazzinelli RT (2002) Glycosylphosphatidylinositol-anchored mucin-like glycoproteins from Trypanosoma cruzi bind to CD1d but do not elicit dominant innate or adaptive immune responses via the CD1d/NKT cell pathway. *J Immunol* **169**(7)**:** 3926-3933

Radwanska M, Guirnalda P, De Trez C, Ryffel B, Black S, Magez S (2008) Trypanosomiasis-induced B cell apoptosis results in loss of protective anti-parasite antibody responses and abolishment of vaccine-induced memory responses. *PLoS Pathog* **4**(5)**:** e1000078

Ranasinghe S, Rogers ME, Hamilton JG, Bates PA, Maingon RD (2008) A real-time PCR assay to estimate Leishmania chagasi load in its natural sand fly vector Lutzomyia longipalpis. *Trans R Soc Trop Med Hyg* **102**(9)**:** 875-882

Raper J, Portela MP, Lugli E, Frevert U, Tomlinson S (2001) Trypanosome lytic factors: novel mediumtors of human innate immunity. *Curr Opin Microbiol* **4**(4)**:** 402-408

Redpath MB, Windle H, Nolan D, Pays E, Voorheis HP, Carrington M (2000) ESAG11, a new VSG expression site-associated gene from Trypanosoma brucei. *Mol Biochem Parasitol* **111**(1)**:** 223-228

Retzer MD, Kabani A, Button LL, Yu RH, Schryvers AB (1996) Production and characterization of chimeric transferrins for the determination of the binding domains for bacterial transferrin receptors. *J Biol Chem* **271**(2)**:** 1166-1173

Reuner B, Vassella E, Yutzy B, Boshart M (1997) Cell density triggers slender to stumpy differentiation of Trypanosoma brucei bloodstream forms in culture. *Mol Biochem Parasitol* **90**(1)**:** 269-280

Reyes MB, Pollevick GD, Frasch AC (1994) An unusually small gene encoding a putative mucin-like glycoprotein in Trypanosoma cruzi. *Gene* **140**(1)**:** 139-140

Richardson JM, Morrison LS, Bland ND, Bruce S, Coombs GH, Mottram JC, Walkinshaw MD (2009) Structures of Leishmania major orthologues of macrophage migration inhibitory factor. *Biochem Biophys Res Commun*

Rifkin MR (1978) Identification of the trypanocidal factor in normal human serum: high density lipoprotein. *Proc Natl Acad Sci U S A* **75**(7)**:** 3450-3454

Robertson M (1912) Notes on the Polymorphism of Trypanosoma gambiense in the Blood and Its Relation to the Exogenous Cycle in Glossina palpalis. *Proc R Soc B* **85:** 527-539

Robertson M (1913) Notes on the life-history of Trypanosoma gambiense, with a brief reference to the cycles of Trypanosoma nanum and Trypanosoma pecorum in Glossina palpalis. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences* **203:** 161-184

Roditi I, Schwarz H, Pearson TW, Beecroft RP, Liu MK, Richardson JP, Buhring HJ, Pleiss J, Bulow R, Williams RO, et al. (1989) Procyclin gene expression and loss of the variant surface glycoprotein during differentiation of Trypanosoma brucei. *J Cell Biol* **108**(2)**:** 737-746

Ross R, Thompson D (1910) A Case of Sleeping Sickness Studied by Precise Enumerative Methods: Regular Periodical Increase of the Parasites Disclosed. *Proc R Soc Lond B* **82:** 411-415

Rotureau B, Morales MA, Bastin P, Spath GF (2009) The flagellum-MAP kinase connection in Trypanosomatids: a key sensory role in parasite signaling and development? *Cell Microbiol*

Rudenko G, Bishop D, Gottesdiener K, Van der Ploeg LH (1989) Alpha-amanitin resistant transcription of protein coding genes in insect and bloodstream form Trypanosoma brucei. *EMBO J* **8**(13)**:** 4259-4263

Ruepp S, Furger A, Kurath U, Renggli CK, Hemphill A, Brun R, Roditi I (1997) Survival of Trypanosoma brucei in the tsetse fly is enhanced by the expression of specific forms of procyclin. *J Cell Biol* **137**(6)**:** 1369-1379

Sagerstrom CG, Sun BI, Sive HL (1997) Subtractive cloning: past, present, and future. *Annu Rev Biochem* **66:** 751-783

Salmon D, Geuskens M, Hanocq F, Hanocq-Quertier J, Nolan D, Ruben L, Pays E (1994) A novel heterodimeric transferrin receptor encoded by a pair of VSG expression site-associated genes in T. brucei. *Cell* **78**(1)**:** 75-86

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. *Exp Parasitol* **109**(3)**:** 188-194

Sbicego S, Vassella E, Kurath U, Blum B, Roditi I (1999) The use of transgenic Trypanosoma brucei to identify compounds inducing the differentiation of bloodstream forms to procyclic forms. *Mol Biochem Parasitol* **104**(2)**:** 311-322

Schaller E, Macfarlane AJ, Rupec RA, Gordon S, McKnight AJ, Pfeffer K (2002) Inactivation of the F4/80 glycoprotein in the mouse germ line. *Mol Cell Biol* **22:** 8035-8043

Schell D, Borowy NK, Overath P (1991a) Transferrin is a growth factor for the bloodstream form of Trypanosoma brucei. *Parasitol Res* **77**(7)**:** 558-560

Schell D, Evers R, Preis D, Ziegelbauer K, Kiefer H, Lottspeich F, Cornelissen AW, Overath P (1991b) A transferrin-binding protein of Trypanosoma brucei is encoded by one of the genes in the variant surface glycoprotein gene expression site. *EMBO J* **10**(5)**:** 1061-1066

Schell D, Evers R, Preis D, Ziegelbauer K, Kiefer H, Lottspeich F, Cornelissen AW, Overath P (1993) A transferrin-binding protein of Trypanosoma brucei is encoded by one of the genes in the variant surface glycoprotein gene expression site. *EMBO J* **12:** 2990

Schleifer KW, Mansfield JM (1993) Suppressor macrophages in African trypanosomiasis inhibit T cell proliferative responses by nitric oxide and prostaglandins. *J Immunol* **151**(10)**:** 5492-5503

Schwartz KJ, Peck RF, Tazeh NN, Bangs JD (2005) GPI valence and the fate of secretory membrane proteins in African trypanosomes. *J Cell Sci* **118**(Pt 23)**:** 5499-5511

Shapiro SZ, Naessens J, Liesegang B, Moloo SK, Magondu J (1984) Analysis by flow cytometry of DNA synthesis during the life cycle of African trypanosomes. *Acta Trop* **41**(4)**:** 313-323

Sharma R, Peacock L, Gluenz E, Gull K, Gibson W, Carrington M (2008) Asymmetric cell division as a route to reduction in cell length and change in cell morphology in trypanosomes. *Protist* **159**(1)**:** 137-151

Shental-Bechor D, Levy Y (2008) Effect of glycosylation on protein folding: a close look at thermodynamic stabilization. *Proc Natl Acad Sci U S A* **105**(24)**:** 8256-8261

Shi M, Wei G, Pan W, Tabel H (2006) Experimental African trypanosomiasis: a subset of pathogenic, IFN-gamma-producing, MHC class II-restricted CD4+ T cells mediumtes early mortality in highly susceptible mice. *J Immunol* **176**(3)**:** 1724-1732

Shiflett AM, Bishop JR, Pahwa A, Hajduk SL (2005) Human high density lipoproteins are platforms for the assembly of multi-component innate immune complexes. *J Biol Chem* **280**(38)**:** 32578-32585

Silverman JM, Chan SK, Robinson DP, Dwyer DM, Nandan D, Foster LJ, Reiner NE (2008) Proteomic analysis of the secretome of Leishmania donovani. *Genome Biol* **9**(2)**:** R35

Sonnhammer EL, Eddy SR, Durbin R (1997) Pfam: a comprehensive database of protein domain families based on seed alignments. *Proteins* **28**(3)**:** 405-420

Sonnhammer EL, von Heijne G, Krogh A (1998) A hidden Markov model for predicting transmembrane helices in protein sequences. *Proc Int Conf Intell Syst Mol Biol* **6:** 175-182

Sternberg MJ, Mabbott NA (1996) Nitric oxide-mediumted suppression of T cell responses during Trypanosoma brucei infection: soluble trypanosome products and interferon-gamma are synergistic inducers of nitric oxide synthase. *Eur J Immunol* **26**(3)**:** 539-543

Steverding D (2006) On the significance of host antibody response to the Trypanosoma brucei transferrin receptor during chronic infection. *Microbes Infect* **8**(12-13)**:** 2777-2782

Steverding D, Stierhof YD, Chaudhri M, Ligtenberg M, Schell D, Becksickinger aG, Overath P (1994) Esag-6 and Esag-7 Products of Trypanosoma-Brucei Form a Transferrin-Binding Protein Complex. *European Journal of Cell Biology* **64**(1)**:** 78-87

Steverding D, Stierhof YD, Fuchs H, Tauber R, Overath P (1995) Transferrin-binding protein complex is the receptor for transferrin uptake in Trypanosoma brucei. *J Cell Biol* **131**(5)**:** 1173-1182

Stewart CN, Jr., Via LE (1993) A rapid CTAB DNA isolation technique useful for RAPD fingerprinting and other PCR applications. *Biotechniques* **14**(5)**:** 748-750

Szoor B, Wilson J, McElhinney H, Tabernero L, Matthews KR (2006) Protein tyrosine phosphatase TbPTP1: A molecular switch controlling life cycle differentiation in trypanosomes. *J Cell Biol* **175**(2)**:** 293-303

Tachado SD, Gerold P, Schwarz R, Novakovic S, McConville M, Schofield L (1997) Signal transduction in macrophages by glycosylphosphatidylinositols of Plasmodium, Trypanosoma, and Leishmania: activation of protein tyrosine kinases and protein kinase C by inositolglycan and diacylglycerol moieties. *Proc Natl Acad Sci U S A* **94**(8)**:** 4022-4027

Tait A, Macleod A, Tweedie A, Masiga D, Turner CM (2007) Genetic exchange in Trypanosoma brucei: evidence for mating prior to metacyclic stage development. *Mol Biochem Parasitol* **151**(1)**:** 133-136

Tasker M, Wilson J, Sarkar M, Hendriks E, Matthews K (2000) A novel selection regime for differentiation defects demonstrates an essential role for the stumpy form in the life cycle of the African trypanosome. *Mol Biol Cell* **11**(5)**:** 1905-1917

Taylor JE, Rudenko G (2006) Switching trypanosome coats: what's in the wardrobe? *Trends Genet* **22**(11)**:** 614-620

Tetley L, Turner CM, Barry JD, Crowe JS, Vickerman K (1987) Onset of expression of the variant surface glycoproteins of Trypanosoma brucei in the tsetse fly studied using immunoelectron microscopy. *J Cell Sci* **87 (Pt 2):** 363-372

Theodosiou A, Ashworth A (2002) MAP kinase phosphatases. *Genome Biol* **3**(7)**:** REVIEWS3009

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**(22)**:** 4673-4680

Thon G, Baltz T, Giroud C, Eisen H (1990) Trypanosome variable surface glycoproteins: composite genes and order of expression. *Genes Dev* **4**(8)**:** 1374-1383

Timms MW, van Deursen FJ, Hendriks EF, Matthews KR (2002) Mitochondrial development during life cycle differentiation of African trypanosomes: evidence for a kinetoplast-dependent differentiation control point. *Mol Biol Cell* **13**(10)**:** 3747-3759

Tomlinson S, Raper J (1998) Natural human immunity to trypanosomes. *Parasitol Today* **14**(9)**:** 354-359

Triggs VP, Bangs JD (2003) Glycosylphosphatidylinositol-dependent protein trafficking in bloodstream stage Trypanosoma brucei. *Eukaryot Cell* **2**(1)**:** 76-83

Turner CM, Barry JD, Maudlin I, Vickerman K (1988a) An estimate of the size of the metacyclic variable antigen repertoire of Trypanosoma brucei rhodesiense. *Parasitology* **97 (Pt 2):** 269-276

Turner CM, Barry JD, Vickerman K (1988b) Loss of variable antigen during transformation of Trypanosoma brucei rhodesiense from bloodstream to procyclic forms in the tsetse fly. *Parasitol Res* **74**(6)**:** 507-511

Turner CM, McLellan S, Lindergard LA, Bisoni L, Tait A, MacLeod A (2004) Human infectivity trait in Trypanosoma brucei: stability, heritability and relationship to sra expression. *Parasitology* **129**(Pt 4)**:** 445-454

Turner CM, Sternberg J, Buchanan N, Smith E, Hide G, Tait A (1990) Evidence that the mechanism of gene exchange in Trypanosoma brucei involves meiosis and syngamy. *Parasitology* **101 Pt 3:** 377-386

Tyler KM, Higgs PG, Matthews KR, Gull K (2001) Limitation of Trypanosoma brucei parasitaemia results from density-dependent parasite differentiation and parasite killing by the host immune response. *Proc Biol Sci* **268**(1482)**:** 2235-2243

Tzou P, De Gregorio E, Lemaitre B (2002) How Drosophila combats microbial infection: a model to study innate immunity and host-pathogen interactions. *Curr Opin Microbiol* **5**(1)**:** 102-110

Urwyler S, Studer E, Renggli CK, Roditi I (2007) A family of stage-specific alanine-rich proteins on the surface of epimastigote forms of Trypanosoma brucei. *Mol Microbiol* **63**(1)**:** 218-228

Urwyler S, Vassella E, Van Den Abbeele J, Renggli CK, Blundell P, Barry JD, Roditi I (2005) Expression of procyclin mRNAs during cyclical transmission of Trypanosoma brucei. *PLoS Pathog* **1**(3)**:** e22

Vaidya T, Bakhiet M, Hill KL, Olsson T, Kristensson K, Donelson JE (1997) The gene for a T lymphocyte triggering factor from African trypanosomes. *J Exp Med* **186**(3)**:** 433-438

Van Den Abbeele J, Claes Y, van Bockstaele D, Le Ray D, Coosemans M (1999) Trypanosoma brucei spp. development in the tsetse fly: characterization of the post-mesocyclic stages in the foregut and proboscis. *Parasitology* **118 (Pt 5):** 469-478

Van der Ploeg LH, Schwartz DC, Cantor CR, Borst P (1984) Antigenic variation in Trypanosoma brucei analyzed by electrophoretic separation of chromosome-sized DNA molecules. *Cell* **37**(1)**:** 77-84

Van der Ploeg LHT, Valerio D, De Lange T, Bernards A, Borst P, Grosveld FG (1982) An analysis of cosmid clones of nuclear DNA from Trypanosoma brucei shows that the genes for variant surface glycoproteins are clustered in the genome. *Nucleic Acids Res* **10**(19)**:** 5905-5923

van Engeland M, Nieland LJ, Ramaekers FC, Schutte B, Reutelingsperger CP (1998) Annexin V-affinity assay: a review on an apoptosis detection system based on phosphatidylserine exposure. *Cytometry* **31**(1)**:** 1-9

van Luenen HG, Kieft R, Mussmann R, Engstler M, ter Riet B, Borst P (2005) Trypanosomes change their transferrin receptor expression to allow effective uptake of host transferrin. *Mol Microbiol* **58**(1)**:** 151-165

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

Vanhamme L, Pays E (1995) Control of gene expression in trypanosomes. *Microbiol Rev* **59**(2)**:** 223-240

Vanhamme L, Renauld H, Lecordier L, Poelvoorde P, Van Den Abbeele J, Pays A (2004) The Trypanosoma brucei reference strain TREU927/4 contains T. brucei rhodesiense-specific SRA sequences, but displays a distinct phenotype of relative resistance to human serum. *Mol Biochem Parasitol* **135:** 39-47

Vanhollebeke B, De Muylder G, Nielsen MJ, Pays A, Tebabi P, Dieu M, Raes M, Moestrup SK, Pays E (2008) A haptoglobin-hemoglobin receptor conveys innate immunity to Trypanosoma brucei in humans. *Science* **320**(5876)**:** 677-681

Vanhollebeke B, Nielsen MJ, Watanabe Y, Truc P, Vanhamme L, Nakajima K, Moestrup SK, Pays E (2007) Distinct roles of haptoglobin-related protein and apolipoprotein L-I in trypanolysis by human serum. *Proc Natl Acad Sci U S A* **104**(10)**:** 4118-4123

Vassella E, Kramer R, Turner CM, Wankell M, Modes C, van den Bogaard M, Boshart M (2001) Deletion of a novel protein kinase with PX and FYVE-related domains increases the rate of differentiation of Trypanosoma brucei. *Mol Microbiol* **41**(1)**:** 33-46

Vassella E, Oberle M, Urwyler S, Renggli CK, Studer E, Hemphill A, Fragoso C, Butikofer P, Brun R, Roditi I (2009) Major surface glycoproteins of insect forms of Trypanosoma brucei are not essential for cyclical transmission by tsetse. *PLoS ONE* **4**(2)**:** e4493

Vassella E, Reuner B, Yutzy B, Boshart M (1997) Differentiation of African trypanosomes is controlled by a density sensing mechanism which signals cell cycle arrest via the cAMP pathway. *J Cell Sci* **110 (Pt 21):** 2661-2671

Vickerman K (1965) Polymorphism and mitochondrial activity in sleeping sickness trypanosomes. *Nature* **208**(5012)**:** 762-766

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

Vickerman K (1985) Developmental cycles and biology of pathogenic trypanosomes. *Br Med Bull* **41**(2)**:** 105-114

Vickerman K, Luckins AG (1969) Localization of Variable Antigens in the Surface Coat of Trypanosoma brucei using Ferritin Conjugated Antibody. *Nature* **224:** 1125-1126

Webb H, Carnall N, Vanhamme L, Rolin S, Van Den Abbeele J, Welburn S, Pays E, Carrington M (1997) The GPI-phospholipase C of Trypanosoma brucei is nonessential but influences parasitemia in mice. *J Cell Biol* **139**(1)**:** 103-114

Webster P, Russell DG (1993) The flagellar pocket of trypanosomatids. *Parasitol Today* **9**(6)**:** 201-206

Wei XQ, Charles IG, Smith AB, Ure J, Feng GJ, Huang FP, Xu D, Muller W, Moncada s, Liew FY (1995) Altered immune responses in mice lacking inducible nitric oxide synthase. *Nature* **375**(6530)**:** 408-411

Weiden M, Osheim YN, Beyer AL, Van der Ploeg LH (1991) Chromosome structure: DNA nucleotide sequence elements of a subset of the minichromosomes of the protozoan Trypanosoma brucei. *Mol Cell Biol* **11**(8)**:** 3823-3834

Welburn SC, Arnold K, Maudlin I, Gooday GW (1993) Rickettsia-like organisms and chitinase production in relation to transmission of trypanosomes by tsetse flies. *Parasitology* **107 (Pt 2):** 141-145

Welburn SC, Coleman PG, Maudlin I, Fevre EM, Odiit M, Eisler MC (2006) Crisis, what crisis? Control of Rhodesian sleeping sickness. *Trends Parasitol* **22**(3)**:** 123-128

Welburn SC, Maudlin I (1999) Tsetse-trypanosome interactions: rites of passage. *Parasitol Today* **15**(10)**:** 399-403

Welburn SC, Maudlin I, Ellis DS (1989) Rate of trypanosome killing by lectins in midguts of different species and strains of Glossina. *Med Vet Entomol* **3**(1)**:** 77-82

Wirtz E, Clayton C (1995) Inducible gene expression in trypanosomes mediumted by a prokaryotic repressor. *Science* **268**(5214)**:** 1179-1183

Wirtz E, Leal S, Ochatt C, Cross GA (1999) A tightly regulated inducible expression system for conditional gene knock-outs and dominant-negative genetics in Trypanosoma brucei. *Mol Biochem Parasitol* **99**(1)**:** 89-101

Xong HV, Vanhamme L, Chamekh M, Chimfwembe CE, Van Den Abbeele J, Pays A, Van Meirvenne N, Hamers R, De Baetselier P, Pays E (1998) A VSG expression site-associated gene confers resistance to human serum in Trypanosoma rhodesiense. *Cell* **95**(6)**:** 839-846

Young R, Taylor JE, Kurioka A, Becker M, Louis EJ, 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* **9:** 385

Zarlenga DS, Higgins J (2001) PCR as a diagnostic and quantitative technique in veterinary parasitology. *Vet Parasitol* **101**(3-4)**:** 215-230

Ziegelbauer K, Quinten M, Schwarz H, Pearson TW, Overath P (1990) Synchronous differentiation of Trypanosoma brucei from bloodstream to procyclic forms in vitro. *Eur J Biochem* **192**(2)**:** 373-378

Zilberstein D, Shapira M (1994) The role of pH and temperature in the development of Leishmania parasites. *Annu Rev Microbiol* **48:** 449-470

Appendix A: Buffers and Recipes

Western blot solutions

[10x] Running Buffer	
Tris-HCl, pH8.3	0.25 M
Glycine	1.92 M
SDS	1%
Laemmli Sample Buffer	
Tris-HCl, pH6.8	62.5 mM
SDS	2 %
Glycerol	10 %
Bromophenol blue	20 μg
Coomassie Blue Stain	
Methanol	50%
Acetic Acid	10%
Coomassie Brilliant blue R250	0.1 % w/v
Destain Solution	
Methanol	40%
Acetic Acid	10%
Water	50%
[10x] Blot Transfer Buffer Stock	
Tris (do not pH)	0.025 M
Glycine	0.15 M
1× Completed Transfer Buffer	
SDS	0.2 %
[10x] Transfer Buffer Stock	10 %
Methanol	20 %

Ponceau Stain

Ponceau S	0.4%
TCA	3%

Trypanosome transfection

ZPFM

NaCl	132 mM
KCl	8 mM
Na_2HPO_4	8 mM
KH_2PO_4	1.5 mM
$MgOAc.4H_2O$	0.5 mM
$CaOAc/Cl_2$	90 μM
Glucose (ZPFMG)	0.5%

Immunoprecipitation

IP lysis buffer

Tris HCl	10 mM
NaCl	150 mM

pH 8.0

Complete with Roche EDTA-free Protease inhibitor cocktail, one tablet per 25ml.

Myristate labelling

Modified RPMI medium

Commercially available RPMI medium plus:

Na-HEPES	25 mM
L-methionine	0.15 μg/ml
L-cysteine	50 μg/ml
L-tyrosine	20 μg/ml
L-valine	20 μg/ml

Glucose	10 mg/ml
Fatty acid-free BSA	1 mg/ml

Immunofluorescence solutions

DAPI working stock

4',6-diamidino-2-phenylindole (DAPI)	10 μg ml-1

MOWIOL Mounting Medium

Glycerol	25 % w/v
MOWIOL	10 % w/v
Tris pH8.5	0.1 M

Medium required heating at 50 °C to dissolve MOWIOL. Aliquots were stored at -20 °C.

Voorheis's modified PBS (vPBS)

NaCl	137 mM
KCl	3 mM
Na2HPO4	16 mM
KH2PO4	3 mM
Sucrose	46 mM
Glucose	10 mM

pH7.6

Phosphate Buffer Solution (PBS)

NaCl	137 mM
KCl	3 mM
Na2HPO4	16 mM
KH2PO4	3 mM

pH7.6

Miscellaneous solutions

TAE 1×

Tris-Acetate	40 mM

EDTA	1 mM
pH 8.0	
TBS 10x	
NaCl	140mM
Tris	10mM
TE	
TrisCl	10 mM
EDTA	1 mM

Small scale plasmid preperation solutions

Solution I	
glucose	50 mM
EDTA	10 mM
Tris-HC1 (pH 8.0)	25 mM
Solution II	
NaOH	0.2 M
Sodium dodecyl sulfate (SDS).	1%
Solution III	
Potassium acetate	3 M
pH 5.2	

Tsetse fly DNA extraction buffers

TNES-Urea	
Tris-HCl	10 mM
NaCl	125 mM
EDTA	10 mM
SDS	0.5%

Urea	4M
CTAB	
CTAB (Sigma)	2%
NaCl	1.4 M
EDTA	20 mM
Tris HCl	100 mM

Appendix B: Primers

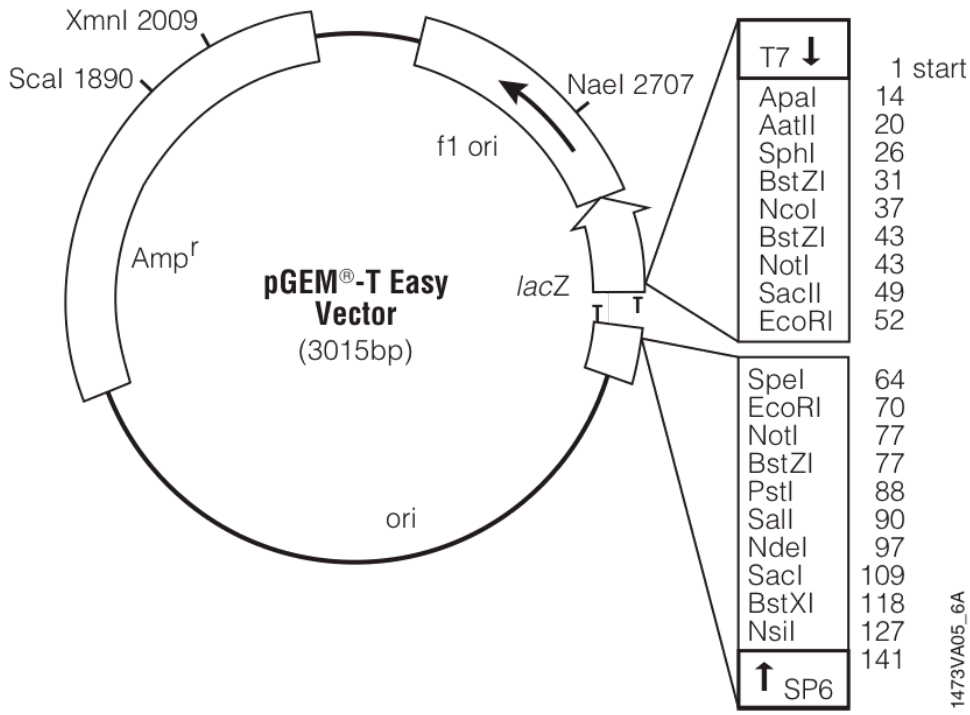
Name of primer	Sequence
TaqMan Forward	TAGCGGCCACGAAAATGA
TaqMan Reverse	CCAGTTCCCCTAGCTTGGTT
TaqMan Probe	6-FAM-CAGCAATAGAAAAGCTCA-MGB
EQ Forward A	AAGCTTATGCACCGGCTTGCAACAGTTC
EQ Reverse A	GGATCCTGGTTAGTATGGACCTCTGTTCCCATATTCTGAGTCCAG
EQ Forward B	ATACTAACCAGGATCCACTTGACCCTTCGGTTTCGCGGCACC
EQ Reverse B	GGATCCTTAAGAGTACATGAGGATTAAT
K69 Forward A	CTCGAGATGTTGAGTTTCAGAACGGC
K69 Reverse A	GGATCCTGGTTAGTATGGACCTCCACGCCTGTAGGCTGAGG
K69 Forward B	ATACTAACCAGGATCCACTTGACACTCCAGTTTCCCCAACCC
K69 Reverse B	GGATCCTTAAAAACTCATGAGACCC

Ty-tag sequence

GluValHisThrAsnGlnAspProLeuAsp

EVHTNQDPLD

Appendix C: pGEM-T Easy Vector Map



Appendix D: Antibody concentrations

Western blotting

α-ESAG9-K9	1:100
α-ESAG9-K69	1:250
BB2	1:5
α-Tubulin	1:5,000
α-VSG-221	1:10,000
α-CRD	1:60

Immunofluorescence

α-ESAG9-K9	1:250
BB2	1:50
α-EP procyclin	1:500
α-BiP	1:500

Immunoprecipitation

BB2	1:10
α-TbZFP3	1:500

FACS analysis antibodies and suppliers (all antibodies used at 1:100)

Rat Ig2a isotype control, PE-conjugated, eBioscience

Rat Ig2a isotype control, FITC-conjugated, BD Pharmingen

Rat Ig2a isotype control, Biotin-conjugated, BD Pharmingen

α-CD4, FITC-conjugated, BioLegend

α-CD8a, PE-conjugated, BD Pharmingen

F4/80, alexa-flour 488-conjugated, Caltag Laboratories

Siglec-F, PE-conjugated, BD Pharmingen

Gr-1, Biotin conjugated, BD Pharmingen

B220, PerCP-conjugated, BD Pharmingen

APC Annexin V, BD Pharmingen

Appendix E: Alignment of ESAG9 3'UTRs

```
CLUSTAL 2.0.10 multiple sequence alignment


Tb927.1.5080       AAACCCGTTGCT-ACAAGTCCGTGCATAAGTGTGCAAATGCCCCAGTAAGGCGGCGGCGATGC  59
Tb11.1000          AAACTCATTGCTTGCGGATTCGTACATGAGTGTACAAATACTCAGTAAATCGGGCGTTGT  60
Tb927.5.4620       AAAATCATTACC-GCAATTCCATGGAATACTGTGAAAACGTCTATTGAGATTAACACCGC  59
Tb427.BES122.10    AAAATCATTACC-GCAATTCCATGGAATACTGTGAAAACGTCTATTGAGGTGAACACCGC  59
Tb09.v1.0330       AACTCATTGCC-CCATATCCGTGAAAAAGTGTGCAAGCTTTCAGTGAAATTGGTAATAC  59
Tb09.160.5400      ATGTCTGTTGCT-TTACGCTGATGCATCAGGTCACAACTGCTGAATAAAGTGAACGCCGC  59
Tb927.1.170        AAGGCTATCACC-GCGGATCCCATCGCGTCGGCATGTGAATGTCTAATAATGCTACCGATGT  59
Tb09.160.5430      GCACCAGTTTCT-GCAAACCCATATATGCACTTAAAATCGTTTAACTGAGCAACAAGCGT  59
Tb927.1.5220       AAAATCATTACT-TCAAATCTGTGAACGGGGATGCACACACCTAATAAGGGCGGGGATGC  59
Tb927.5.120        ----------------GGACCGTCGCATAACTACGCAACTATCTAATGAGATCAGAAAAGT  44
                                                                                


Tb927.1.5080       GTGACACAGGGGCT--GGACGGAGAGTCAGTAACTAAATCACAGCTTCGTTTAC-GAGAG  116
Tb11.1000          ATAACACAAGAGCT--ACCGCGGAAGTCGGTAGCTAAATCATAACCACTTTTAC-GAAAG  117
Tb927.5.4620       ATGATGAGGAGGAA--GAGTTGAAAGTTGGAACTAGCTTGCAACTGCTGAATAAGAAGA  117
Tb427.BES122.10    ATGATGCGGAGGAA--GAGTTGAAAGTTGGAAGCTATCTTGCAACTACTTTCACTAGGGA  117
Tb09.v1.0330       ATAACACAAGGGAC--AGATTGAGAATCAGTGGCCAAATGCCAACTTCTTTTAC-AAGAG  116
Tb09.160.5400      ATAGGTCGAAGTA----AGGCTGTGAATCAGTAACTTGCAAGTGTTTTCATTAAGGG  116
Tb927.1.170        GTAACGCATGGGTC--AGAATGGAAATTGGTACCTAACTCGAAACTTTCTTTTAC-AAGAA  116
Tb09.160.5430      CAAGTACAAAG-AC--TGGTGAAAAATACGTAACTACCTTGCAGCAGTTATTACAAAGG  116
Tb927.1.5220       ACAATGTGCATACTTCAGCGCGAGAATCAGTAACTATTCTACAGCTTCGTTCAC-GGGAT  118
Tb927.5.120        ATGACACGGGTTC----GAATGGGAAACAGTAACCAATTAACAACTTCTTTTAC-GAGAA  99
                                                                                


Tb927.1.5080       AAACACGTGTGAAGCGTGATAACGACATAATTTTTGAAAGGACAGTGTGACAACAGGCGT  176
Tb11.1000          AATCACTTGTGAGGCATCACAACGGCATAATATTT-AAGGGACGACGTTGGAAGAAGGTA  176
Tb927.5.4620       TCATGTGGATGGAACGTTGTGGACA----------------------GACAAGATAAGACGATC  159
Tb427.BES122.10    TCATGTGGATGGAACGTTGTGGCGA----------------------GACATGCTAAGAGGATC  159
Tb09.v1.0330       ACACACTTCCGGAACGTTGCAACCACATACTATTTAAAATGGCGTTGCTGCAAAAGGTGT  176
Tb09.160.5400      ATATGTGCATGGGCTGTTGTGGCAA----------------------GTCATTGTAAGATGGTC  158
Tb927.1.170        AAGCAGATGTGGAGCTACACAACATCATAATATTTTAGAGGATGATGCTGCAAGATGGGA  176
Tb09.160.5430      GCTTGTTTAAGAAGTCTATAGCGGCAT----------------------TGAATGC--GATGGAT  158
Tb927.1.5220       AAACACTTATGGAACGTACTAAAGACATAATTTCCAGAAGGACGATGTTGCAACAGATGT  178
Tb927.5.120        CAAAATTTTTGTGGTGTTTTAACGAAATAACACTTTAAAGGGGAATGCTATAACAGGGAT  159
                                                                                


Tb927.1.5080       GAATAACCCTAATGTAATCATGA-GTAACGTGCAAGAGTGGAGTGAATGTCCTGAAAGGG  235
Tb11.1000          GTATAATTCCGATGAAATCATAA-GTAAGATGGAATGGTGAAGTGAATGTCCTAAGTA  235
Tb927.5.4620       GAATAACACTAATGTAATCATAA-GTAGCGTACATGAGTAAAACCAGTAACTGGAATGGA  218
Tb427.BES122.10    GAATAGCTCCAATGTAATCATAA-GTAGAGTACATGAGTAAAACCAGTAACTGGAATGGA  218
Tb09.v1.0330       GGATAGCATCAACATATTTATCG-GTAAATGTGCAAAACTAAGGAGAAATAGTTGGAATGGA  235
Tb09.160.5400      GAATAACTCTGAATTAGTAAGAG-CGAACATGCTAAAATTAAA-----AGTCGCAAGAGA  212
Tb927.1.170        TGATAACCATAATATGAGTAGCA-TTCAAAGGCAAAGTTAATG-----ATTTTAGGGGGT  230
Tb09.160.5430      GAATAACTTTAATATAAAAATTAACATACATGTGTAATTCGAATAATGTGAAGACA  218
Tb927.1.5220       AGATAGCTCCAGCATAATCATAA-GGAAAATGAAAAAGTAGAGCGAAC-GATTTTTGTGG  236
Tb927.5.120        GAACGACTCTGCCCTAATCATGA-GTAAAATGAAAAGGTAAAACGAAAAGACAGAAGGGA  218
                                                                                


Tb927.1.5080       AGGGAACCAAGTTAAA-GTTCGTAATCTCATACAG----------------------CGAA  273
Tb11.1000          TGAAAATAAAAAAAATATTTTTAAAGTCATAA----------------------TAA  271
Tb927.5.4620       GGAAAAGCAGAATAAA-GTTGT-TAGCGCAAGAAGCAA---AAAATGTTCAC--ATTAAG  271
Tb427.BES122.10    GGAAAAGCAGAATAAA-GTTAG-GCCCGCAAGGGG-----AAAACGTCGAC--ATTGGG  268
Tb09.v1.0330       AAAAAT-CAAGTTAAA-ATTCTCGAGTGTATAAGGTAA---AATGCACTAAT--AATAAA  288
Tb09.160.5400      GGAAAA-GAAATTAAG-GCTATTTATCGTGTAAGCGGA---AAAGTACAACC--ATTAAA  265
Tb927.1.170        AAAGAA-TAAATAAAT-GATTTTAGGCATGTAAAGCAT---AGGGCGCCACC--ACTCAA  283
Tb09.160.5430      GGAACTGTAACTCAAT-ATTTTAAAGCGTATAAATGGCCCAAAATAGCGTTGTAAAAAA  277
Tb927.1.5220       AAGAAGATAGGTTAAT-GGCTTCCAGACCAGAAGGGGA----------AAGCCTATTAAA  285
Tb927.5.120        GGGAACGTAACTTAAA-GGTTTTGAGTGCGTAAAG----------AATACCATTAAT  264
                                                                                


Tb927.1.5080       AATGACACAATCGCAAGGGAATAAGCAGTAGACGGGTTC--GATTAAACTAATCGGGTA-  330
Tb11.1000          AGTAAACAGTTCGCAAGCT------CAATAAAATGATC-----------ATGGATAAGAAA-  315
Tb927.5.4620       AACGACACGAG-GCAAGGTGAGAAACAGGCGACAACCT---TAACAATATTATTATTAAT  327
Tb427.BES122.10    AAATACACCAATACAAAACTGAAGTCCGGCAAGCGAAGT---AAAAAATGCGGTCATGAAC  325
Tb09.v1.0330       AGGACTATAAATACGATATCAAAAGAGTAAACGAACT--GAAAAATGTTACTGCAAAT  345
Tb09.160.5400      AAATGCAGAAATGTGATACACAAACGATTTACGATCT---CAAATAATTGATTTTAAAT  322
Tb927.1.170        AGCACCGCCAGTACGAAATCAGAAACAACAAGTAACTCAAGAAAAATGTGATAGTAAAT  343
Tb09.160.5430      AAAAGCCCCACGACAAAAACAGAAACGGTCAATGATCT---TACGAATGTTTTTATTAA-  333
Tb927.1.5220       GGCAGACAGGGTAACAAACGAGAAACAGTTAGAGCATC---------ACAACATCTTTGAT  338
Tb927.5.120        GAGATCACAAGTATCAA---AAAAACCATGGACAGTTCC---------ACAGGACCGTTGAT  314
                                                                                


Tb927.1.5080       --------AAACTCAGGTTC-AATACTATTGGGAAGTTTGTGGTTGATGA---ACTTTT  377
Tb11.1000          --------GAGCTTAACTTT-AATGGTATTAAGGAAGCGGTGATTATGG---AGTTTT  362
Tb927.5.4620       TAAACAAGGAAATTTAACTTC-AAAAGTATTAAGGACATCGTAGCATTGAA----TTTTTT  382
Tb427.BES122.10    GCAACCGAGCAACTTACTTTC-AAGAGTATTGTGAAGACAGTAGAATTTAAAATATTTTT  384
Tb09.v1.0330       GAGGTAAGAAAATTTAACTTC-ATTGGTATTAAGAAAG-TAACAC-ACAGTTGAATTTTT  402
Tb09.160.5400      AGGAAAAGGAAATATGAGGTT-AATACTATCAAAAAAC-CGCAAT-GAGGAGTTTCTTTA  379
Tb927.1.170        GGGTTAAGGAAACCTAACTTC-AGTGGTACTGGGGAGGTGTAACCGAAGATTTTTTTT  402
Tb09.160.5430      -----AATGAACTTAACCTT-AATGGAACTGAGAAATCCGCATCACATGAAGTATTCAA  386
Tb927.1.5220       GAGGCAGGGAAAATAAATTC-GATTGTGTTGAAGAG-TCCATTCCCGT------TTTTC  390
Tb927.5.120        AAGGCAAAGAATCTAAAATTTTAACGGTGTTGGAGAGATCTTCTTCAGTGAACTATTTTT  374
                                                                                


Tb927.1.5080       C----------AAAAGTGTAATGGACGTTGCTGCATATCAAAAGTTAACCTTATTGCGA  426
Tb11.1000          T----------AAAAGTGCAGTGAAAGATGCCGTGTGGCGAAAGTTAGAGCCCTCG-AA  410
Tb927.5.4620       T----------TTTTAGG----------GAACGTGATAAGAGCCACTTCACACCCTG  419
Tb427.BES122.10    T----------CTTTAGTTAAAT------GGACAAAGTAAGAACCACGTTGCATGCAG  426
Tb09.v1.0330       C----------AAAAATG----------TAATAGAAGTTAGTG------TGA  428
Tb09.160.5400      A----------AGATGTA----------TAATGAAGTCACTGAACACGCAA  411
Tb927.1.170        T----------TAATGTAAGGAAATTT--GCTGCATATCCAAAGTCAGCGCTTTGCAA  449
Tb09.160.5430      A----------AAAAAAGGACCACTTTGGTTCCGAATCGGAAACCGTCACA---GCGA  432
Tb927.1.5220       TTTTGCATTCAAAAAAAAAACATAATAGGAATCGCTTGGGCGATACATTATGA  450
Tb927.5.120        TTTTGGTTCTAAAAATTATA--AAGAGCTGCCCCGTGCCTAGAAGAAGGCACATTCTA  432
                                                                                


Tb927.1.5080       GGATGCTGATGGCAGTCGTGAGGGCAAGCTGGATGAAATCGTAAATTTTTAATTGTGGGGA  486
Tb11.1000          AAGTGCCGAAGGCAGTCGTAAAGATGAACTTAACGGAATCGTTATTAAATTTTGGTGA  470
Tb927.5.4620       GAGT---AAACTTGATACCAGAGACGACTAACGAAATCGTAGCGTCACGTTATGGGGGGGG  476
Tb427.BES122.10    GGA---TTA----------GCTTAATAGCATTGCAAATTTAAATCTTCTAA  467
Tb09.v1.0330       GGG---TTAACATACT---GCAGAGCGGCTTAACGGAAACGTAGGCTTGCGCTGTGGA  465
Tb09.160.5400      AGGGAGTGAGGATAGTCGTAAATTTAAATTCCGAT  509
Tb927.1.170        AGAT-GTAAATGGGTTAACGAGGAGGGTCTTAGTGAAATTTTAAATTGTGGGGA  491
Tb09.160.5430      AGAGCTAAATGTAGCT-TACGAAGAATACTATGCAGATCACAAGCTAATATGAAGAGAG  509
Tb927.1.5220       AGAGAGCGAAAAATGTT-TAAAGCAGGACTTATTTAGATCATTAACTTAAATGAGGA---  488
Tb927.5.120        AGAGAGCGAAAAATGTT-TAAAGCAGGACTTATTTAGATCATTAACTTAAATGAGGA---  488
                                                                                


Tb927.1.5080       ATTTAGAGACCACCTTGTTAATCA----------  510
Tb11.1000          CGATATTAATCGTTTGTTAACAATACATAAGACCTTACCGATAAA--TTTAATATATAT  528
Tb927.5.4620       GGGCACTGAATACTTTGACGGCCAAACCTGATATCATATCGGTAAC--CCTAAGATATGC  534
Tb427.BES122.10    ----------  
Tb09.v1.0330       AGAAGCTAACTATCTCTTTAAGCATGTGTGAAATCCTATCGATAAA--CTTTAAGATATTT  525
Tb09.160.5400      AGAGAGTGATTGTTTCAGCAGTCAAATGTGAGATTATATTTCTAAC--CTAAAAATGTGT  523
Tb927.1.170        AAACACAGGCCGCTTCGTTGACAATACGTAGGACCCCATCGATACG--CTTAAGGTATTT  567
Tb09.160.5430      AGAAACTGAGAAATTCAACGACTAAATACGGGATTAAGTTATAAC--CATAAGACATTC  549
Tb927.1.5220       AAGCACAGAACGGTTCGGCGGCTAGAATTAGGGGTTCTTCGGTAAC--TTGAAGACCTTT  567
Tb927.5.120        AAACGCAGACGATTGCGGTAACTAAATTTAGGGTTCTTATCGGTAATCATAAAAAAGAAT  548
                                                                                


Tb927.1.5080       ----------  
Tb11.1000          TTTAACTGATGTTGGTAGCTGC--TAAAAGAATATAGATTTTAATGCTGATAAATTGTT  586
```

```
Tb927.5.4620        TT-AACTGATGCGAATAACAGA-AAACATGCATAAAGGATATGACCGACGAATAATCAGT 592
Tb427.BES122.10     ------------------------------------------------------------
Tb09.v1.0330        TTTAGTGTGTGTGAATATCGGATAAAAACAT-GAATGATCCAAATGC-TAATAAATTGCT 583
Tb09.160.5400       TT-AATCAATGCGGATAACTG--AAGCGCAT-AAAGGATTTGAAGGGACGAATACTCAAT 579
Tb927.1.170         TTTAGCATGCGTAAATA--------AAACAT-AAACAGTATTAATCC-TTATGAATTAAC 617
Tb09.160.5430       TTTGATACGGGTGATGGACGA----AAACATTAAATGATTTCGATGA-TGAGTAATCGAC 604
Tb927.1.5220        TTAAACTGATACTAATAA--------AAAATATATATTGAAAAAAAAGTA--TAATGAAT 617
Tb927.5.120         TTTAATTGGTGCAAATAAT------GAGGGCATAAGTCATTGAAACAGTAAGCATTCAAT 602


Tb927.1.5080        ------------------------------------------------------------
Tb11.1000           AAGTAAACAAAAATTTCGGGGTGCGGATATGTGAGTAATTTAAATCAAA--CATTTTACG 644
Tb927.5.4620        GAACGAGCGAATTGCACAAAATAAGGATATGTGAAGACTTTAAATCAAA--CGGGAGAAG 650
Tb427.BES122.10     ------------------------------------------------------------
Tb09.v1.0330        AAGTGCACGAAATTTACGAGGAATTGATATGCTAGAAAGGTCAAT-AAC--GGTAAATAG 640
Tb09.160.5400       AAGTGAGTGAATTGCTAAAGGTAAAGATATGTAAAGAACTTAAGTCAAT--TGGAAACAA 637
Tb927.1.170         AAGTAAATAAAATTTACGGTGTGGCGATACGTGAGGAAGTTAAATTAAG--AAAAAACCG 675
Tb09.160.5430       GAATCAATGACTTGTACAAGGTACAGGTATGCGAAAAAATTAAATTAAT--CAAGAGTGT 662
Tb927.1.5220        GGGCGGAGCGCATGCACATGGTCCGGGTACATGAAAAGTTATAACTAAGTGTTGAAAAAA 677
Tb927.5.120         GAGTGAGAGAATCGCACAGGGTACAGATACGTCGAAGACTTGAATAAA---TTGATAATA 659


Tb927.1.5080        ------------------------------------------------------------
Tb11.1000           A--ACACACGTCGATACAAACGAAGGTGAATGGTAATGGTAACA--------ACATCTC- 693
Tb927.5.4620        ATAATGAAAGCTGGAAGATATGAAGGAGAGAATTAACAA---------- ------------ 689
Tb427.BES122.10     ------------------------------------------------------------
Tb09.v1.0330        AACACAAATGCACGCGGATGTGAAGGAAGGAATTAACGAAGATGGTAACAACACCTCCA- 699
Tb09.160.5400       GTTTCAAAAGCGTGATAATCTGAACGAGGAAATTCATGACGGTAAAAACGGTGCCTCAC- 696
Tb927.1.170         TTAATAAAAGCAGGCAGGTGCTGAGGAAAGGATTAGGGAAGGTT------ACACATCAC- 728
Tb09.160.5430       G----GGAAGTGGACGAATGTGCATGAGGCGATGAATGATTGATACG-CAACACTTGAG- 716
Tb927.1.5220        ACTCTAAAAGCAGGTAATCGTAACGAAAGAGGTGAGGGATAGCCAAAGCGGTGCCTGAA- 736
Tb927.5.120         GTCGCAAAAGCAGGTAAGTGTATGGAAAGCGATGGGGGTTGCTTTAGGCAGTACCCGAAA 719


Tb927.1.5080        ------------------------------------------------------------
Tb11.1000           AGTTATACAGTTAATTAATG---AACATTGCTCATTGAAGC----------GACCCCTT 739
Tb927.5.4620        -----------------ACA---TAACTTATTTTTTAAAA-----------AGGCGCCC 717
Tb427.BES122.10     ------------------------------------------------------------
Tb09.v1.0330        CTACAATGATTTAATTAACG---AACACAGCTCATTGGAGCGAGCACCACTGGGGTGTCA 756
Tb09.160.5400       CGATGTTTTTTCATTTAAGA---AACGCAACTTGTTTAAA-----------GGGTG-CA 740
Tb927.1.170         TGACATTATGTTAAATAACA---AGCACAACTTGTTGAAGCG----------AGCCCTCA 775
Tb09.160.5430       CAATTTATGATTAATTAAAG---GAGGTAGCTTATTGAAAC----------AGACACCT 762
Tb927.1.5220        CGAAATGTGATTAATTAAAATAAAACCCAACTTATTTTCA-------------ATAGCGC 783
Tb927.5.120         TGAGCTGCAATTAATTTTAA-AAAACATAACTT-CTTCAA-------------AGATTCT 764


Tb927.1.5080        ------------------------------------------------------------
Tb11.1000           CCTTGGATGCCAATGTTTTGATTGTTTCTGCAAACGCTCAACCACGCCTCATTTAATGAA 799
Tb927.5.4620        TTACATATATTTTCTGTGTTTGTTTTTTGCCGCT-TTATCGTCGCAGCACATTGAATTAA 776
Tb427.BES122.10     ------------------------------------------------------------
Tb09.v1.0330        TTGGTTATATATATATATATTTTTTTTTGTAAGCGTTGATTGCAAAACCCATTTGACTAA 816
Tb09.160.5400       TTGCTTATATATTTTTTTCATTACTTTTGCGGGT-TTGCTTCCACGACAGATTTAATTAA 799
Tb927.1.170         CTAAGGTGTTATTAGTTTCTTTCTTTTTGCAAGGACTCAAC-CATAGCTTATTTAATGAA 834
Tb09.160.5430       CTCCGGGTATTTTCTTTGATATTTTCTTC------TTGTCGCCACAGTACATTTGATTAA 816
Tb927.1.5220        CCCTGTGTGTCTTTGCTTTAATAGTCTTTATGAGAGTTAAACCACAACCCATTTAATTAA 843
Tb927.5.120         CACGGTGCGTATTTGATTTAATTGTTTTTGCAAAAGGTTATCCACAACCTATTTAATTAA 824


Tb927.1.5080        ------------------------------------------------------------
Tb11.1000           GCAATTGTTGTGTAATGGCGGCAAATATCCTCAACCAT-CATACACT-----CTTAAGCT 853
Tb927.5.4620        ATCAGGAGCGCATCCTTGCTGGAAAAAGCATTAAAAGTTCGTATACT-----TTTAAGGG 831
Tb427.BES122.10     ------------------------------------------------------------
Tb09.v1.0330        G--ATAGTAGCACAATTGCTGCAAACTGTATTGAAACT-CGCA--AA-----CTTAAGCG 866
```

```
Tb09.160.5400       ACCACAAGAGCATGCGTGCTGCAAATAATCTTAGCCGT-CGCATTAT-----TCTGAGCT 853
Tb927.1.170         GTAAGCGGTGTTTAGTGGTAGCAAACAGTCACAACCTT-CGTACACT-----CTTGAGCG 888
Tb09.160.5430       ACCATGACTGCATAGTTGCTTTTAACAGGGCTAGACGT-AACCCTAA-----CCGCAGAA 870
Tb927.1.5220        CTGGACGGAGAATTCGTAAGGCAGTAATTCTAACATTTACATACTTCAACTGCCGGGACA 903
Tb927.5.120         CAGGGCAATTCGCGCTTGAGTTCGCGGAGTTGACCCTCACATGCCCTGAACAGTGAAACA 884


Tb927.1.5080        ------------------------------------------------------------
Tb11.1000           C--------CAAACATATTAAT---ACTGCAACAAGCATAAGCA--GGTAAGTGCGTTGC 900
Tb927.5.4620        AA-------TAATAATAATAAACA-ATGCCTTAACGAAGAAACT--GGCAAATACGAAGC 881
Tb427.BES122.10     ------------------------------------------------------------
Tb09.v1.0330        ---------CAGAAATAATAAA---ATTTTAACGAGAAGACGCT--GAAAAATACATTGC 912
Tb09.160.5400       ---------TAATAATAATAGGTGCACTGTAACATGAAGGCGCT--GGTAAATGGTCTGC 902
Tb927.1.170         A--------CAGACACAGTAAT---GCTTCAAAAGGAAG---CT--GGTAAATGCAGTGC 932
Tb09.160.5430       CG-------CAGTAATGGTAAA---ATCGTGACTAGGAGACACT--GGTAGATGCATTGC 918
Tb927.1.5220        GA------ATAAAAATAATGATACGACATCAAACAGGAGATATTTTGACAAATGTGCTAT 957
Tb927.5.120         TAGAGATAATAGTAGTAGTGATGCAACCGCAAAAGGAAGGTGCTTTAGTAAATGCACTAT 944


Tb927.1.5080        ------------------------------------------------------------
Tb11.1000           GGAGCCTCTTTAATATTTTGGTTG--TTCACGACATACATATTTCTACAA-CTTGTGCTT 957
Tb927.5.4620        AAAGGCTCTTTATTTTGGTTGTTAACTTCATGATGTGCGCAATTCCACAGGCGTATGTTG 941
Tb427.BES122.10     ------------------------------------------------------------
Tb09.v1.0330        TAGCACTTTTTCTTTTGGTTGATTACTTTATTATTTGCATAATTCTAC----GTGTACTG 968
Tb09.160.5400       -AGTTTTCTTTGTTTCAGTGGTTGAATTCACGATATACATGATTCTACAAATGCATGCTA 961
Tb927.1.170         GATTTTTTTAAATTTTGTCATTATACTTCACGATATACATGATTCTACAAATACATGCTG 992
Tb09.160.5430       GAGCGCTTTTATTTTCGA-AGCTGATTTTTCCAT----AAAATTCTACGG-TATGTGTCG 972
Tb927.1.5220        GAGCTGTTTTCACTCCTATGATTGGCTGTAT-ATTTGTGGATGTCAACAA-TGTTTGACA
1015
Tb927.5.120         GGACT---TTTATCTCTTTGATTAATTGAAT-ATAAGCGGATGTC---AA-TGTTTAGCA 996


Tb927.1.5080        ------------------------------------------------------------
Tb11.1000           GTAAGGCGGGCCTGAAAAAGGATATTACCTCCTCTTTTGCCTTCTCCATCAGGATCTAAA
1017
Tb927.5.4620        TTATTGCGGATCTGGAAACTGATATT-TATCCTCTCTGTTTTATTTT-TCCATATTGAGT 999
Tb427.BES122.10     ------------------------------------------------------------
Tb09.v1.0330        GCAAGATTAACCTGGCCGCGGATATACTATCTTTTTTTGACACT---------ACTGGAC
1019
Tb09.160.5400       GTAAGGTGAACCTGATCGCAGATAATGTGTCTTCTTTACCCCTTCCC-CAGTGACCATGT
1020
Tb927.1.170         GTAAGGTGAACCTGAACGCAGACAT--TATCTTTTCTTTCCTTTACC-CCACGACTATGT
1049
Tb09.160.5430       GCACTGCGGACCCGAAAACTGCTAT--TATC-----------------------------
1001
Tb927.1.5220        GTGATAGAAGTCAACAATTGAATAT--TATTCTGCTATTCCAGTT--ATAACTGATAGTT
1071
Tb927.5.120         CTGATAAGAGGCAAGTTTTGGTTAT--TATTTCTCTATTTCTGTTGAACGGTTTTCAAAG
1054


Tb927.1.5080        ------------------------------------------------------------
Tb11.1000           ATTTTCATTGCGGTTATTGGTTATATCTTAGTTTATGATTTGTTTTTCTTTATGC-----
1072
Tb927.5.4620        TTTGAATTTTTTTAACCA----TTTATTGTCTTCATTATGCGCTGTTGTACTGTT-----
1050
Tb427.BES122.10     ------------------------------------------------------------
Tb09.v1.0330        TTTAAGATTATAACTTTA---CTCTACTTATCGCTTGT-TAATGAGTACTTCAAC-----
1070
Tb09.160.5400       CTTAACGTTATCGTTTCAAGCCGTCATGTTTAGTTTGTGTTTTGTGTTTCTTGTA-----
1075
Tb927.1.170         CTTAGCGTTATCGTTTCTGGCTGCTGTGTTTGGTTTGTGTTTTGTGTTTCCTGTA-----
1104
```

```
Tb09.160.5430   ------------------------------------------------------------
Tb927.1.5220    TTCAAACTCTTATTTTATAAGGGTATCAGCTTAGGCGTTGTTGTTGTTGGTGGTAA-----
1126
Tb927.5.120     CTTTTGTTTATTTTTCAAGAAGTTGATTTCTCATTCGTTTACATGATAATAAGAATCTGA
1114


Tb927.1.5080    ------------------------------------------------------------
Tb11.1000       ---GCATGATACCGTATGTTTCACCTACACCGAAAGTGAATGTATTATTTGATGCAGTTT
1129
Tb927.5.4620    ---TCGTAACTTTGGTAATGAGTTGAATATGAAGAGCAACTACATTACTTTAACCAAGTT
1107
Tb427.BES122.10 ------------------------------------------------------------
Tb09.v1.0330    ---TTACG-CGTAATAAATCTCGTGAATACCAATA-TAAGCGTATTACTTGAAGCTAATT
1125
Tb09.160.5400   ---GAGTGGCGTAGTAAGTGGCACAATTATCAAAAGCAAATGTATTCCTTGAAGCGGTTT
1132
Tb927.1.170     ---GCATGACATTGTAAGTGACATGAATACCAAAAGCAAATGTGTTACTTGAAGCAAATT
1161
Tb09.160.5430   ------------------------------------------------------------
Tb927.1.5220    -GTTTTTGGCGTTCCCTAAAAGAGTATCTCCGGTTGCAACACTTTTGGCTCACAGTGTCA
1185
Tb927.5.120     AGTAGTTCGCTTTCCACAAAATAATCTCTTCGGTTGCAAACGTTTTAGCTCGCAGCGACA
1174


Tb927.1.5080    ------------------------------------------------------------
Tb11.1000       CTTCACACATGTGACTAGGGTTGCTGCTGGGGGAGAGATGTGACCCCTATTAGCATGTGA
1189
Tb927.5.4620    CTCAATATGAGCAATCGGAGTTTTCCGTAAGGTAGCGTCGCACGACTCAGTGGCATGTTG
1167
Tb427.BES122.10 ------------------------------------------------------------
Tb09.v1.0330    GGTTATGCCAGTAACTGTAGTTTCTGGCGAAAGTGAGGTGCCCAATCAAGCAGAGTGTTA
1185
Tb09.160.5400   GTTAGTGCCACTAACTGGGATTGCTGGCGAAGGTGAAGTGCCCATTCAAGTGAAGTGTGA
1192
Tb927.1.170     TTTAATATCGGCAGTTGGGGTTTCTGGCGAAGGTGAAGTGCCCATACAAGCGCAGTGTGA
1221
Tb09.160.5430   ------------------------------------------------------------
Tb927.1.5220    AGCGATGTTTGGGAGAAAATTCGTAGAGTATTGTGAAACTTCTGT-GCGTGATGGT--CG
1242
Tb927.5.120     TGCGAGGATGAAGAGGATCCTGGCGGGGTATCGTGAAATCTGTGAAGAGTGATGGTTGCG
1234


Tb927.1.5080    ------------------------------------------------------------
Tb11.1000       ATATTTACTAAGTTGCAAAATGAGCGAGAATGCACGCACCGACACGATCACAACCGCGGA
1249
Tb927.5.4620    CGCCATTTTCTTCCGAGTCAGAGAACAACAACGAGACTG----GTCGATGGAGTTTAAAT
1223
Tb427.BES122.10 ------------------------------------------------------------
Tb09.v1.0330    GACCGTACTCTTCCGAGGTAGAAACCTGCAACAAGACTA----GCAGAGACTGTCTGGAC
1241
Tb09.160.5400   GACCTTGCTCTTC-----------------------------------------------
1205
Tb927.1.170     GGCCGTACTCTTTTGAGGCAGAAACCTGAAACAGGACAG----GCAGAGACTGCTTGGAC
1277
Tb09.160.5430   ------------------------------------------------------------
Tb927.1.5220    GGAGACCTCATCATGTCGCAAAAGAAAACATACACTCATGATATTTTTGTTATTTAAAGA
1302
Tb927.5.120     GCGAATTTCATGATTTCAAGAAAAAAACCATGCGTTCGAGGTATTTGTGTGATTACAGTG
1294
```

```
Tb927.1.5080    ------------------------------------------------------------
Tb11.1000       CACAAACGCAAAGCCAAGCGCAGGGGAAATCACACCAAGGAAGGGATGAGGCGAACCACA
1309
Tb927.5.4620    AGATTGGATGACACATGATAGTGCTGTGCGCACAGGACACAGGTGATTCAATAAAATTGA
1283
Tb427.BES122.10 ------------------------------------------------------------
Tb09.v1.0330    CAATTGTACGACCCAAGGAAGTGCCGTAACCATGGGATACAGGCAATTCAATTCTTAGGA
1301
Tb09.160.5400   ------------------------------------------------------------
Tb927.1.170     AACGTATTTGACCAAAGGAATTGTCGTGACGACGGGATACAGGCAATTCAGTCTTTCGGA
1337
Tb09.160.5430   ------------------------------------------------------------
Tb927.1.5220    GATTGGCATGACACAACGAAATTGTAGAAGTGGATATGGGACTTCGAGGGAAAGCTTGGA
1362
Tb927.5.120     CGTTGCCATTTACCAACCGACATAAAGAGATGGAAGCGAGAATTTGGCAGAATGCATTGA
1354


Tb927.1.5080    ------------------------------------------------------------
Tb11.1000       AAAAAAATGAGATCCCAAATTCACAAACATCTACCAGTACCGACACAAATGCCACTACAA
1369
Tb927.5.4620    AGCTTATTGAGTGCGTGGGA-ACAGACAGAGAAGCATTAAGGACATTCCGCTTGGGATTG
1342
Tb427.BES122.10 ------------------------------------------------------------
Tb09.v1.0330    GCGCGCCGGATACTGTGGAACAGGGAAGGGGGAAACAAGAATGCGTTTGCATTTGCACAG
1361
Tb09.160.5400   --------------------------------GGAGACGAGAAAGTGTTTGCATTTGCACAG
1235
Tb927.1.170     GCACGCTGGATGAACTGGAATGGGGAAGAGGGAGACACGAAAGTGCTTGCGTCTGCACGG
1397
Tb09.160.5430   ------------------------------------------------------------
Tb927.1.5220    GTGCTAGCATCAA-GCAGGGGGTTAGGGTTATGTCTAGGCTTCGTGGGATTCTAATATAA
1421
Tb927.5.120     ATGTTAGCGGTAA-ATAGGAAGTGGGGTTTACTATTACGTGTTCTGTTATTCTA-CAGCA
1412


Tb927.1.5080    ------------------------------------------------------------
Tb11.1000       CAATACCACCAAAAGAACAAAAGGGTTCCACGAAAATTAATTCACATTTTATATGGAGAC
1429
Tb927.5.4620    TTGCAGGCGAGGATGAGTTACGGAACAGGGGCATGGTTGCTCTTAACCTTCGTAGGTGAT
1402
Tb427.BES122.10 ------------------------------------------------------------
Tb09.v1.0330    TGGTGGAAACAAAGGAAGTAGTGTTTAAGCCTGAAAGCAACCAAAAGTTGTAGCGACTGT
1421
Tb09.160.5400   CGCCAGAAACACGGGAAGGGGAAAT-AAACCTGCAAGCAGCCAGATGTTGCGGTGGCAGT
1294
Tb927.1.170     CGGTGAAAACGGAGGAAGGAATGTTTAAGCCTGAAAAAAGCTGCAGATTGTAGCGGGTGT
1457
Tb09.160.5430   ------------------------------------------------------------
Tb927.1.5220    AGTGTGATGTGAGGAAAAGCGTTTTTTGTTGCTCAGAGAGCGGTAAAATTCAATAGCAGC
1481
Tb927.5.120     GGCAATATGTCACGGCAATAGTTTTTTGTTCCGCAGGAAACACTTATGTTTGACCGCAAA
1472


Tb927.1.5080    ------------------------------------------------------------
Tb11.1000       TTTTTGTGTCGTTTCGATAAACCTGTCCACTACTCATTTAATTTCATTTTTGACATTAAG
1489
Tb927.5.4620    TAGGAAACGACAAGTGTGAAGTCAACAAAACGGCTTCGTGGCAGAAGTGGTGACTCAAGA
1462
Tb427.BES122.10 ------------------------------------------------------------
```

```
Tb09.v1.0330        AGACAAGC-AGTATCCTGGAAAGAAGCTGCTCAGAACTCTCACTCACAAATCTTAACCGA
1480
Tb09.160.5400       AGACAGGT-AGTTTCCTCAACAAAGATTGCTCACAACCCTCAAACACAAATCTTACCCGA
1353
Tb927.1.170         AGACAATCCAGTATCCCAGAAAGAAGCTGCTCAGAACCCTCAATCACAAAACACACTCGA
1517
Tb09.160.5430       ------------------------------------------------------------
Tb927.1.5220        TTCTACTCTAAGG---TGCGGCGTTTCTCCACACTTGT----CCCTTTTTAGATCAGA-G
1533
Tb927.5.120         TGCTTTTTCAGGGAAGCATTGTATTTCTATACACTTGTGTGTCTCTTTTCTCATAAAAAG
1532


Tb927.1.5080        ------------------------------------------------------------
Tb11.1000           AAAAATAAAAAAAGATGCGAACGGAAAAAAGACAGAAAG--CAGGATGGAGATGGCGATG
1547
Tb927.5.4620        AAGCGGCAAAAATGCTGTGGAGACGCGGTTGAGCACATT--TGCCACTATAGCTAACATG
1520
Tb427.BES122.10     ------------------------------------------------------------
Tb09.v1.0330        CACTTCTGATACTAGTTGAGTAGGACCAGAGAGAATGTC--TGAGCGCACCATTTGAATA
1538
Tb09.160.5400       CATTTCTGACATTGGTTGAGTAGGACCAGGGAGAATGTC--TGAGCGCACCTTTTTAATA
1411
Tb927.1.170         AACTTCTGATGCTGGTTGAGAGGGACTAGAGAGAATGTC--TGAGCGCACTATGGGAATA
1575
Tb09.160.5430       ------------------------------------------------------------
Tb927.1.5220        CAGCGTTGGTACTATGTAGGTATCCAGACTTTGTACAGCGCCGGAGGCTTTGTGCGTGTG
1593
Tb927.5.120         TAATATTGAGACCATGTAGTGTCTCACAGCTTGTACCACCTTGGAGTCATTGTGTGATTT
1592


Tb927.1.5080        ------------------------------------------------------------
Tb11.1000           GACTTACGTGTTTTTTCTACACATTTACTGGGGTTCTTAAGGTGGATAACTTTAACACCT
1607
Tb927.5.4620        TGTGCAATAAAATGCTATCTGACGCTCAGTCA------AAGGTCGGACATCCAGTAGACC
1574
Tb427.BES122.10     ------------------------------------------------------------
Tb09.v1.0330        ATTCGAATGCATGAGCAGGCAGCGGGGGAAAAGGGCGTACGAGAACGTGATTTTCAACCT
1598
Tb09.160.5400       ATTCGAATGCATGAGCAGGCAGCGAGAGGCAATGGCGTACGAGAGCGCGTTTCTCACCCT
1471
Tb927.1.170         ATCCGAATGCATGAGCAGGCAGCGAGGGGCAGTGGCGTACGAGAGCGTGTTTCGCAGCCT
1635
Tb09.160.5430       ------------------------------------------------------------
Tb927.1.5220        TGCTTGTCAGCTGCACTTGCGTCACCCAAATA-CTTGCTTTGCTGATGGTGTTCAAACCA
1652
Tb927.5.120         TATTT--TAACTATATTTCAGCCAATACCTTGTCTTTCCTTCTTCGGCGTCTTCTAGCCC
1650


Tb927.1.5080        ------------------------------------------------------------
Tb11.1000           CAAAAAGGAAAATAAACCACTCGCATCTGCGAACTGCAATAAAGGTTAAGCAAAGCGAAA
1667
Tb927.5.4620        AACCGACAAATTGAACCTTCCAAGCGCATTTTACACATCACTGGTTGAAGAATTAGCGTA
1634
Tb427.BES122.10     ------------------------------------------------------------
Tb09.v1.0330        CCTCAGCAGAAGGTAAAAACAACAAACCCTCTGGGGCTGGCTAGTCTGCGTGGCGAGGTG
1658
Tb09.160.5400       TCCTTGCAGGAGGTAAAAACAAAGAGCCTTCCTAGACGGTCCACTCTATATAGCAGACTG
1531
Tb927.1.170         GCCTTGCAGGAGGTAAATATAACAAACCCCTTGGGACTGGTTAGTCTGCGTAGCAAGGTG
1695
```

```
Tb09.160.5430       ------------------------------------------------------------
Tb927.1.5220        CTTGTTTGACGAACGCAAACGCAATTTTACTTATTTTGTCACCGATAAATAATTTTACT-
1711
Tb927.5.120         TTTGTTTAGATAAAGGGG--GCCATTTT-TTCATTTTCTCACTAAGGAAAAATTATACCG
1707


Tb927.1.5080        ------------------------------------------------------------
Tb11.1000           GCGACAGGTAGACAAACAAAGATAAATAGTGAAAAAACATTATAGCCTTTTGGATTCAAA
1727
Tb927.5.4620        CTTGGACCTC-AAACATGGATGCGCGTTACATGTTGATGGCGCCAAAAAATCCACAGTCA
1693
Tb427.BES122.10     ------------------------------------------------------------
Tb09.v1.0330        ACCGAAATAA-CTTTGCGGAATGATCTACCTAATGAACAGTATGCAGTTGCTTTGGTGCG
1717
Tb09.160.5400       ACCAAAAAAAACTTTGCGGAATGATCTACCTAATGAACAGTATGCAGTTGCTTTGGCGCG
1591
Tb927.1.170         GTCGAAATAA-CATTGCGGAATGATCTACCGAAGGGACAGTGTGCAGTTCCCTTGGTGTG
1754
Tb09.160.5430       ------------------------------------------------------------
Tb927.1.5220        -CCTTAAGGCGTAGTATTTCTATCGAAAAGGGGCAAGT--TAAGTAAAAAATACTTTACC
1768
Tb927.5.120         GTGTTTGAGTGCAGCCGTTCTTGTGTTTATGAGAGAGTGGTATGCACAATAAAATTTCTC
1767


Tb927.1.5080        ------------------------------------------------------------
Tb11.1000           AAAAAAAAACAAATGTTGGTATTGTTGTTATAGAATATGGAAAAACAAGTATGTTAAGGC
1787
Tb927.5.4620        GCGGCGGGGAACAAAGTAAGGACCTCTCTAACCGCGTTTGGCGGAAAACCAATACGTTGC
1753
Tb427.BES122.10     ------------------------------------------------------------
Tb09.v1.0330        GTGGCACTCAATTGGGCGAGACTTGTTGAGTCAGGAAATAGG-GTTGACGGAACGAAGAC
1776
Tb09.160.5400       CTGGCACTCGATTGGGTGAGACTTGTTGAGTCAGGAAAAAGGAGTTGACGGAACGAAGAC
1651
Tb927.1.170         ATGGCACTCAGCTAGTCAATTCTTGTGGAGTTAGGAAA---GGACTAAGGAGACGGAGAC
1811
Tb09.160.5430       ------------------------------------------------------------
Tb927.1.5220        TGAGTTT-CTGTTGGCTATCATATGCAACATTGTGGCCTAGAAAAGCTAA-------GGT
1820
Tb927.5.120         TTAATCTATTGCTAATGGTAATCCTATCCATTGTGCCCTGGCGATGCCGGCCATCTTAAC
1827


Tb927.1.5080        ------------------------------------------------------------
Tb11.1000           AAACGCGTTGCACT---CTCACAAACCCTGAATTTTGCAAAACTTACAATTGGGCCTGTT
1844
Tb927.5.4620        ACAAAGGGGAAAAG---TACACAATGAG-AACCGTTGCGAGCTACGGCGCTTTGTGCCGT
1809
Tb427.BES122.10     ------------------------------------------------------------
Tb09.v1.0330        GCCAATGCAATAGC---TACCCACCTTGTGGTCTATATTGGGCCCATTGAAATCTGCTCA
1833
Tb09.160.5400       GCCAATGCAATAGC---TACCCACCTTGTGGTCTATATTGGGCCCATTGAAATCTGCTCA
1708
Tb927.1.170         ACCAATGCAGCCAG---GTCTCGCTGTGTAGACCACATTAGACCCGTCTGAATACGTCCA
1868
Tb09.160.5430       ------------------------------------------------------------
Tb927.1.5220        GTGAGGATAAAGTT---TAACGTGAAGGAGATTTGTTG--AAGTCTACG----------
1864
Tb927.5.120         GTGGTGCCAAGGTCCAGTAACAGAAGAGAAATCGATTGGGAAGCCAAGAGCCAGCAGCGT
1887
```

```
Tb927.1.5080        ------------------------------------------------------------
Tb11.1000           CAAAAACAAGAGACAGATGTAGCAAAACCTATCAGTGAAACTTGGTTAAAGCATTTAAGA
1904
Tb927.5.4620        GATATCCAAACCTAGTGTCAACAGGGTGGGTTTATCATAGTGACCTTTATTGTTGGTCAC
1869
Tb427.BES122.10     ------------------------------------------------------------
Tb09.v1.0330        AAGGTTATAAAACGCTGTTAGACTG-ACATGC----ATTACAAGTGTTATTGTCACAACA
1888
Tb09.160.5400       AAGGTTATAAAACGCTGTTAGACTG-ACATGC----ATTACAAGTGTTATTGTCACAACA
1763
Tb927.1.170         AGAGGTATAAAGGGGTTCCAGGTTGTAGGTGGTGGAATAGCAAGCCATGTTTTTACTGAA
1928
Tb09.160.5430       ------------------------------------------------------------
Tb927.1.5220        --CTTCAGTAAGAAACGTGGGCAT-TTTTAACTGTA--AGTATACGTTTAATAAAACATT
1919
Tb927.5.120         TCCTTTCATGGGGAACACTGCTTTGCTCCGGCTACGGCATCATACAGCACAGGGATCAGC
1947


Tb927.1.5080        ------------------------------------------------------------
Tb11.1000           AGAGAGAGAGAAAAAAAATAGTTATTGCTGCAACAGCCAACTGAAACGACGATTTAACCC
1964
Tb927.5.4620        CCGCCTCCGCAGTTGCTGCCATGGTATCTGGCTCCACCCTTTTCCACTCTCGTTTCAACG
1929
Tb427.BES122.10     ------------------------------------------------------------
Tb09.v1.0330        GTGGGGACACAGCGAGAAGCATTCATGGCACTCCGTTTGTAATTGATTTGGTCACGGATG
1948
Tb09.160.5400       GTGGGGACACAGCGAGAAGCATTCATGGCACTCCGTTTGTAATTGATTTGGTCACGGATG
1823
Tb927.1.170         ATGGGGATAAGTAGAGAAGCATTAATGGCACTTCATTTGGGACCGGCGTAAGCACGTATT
1988
Tb09.160.5430       ------------------------------------------------------------
Tb927.1.5220        GATGTGTCA--GAACCGTCCTCGTCCAGTAATAGCTCCACTATTTTTTATTTCCTCCGCA
1977
Tb927.5.120         AGCGTCTTGCTGGGACACCGTTTTTCATTTGTCGGTCCCTGGGCACGTGCCAGCGTGCCA
2007


Tb927.1.5080        ------------------------------------------------------------
Tb11.1000           AACCACATTGGAGCGGCAAAGGACTGCAGATTCACGATAAAGGCAGCACTACAAGCACTG
2024
Tb927.5.4620        GAACTTATGGAATCTGCGGTTCGCTGCTTCTCTGCACGTCTCACTGACACCTTGGTTAAC
1989
Tb427.BES122.10     ------------------------------------------------------------
Tb09.v1.0330        GGATTCACAAAAGGGACATGACGGGTCTCAACCTTCCTAAATGATAACAGCAAAAGAAAC
2008
Tb09.160.5400       GGATTCACAAAAGGGACATGACGGGTCTCAACCTTCCTAAATGATAACAGCAAAAGAAAC
1883
Tb927.1.170         GGATGCTCAACAGGGACATGGTGGGTCTCACCCTTCTTAAATGATAATAGCAGGAGAAAG
2048
Tb09.160.5430       ------------------------------------------------------------
Tb927.1.5220        CTGAAGGTA--AGCTATGTTGACGACACAATGGGAAATGATGTGGTGGTGATTAGGGTAG
2035
Tb927.5.120         TCAGCAGCATCATCCGCACTAAGATGCTGCTGTCCGGTGATGTGGAAGAGAATCTCAGCC
2067


Tb927.1.5080        ------------------------------------------------------------
Tb11.1000           TTCAGGTCTATAGTTAAATTTAGAAAAAAAAAGATACCAAAAAAACTAGATGTATGTTTT
2084
Tb927.5.4620        ACCAGTAAT-----------------------------------------------------
1998
```

```
Tb427.BES122.10     ------------------------------------------------------------
Tb09.v1.0330        AGGTAATGTCAAGGATCCAGCATTGTAACAAGAATACCAATTAAGTAAAACGGGAGCGAA
2068
Tb09.160.5400       AGGTAATGTCAAGGATCCAGCATTGTAACAAGAATACCAATTAAGTAAAACGGGAGCGAA
1943
Tb927.1.170         AAGCAAATTCAAGGATACAGGGTTGTGAAGGGAGTACAGATT--GGAGAACAGGAGCAAC
2106
Tb09.160.5430       ------------------------------------------------------------
Tb927.1.5220        CGTTTTTCCATGGGGTTAAAT-----------------------------------------
2056
Tb927.5.120         AGTTATTGCGCGGGATGCAGTGTAACTGCGCCGGGCTATCTCAAGGAAAGAGATTAGCAC
2127


Tb927.1.5080        ------------------------------------------------------------
Tb11.1000           TTTTCATTTAATTGGTAGGAAAGTGAAGCCACCAAAAGACATATTGTTTAAGCTGATTTT
2144
Tb927.5.4620        ------------------------------------------------------------
Tb427.BES122.10     ------------------------------------------------------------
Tb09.v1.0330        ATGGTAGAGGATGTGCAGTTGACTACGTTTGATAAGGTGGTACAAACGTTTGCAGTAAAT
2128
Tb09.160.5400       ATGGTAGAGGATGTGCAGTTGACTACGTTTGATAAGGTGGTACGAAACACTGCGATAAAT
2003
Tb927.1.170         AAGGTGAAGGATGTAACGTTGGCAACATTTGACAAGTTTGTACACAAACATGCAATAAAA
2166
Tb09.160.5430       ------------------------------------------------------------
Tb927.1.5220        ------------------------------------------------------------
Tb927.5.120         TCCACAAAACCCTTGTTGATGAGCGGATTAACTTTTTTCTGTTGAGCGAGACAAGGATGA
2187


Tb927.1.5080        ------------------------------------------------------------
Tb11.1000           TATTTGTGTAACCCCGAAGTTTAAGGTATATGGCAAACTGGGAATCGTGCAAGGAAGCTG
2204
Tb927.5.4620        ------------------------------------------------------------
Tb427.BES122.10     ------------------------------------------------------------
Tb09.v1.0330        TGTTATCGTAGACTCACGACCAAAGGCAGTGCATCTCTGAGAATTACCAATAAACTGCAC
2188
Tb09.160.5400       TGCTACAGTAAGCTCAACAGCAAGGGCAGAACATCTTTCGGAATAGCTAGTAAAATGCAC
2063
Tb927.1.170         TGTTGTCTCAGGCTGGGAACCAAGGGCAGAACATCTTGCGGAATAGCTAGTAAAATGCAC
2226
Tb09.160.5430       ------------------------------------------------------------
Tb927.1.5220        ------------------------------------------------------------
Tb927.5.120         CACCTGGAGAGGCGGTTTGCTTTAGTGTTGCTGGCTACCAACATCACGGAATAGCTCGTA
2247


Tb927.1.5080        ------------------------------------------------------------
Tb11.1000           AGGGGCGTTAAAAAGGATTGTAGGTGAGGAAGGCGTTGGAGTAATACATCAAAAATTAAG
2264
Tb927.5.4620        ------------------------------------------------------------
Tb427.BES122.10     ------------------------------------------------------------
Tb09.v1.0330        -GCCAAATCGCATCCCATAA---GTGACCGGTTAAATCAATTGAGGGTTTAGAATTCTGC
2244
Tb09.160.5400       TATCCTAGTGCACTTTATTA---TTG-CCGGTTGAATAGTTTGAAGGGCTGGTTTTTTTG
2119
Tb927.1.170         TTTCCTAGTGCACTTTATTACACATGGCTAGTTGAATAGTTTGAAGGGCTGAGCTCTTGG
2286
Tb09.160.5430       ------------------------------------------------------------
Tb927.1.5220        ------------------------------------------------------------
Tb927.5.120         ACTGCAAAGGAGGTGGTGTATCAATACTAGTGAGGGAGGACCTACCAGTCGAGACCGGTA
2307
```

```
Tb927.1.5080        ------------------------------------------------------------
Tb11.1000           TTAAAAGGCGCTGAAGGCGGCAGAAGGGGCCCTCATAGTCTTTGAGAGACCTCTTCATGG
2324
Tb927.5.4620        ------------------------------------------------------------
Tb427.BES122.10     ------------------------------------------------------------
Tb09.v1.0330        AATGGACGTGTGCTACTCACTACTCACGGCAAGAAAACCCACACCGTTAGTAATGGGAAT
2304
Tb09.160.5400       CATGGTCGTGTATGAAAAGCCGCTGGAATCAGAAAAAAC-ACAGTGTTAGTAATGGAAAC
2178
Tb927.1.170         CATGGGTGTGTATGAGACGCCGCTGGCACCAGAAAACCC---AGAGTTAGTAATGGGAAC
2343
Tb09.160.5430       ------------------------------------------------------------
Tb927.1.5220        ------------------------------------------------------------
Tb927.5.120         TGGCCGTTGTTGGTCGCATTGAACAAGTGCATGCAACAATTTACCTTGCACGTGGAGCGG
2367


Tb927.1.5080        ------------------------------------------------------------
Tb11.1000           -GAGGAAACAGTCGGAATGACATGTGTTTAGTTCGTGTGTATTTGCTCTAATTCGGAATT
2383
Tb927.5.4620        ------------------------------------------------------------
Tb427.BES122.10     ------------------------------------------------------------
Tb09.v1.0330        CATAGCAATAATCTCCCCTATTTGTGCTTGACTGAAAGCCGATGTGTCGAGTAAGGAGGA
2364
Tb09.160.5400       CGTTGCAGTAGTCTCCCATATTTGTGTTTGGCTGGAAGCCGATGTGTCGAGCAAGGAAGA
2238
Tb927.1.170         CGTGGCAGTAGTCTCCCCTATTTGTGGTTGACTGAGAGCCAATGCGCTGAACAATGAGGA
2403
Tb09.160.5430       ------------------------------------------------------------
Tb927.1.5220        ------------------------------------------------------------
Tb927.5.120         CGCTGACTGTCACGTCGGCATACATCCCACCAAAACACACCTTCACGGCAACTGACCTAG
2427


Tb927.1.5080        ------------------------------------------------------------
Tb11.1000           AAATAATTGTTGTTCGCCATCTAACATATA------------------------------
2413
Tb927.5.4620        ------------------------------------------------------------
Tb427.BES122.10     ------------------------------------------------------------
Tb09.v1.0330        AAATATGTGCGGTGATAAGTGTTGCAAATGACGGCACTGTGTTCTGTCACATTTAGAGTA
2424
Tb09.160.5400       AAATATGTGCGGTGATAAGTGTTGCAAATGAAGGCACTGTGTTCTGTCACATTTAGAGTA
2298
Tb927.1.170         AAATACGTGCAGTGATAAGTGTTACAAGTGACGGCGCTGCGTTCCGTCACATTCAGAGTA
2463
Tb09.160.5430       ------------------------------------------------------------
Tb927.1.5220        ------------------------------------------------------------
Tb927.5.120         ATACGCTTCTGACGACTGACGGTGCCCAGCTCATCGGTGCAGACGCTAACGCACACGCGT
2487


Tb927.1.5080        ------------------------------------------------------------
Tb11.1000           ------------------------------------------------------------
Tb927.5.4620        ------------------------------------------------------------
Tb427.BES122.10     ------------------------------------------------------------
Tb09.v1.0330        ACGGTTCCAGTAAGGTAGAGTTATCATGTTTGCCTCTACTACTGGCCATCCATCTTCCGT
2484
Tb09.160.5400       ACGGTTCCAGTAAGGTAGATTTAACATGTTTGCCTCTACTACTGGCCATCCATCTTCCGT
2358
Tb927.1.170         ACGGTTCCAGTAAGGTAAAATTGTCATGTTTGCCTCTGCTATTGGTCATCCACCTTCCGT
2523
```

```
Tb09.160.5430       ------------------------------------------------------------
Tb927.1.5220        ------------------------------------------------------------
Tb927.5.120         TGGCATGGGATCGCGCGAGCCCACCAAATACCAAGGGTGAAACCCTCACACAGTGGTGCA
2547


Tb927.1.5080        ------------------------------------------------------------
Tb11.1000           ------------------------------------------------------------
Tb927.5.4620        ------------------------------------------------------------
Tb427.BES122.10     ------------------------------------------------------------
Tb09.v1.0330        GCTTGTTGCCACGTTATCCAACATTACCTCTTTTCTACCTACTTTTCAACGAAATCTATG
2544
Tb09.160.5400       GCTTGTTGCCACGTTATCCAACATTACCTCTTTTCTACCCACATTTCAACGAAATCCACA
2418
Tb927.1.170         ACTTGTTGTCATGCGATACAACA----CCCATTTCTATCCACGTTTCAACGCAATCTATG
2579
Tb09.160.5430       ------------------------------------------------------------
Tb927.1.5220        ------------------------------------------------------------
Tb927.5.120         TTGACAACCAGTTTCTGGTCTGCAACACCGGTGAGTGCACCAGGTACGCGCGCCACCACG
2607


Tb927.1.5080        ------------------------------------------------------------
Tb11.1000           ------------------------------------------------------------
Tb927.5.4620        ------------------------------------------------------------
Tb427.BES122.10     ------------------------------------------------------------
Tb09.v1.0330        GCAACTGTGCTTCGGTACTGCTCGACACATCTCGATAGCGCCGTGGCTAATAGTAGTAAT
2604
Tb09.160.5400       ACAACTGTGATTCGGTACTGCTCGACACATCTCGATAGCGCCGTGGCTAATACTACTAAT
2478
Tb927.1.170         GCAACTGTACTTCGCTGCTACCCGATATGTCTCGATAACACGGTGGCTAATATTAGTAAT
2639
Tb09.160.5430       ------------------------------------------------------------
Tb927.1.5220        ------------------------------------------------------------
Tb927.5.120         GAGAGTCCACCCCTGATGTGACACTGTCAAGGAATTGCACAGTGTACACGTGGACATCGC
2667


Tb927.1.5080        ------------------------------------------------------------
Tb11.1000           ------------------------------------------------------------
Tb927.5.4620        ------------------------------------------------------------
Tb427.BES122.10     ------------------------------------------------------------
Tb09.v1.0330        TTTACTGTGTAGCAAATACTTTCGGATACGTTCCCCCTACTCCATAATAAAATACATGCA
2664
Tb09.160.5400       TTTACTGTGTAGCAAATACTTTCGGATACGTTCCCCCTACTCCATAATAAAATACATGCA
2538
Tb927.1.170         TTTACCGTGCAGAAAATACTTTCGGACACGTTCCTCTCACTCCAAGATAAAATTCATGCA
2699
Tb09.160.5430       ------------------------------------------------------------
Tb927.1.5220        ------------------------------------------------------------
Tb927.5.120         TGTATTCTCCCGATAGCGATCACCATAACATATTTTTCGAAGTTATCGTTGGAGACGACA
2727


Tb927.1.5080        ------------------------------------------------------------
Tb11.1000           ------------------------------------------------------------
Tb927.5.4620        ------------------------------------------------------------
Tb427.BES122.10     ------------------------------------------------------------
Tb09.v1.0330        AGACGTTGTATGTCCGTGCTCACATGTTTCGTGTTATTGGTCTTCTATTTCGTTGGAGTG
2724
Tb09.160.5400       AGACGTTGCATGTCCGTGCTCACATGTTTCGTGTTATTGGTCGTCTATTTCGTTGGAGTG
2598
Tb927.1.170         AGACTTTGCTTGTGCAGACCCACATGTTTCGTGTCACTAATCATCCATTTCGTTGGAGTG
```

```
2759
Tb09.160.5430       ------------------------------------------------------------
Tb927.1.5220        ------------------------------------------------------------
Tb927.5.120         CAGATGCACTGAGTTGCCCGCGGCTTCGAAAGCCCATGTACGCATGGCTAAAAGCTAACT
2787


Tb927.1.5080        ------------------------------------------------------------
Tb11.1000           ------------------------------------------------------------
Tb927.5.4620        ------------------------------------------------------------
Tb427.BES122.10     ------------------------------------------------------------
Tb09.v1.0330        TCTTTGCTT-GAGTTATCTTAATTGTGAGTTGTGATAACTGTGGCTGTC-TCCACCATCT
2782
Tb09.160.5400       TCTTTGCTT-GAGTTATCTTAATTGTGAGTTGTGATAACTGTGGCTGTC-TCCACCATCT
2656
Tb927.1.170         TCCCTTCTTTGTGTTATCTTAATTGTGAGTTGTGATAACCGTGGCCTTCGTCCACTATCT
2819
Tb09.160.5430       ------------------------------------------------------------
Tb927.1.5220        ------------------------------------------------------------
Tb927.5.120         GGAGGAATTTCCGTCACAAGATTGACGAGCCCTGCAGGAAAAATTGGTAGAGGGAAGAAC
2847


Tb927.1.5080        ------------------------------------------------------------
Tb11.1000           ------------------------------------------------------------
Tb927.5.4620        ------------------------------------------------------------
Tb427.BES122.10     ------------------------------------------------------------
Tb09.v1.0330        GTATTTTGTGGGGGAATGGAGTGTGTGAATGCTCTCTGTATTTGCTGTAAGGATGGAGTG
2842
Tb09.160.5400       GTATTTTGTGGGGGAATGGAGTGTGTGAATGCTCTCTGTATTTGCTGTAAGGATGGAGTG
2716
Tb927.1.170         GTATTTT--GGGGTGGCGGGGTTAGTGAATTTTCTCTGTATGTGAAATAAGGAAGAAGTA
2877
Tb09.160.5430       ------------------------------------------------------------
Tb927.1.5220        ------------------------------------------------------------
Tb927.5.120         GTCAACACCCCGGAACAGAAATTGTGCTCCGCCATCCGCGTTGCGACGAAGGTCTCCGTC
2907


Tb927.1.5080        ------------------------------------------------------------
Tb11.1000           ------------------------------------------------------------
Tb927.5.4620        ------------------------------------------------------------
Tb427.BES122.10     ------------------------------------------------------------
Tb09.v1.0330        TTTGATTTTGAGCTAACCCAAGTAGCGTTTTTGCCGTTTGTTTTTTTGCTAATGTAACGAA
2902
Tb09.160.5400       TTTGATTTTGAGCTAACCCAAGTAGCGTTTTTGCCGTTTGTTTTTTTGCTAATGTAACGAA
2776
Tb927.1.170         TTCAGTTTCGTGCTCACATGAGTAGCG---TTGTAGTTTTTTTTTTTGCTTATGTAAAGAA
2934
Tb09.160.5430       ------------------------------------------------------------
Tb927.1.5220        ------------------------------------------------------------
Tb927.5.120         CCCCGTGGCTGCAGAGCAACGCCACCACATTGGACACCTGAGCTCGCGAAACTCGATGAA
2967


Tb927.1.5080        ------------------------------------------------------------
Tb11.1000           ------------------------------------------------------------
Tb927.5.4620        ------------------------------------------------------------
Tb427.BES122.10     ------------------------------------------------------------
Tb09.v1.0330        CTTGTGTGGAGTAAAGGTGATGAGAGTTCACTGAGAGGATAAATGGTTACGTATGTGACC
2962
Tb09.160.5400       CTTGTGTGGAGTAAAGGTGATGAGAGTTCACTGAGAGGATAAATGGTTACGTATGTGACC
2836
```

```
Tb927.1.170         CTTGTGTGGAGTAAGGGTGATGGGAGTTCACTGACAGGATTGATGATTACGTATGTGACT
2994
Tb09.160.5430       ------------------------------------------------------------
Tb927.1.5220        ------------------------------------------------------------
Tb927.5.120         GAGATAGCTGGATGCGGACCCTCCCACCGAAGGAAAAAGTTGGTAGCCACGCGTAAGCAG
3027


Tb927.1.5080        ------------------------------------------------------------
Tb11.1000           ------------------------------------------------------------
Tb927.5.4620        ------------------------------------------------------------
Tb427.BES122.10     ------------------------------------------------------------
Tb09.v1.0330        TCCTTTTTTGTCACCAGCGTATCTATAAGAATGGTGACATCTTTGATTGTGGTAAATAAT
3022
Tb09.160.5400       TCCTTTTTTGTCACCAGCGTATCTATAAGAATGGTGACATCTTTGATTGTGGTAAATAAT
2896
Tb927.1.170         TCCTTTTTTTCCACCTGCCAGTCTATAAATATGGTGATTATTTTGCGTGTGGTAAATGAT
3054
Tb09.160.5430       ------------------------------------------------------------
Tb927.1.5220        ------------------------------------------------------------
Tb927.5.120         ATCCTGGACCGTACCACAAAGAAGAGATGGAGCACGCTATGCTCCAGACTTGCGGTGTCA
3087


Tb927.1.5080        ------------------------------------------------------------
Tb11.1000           ------------------------------------------------------------
Tb927.5.4620        ------------------------------------------------------------
Tb427.BES122.10     ------------------------------------------------------------
Tb09.v1.0330        TCTAACGAGGGTGGTGACAAAGAGGGTCTCTTGTTATTCTGGAGCCGTTTTTTTT--GTG
3080
Tb09.160.5400       TCTAACGAGGGTGGTGACAAAGAGGGTCTCTTGTTATTCTGGAGCCGTTTTTTTT--GTG
2954
Tb927.1.170         ATAAATGATGGTGGAGACAAAGAGGCTCTCTTGTTATTGTGGAGTTGTTTTTTTTTTGTG
3114
Tb09.160.5430       ------------------------------------------------------------
Tb927.1.5220        ------------------------------------------------------------
Tb927.5.120         GACCGCTGCAGTTGGCACATTGTCAAGAAAGTATATGCGCCACGACCACTAACCACACCG
3147


Tb927.1.5080        ------------------------------------------------------------
Tb11.1000           ------------------------------------------------------------
Tb927.5.4620        ------------------------------------------------------------
Tb427.BES122.10     ------------------------------------------------------------
Tb09.v1.0330        TTGGTGACTGACATTATATCCCGTAACGTCACCGTATAA-CAAGGGCTTCAATAATAAGT
3139
Tb09.160.5400       TTGGTGACTGACATTATATCCCGTAACGTCACCGTATAA-CAAGGGCTTCAATAATAAGT
3013
Tb927.1.170         TTGGTGACTGAAATTATATCCCGTAATGTCACCGTATCATCAAGGGCTTCAATAATAAGT
3174
Tb09.160.5430       ------------------------------------------------------------
Tb927.1.5220        ------------------------------------------------------------
Tb927.5.120         GCTGTACTTGTTGATAACGCGGCCATCACGGACTACCGTCAAGCTGAGAGGTTCAGTAAA
3207


Tb927.1.5080        ------------------------------------------------------------
Tb11.1000           ------------------------------------------------------------
Tb927.5.4620        ------------------------------------------------------------
Tb427.BES122.10     ------------------------------------------------------------
Tb09.v1.0330        GTCCACTGATGGTATTTTTAAGATACAACTTTTATATTAAACAATGTGTCGTCTAAGAAA
3199
Tb09.160.5400       GTCCACTGATGGTATTTTTAAGATACAACTTTTATATTAAACAATGTGTCGTCTAAGAAA
```

```
3073
Tb927.1.170          GTCCAATGACGGTATTTTAAAGATACAACTTTTATATTAACTAATGTACTGTTTAAGAAA
3234
Tb09.160.5430        ------------------------------------------------------------
Tb927.1.5220         ------------------------------------------------------------
Tb927.5.120          CTGTACTCGTCCCGCGCAAGAAGGAACCCCGACTCACACCCACCGGCACCAATAAAGACG
3267


Tb927.1.5080         ------------------------------------------------------------
Tb11.1000            ------------------------------------------------------------
Tb927.5.4620         ------------------------------------------------------------
Tb427.BES122.10      ------------------------------------------------------------
Tb09.v1.0330         ATGGAGGACGGGAAGCTGCACATGCTCACAAAAAGAGTAACGAAGAGTATTGTCAATGTG
3259
Tb09.160.5400        ATGGAGGACGGGAAGCTGCACATGCTCACAAAAAGAGTAACGAAGAGTATTTTCAATGTG
3133
Tb927.1.170          ATGGGAAGCAGGAAACTGTCCATGCTTACAAAGAGAGGACCAAAGAGTATTGTCAATATG
3294
Tb09.160.5430        ------------------------------------------------------------
Tb927.1.5220         ------------------------------------------------------------
Tb927.5.120          ATAGCGAGTGAGTTCAGTCCCATCACGATGGCTGAACTACGGAGATCGATCAAACTGCTA
3327


Tb927.1.5080         ------------------------------------------------------------
Tb11.1000            ------------------------------------------------------------
Tb927.5.4620         ------------------------------------------------------------
Tb427.BES122.10      ------------------------------------------------------------
Tb09.v1.0330         CTTATGAAAAGAGCGGGATATGCTAAATGGGACTCAGAAGCAGAGTTTAAAGGAGTTGGT
3319
Tb09.160.5400        CTTATGAAAAGAGCGGGATATGCTAAATGGGACTCAGAAGCAGAGTTTAAAGGAGTTGGT
3193
Tb927.1.170          TGTATGAAAGGAGGAGGATATTCCAAATGGGGCTCAGTTGCAAAGTTTT-----GTT--C
3347
Tb09.160.5430        ------------------------------------------------------------
Tb927.1.5220         ------------------------------------------------------------
Tb927.5.120          CCGAGTGGATCCGCAGCCGGACCTGATTGCTTATACAACGAGGCACTGCAACATCTCGGC
3387


Tb927.1.5080         ------------------------------------------------------------
Tb11.1000            ------------------------------------------------------------
Tb927.5.4620         ------------------------------------------------------------
Tb427.BES122.10      ------------------------------------------------------------
Tb09.v1.0330         ACAGGATGTGAAAAAAAAATAGTATGAAAGTGCAGGTACACATTAGGGTAGTGGATAAGT
3379
Tb09.160.5400        ACAGGATGTGAAAAAAAAATAGTATGAAAGTGCAGGTACACATTAGGGTAGTGGATAAGT
3253
Tb927.1.170          ATTGGATGAAGGTAACACAT-GTGTGCAAAT-CAAGCACCTAATCTGTTGGCGAGTAAA-
3404
Tb09.160.5430        ------------------------------------------------------------
Tb927.1.5220         ------------------------------------------------------------
Tb927.5.120          AGAACAGCGCTGAATGTTGTTCTGAGGCTATTCAATGAGGGCCTACGAACGGGAGTCGTG
3447


Tb927.1.5080         ------------------------------------------------------------
Tb11.1000            ------------------------------------------------------------
Tb927.5.4620         ------------------------------------------------------------
Tb427.BES122.10      ------------------------------------------------------------
Tb09.v1.0330         CACTAACGAGGAAATACAGAAAATAGAAGCGCAACTCAGTTAGAAAATTGGGACAAGCCA
3439
```

```
Tb09.160.5400        CACTAACGAGGAAATACAGAAAATAGAAGCGCAACTCAGTTAGAAAATTGGGACAAGCCA
3313
Tb927.1.170          -GCCATTAAGTGACAACGGAAGCTAATAGAGC----CAATGAAGGATGCGGAAAAACATA
3459
Tb09.160.5430        ------------------------------------------------------------
Tb927.1.5220         ------------------------------------------------------------
Tb927.5.120          CCGCCTGCATGGAAGACTGGTGTTATCATCCCCATCCTGAAGGCCGGAAAAAAGGCGGAG
3507


Tb927.1.5080         ------------------------------------------------------------
Tb11.1000            ------------------------------------------------------------
Tb927.5.4620         ------------------------------------------------------------
Tb427.BES122.10      ------------------------------------------------------------
Tb09.v1.0330         GAATGGTAGTTAAGGTAGTGGAGGTTAATTTGGTGGAGGAAGATATTGGAGAAACAAA-A
3498
Tb09.160.5400        GAATGGTAGTTAAGGTAGTGGAGGTTAATTTGGTGGAGGAAGATATTGGAGAAACAAA-A
3372
Tb927.1.170          GCATAAATCTGCAAGTACCCAAAGAGAG---AATAGGTAAGCTTGTTGAGGAAATAAAGA
3516
Tb09.160.5430        ------------------------------------------------------------
Tb927.1.5220         ------------------------------------------------------------
Tb927.5.120          GACCTCGATTCTTACAGGCCTGTGACGC--TCACGAGCTGTCTCTGCAAAGTCATGGAGC
3565


Tb927.1.5080         ------------------------------------------------------------
Tb11.1000            ------------------------------------------------------------
Tb927.5.4620         ------------------------------------------------------------
Tb427.BES122.10      ------------------------------------------------------------
Tb09.v1.0330         CAGTTGTAAAGGAACTTTATGTTGATATCCAGCATGACGAGAGGGAACTCATGTAAATGT
3558
Tb09.160.5400        CAGTTGTAAAGGAACTTTATGTTGATATCCAGCATGACGAGAGGGAACTCATGTAAATGT
3432
Tb927.1.170          AAACAGCGGAAGGACGTAGT-TGAAAAATTGGAGGAACTGGTATGGTTATTCGTAAAGTG
3575
Tb09.160.5430        ------------------------------------------------------------
Tb927.1.5220         ------------------------------------------------------------
Tb927.5.120          GCATAATTGCCGCGAGGCTTAGAGACACTGTTGAGTCCCAGCTGACGCCGCAGCAGTCAG
3625


Tb927.1.5080         ------------------------------------------------------------
Tb11.1000            ------------------------------------------------------------
Tb927.5.4620         ------------------------------------------------------------
Tb427.BES122.10      ------------------------------------------------------------
Tb09.v1.0330         GTTTTGGAATGTGATTTACTCTTTGTTGGGT--CCTGTTGTAGTTGGTTTGAGCCAGGCA
3616
Tb09.160.5400        GTTTTGGAATGTGATTTACTCTTTGTTGGGT--CCTGTTGTAGTTGGTTTGAGCCAGGCA
3490
Tb927.1.170          GGTTTATTAAATGCATAAATTAGAGTGAGGTAACCAATGGGAAAGCATTTGTGA-GGGCG
3634
Tb09.160.5430        ------------------------------------------------------------
Tb927.1.5220         ------------------------------------------------------------
Tb927.5.120          GCTTTCGCCCCGGATGCTCAACGCTCGAACAACTCCTGCACGTCCGCGCTGCCCTCTGCC
3685


Tb927.1.5080         ------------------------------------------------------------
Tb11.1000            ------------------------------------------------------------
Tb927.5.4620         ------------------------------------------------------------
Tb427.BES122.10      ------------------------------------------------------------
Tb09.v1.0330         CTTATTAATTTCAATATTTTTCATTGTAT---ACGTCTGT-TGAGTGCGTCTACCAATAG
```

```
3672
Tb09.160.5400        CTTATTAATTTCAATATTTTTCATTGTAT---ACGTCTGT-TGAGTGCGTCTACCAATAA
3546
Tb927.1.170          ---GTTAATGTTGATATCCGTCTTGGGATGGTACGCTCATCTGACTGTGTGTATTTAGGT
3691
Tb09.160.5430        ------------------------------------------------------------
Tb927.1.5220         ------------------------------------------------------------
Tb927.5.120          GTCCCGCGCACCAATATCGTACGGGTGCTGTATTTGTTGACTACGAGAAGGCATTCGATA
3745


Tb927.1.5080         ------------------------------------------------------------
Tb11.1000            ------------------------------------------------------------
Tb927.5.4620         ------------------------------------------------------------
Tb427.BES122.10      ------------------------------------------------------------
Tb09.v1.0330         CAATAATGACTGCGATGATGAGGT--------------------------------------
3696
Tb09.160.5400        CAATAATGACTGCGATGATGAGGT--------------------------------------
3570
Tb927.1.170          GCGC----------------------------------------------------------
3695
Tb09.160.5430        ------------------------------------------------------------
Tb927.1.5220         ------------------------------------------------------------
Tb927.5.120          CAGTAGACCACGACAAAATTGCGAGGGAAATGCACAGAATGAAGGTATCACCCCACATTG
3805


Tb927.1.5080         ------------------------------------------------------------
Tb11.1000            ------------------------------------------------------------
Tb927.5.4620         ------------------------------------------------------------
Tb427.BES122.10      ------------------------------------------------------------
Tb09.v1.0330         ------------------------------------------------------------
Tb09.160.5400        ------------------------------------------------------------
Tb927.1.170          ------------------------------------------------------------
Tb09.160.5430        ------------------------------------------------------------
Tb927.1.5220         ------------------------------------------------------------
Tb927.5.120          TGAAGTGGTGCCTATCATTTCTGAGTAACCGAACTGGCAGAGTGAGATTCAAGGAGAAGC
3865


Tb927.1.5080         ------------------------------------------------------------
Tb11.1000            ------------------------------------------------------------
Tb927.5.4620         ------------------------------------------------------------
Tb427.BES122.10      ------------------------------------------------------------
Tb09.v1.0330         ------------------------------------------------------------
Tb09.160.5400        ------------------------------------------------------------
Tb927.1.170          ------------------------------------------------------------
Tb09.160.5430        ------------------------------------------------------------
Tb927.1.5220         ------------------------------------------------------------
Tb927.5.120          TTTCCAGAAGCAGAACATTTGAGCGAGGAGTGCCACAAGGAACTGTCATTGGCCCAATCA
3925


Tb927.1.5080         ------------------------------------------------------------
Tb11.1000            ------------------------------------------------------------
Tb927.5.4620         ------------------------------------------------------------
Tb427.BES122.10      ------------------------------------------------------------
Tb09.v1.0330         ------------------------------------------------------------
Tb09.160.5400        ------------------------------------------------------------
Tb927.1.170          ------------------------------------------------------------
Tb09.160.5430        ------------------------------------------------------------
Tb927.1.5220         ------------------------------------------------------------
Tb927.5.120          TGTTCATTATTGCCATGAACTCGTTGAGCCAACGCCTTGCAGAAGTGCCGTTACTGCAGC
```

```
3985


Tb927.1.5080         ------------------------------------------------------------
Tb11.1000            ------------------------------------------------------------
Tb927.5.4620         ------------------------------------------------------------
Tb427.BES122.10      ------------------------------------------------------------
Tb09.v1.0330         ------------------------------------------------------------
Tb09.160.5400        ------------------------------------------------------------
Tb927.1.170          ------------------------------------------------------------
Tb09.160.5430        ------------------------------------------------------------
Tb927.1.5220         ------------------------------------------------------------
Tb927.5.120          ACGGATTCTTTGCAGACGACCCGACGCTACTTGCGAGGGACACAGAGAGGGATGTCATCA
4045


Tb927.1.5080         ------------------------------------------------------------
Tb11.1000            ------------------------------------------------------------
Tb927.5.4620         ------------------------------------------------------------
Tb427.BES122.10      ------------------------------------------------------------
Tb09.v1.0330         ------------------------------------------------------------
Tb09.160.5400        ------------------------------------------------------------
Tb927.1.170          ------------------------------------------------------------
Tb09.160.5430        ------------------------------------------------------------
Tb927.1.5220         ------------------------------------------------------------
Tb927.5.120          ACCACACACTACAGTGCGGTCTCAACGTGGTGTTACAGTGGTCACAAGAGTACTTCATGT
4105


Tb927.1.5080         ------------------------------------------------------------
Tb11.1000            ------------------------------------------------------------
Tb927.5.4620         ------------------------------------------------------------
Tb427.BES122.10      ------------------------------------------------------------
Tb09.v1.0330         ------------------------------------------------------------
Tb09.160.5400        ------------------------------------------------------------
Tb927.1.170          ------------------------------------------------------------
Tb09.160.5430        ------------------------------------------------------------
Tb927.1.5220         ------------------------------------------------------------
Tb927.5.120          CTGTCAATGTAGCGAAAACAAAGTGCACACTCTTCGGGTGTATAGAGCGCCACCCCCTTA
4165


Tb927.1.5080         ------------------------------------------------------------
Tb11.1000            ------------------------------------------------------------
Tb927.5.4620         ------------------------------------------------------------
Tb427.BES122.10      ------------------------------------------------------------
Tb09.v1.0330         ------------------------------------------------------------
Tb09.160.5400        ------------------------------------------------------------
Tb927.1.170          ------------------------------------------------------------
Tb09.160.5430        ------------------------------------------------------------
Tb927.1.5220         ------------------------------------------------------------
Tb927.5.120          CATTGCAACTGGACGGCGAAAGAATAGGAGCTGACAGGACACCGAAGCTTCTAGGAGTAA
4225


Tb927.1.5080         ------------------------------------------------------------
Tb11.1000            ------------------------------------------------------------
Tb927.5.4620         ------------------------------------------------------------
Tb427.BES122.10      ------------------------------------------------------------
Tb09.v1.0330         ------------------------------------------------------------
Tb09.160.5400        ------------------------------------------------------------
Tb927.1.170          ------------------------------------------------------------
Tb09.160.5430        ------------------------------------------------------------
```