Assessing stumpy formation and stumpy-specific gene expression in *Trypanosoma brucei*

Paula MacGregor

PhD – The University of Edinburgh - 2010

Abstract

During the bloodstream stage of the *Trypanosoma brucei* lifecycle, the parasite exists in two different states: the proliferative slender form and the non-proliferative, transmissible, stumpy form. The transition from the slender to stumpy form is stimulated by a density-dependent mechanism and is important in infection dynamics, ordered antigenic variation and disease transmissibility. The slender to stumpy transition and the contribution of stumpy formation to within-host dynamics have been difficult to analyse, however, because cell-type specific markers have been restricted to imprecise morphological criteria.

PAD1 is a recently identified stumpy-specific protein which acts as a molecular marker for stumpy formation and a functional marker for transmission. Here, the control of stumpy-specific gene expression via the 3'UTR has been analysed, identifying that there are repressive elements in the 3'UTR preventing inappropriate expression during the slender life stage. Further, both pleomorphic and monomorphic transgenic reporter cell lines utilising the *PAD1* 3'UTR have been created that report on stumpy formation *in vitro* and these have been used for the analysis of stumpy-inducing chemical compounds.

Finally, a sensitive and accurate qRT-PCR assay has been developed and optimised that faithfully reports both parasitaemia and stumpy formation throughout host infection. Using a chronic infection rodent model, stumpy levels have been monitored on the basis of conventional morphological and cell cycle assays, as well as by qRT-PCR for *PAD1* expression. The results define the temporal order of events that result in the generation of stumpy forms early in a parasite infection and thereafter describe the dynamics of slender and stumpy forms in chronic infections extending over several weeks. This quantitative data has allowed the mathematical modelling of transmission competence in trypanosome infections, suggesting dominance of transmission stages throughout infection.

Declaration

I declare that all material presented in this thesis is my own work, unless otherwise stated, and has not been submitted for any other degree previously.

Paula MacGregor

Acknowledgements

Firstly, and most importantly, I would like to thank Keith Matthews for his guidance, advice and unwavering support throughout this project. I particularly would like to thank Keith for introducing me to trypanosomes and for providing me with a project that I always found interesting. I also appreciate the numerous opportunities for professional development that were extended to me throughout the duration of my PhD.

I would also like to thank all members of the Matthews laboratory, both past and present, for all of the ideas they have contributed to this project. It has been a pleasure to work beside such a supportive, knowledgeable and interesting group of scientists. I would particularly like to thank Katelyn Fenn for training me when I first started, and for always finding the time to give me advice when I needed it most.

I would like to thank Deborah Hall who, along with Keith Matthews, carried out all animal handling and provided technical (and moral) support; Petra Schneider for advice regarding qRT-PCR methods; and Martin Waterfall for assistance with flow cytometry. I would also like to thank Sarah Reece, my second supervisor, for taking a genuine interest in this project, providing useful ideas and feedback throughout. Nick Savill was a key collaborator for this project, producing the mathematical model and associated figures. The drug screening was carried out in collaboration with the Drug Discovery Unit at the University of Dundee with a SULSA grant. I am also very grateful to the Wellcome Trust for financial support.

On a more personal note, I would like to thank Steph, Catriona, Lowri, Antoine, Laura, Ruth, Olivia, Sarah, Ellie and Jayna who all, in their own way, have provided friendship and laughter that has made my PhD more enjoyable. I would like to express my love and gratitude to my family and friends, who have always been there for me, especially my Dad and Sandra, who I know have always believed in me. Finally, I would like to say a huge thank you to Andrew, for his continued support, love and encouragement and, most importantly, for always being able to make me smile.

Abbreviations

AC	Adenylyl Cyclases
ARE	AU-Rich Element
ATP	Adenosine Triphosphate
BES	Bloodstream Expression Sites
BLAST	Basic Local Alignment Search Tool
bp	Base Pair(s)
BSA	Bovine Serum Albumin
BSF	Bloodstream Form
cAMP	cyclic adenosine monophosphate
CAT	Chloramphenicol Acetyl Transferase
CCA	citrate or *cis*-aconitate
cDNA	complementary DNA
CFP	Cyan Fluorescent Protein
CNS	Central Nervous System
COX	Cytochrome Oxidase
CSF	Cerebrospinal Fluid
DALY	Disability Adjusted Life Years
DAPI	4' 6' Diamino-2-phenyllindole
DFMO	Difluoromethylornithine
dH_2O	distilled water
DIG	digoxigenin
DNA	Deoxyribonucleic Acid
DNDi	Drugs for Neglected Diseases initiative
ELISA	Enzyme Linked Immunosorbant Assay
ESAG	Expression-Site Associated Gene
EtBr	Ethidium Bromide
FACS	Flow Activated Cell Sorting
FAZ	Flagellar Attachment Zone
FC	Flagella Connector
FCS	Foetal Calf Serum
FIND	Foundation for Innovative New Diagnostics
FITC	Fluorescein isothiocyanate
FP	Flagellar Pocket
g	Gravities
GFP	Green Fluorescent Protein
GLM	General Linear Model
GUS	beta-glucuronidase

HAT	Human African Trypanosomiasis
IP	Intraperitoneal
kDNA	kinetoplast DNA
LB Broth	Luria Bertani Broth
MOPS	3-(N-morpholino)propanesulfonic acid
mRNA	messenger RNA
MUG	4-Methylumbelliferyl-β-D-glucopyranosiduronic acid
NCBI	National Centre for Biotechnology Information
noRT	no Reverse Transcriptase
nt	nucleotide(s)
ORF	Open Reading Frame
PAD	Protein Associated with Differentiation
PBS	Phosphate Buffered Saline
PCF	Procyclic Form
PCR	Polymerase Chain Reaction
PDE	Phosphodiesterase
PFR	Paraflagellar Rod
PolI, II and III	RNA polymerase I, II and III
PTRE	Post-Treatment Reactive Encephalopathy
qRT-PCR	Quantitative Reverse Transcription PCR
RFU	Relative Fluorescent Units
RNA	Ribonucleic acid
RNAi	RNA interference
RT-PCR	Reverse Transcription PCR
SIF	Stumpy Induction Factor
SL	Splice leader
SSC	Saline Sodium Citrate
SULSA	Scottish Universities Life Sciences Alliance
TAC	Tripartite Attachment Complex
U	Units
UTR	Untranslated Region
UV	Ultra-Violet
VSG	Variant Surface Glycoprotein
v/v	Volume per volume
WHO	World Health Organisation
w/v	Weight per volume

Table of Contents

1 Chapter 1: Introduction

1.1 Background

Members of the order Kinetoplastida comprise a highly diverged group of eukaryotic organisms which are named after their unique mitochondrial DNA which forms a structure known as the kinetoplast. These organisms are biologically intriguing due to the vast differences in their molecular and cellular biology compared to other eukaryotes, this reflecting their evolutionary divergence. However, they have received particular attention due to a number of the species within this order causing life-threatening and debilitating disease: *Trypanosoma brucei, Trypanosoma cruzi* and *Leishmania* species (such as *L. major, L. infantum* and *L. braziliensis*) are human pathogens causing Human African Trypanosomiasis, Chagas disease, and leishmaniasis respectively.

Human African Trypanosomiasis (HAT), or sleeping sickness, is caused by the protozoan parasite, *Trypanosoma brucei* subspecies *T.b. gambiense* and *T.b. rhodesiense.* The parasite is transmitted between mammalian hosts through the tsetse fly vector. The ability of trypanosomes to infect both domestic and wild animals means that non-human hosts in affected regions can act as a reservoir for the parasite, confounding control of the disease. Moreover, vaccination approaches have not been considered feasible because these parasites evade mammalian immune responses by frequently changing the proteins on their surface in an extreme form of antigenic variation. Due to the severity and prevalence of the disease, as well as the lack of efficient disease management, African trypanosomiasis remains an important public health issue and is considered by the World Health Organisation (WHO) to be a neglected tropical disease (http://www.who.int/en).

Trypanosomes differentiate between distinct life stages in order to prepare for, and adapt to, the different environments they encounter during their life cycle (Vickerman, 1965b; Vickerman, 1985). During the bloodstream stage of the life cycle, trypanosomes exist as either proliferative ‘slender forms’ or non-proliferative, transmissible, ‘stumpy forms’ (Vickerman, 1965b; Vickerman, 1985), with transitional forms between these two types being described as ‘intermediate’ forms.

The proportion of these types changes during the course of infection, this apparently being governed by the density of parasites in the blood (Vassella et al., 1997). The bloodstream trypanosome population is adapted to promote parasite maintenance in the mammalian blood and transmission to tsetse flies. Thus, the proliferation of slender forms contributes to the establishment of parasite numbers in the blood and, through antigen variation, to immune evasion. In contrast, the irreversible division-arrest of stumpy forms acts to control the expansion of parasite numbers in the mammalian bloodstream and so prolongs host survival. This, combined with the preferential survival and the adaptations for differentiation of stumpy forms upon tsetse uptake, increases the probability of disease transmission (Tyler et al., 2001).

As the dynamic balance between slender and stumpy forms in mammalian hosts plays a pivotal role in trypanosome infection and transmission, understanding the molecular events implicit in the slender to stumpy form transition and their subsequent differentiation to tsetse-midgut forms (procyclic forms) is of interest and importance, though poorly understood. The absence of markers distinguishing slender from stumpy forms has hindered our understanding of the transmission biology of trypanosomes, as has the difficulty in successfully generating stumpy forms in culture. Here the analysis of stumpy formation in *Trypanosoma brucei* is revisited with the use of a new molecular marker for the stumpy-life stage, *PAD1* (Dean et al., 2009) which is differentially expressed between slender and stumpy forms at both the mRNA and protein level.

1.2 Human African Trypanosomiasis

Human African Trypanosomiasis is caused by two subspecies of *Trypanosoma brucei*; namely *T.b. gambiense* and *T.b. rhodesiense*. *T.b. gambiense* causes chronic disease in west and central Africa (Figure 1.1), with infections lasting years. *T.b. rhodesiense* causes acute infection in east and southern Africa (Figure 1.1), with infections lasting from weeks to months. There has been recent concern that the regions affected by the two sub-species may be beginning to overlap (Picozzi et al., 2005), potentially complicating drug treatment (see section 1.2.2).

The trypanosome is transmitted between hosts by the tsetse fly from the genus *Glossina*. The distribution of African trypanosomes is restricted to 36 countries in sub-Saharan Africa due to the limited distribution of the tsetse vector (Figure 1.1). The disease is therefore localised into foci where both people and tsetse exist in the same area, which is often in rural populations with limited access to healthcare (Fevre et al., 2008). Overall prevalence is low, as transmission through the tsetse fly is inefficient, with as few as 0.1% of tsetse harbouring mature parasites ready to infect the mammalian bloodstream (Brun et al., 2010). However, in rural foci, the local prevalence can be very high, having a large impact on small communities (Fevre et al., 2008).

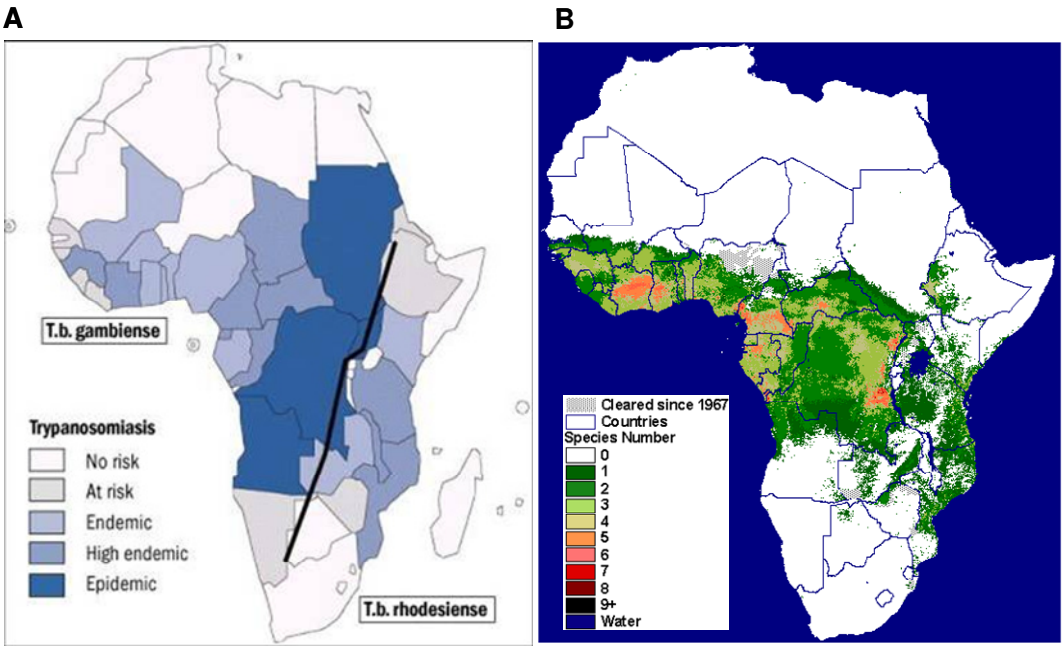


Figure 1.1 The distribution of HAT and the tsetse fly vector in Africa. (A) HAT affects 36 countries in sub-Saharan Africa. *T.b. gambiense* causes HAT in west and north Africa whereas *T.b. rhodesiense* causes HAT in east and south Africa. (B) Various species of the tsetse fly vector are predicted to be distributed throughout sub-Saharan Africa. The distribution of HAT is limited by the distribution of the tsetse vector. Figures from (A) (WHO, 2000) and (B) the Environmental Research Group Oxford (http://ergodd.zoo.ox.ac.uk/tseweb/index.htm), using satellite imagery.

As well as being a human pathogen, *T.b. rhodesiense* is a zoonosis, with animals acting as an important reservoir for disease (Welburn et al., 2001). The parasite can affect wild animals, but it is cattle and other livestock that are the main reservoir for

HAT due to their large numbers and close proximity to humans. The cattle form of trypanosomiasis is known as nagana, meaning 'depressed' in Zulu: It causes weakness, thinning and discharge from the eyes and the animal will die. Although *T.b. gambiense* has been identified in animals, such as domestic pigs (Simo et al., 2006), they are not thought to play as important a role in transmission as for *T.b. rhodesiense.*

Due to the distribution of the disease, it is difficult to estimate the number of cases of HAT per year (Fevre et al., 2008). WHO currently estimates that there are approximately 50,000-70,000 current cases of HAT (WHO, 2006). As well as causing mortality, HAT causes morbidity and the estimated disability-adjusted life years (DALYs) lost through HAT in 2000 were 1,534,845 (WHO, 2003). Along with the effect on human health, the inability to farm cattle in tsetse dense locations and the death of cattle due to nagana means that African trypanosomes have a huge economic effect in affected regions. It is difficult to assess the burden of the disease, which in turn makes it difficult to assess the value of different intervention strategies such as population screening, vector control and treatment of the animal reservoir (Allsopp, 2001; Fevre et al., 2008).

Over recent history, the number of cases of HAT has fluctuated according to the levels of population screening for the disease. As *T.b. gambiense* infection is often undiagnosed, active screening for, and treatment of, otherwise undiagnosed cases can reduce the prevalence of HAT within an area. This is turn reduces the risk of transmission. Indeed, in the early 20th century disease levels were reduced through population screening programmes (Figure 1.2). However, from the 1970s-1990s civil unrest caused surveillance programmes to stop and the displacement of populations infected with the disease, causing the number of HAT cases to rise once again (Barrett, 1999; Brun et al., 2010). Re-introduction of population screening has started to reduce HAT cases once more, demonstrating the value of these programmes (Barrett, 2006; Brun et al., 2010). However, as the disease is that of rural populations, it can often be difficult to access the affected areas and provide the appropriate treatment (see section 1.2.2) (Brun et al., 2010).

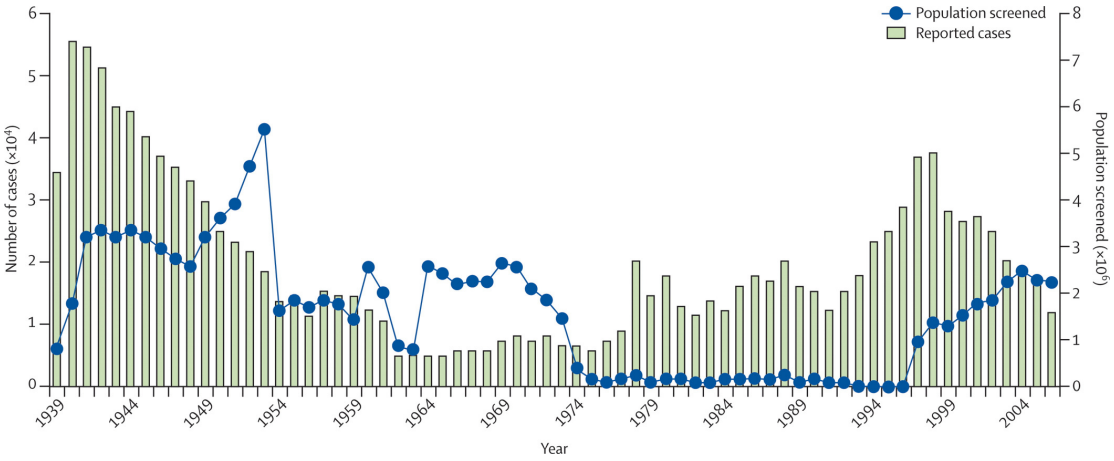


Figure 1.2 The number of cases new cases of *T.b. gambiense* compared to the population actively screened from 1939-2006. The number of cases reporter (bars) decreased throughout the 1940-50's due to increased active screening (circles). After active screening stopped in the 1970s-1990s the number of new cases rose once again. An increase in population screening since 1997 has seen the number of new cases begin to fall again. Figure from (Brun et al., 2010)

1.2.1 Symptoms

Trypanosomes are extracellular parasites, which exist in the blood and lymph of the mammalian host, before crossing the blood-brain barrier into the CSF later in infection. The disease can be divided into two clinical stages correlating with the location of the parasite and the associated symptoms: the early haemolymphatic stage and the later neurological stage. In the early stage of the disease, symptoms include intermittent fever, headache and joint pain. At this stage many cases are undiagnosed due to the lack of specific symptoms, despite availability of effective treatment for this stage of the disease (see section 1.2.2). The later stage of the disease occurs when the trypanosomes move into the CNS.

The blood-brain barrier separates the brain parenchyma from the circulating blood and is formed from the tightly bound epithelial cells that line the cerebral capillaries and selectively prevents the passage of molecules from the bloodstream into the brain. It is thought that the parasites move into the brain through an active process that causes damage to the blood-brain barrier over time and appears to be linked to migration of inflammatory cells into the CNS as reviewed in (Rodgers, 2009). Once

the parasites cross the blood-brain barrier and enter the CNS the neurological symptoms occur. These symptoms include, but are not limited to, the disruption of circadian rhythm, after which the disease is named sleeping sickness, as well as sensory impairment, confusion, motor defects such as slurred speech and lack of coordination as well as psychiatric effects such as aggressive behaviour and irritability as reviewed in (Brun et al., 2010; Kennedy, 2006). For *T.b. gambiense* the infection lasts, on average, 3 years with approximately equal time in the early and late stage of disease (Checchi et al., 2008) whereas *T.b. rhodesiense* infection lasts only weeks or months and so progression to the late stage occurs much faster. Although the disease progresses at different rate, infection with either *T.b. gambiense* or *T.b. rhodesiense* always causes coma and death if untreated.

1.2.2 Current drug treatments for HAT

For the treatment of HAT there are only currently four compounds licensed and used on their own, Pentamidine, Suramin, Eflornithine and Melarsoprol, although a fifth, Nifurtimox, is increasingly used in combination with Eflornithine. All these drugs have associated toxicity and all require hospitalisation for a complex intramuscular or intravenous administration regime. The drug of choice depends on the infecting species and the stage of disease (Figure 1.3). Early stage HAT can be treated effectively with Pentamidine for *T.b. gambiense* and Suramin for *T.b. rhodesiense* with limited side effects (Figure 1.3). However, many cases are undiagnosed at this early stage (see section 1.2) and as such many of the cases presented are in the neurological stage of the disease. For the second stage of *T.b. gambiense* infection, Eflornithine is effective, but is expensive and requires a rigorous administration programme (Figure 1.3). For the second stage of *T.b. rhodesiense* infection, and often in cases where Eflornithine is not available for *T.b. gambiense* infections, the arsenic-based drug, Melarsoprol is used (Figure 1.3). Melarsoprol has severe side effects, including post-treatment reactive encephalopathy (PTRE), which causes coma and death in 5-10% of patients. Despite the risk associated with this drug, HAT will always cause death if untreated and therefore this treatment is still used.

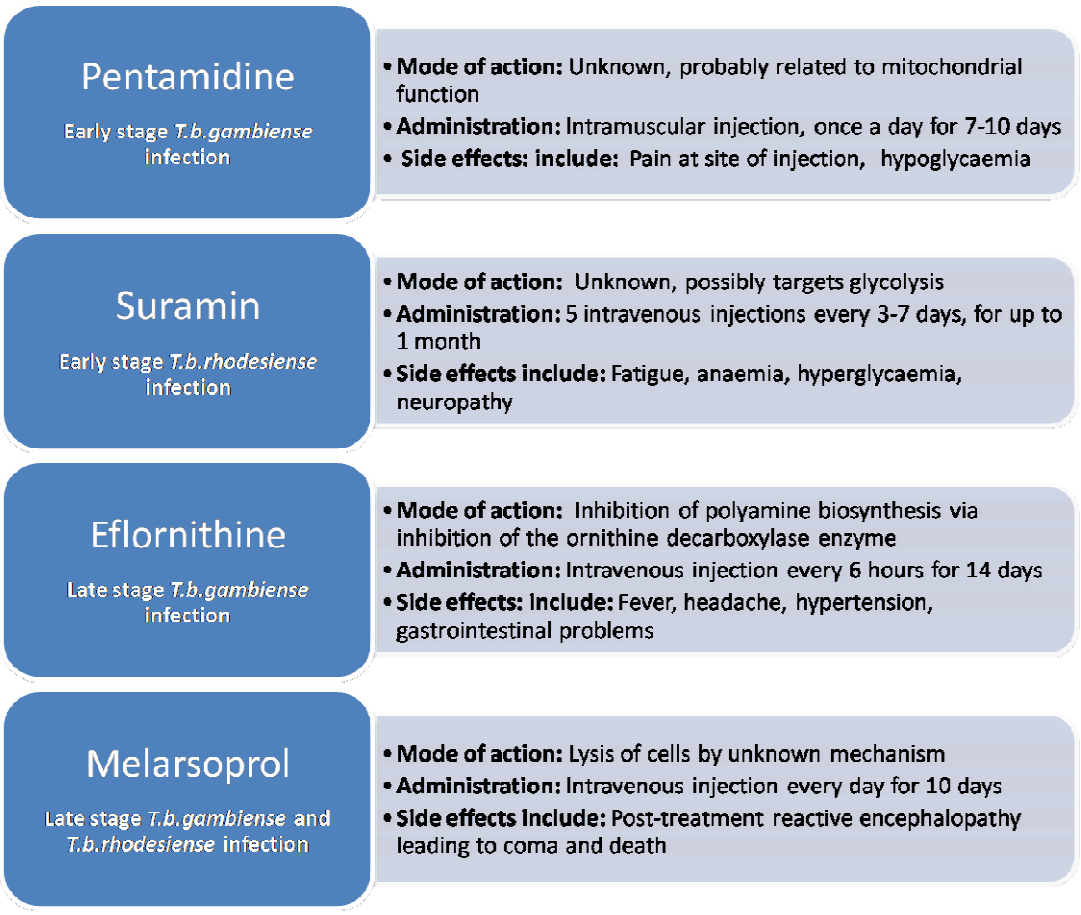


Figure 1.3 Drugs currently licensed for use against HAT. Figure from data gathered in (Barrett et al., 2007; Brun et al., 2010)

1.2.3 New developments in the treatment of HAT

HAT is a disease afflicting a limited number of very poor populations and as such it is not a financially attractive target for pharmaceutical companies (Barrett et al., 2007; Trouiller et al., 2002). Fortunately, WHO currently receive the licensed drugs free of charge from the pharmaceutical companies Bayer and Sanofi-Aventis. WHO then distribute the drugs through Médecins Sans Frontières. However, all current drug treatments have associated toxicity and require hospitalisation for complex treatment regimes, hence, there is a requirement for the production of new, safe, and easy to administer, compounds for the treatment of HAT. It is not thought possible to create a vaccine against HAT, due to antigenic variation (see section 1.3).

Encouragingly, various initiatives to drive the development of new drugs against HAT, and other neglected disease, have been established in recent years. These include, but are not limited to: the foundation of the Drugs for Neglected Diseases initiative (DNDi), which is a non-profit collaborative project for the research and development of new treatments for HAT, among neglected other disease; the Wellcome Trust has established the Drug Discovery Unit at the University of Dundee for the translation of academic research into drugs for clinical trial for neglected disease; The Bill and Melinda Gates Foundation have committed large sums of money to the funding of the improvement of diagnostics and drug treatments for a variety of diseases including HAT; finally, the Foundation for Innovative New Diagnostics (FIND) are improving existing diagnostic methods, and developing new ones, which, in combination with active screening programmes, should help to decrease HAT prevalence.

New potential drugs and drug targets are beginning to emerge. A pro-drug which could be administered orally, pafuamidine maleate (also known as DB289), has undergone clinical trials for use against early-stage HAT (Ansede et al., 2004). Unfortunately, the drug was shown to cause renal disease in a number of healthy volunteers and at least two HAT patients, causing the clinical trial for this drug to be discontinued (Pholig, 2008), although related drugs are now being studied (Wenzler et al., 2009). The NMT (myristoyl-CoA-protein N-myristoyltransferase) gene has been identified as a potential drug target (Price et al., 2003) and inhibitors have been shown to cure trypanosomiasis in mice by acting directly on the protein (Frearson et al., 2010) providing a validated drug target for HAT. Hopefully, in the future, new drugs which can be delivered orally with limited toxicity will be available for the treatment of both early and late-stage HAT.

1.2.4 Vector control

As well as treatment of those affected with HAT, efforts are being made to prevent the transmission of the disease in the first place via vector control (Allsopp, 2001). Spraying the habitats of tsetse with insecticides can be used to kill the flies, although

this method is environmentally undesirable. Fly traps infused with odours that attract the tsetse can reduce the density of tsetse around communities or farms. However, these methods require sustained commitment to the clearing of tsetse from an area, otherwise the tsetse re-populate the cleared area. Historically, sustained commitment has not been observed, largely due to political unrest and financial instability (Allsopp, 2001). Finally, sterile males can be introduced into the environment, but this is unlikely to cause eradication of the tsetse in the large expanse of mainland Africa (Brun et al., 2010). Due to the key role played by tsetse in HAT, vector control should play a key role in disease management given adequate commitment.

1.3 Antigenic Variation

The trypanosome is an extracellular parasite, which, during the bloodstream stage of the lifecycle (see section 1.6.1), exists in the blood and tissue fluids of the host. As such, the parasite is fully exposed to the host immune system. To evade clearance by the host, trypanosomes undergo antigenic variation: at any one time the parasite surface is composed of only one type of surface antigen, the variant surface glycoprotein (VSG) coat. The parasite periodically switches the VSG such that those cells expressing the current antigen will be recognised and cleared by the immune system whereas those parasites expressing a new VSG will avoid antibody clearance and populate the next wave of parasitaemia (see section 1.7, Figure 1.8).

At some of the telomeres of the mega-base chromosomes, there are bloodstream-expression sites (BES) which are polycistronic units containing one VSG gene flanked upstream by 70bp repeats and numerous expression-site associated genes (ESAGs), reviewed in (Rudenko, 2000). Only one BES is expressed at any one time, so that only one VSG is present on the surface at any one time. There are two classes of VSG gene: intact genes and pseudogenes, of which there are approximately 200 and 1000 respectively (Barry et al., 2005; Berriman et al., 2005) located at the end of telomeres on minichromosomes or in subtelomeric regions. Intact genes are activated either by being relocated into the active BES by homologous recombination, using the 70bp repeats, or, for those present in the BESs on the megabase chromosomes,

activation of the BES in which that gene is present. The pseudogenes are able to form mosaic genes by gene conversion (Barry et al., 2005; Berriman et al., 2005). The differential activation rates within and between these two groups of VGS genes, along with the contribution of the slender to stumpy transition, gives rise to ordered antigenic variation which ensures the parasite does not exhaust the VSG repertoire (Lythgoe et al., 2007). Additionally, the capacity to form mosaic genes provides the parasite with a large repertoire of VSG genes that allows survival in the host for many years and makes the possibility of the production of vaccines unlikely.

1.4 The trypanosome cell

Trypanosomes are single cell diploid organisms with many of the organelles existing in just a single copy (Matthews, 2005) (Figure 1.4). The positions of the organelles within the cell are consistent and are shown in Figure 1.4. The shape of the trypanosome cell is defined by the microtubule cytoskeleton (Sherwin and Gull, 1989), of which, all the sub-pellicular microtubules lie in the same orientation with the positive ends at the posterior end of the cell (Robinson et al., 1995), with the exception of a specialist subset of four that form part of the flagellar attachment zone (Robinson et al., 1995; Vaughan et al., 2008) (section 1.4.1).

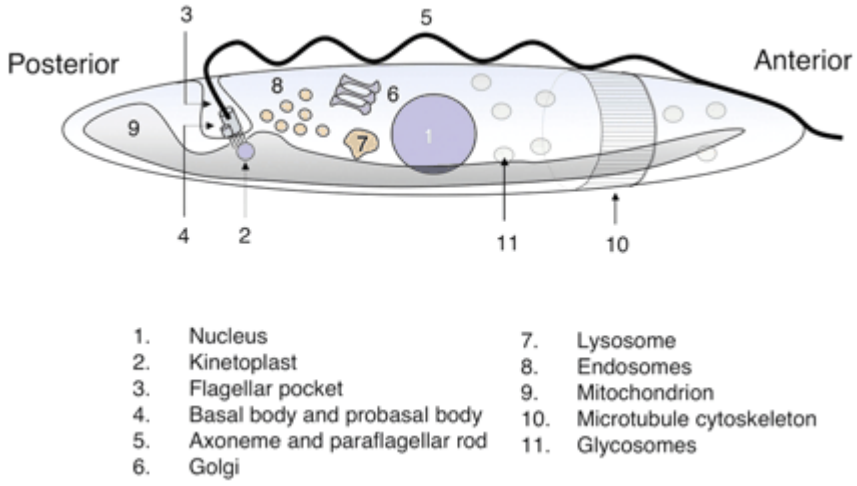


Figure 1.4 Trypanosome cell architecture. The location of the major structural features of the trypanosome cell. Figure from (Matthews, 2005).

1.4.1 The flagellum

Trypanosomes have a single flagellum which is required for cell motility (Bastin et al., 1998) and plays a role in the cell cycle and morphogenesis (Kohl et al., 2003; Moreira-Leite et al., 2001; Robinson et al., 1995). The flagellum is composed of both the 9+2 axoneme and the paraflagellar rod (PFR) which is required for motility (Bastin et al., 1998; Sherwin and Gull, 1989). The flagellum is attached to the cell body along its length at the flagellum attachment zone (FAZ), following a left-handed helix (Kohl et al., 1999; Sherwin and Gull, 1989). The FAZ is a cytoskeletal structure which lies under the plasma membrane from the basal body to the anterior of the cell and which plays a role in cytokinesis (Kohl et al., 2003) (see section 1.5.1). At the base of the flagellum is the basal body which is connected to the kinetoplast by a filament structure called the tripartite attachment complex (TAC) which allows for segregation of the daughter kinetoplasts by the basal bodies during the cell cycle (see section 1.5.1) (Ogbadoyi et al., 2003; Robinson and Gull, 1991).

1.4.2 The flagellar pocket

The flagellum exits the cell body at the flagellar pocket, which is an invagination of the plasma membrane located at the posterior of the cell (Figure 1.4). In trypanosomes the flagellar pocket is the only site of endocytosis and exocytosis for the cell (Field and Carrington, 2009). Due to this, the flagellar pocket is in close vicinity to the single Golgi and to the endoplasmic reticulum. Endocytosis plays an important role in protecting the trypanosome in the mammalian blood as they undergo antibody clearance by endocytosis of VSG-bound antibodies. These bound antibodies move to the flagellar pocket at the posterior of the cell, by hydrodynamic forces as the trypanosome swims forward (Engstler et al., 2007). This helps to protect the cells against complement-lysis. Indeed, in stumpy forms, where the rate of endocytosis is greater, antibodies are cleared faster than in slender forms, allowing stumpy forms to persist for longer in the host which may be important for transmission (see section 1.7.1) (Engstler et al., 2007).

1.4.3 The mitochondrion and kinetoplast

Trypanosomes have a single mitochondrion whose activity is considerably different between life stages. In slender bloodstream forms it is relatively repressed, with no cristae present. The cells obtain their energy entirely from glycolysis using glucose as the main energy source since it is abundant in the mammalian bloodstream (Priest and Hajduk, 1994). Unusually, glycolysis occurs in an organelle called the glycosome and not in the cytosol. Stumpy form trypanosomes also obtain their energy from glucose, but the mitochondrion is elaborated in preparation for the differentiation to the procyclic life stage (see section 1.6). Procyclic forms have a fully developed mitochondrion as they produce ATP via oxidative phosphorylation using the amino acid proline as the main energy source in the tsetse midgut (Priest and Hajduk, 1994).

The mitochondrial genome in trypanosomes is called the kinetoplast. The kinetoplast is a network of DNA with interlocked maxi and mini circles. The maxi circles express mitochondrial genes, with many of these genes requiring RNA editing to form mature mRNAs (Benne, 1994; Benne et al., 1986): the mini circles encode guide RNAs for such editing (Sturm and Simpson, 1990). Dyskinetoplasty, i.e. loss of the kinetoplast, can arise in bloodstream forms, both naturally (i.e. *T.evansi* and *T.equiperdum*) and by chemical mutagenesis (Schnaufer et al., 2002). These cells are unable to survive as the procyclic life stage due to the requirement for mitochondrial function at this stage.

1.5 The trypanosome cell cycle

The trypanosome cell cycle requires the replication and segregation of both the nuclear and mitochondrial DNA. Additionally, many of the organelles exist in single copy and therefore these also need to be duplicated and segregated equally to create two viable daughter cells (Sherwin and Gull, 1989). As such the cell cycle is tightly controlled (Robinson et al., 1995; Woodward and Gull, 1990) (Figure 1.5). The cell cycle in trypanosomes in closely linked with differentiation between life stages (Matthews and Gull, 1994a; Matthews and Gull, 1994b).

1.5.1 Events during the trypanosome cell cycle

During the trypanosome cell cycle, both the nuclear DNA and the kinetoplast must be replicated and segregated although these events are not synchronised between the two (Figure 1.5A). The first step of the cell cycle is the kinetoplast entering S phase (Woodward and Gull, 1990). There is elongation and maturation of the pro-basal body into a new mature basal body, posterior to the original basal body, out of which the new flagellum forms (Sherwin and Gull, 1989; Woodward and Gull, 1990). In procyclic forms the new flagellum tracks along the side of the old flagellum and remains physically associated with the old flagellum at the distal tip of the new flagellum by the flagella connector (FC) (Moreira-Leite et al., 2001) which is produced either before or concurrently with the formation of the new flagellum (Briggs et al., 2004). This structure dictates the location of the new flagellum and the associated cytoskeletal structures such as the FAZ. Thus, the FC provides a method of cytotaxis for cell morphology during the trypanosome cell cycle (Moreira-Leite et al., 2001). Despite the important role played by this structure in procyclic forms, a flagellar connector is not present in bloodstream forms and the new flagellum does not follow the path of the old flagellum as closely (Briggs et al., 2004).

While the flagellum continues to elongate, the nuclear DNA enters S phase (Woodward and Gull, 1990), then the basal bodies begin to separate. As the basal bodies are attached to the kinetoplast, via the tripartite attachment complex (Ogbadoyi et al., 2003; Robinson and Gull, 1991), the movement of the basal bodies causes segregation of the two kinetoplasts (Ogbadoyi et al., 2003; Robinson and Gull, 1991). This also results in the production of a second flagellar pocket (Moreira-Leite et al., 2001). Kinetoplast segregation can occur before the end of nuclear S-phase or during G2 (Siegel et al., 2008). As kinetoplast segregation occurs prior to mitosis, the cells are intermittently 2K1N (2 kinetoplasts, 1 nucleus).

The nucleus then undergoes mitosis, without the breakdown of the nuclear envelope. There is a mitotic-entry checkpoint where the nuclear DNA must have successfully undergone S phase before mitosis can begin (Ploubidou et al., 1999). There does not, however, appear to be a mitosis-dependent cytokinesis checkpoint in trypanosomes,

with cytokinesis progressing even if mitosis fails (Ploubidou et al., 1999; Robinson et al., 1995). Rather it appears that correct kinetoplast segregation is the required cell-cycle checkpoint for entry into cytokinesis (Ploubidou et al., 1999). A cleavage burrow forms from the anterior end of the cell, at the end of the FAZ (Kohl et al., 2003; Robinson et al., 1995), and moves towards the posterior end between the two flagella which separates all the duplicated organelles into the two daughter cells (Sherwin and Gull, 1989; Vaughan and Gull, 2003; Woodward and Gull, 1990).

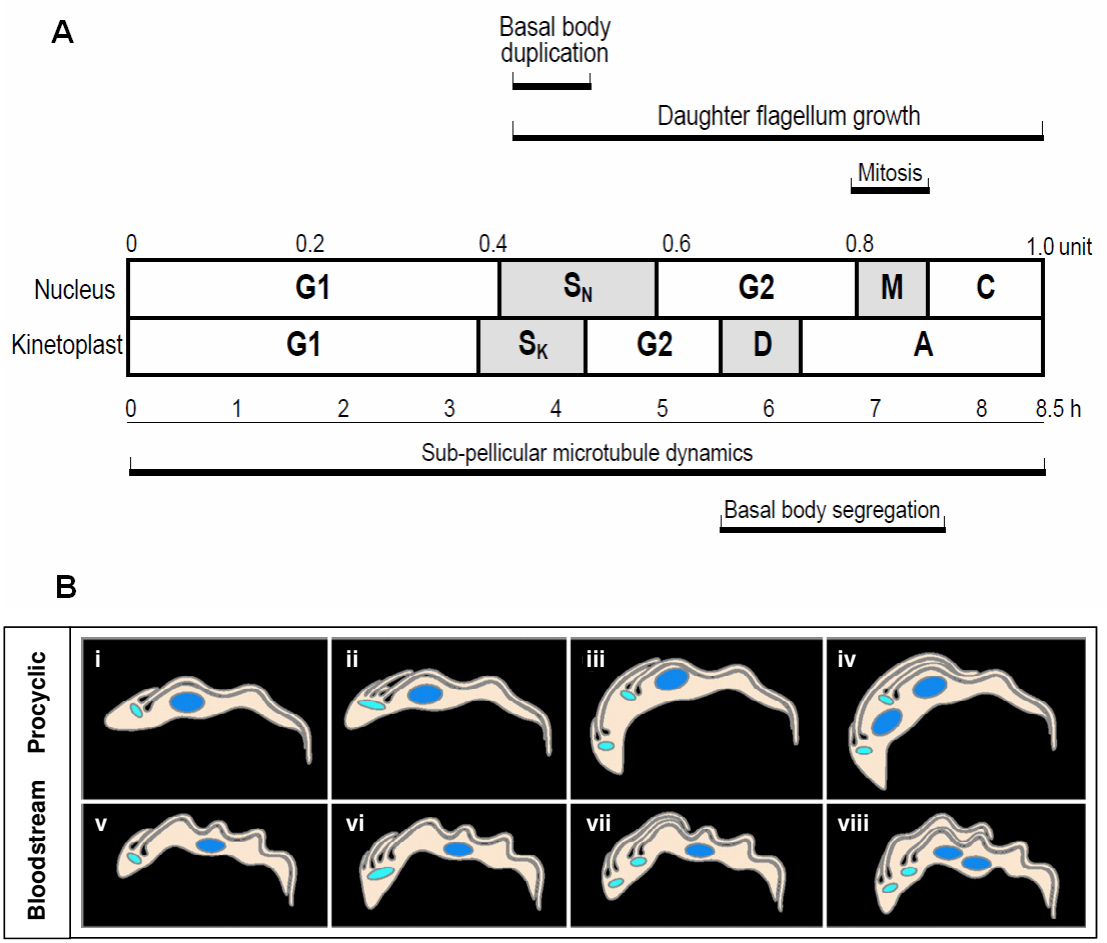


Figure 1.5 The order of events during the *T.brucei* cell cycle. (A) A summary of the order of key events during the cell cycle. (S_N, nuclear S phase; M, mitosis; C, cytokinesis; S_K, kinetoplast S phase; D, kinetoplast division; A, kinetoplast segregation). (B) The cell cycle in (i-iv) procyclic forms and (v-viii) bloodstream forms. The segregation of the kinetoplast is more limited in bloodstream forms than procyclic forms (compare iii with vii). In procyclic forms the duplicated nucleus and kinetoplast form a KNKN configuration as in (iv) where as in bloodstream forms they form a KKNN configuration as in (viii). Figure from (A) (Ploubidou et al., 1999) and (B) modified from (Briggs et al., 2004).

Although the cell cycle has been characterised in great detail, most of the analysis has been carried on procyclic forms. There are, however, differences in the cell cycle at different life-stages. In bloodstream forms, as previously mentioned, there is no FC, and the daughter kinetoplasts and nuclei are in a KKNN arrangement rather than a KNKN arrangement as seen in procyclic forms (Briggs et al., 2004) (Figure 1.5). It may be that in the procyclic forms the FC is required due to the large migration required between segregated daughter basal bodies and kinetoplasts, whereas it is not required in the bloodstream form where there is much less segregation between the kinetoplasts (Briggs et al., 2004). Alternatively, it may be that the bloodstream forms use some other similar structure in place of the FC (Briggs et al., 2004).

The cell cycle is intimately linked with differentiation in trypanosomes (Matthews and Gull, 1994a; Matthews and Gull, 1994b). At six points throughout the life cycle the parasite must differentiate between cell cycle arrested and proliferative forms or vice versa (see section 1.6). Additionally, asymmetric cell division facilitates the changes in cell morphology in at least one of these differentiations (section 1.6.2) (Sharma et al., 2008; Van Den Abbeele et al., 1999).

1.6 The trypanosome life cycle

Trypanosomes encounter many different environments throughout their life cycle: they pass from the blood of the mammalian host to the midgut of the tsetse vector in a tsetse bloodmeal, they then migrate from the midgut to the salivary glands before entering a new mammalian host. The trypanosome must differentiate between different life stages, pre-adapted or adapted for survival and function in each environment (Vickerman, 1965b; Vickerman, 1985). During differentiation the parasite undergoes various morphological changes: there are changes in the size and shape of the parasite; changes in the relative positions of the kinetoplast and the nucleus; the length and positioning of the flagellum (Vickerman, 1965b; Vickerman, 1985), (Figure 1.6). There are also major changes in gene expression, energy metabolism and cell cycle status, with many of the life-stages being cell cycle arrested.

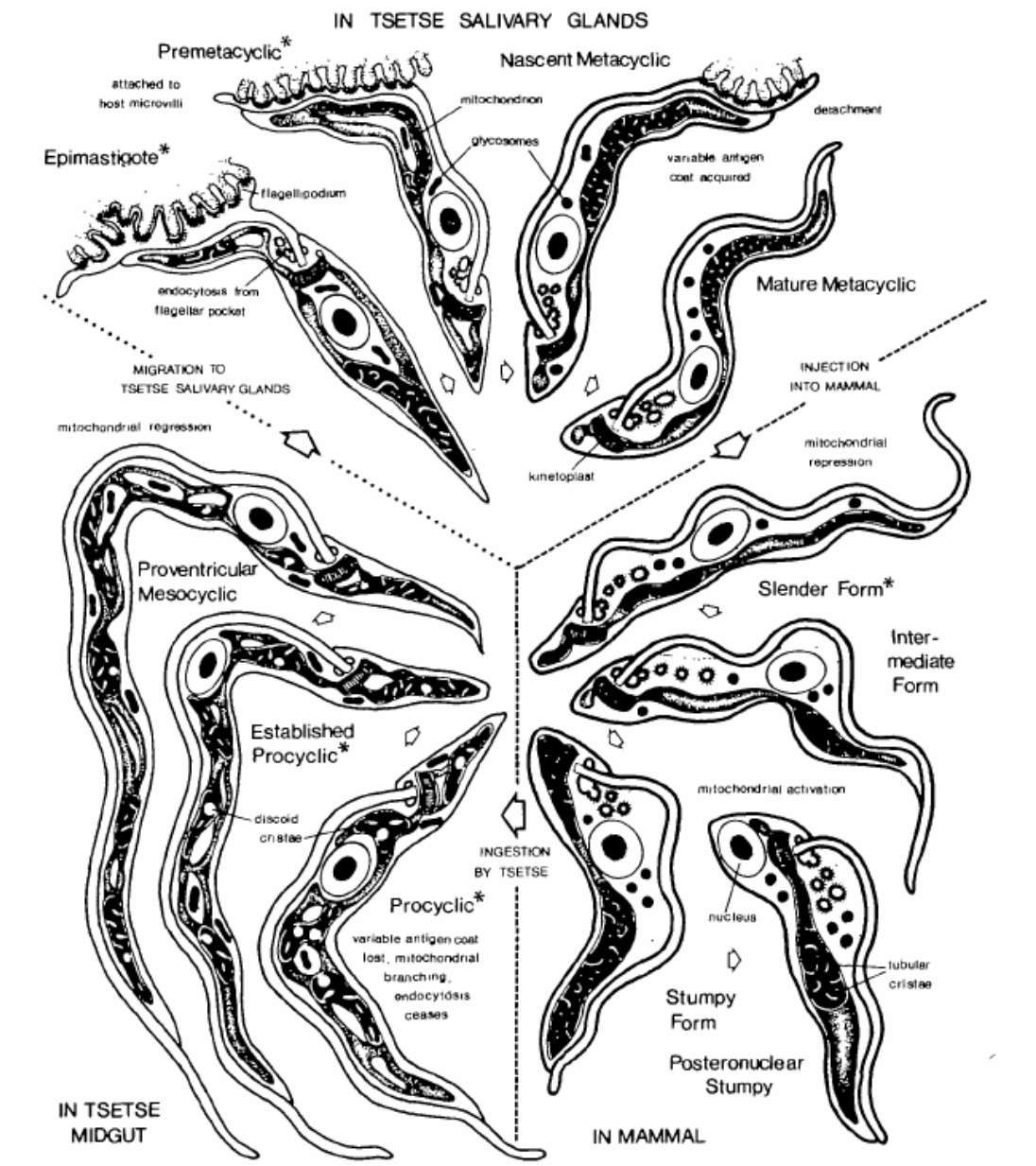


Figure 1.6 Schematic diagram of the life cycle of *Trypanosoma brucei.* The developmental cycle of the trypanosome in the mammalian host and tsetse vector is depicted showing changes in the morphology, VSG expression, proliferation, location, and the mitochondrion. Division occurs in forms labelled with *. (Figure from Vickerman, 1985.)

1.6.1 The bloodstream stages

During the bloodstream stage of the life cycle, trypanosomes exist as either slender forms or stumpy forms (Vickerman, 1965b; Vickerman, 1985), with transitional forms between these two types being described as intermediates (Figure 1.6). The

slender forms are proliferative and ensure maintenance of parasitaemia within the host. In order to evade the host immune system, the slender forms undergo antigenic variation (see section 1.3). In the glucose-rich mammalian bloodstream, slender forms obtain their energy by glycolysis. Therefore, in this stage, the mitochondrion is relatively repressed (Vickerman, 1965b). At high parasitaemia, a proportion of the slender forms differentiate to the non-proliferative stumpy forms. Although previously subject to some controversy (Bass and Wang, 1991), the stumpy forms are the transmissible form, pre-adapted to life in the tsetse vector (Dean et al., 2009). Differentiation of the slender to stumpy forms shall be discussed in section 1.7. As the stumpy forms are pre-adapted to life in the tsetse vector, elaboration of the mitochondrion is observed, in preparation for the change in energy metabolism in the new environment (Priest and Hajduk, 1994; Vickerman, 1965b). These events occur in natural pleomorphic trypanosomes. In the laboratory however, monomorphic cells which are unable to undergo stumpy formation have been created due to serial passage. Monomorphic cells are a useful tool in the laboratory: they are easier to culture than pleomorphic cells, they have a reduced rate of antigen switching, allowing antigenic variation to be studied in close detail, and they can be induced to differentiate to the procyclic life stage *in vitro* allowing the molecular events in the bloodstream to procyclic differentiation to be studied in the laboratory.

1.6.2 The tsetse stages

When trypanosomes are taken into the tsetse midgut they differentiate into procyclic forms. The differentiation from bloodstream to procyclic forms is well characterised and will be discussed in section 1.8. The procyclic forms move from the endoperitrophic space through the peritrophic membrane and into the ectoperitrophic space. The procyclic form uses proline as the main source of energy and as such, there is elaboration of the mitochondria (Priest and Hajduk, 1994).

The ectoperitrophic space becomes dense with parasites and they migrate to the proventriculus and as they do, differentiate to the proventricular mesocyclic form. This cell type is elongated and cell cycle arrested. The parasite then differentiates to

the epimastigote form via an asymmetric cell division (Sharma et al., 2008; Van Den Abbeele et al., 1999), which creates a short epimastigote and another long daughter cell which is thought to die. The epimastigote form moves to the salivary glands and attaches to the epithelial cells via the flagellum.

Epimastigotes differentiate to proliferative premetacyclics and then to attached, cell-cycle arrested, metacyclics which express the VSG coat. The nascent metacyclic become detached to form free mature metacyclics ready for injection into a new mammalian host. It is while in the salivary glands that the trypanosome undergoes meiosis (Bingle et al., 2001), but this stage of the life-cycle is not yet well characterised.

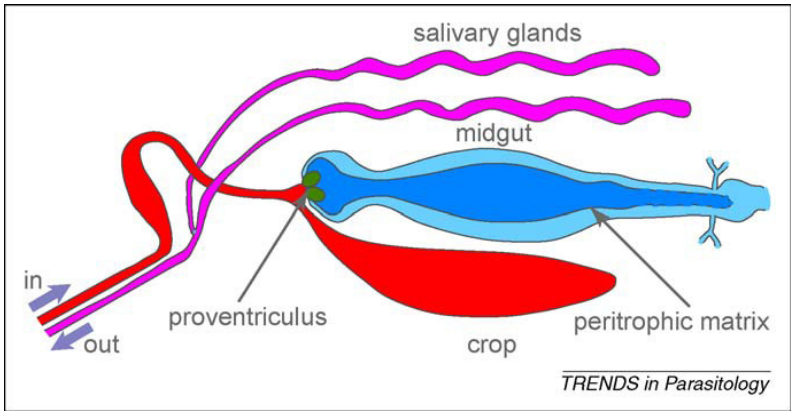


Figure 1.7 Internal organs of a tsetse fly. Proliferative procyclic forms enter the midgut before migrating to the proventriculus where they differentiate to epimastigotes. They then migrate from the proventriculus to the salivary glands where they differentiate to the metacyclic form. Figure from (Sharma et al., 2009).

1.7 The slender to stumpy transition

The slender to stumpy transition plays an important role in trypanosome infection dynamics: it is involved in prolonged host survival, ordered antigenic variation (see section 1.3) and transmission to the tsetse vector (see section 1.7.1).

During the course of a trypanosome infection, the parasitaemia fluctuates due to two factors: antigenic variation (section 1.3) and the slender to stumpy transition (Figure 1.8) (Tyler et al., 2001). As the immune system recognises and clears all parasites expressing the current VSG the parasitaemia will fall. The small number of cells

which have switched to a new antigen type will then proliferate to form the next wave of parasitaemia.

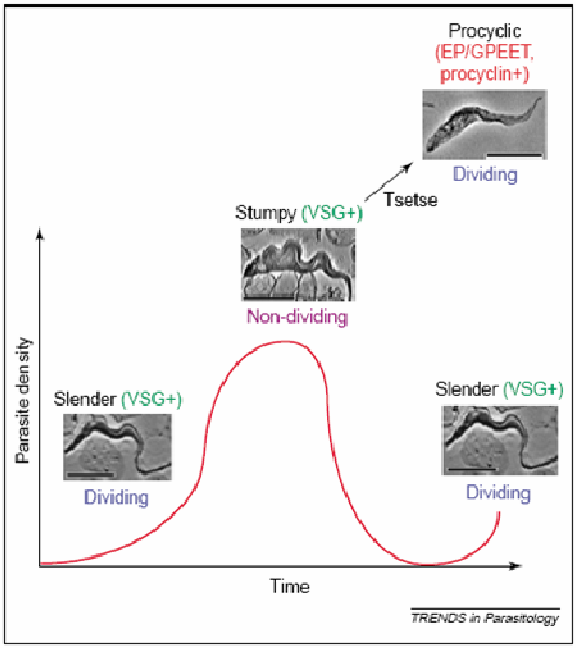


Figure 1.8 The slender to stumpy transition plays a key role in infection dynamics. Slender forms proliferate in the bloodstream causing parasitaemia to rise. As the density of the parasites increase a proportion of slender cells differentiate into the non-proliferative stumpy forms. The stumpy forms are the transmissible form, able to establish a tsetse midgut infection if taken up in a blood-meal. The slender to stumpy transition and antigenic switching both contribute to a fall in parasitaemia. The parasitaemia then rises again as slender cells, expressing a new antigen variant, proliferate. The pattern then continues throughout the infection. By preventing the uncontrolled growth of the slender forms the parasitaemia does not rise to lethal levels in the host, thus prolonging host survival. Figure from (Matthews et al., 2004).

However, oscillations in parasitaemia can occur independently of the host immune system due to stumpy formation (Tyler et al., 2001). As the parasitaemia in the blood increases, a proportion of slender cells differentiate to stumpy forms in a density dependent manner (see section 1.7.3). As the stumpy forms are cell cycle arrested, the parasitaemia peaks and then falls as these non-proliferative stumpy cells die out. The combination of antigenic variation and stumpy formation prevents the parasitaemia from rising uncontrollably and therefore prolongs host survival. Indeed, where mice are infected with monomorphic cells, which are unable to undergo

stumpy formation, the parasitaemia rises uncontrollably and is lethal to the host (Tyler et al., 2001).

1.7.1 The role of stumpy formation in transmission competence

Stumpy cells persist longer than slender cells in the face of the developing immune response in the mammalian bloodstream (Balber, 1972; McLintock et al., 1993). This is probably not caused by antigenic variation: being irreversibly arrested in the bloodstream stumpy forms are unlikely to be able to undergo the DNA recombination events important for antigen switching. Rather, they withstand antibody clearance more effectively than slender cells because they preferentially clear bound antibody by hydrodynamic flow, whereby parasite-bound immunoglobulin is swept backward by the swimming action of the trypanosome in the blood and is then internalised and degraded (Engstler et al., 2007). This may ensure that there are stumpy forms present throughout the infection to ensure transmissibility, given that tsetse bites can be a rare event with seasonal influences (Baylis, 1997).

T.b.gambiense infections last, on average, 3 years (Checchi et al., 2008), hence, each infected individual acts as a reservoir for the disease for a substantial length of time. For *T.b.rhodesiense*, which kills humans much faster, cattle and other animals can harbour infection, providing a disease reservoir (see section 1.2). Due to the fluctuations in parasitaemia and in the proportions of poorly transmissible slender forms and transmissible stumpy forms it may be that there are fluctuations in transmission capacity. Given that tsetse bites are rare, and show seasonality (Baylis, 1997), any variations in transmission capacity could potentially play an important role in progression of the disease in endemic areas. A recent analysis of the relationship between parasite morphology and tsetse infectivity did not reveal a correlation between stumpy forms and infectivity (Janelle et al., 2009), however, this may reflect the requirement of only one parasite for tsetse infectivity (Maudlin and Welburn, 1989). Indeed, in cattle there are periods during chronic infection, where the parasitaemia is sufficiently low that feeding tsetse do not become infected (Van

den Bossche et al., 2005). This, combined with the likelihood of co-infection with different trypanosome strains during the course of a chronic infection (Balmer and Caccone, 2008), generates the potential for complex infection dynamics with important possible implications for disease spread. Further data on the dynamics of proportions of slender and stumpy parasites relative to the total parasitaemia over a chronic infection would help to better understand the role of this transition in transmission and may allow the consequences of potential therapeutic strategies to be predicted (see section 1.12).

1.7.2 The temporal order of events

Differentiation between life cycle stages and the progression of the cell cycle are intimately linked in trypanosomes (Matthews and Gull, 1994b; Szoor et al., 2006). Slender forms are proliferative, yet stumpy forms are arrested in G0/G1. The slender to stumpy transition, therefore, requires exit from the cell cycle as well as a series of morphological changes and stumpy-specific gene expression including the changes necessary for mitochondrial elaboration. Both experimental and mathematical studies have been carried out in an attempt to elucidate the temporal order of events during this transition. From both forms of analysis it appears that the commitment of a slender cell to differentiate to a stumpy form precedes at least one final cell division, if not more, which in turn precedes morphological change (Matthews and Gull, 1994a; Savill and Seed, 2004; Tyler et al., 1997). These analyses have been hindered by the lack of molecular markers for the stumpy life stage (see section 1.9).

1.7.3 Stumpy induction factor

The slender to stumpy transition can occur independently of the host (Seed and Sechelski, 1988; Vassella et al., 1997) and correlates with increased cell density (Hesse et al., 1995; Reuner et al., 1997; Seed and Black, 1997; Vassella et al., 1997), with parasite-parasite signalling operating through a form of quorum sensing. This is proposed to occur through the production by slender forms of a signal, termed Stumpy Induction Factor (SIF), which accumulates and stimulates stumpy formation as parasite numbers increase (Vassella et al., 1997). The identity of SIF, and the

molecular mechanisms of its reception are currently unknown. However, SIF is known to be able to pass through a filter with a 500Da cut-off and so is proposed to be a small metabolite or pheromone-like molecule. It is also known to be stable at 37°C for at least 27 days and able to resist degradation by heating at 90°C for 15 minutes (Vassella et al., 1997).

It was thought that the SIF signal was transmitted via the cAMP pathway, although the cAMP signalling pathway in trypanosomes is poorly understood (see Box 1). It was demonstrated that cAMP levels varied throughout the bloodstream stage of the lifecycle, with cAMP levels, per cell, increasing with parasitaemia during rat infections with pleomorphic cell lines (Mancini and Patton, 1981). The cAMP levels then fell as the population differentiated into the stumpy forms. In the second peak of parasitaemia, the cAMP levels again began to rise as the slender forms proliferated, suggesting that cAMP level was somehow related to morphological change. A monomorphic cell line, by contrast with the pleomorphic cells, maintained a low level of cAMP throughout the infection, regardless of cell density, supporting this hypothesis.

When it was recognised that stumpy formation was induced by the parasite-derived SIF, it was shown that cell permeable cAMP, as well as the phosphodiesterase inhibitor, etazolate, could mimic the action of SIF in a pleomorphic cell line (Vassella et al., 1997). This provided further support for the theory that cAMP is involved in the transmission of the SIF signal. Indeed, later, it was also shown that cell permeable cAMP was able to induce stumpy or 'stumpy-like' formation in a monomorphic cell line (Breidbach et al., 2002), suggesting that monomorphs are able to differentiate to stumpy forms, but are unable to perceive SIF, and that the addition of cAMP into the cell was bypassing the need for the SIF receptor.

Box 1: cAMP signalling in trypanosomes

In the classical mechanisms of cAMP signalling in eukaryotes, cAMP is synthesised by adenylyl cyclases (ACs) from ATP and degraded into AMP by phosphodiesterases (PDEs). cAMP acts mainly via the effector protein, protein kinase A, which in turn phosphorylates transcription factors and therefore results in changes in gene expression. Clearly, given the lack of transcriptional control in trypanosomes (see section 1.10), this is highly unlikely to be the mechanism by which cAMP acts in trypanosomes (Seebeck et al., 2001).

Numerous ACs have been identified in trypanosomes but the ligands of these are yet to be described (Alexandre et al., 1996). Five genes encoding four families of PDE (A-D) have also been identified (Kunz et al., 2006). RNAi against the both of the two PDEB genes causes cell death in bloodstream forms but not in procyclic forms correlating with a reduced increase in intracellular cAMP in procyclic forms compared to bloodstream forms upon RNAi induction (Oberholzer et al., 2007).

Although these components of the classical cAMP signalling pathway have been identified, it is not clear how cAMP signalling occurs in trypanosomes due to the lack of identification of effector molecules (Laxman and Beavo, 2007; Seebeck et al., 2001). Potential cAMP binding proteins have been identified by bioinformatic analysis, but these are yet to be verified (Laxman and Beavo, 2007).

However, this hypothesis was thrown into doubt when more recent evidence demonstrated that the hydrolysis products of the readily hydrolysable analog of cAMP, 8-pCPT-2'-O-Me-cAMP, have anti-proliferative effects and can also induce stumpy-formation, but with greater potency that cAMP (Laxman et al., 2006). Indeed, hydrolysis resistant cell permeable cAMP did not cause stumpy formation (Laxman et al., 2006), arguing against a conventional cAMP signalling pathway in stumpy induction (Laxman and Beavo, 2007; Laxman et al., 2006). Moreover, RNAi knockdown of the *T.brucei* PDEB family causes an increase in cAMP without causing stumpy formation (Oberholzer et al., 2007). Additionally, it was demonstrated that etazolate is a poor inhibitor of *T.brucei* PDEs and therefore the

anti-proliferative effect is not likely caused by an increase in intracellular cAMP (Laxman and Beavo, 2007). Interestingly, etazolate is an adenine/adenosine analog and it is the hydrolysis product, 8-pCPT-2'-O-Me adenosine (also an adenosine analog) that is suggested to be causing the anti-proliferative and stumpy-inducing effects of the hydrolysis products of cAMP (Laxman et al., 2006). The authors propose that the adenosine analogs could be targeting either cyclin-dependent kinases, which play a role in the cell cycle (Hammarton, 2007) or mitogen-activated protein kinases, which have been implicated in stumpy formation (Domenicali Pfister et al., 2006) and will be discussed below. Overall, it is not yet clear what role cAMP or its hydrolysis products plays in stumpy formation, but it is unlikely to be due to a classic cAMP signalling mechanism given the current data.

The trypanocidal drug, eflornithine (or DFMO) has been reported to cause stumpy formation (de Gee et al., 1984) and additional compounds such as troglitazone (Denninger et al., 2007), and Z-Phe-Ala-CHN_2 (Scory et al., 2007) have also been proposed to mimic SIF activity *in vitro,* inhibiting cell growth and generating morphologically "stumpy-like" forms. The biological significance of these different treatments, however, has been difficult to assess in the absence of a functional molecular marker for stumpy forms. With DFMO treatment, although cells appear morphologically stumpy and have increased mitochondrial activity, there is also an increase in the number of multinucleated cells (de Gee et al., 1984) implying an unphysiological transformation. Indeed, in the case of troglitazone, the "stumpy" cells were not, unlike naturally occurring stumpy forms, arrested in G0/G1 (Denninger et al., 2007).

It has been shown that kinases are likely to play a role, downstream of SIF, in the stumpy induction pathway. Firstly, knock-out of the protein kinase, ZFK (Zinc Finger Kinase) in pleomorphic bloodstream cells resulted in decreased cell growth and increased slender to stumpy differentiation in culture, although no effect on parasitaemia was observed in mice infections (Vassella et al., 2001b). Similarly, TbMAPK5 knock-out resulted in increased stumpy formation in pleomorphic cell lines and this effect was observed both in culture and during mouse infections

(Domenicali Pfister et al., 2006). Interestingly, monomorphic TbMAPK5 knock-out cells grew normally and did not undergo stumpy formation. This suggests that the TbMAPK5 is involved in the stumpy induction pathway, perhaps either by affecting the sensitivity of the cell to SIF, or by affecting the downstream pathway (Domenicali Pfister et al., 2006).

1.8 The bloodstream to procyclic differentiation

When a tsetse fly ingests a trypanosome-infected blood-meal, both slender and stumpy form parasites enter the tsetse midgut. Although subject to some controversy (Bass and Wang, 1991), it is generally thought that stumpy forms are pre-adapted to life in the tsetse midgut and only they differentiate efficiently to the tsetse-midgut procyclic form (Dean et al., 2009; Vickerman, 1965b). Cytological evidence based on *in vitro* differentiation experiments suggests that bloodstream forms can only transform to procyclic forms from a given point in the cell cycle, namely within the G1, or the quiescent G0, phase (Ziegelbauer et al., 1990). As stumpy forms are uniformly arrested as a population in G0/G1 and can accumulate to near homogeneity in experimental infections, this allows for their efficient and synchronous differentiation to procyclic forms. In contrast, proliferative slender cell populations are asynchronous in their cell-cycle and must reach G0/G1 before differentiation (Matthews and Gull, 1994b). Additionally, as cells enter the tsetse midgut, they are exposed to harmful proteases and changes in their environmental pH. The stumpy forms are more resistant than slender forms to the effects of such stresses (Nolan et al., 2000) and as such are more likely to endure the transition from the bloodstream to the tsetse midgut. Moreover, stumpy forms have an elaborated mitochondrion compared to slender forms, enabling them to quickly begin oxidative phosphorylation as they enter the glucose-poor tsetse midgut (Priest and Hajduk, 1994; Vickerman, 1965b).

The molecular mechanism by which bloodstream trypanosomes perceive their change in environment and initiate differentiation to procyclic forms once in the tsetse fly has been elusive, until recently. It has long been known that the use of high

concentrations (>3mM) of citrate or *cis*-aconitate (CCA) could induce differentiation of bloodstream form trypanosomes to procyclic forms *in vitro* (Brun and Schonenberger, 1981; Czichos et al., 1986; Hunt et al., 1994). The biological relevance of this signal was in doubt, however, as these levels exceed the concentrations of CCA that a trypanosome would be exposed to in the tsetse midgut (Hunt et al., 1994). However, in 2004 Engstler and Boshart discovered that a drop in temperature from 37°C to 20°C could significantly increase trypanosome sensitivity to CCA, potentially matching levels present within the tsetse blood-meal (Engstler and Boshart, 2004). Since, this temperature drop was compatible with that encountered by trypanosomes as they entered tsetse flies feeding at the cooler ambient temperatures of dusk or dawn, CCA could be considered physiologically relevant. Since temperature drop also allows some molecules regulated access to the parasite surface, Engstler and Boshart proposed a model whereby an unknown surface protein responsible for the reception of the differentiation signal would be retained within the cell in the bloodstream (i.e. when at 37°C) and then trafficked to the cell surface in response to the cooler temperatures associated with tsetse uptake (Baylis, 1997; Engstler and Boshart, 2004).

The molecular basis of this model, however, was unknown until a gene family encoding surface carboxylate transporters known as *PAD* proteins (Proteins Associated with Differentiation) was recently identified (Dean et al., 2009). Indeed, two of the encoded *PAD* proteins (PAD1, PAD2) were shown to be expressed in stumpy forms but not in slender forms and were found to be required for perception of CCA at physiological concentrations (Dean et al., 2009). Matching the predictions of Engstler and Boshart, PAD2 was also demonstrated to be confined to the parasite's flagellar pocket region at 37°C but released to the cell surface at 20°C, thereby demonstrating the redistribution of the transporter in response to cold shock (Dean et al., 2009; Engstler and Boshart, 2004). Such thermoregulatory events are increasingly recognised as being important in insect-borne parasites, including in leishmania and malaria (Fang and McCutchan, 2002; Zilberstein and Shapira, 1994).

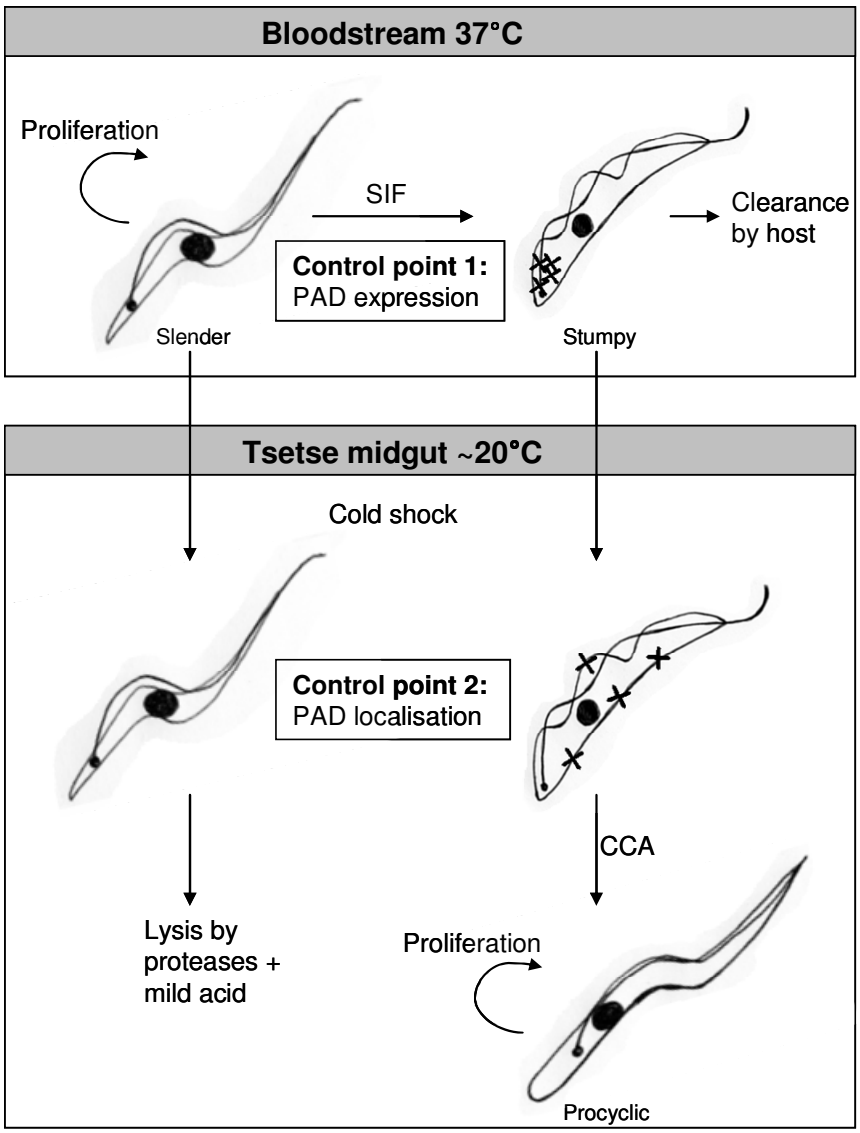


Figure 1.9 PAD protein expression and localisation act as control points for differentiation. Slender cells proliferate in the bloodstream of the mammalian host. As parasite density increases, slender cells produce SIF which induces production of stumpy forms. Stumpy forms express PAD proteins (denoted by X) whereas slender forms do not. This ensures only the transmissible stumpy forms are able to detect the differentiation signal. Upon ingestion in a tsetse blood-meal and exposure to a drop in temperature there is up-regulation and a relocation of at least one PAD protein (PAD2) to the cell surface. Retention of PAD2 within the cell prior to cold shock ensures stumpy forms do not differentiate prematurely. Stumpy forms then differentiate to procyclic forms synchronously in response to CCA. Slender forms do not perceive the signal and are sensitive to proteolytic and potential pH stress in the tsetse midgut and therefore do not survive. Figure from (MacGregor and Matthews, 2010).

The differential expression and localisation of the PAD proteins provides two levels of control during the differentiation to procyclic forms (Figure 1.9). First, differential expression of PAD proteins between slender and stumpy forms ensures only the pre-adapted transmissible stumpy forms are able to receive the differentiation signal. Second, the differential localisation between stumpy cells at 37°C and 20°C ensures that only cells in the appropriate environment (i.e. the tsetse midgut) are able to receive the signal. As citrate is present in blood at ~130µM (Jacobs and Lee, 1964), this secondary control point could be crucial to prevent premature differentiation to procyclic forms whilst within the host, this signal only being detected upon temperature reduction in the tsetse blood-meal.

Before the discovery of PAD proteins, it was already established that the activity of a protein tyrosine phosphatase *Tb*PTP1, inhibits the differentiation of stumpy forms whilst in the bloodstream, this block being removed upon uptake by the tsetse fly (Szoor et al., 2006). Recently, it was demonstrated that *Tb*PTP1 acts via a downstream phosphatase *Tb*PIP39, which stimulates differentiation (Szoor et al., 2010). A model for the early events in bloodstream to procyclic differentiation has been proposed (Figure 1.10) whereby *Tb*PTP1 acts downstream of the PAD proteins as an effector of the CCA signal (Szoor et al., 2010). Specifically, CCA inhibits activation of *Tb*PTP1 which in turn prevents the dephosphorylation of *Tb*PIP39 by *Tb*PTP1. Then, the active, phosphorylated form of *Tb*PIP39 is targeted to the glycosome where it promotes differentiation. Similarly, exposure to proteases (Hunt et al., 1994; Sbicego et al., 1999) or pH stress (Rolin et al., 1998) in the tsetse midgut might also trigger differentiation, either in a complementary manner or as independent cues (Rolin et al., 1998). Although the role of these other environmental cues is yet to be fully understood, the discovery of the PAD proteins provides the first molecular insight into the environmental sensing mechanisms used by the stumpy form to perceive transmission and so initiate differentiation.

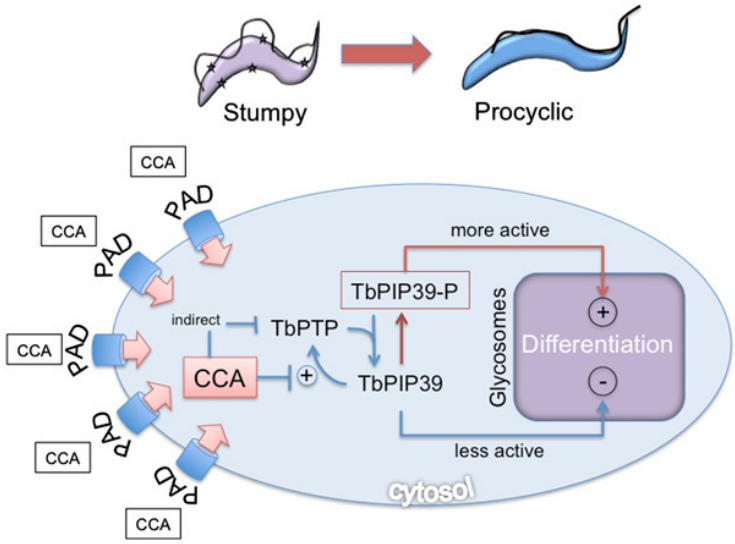


Figure 1.10 A model for the early events in bloodstream to procyclic form differentiation. The CCA signal is transported into the cell via the PAD proteins which are expressed on the surface of stumpy form cells. CCA inhibits the activation of *Tb*PTP1 by *Tb*PIP39 (as well as by indirect mechanisms). Once *Tb*PTP1 is inactivated, *Tb*PIP39 is phosphoylated and moves to the glycosome, promoting differentiation. Without the CCA signal, *Tb*PTP1 is active and dephosphorylates *Tb*PIP39, preventing differentiation and acting as a molecular block to differentiation. Figure from (Szoor et al., 2010).

1.9 Markers for the stumpy life stage

One of the key reasons that the differentiation from bloodstream forms to procyclic forms has been so well studied is that there are clear molecular markers to distinguish each developmental stage. Furthermore, the development between the bloodstream and procyclic forms can be easily mimicked *in vitro* and occurs synchronously. Contrasting with this, the absence of markers distinguishing slender from stumpy forms has hindered our understanding of the transmission biology of trypanosomes, as has the difficulty in successfully generating stumpy forms in culture and the lack of synchronicity in which the population differentiates.

In earlier studies, various markers have been used to identify stumpy forms, including morphology, mitochondrial activation, cell-cycle arrest and the capacity of cells to differentiate synchronously to the procyclic form (Breidbach et al., 2002; Denninger et al., 2007; Laxman et al., 2006; Reuner et al., 1997; Tyler et al., 1997). These assays, however, are far from ideal. For example, analysis by morphology is

subjective and time consuming and is complicated by the presence of morphologically intermediate forms in the population. Similarly, differentiation analyses are time-dependent and require population assays, rather than the analysis of individual cells. Finally, mitochondrial activation, often visualised through the activity of dihydrolipoamide dehydrogenase in the diaphorase assay, may also be activated under non-physiological stress conditions.

With the identification of PAD1 as a surface molecule that identifies cells that are functionally competent for transmission (Dean et al., 2009), as well as developments in parasite culture and transfection (Vassella and Boshart, 1996), the route has been opened up to dissect trypanosome transmissibility at the level of individual cells as well as in the infecting parasite population (see section 1.12)

1.10 Gene expression in African trypanosomes

In the 26Mb *T.brucei* genome there are 11 mega-base size chromosome and numerous smaller intermediate and mini-chromosomes, together encoding over 9000 predicted genes (Berriman et al., 2005). The architecture of the trypanosome genome is unusual compared to other eukaryotes in that genes are arranged into large polycistronic units which seem to lack RNA Polymerase II (PolII) promoters (Figure 1.11). Indeed, gene expression in trypanosomes is unusual in that it is controlled almost entirely post-transcriptionally.

1.10.1 Transcription

In *T.brucei* RNA polymerase I (PolI), as in other eukaryotes, transcribes the ribosomal RNA genes. However, unusually, PolI also transcribes some protein coding genes, namely the bloodstream VSG genes and the procyclic form procyclin genes, which are the major surface proteins for the respective life stages. There are PolI promoters, both in the rRNA loci and upstream of the VSG and procyclin genes (Figure 1.11). RNA polymerase III (PolIII) in trypanosomes transcribes the tRNA as well as the small nuclear RNA and has promoters similar to those seen in other eukaryotes.

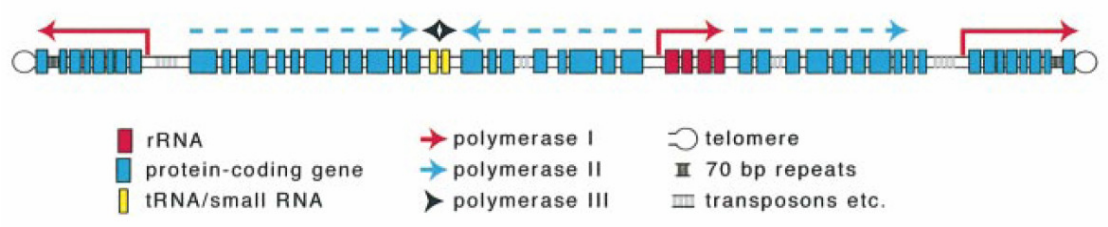


Figure 1.11 A schematic representation of a generic *T.brucei* chromosome. RNA polymerase I transcribes the ribosomal RNA (rRNA) genes and the mRNA for the major surface protein genes for the bloodstream and procyclic life stages. RNA polymerase II transcribes the majority of the protein coding genes, although there is a lack of RNA polymerase II promoters. RNA polymerase III transcribes the tRNA and small nuclear RNA. This diagram does not represent any one chromosome, but rather illustrates typical features of the genomic architecture of the trypanosome. Figure from (Clayton, 2002).

T.brucei PolII transcribes the mRNA (Figure 1.11) as well as the splice leader (SL) RNA (see section 1.10.2). For the SL RNA there is a PolII promoter (Luo et al., 1999). However, for all of the PolII transcribed protein coding genes there are no discernable promoters. Indeed, as already mentioned, the vast majority of protein coding genes are arranged in polycistronic units where all the genes within a unit lie in the same orientation and are co-transcribed. Polycistronic units are separated by strand switch regions, where the genes in the neighbouring unit then lie in the opposite direction. It appears that it is at these divergent and convergent strand switch regions that PolII initiates and terminates transcription, respectively. Recently, it was demonstrated that there are specific chromatin modifications at these probable PolII transcription start and stop sites, such that there is an open chromatin structure at sites of transcription initiation (Siegel et al., 2009). Hence, trypanosome PolII transcription initiation is dependent on chromatin structure.

Genes within a polycistronic unit most often do not represent families of genes with either common function or common patterns of expression, with gene expression controlled largely at the post-transcriptional level (see section 1.11), although some small clusters of differentially expressed genes have recently been identified (Veitch et al., 2010). Once the polycistronic unit is transcribed the pre-mRNA must be processed in order to resolve the transcript into individual mRNAs.

1.10.2 mRNA processing

Once the protein coding mRNA has been transcribed, the transcript must be processed into mature mRNAs (Figure 1.12).

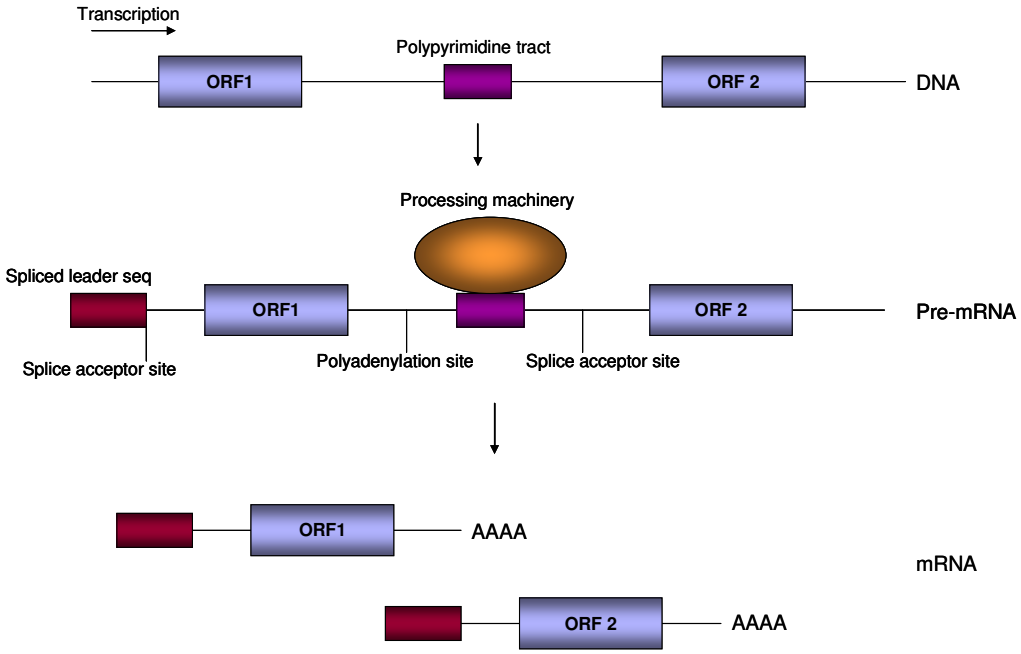


Figure 1.12 mRNA processing in trypanosomes. Transcription occurs in polycistronic units which are required to be processed into independent mRNAs. Trans splicing of the downstream ORF is coupled to polyadenylation of the upstream ORF. A pyrimidine-rich motif between the two is important for controlling the location of this process (Matthews et al., 1994). AAAA, represents the polyA tail. Figure modified from (Vanhamme and Pays, 1995).

First, the SL RNA is modified co-transcriptionally (Mair et al., 2000) such that the transcript then has a hyper-modified 5' end (Bangs et al., 1992; Perry et al., 1987). The first 39-bp of the processed SL RNA is then added to the 5'end of the mRNA transcripts via trans-splicing: this resolves the large polycistronic transcript into the individual mRNAs, each with a 5'cap (Figure 1.12) (De Lange et al., 1984; Milhausen et al., 1984; Parsons et al., 1984). This occurs downstream of a pyrimidine rich motif (Matthews et al., 1994). In trypanosomes there is no known polyadenylation signal. The polyadenylation is coupled to the addition of the 5' cap of the downstream OFR and is dependent on the location of the polypyrimidine motif (Matthews et al., 1994; Ullu et al., 1993). Additionally, for a given mRNA, there can

be more than one polyadenylation site relatively close to one another, reflecting the lack of a precise polyadenylation signal. Indeed, recent genome-wide analysis of polyadenylation sites in bloodstream and procyclic life-stages revealed that, for transcripts with high coverage, there were as many as 10 polyadenylation sites, on average, for each gene (Siegel et al., 2010).

1.11 Post transcriptional control of gene expression

During the trypanosome life cycle, the parasite must differentiate in order to adapt to the varying environments in which it must survive. These differentiation events are associated with changes in large-scale changes gene expression, as has been demonstrated in genome-wide analysis of mRNA abundance between life-stages and during differentiation (Jensen et al., 2009; Kabani et al., 2009; Queiroz et al., 2009; Siegel et al., 2010). Due to the lack of transcriptional control in *T.brucei,* gene expression is controlled largely at the post-transcriptional level by both mRNA stability and translational control.

1.11.1 mRNA stability and translational control

In other eukaryotes, mRNA degradation is generally initiated by deadenylation and subsequent decapping resulting in 5'-3' cytoplasmic exonuclease activity and 3'-5 degradation by the exosome, reviewed in (Clayton, 2002). In trypanosomes, there are two main pathways for mRNA degradation. The first, similar to other eukaryotes, involves deadenylation and then decapping, which results in both 5'-3' and 3'-5' exonuclease activity by the cytoplasmic exonuclease, XRNA, and the exosome, respectively (Estevez et al., 2001; Haile et al., 2003; Irmer and Clayton, 2001; Li et al., 2006). The second pathway involves the rapid degradation of highly unstable mRNAs dependent on XRNA, without the requirement for deadenylation (Irmer and Clayton, 2001; Li et al., 2006; Schwede et al., 2009).

As well as by mRNA stability, post-transcriptional gene expression is controlled by translation. One mechanism by which translation controls gene expression in trypanosomes is via translational selection through codon bias, where highly

expressed genes use favoured codons. Indeed, a strong correlation has been demonstrated between codon usage and gene expression, suggesting that codon bias is a key contributor to differential gene expression in trypanosomes (Horn, 2008). For stage regulated gene expression, developmental control can be determined by RNA-binding proteins causing the association of bound mRNAs with the translational apparatus, such as in the case of *Tb*ZFP3 binding to EP1 mRNA (see section 1.11.3) (Paterou et al., 2006; Walrad et al., 2009). On a more global scale, in stumpy form trypanosomes, there is decreased protein synthesis compared to slender form trypanosomes (Brecht and Parsons, 1998). This is due to decreased association of mRNA with ribosomes during the stumpy life stage, likely due to decreased initiation of translation by some unknown mechanism (Brecht and Parsons, 1998). How some genes, such as *PAD1*, escape this global repression of expression to be upregulated in stumpy forms is unknown.

1.11.2 *Cis*-acting signals in gene regulation

In trypanosomes, regulation of transcript-specific gene expression is controlled predominantly by *cis*-acting signals in the 3'UTR of the mRNA which recruit regulatory RNA-binding proteins and result in differential gene expression between life stages. Many procyclic-specific genes have destabilising AU-rich elements (AREs) in the 3'UTR, which cause them to be highly unstable during the bloodstream life-stage, thus preventing expression (Mayho et al., 2006). For example, cytochrome oxidase (COX) complex genes are developmentally regulated in that they are expressed during the procyclic but not bloodstream life-stage. An analysis of the control of COX gene expression revealed that regulatory regions in the 3'UTRs control translation efficiency, and mRNA stability, in a stage-specific manner (Mayho et al., 2006). The effects of 3'UTR regulatory signals will be transmitted through the actions of RNA binding proteins (section 1.11.3). As well as linear sequences, secondary structures in the 3'UTRs likely contribute significantly to the regulation mRNA stability, with the RNA binding proteins binding to these secondary structures (Clayton, 2002). This complicates the identification of stage-

specific regulatory signals due to difficulties in predicting the secondary structure of RNA molecules.

The majority of protein coding genes are transcribed by PolII (section 1.10.1), but the most well studied 3'UTR is that of the procyclic genes, the major surface proteins of the procyclic life-stage, which are transcribed by PolI (Biebinger et al., 1996). It may be that PolI is utilised for transcription of the major surface proteins in order to obtain very high levels of protein expression (Biebinger et al., 1996; Clayton, 2002). There are four procyclin genes (EP1, EP2, EP3 and GPEET) which are differentially expressed within the procyclic life-stage as they have similar, but distinct 3'UTRs (Vassella et al., 2001a). The EP1 procyclin 3'UTR is composed of three regulatory stem loops: In bloodstream forms, Loop I contains a positive element, Loop II contains a negative element and Loop III contains a 16mer which acts as an anti-repressor of the Loop II element (Furger et al., 1997; Hehl et al., 1994). During the bloodstream life stage, EP1 procyclin mRNA is rapidly degraded due to the presence of an ARE in the Loop II region of the 3'UTR (Furger et al., 1997; Irmer and Clayton, 2001; Mayho et al., 2006).

1.11.3 *Trans*-acting factors in gene regulation

Regulatory signals in the 3'UTR of mRNAs are transmitted through the action of *trans*-acting RNA binding proteins. There are a large number of potential RNA binding proteins encoded in the trypanosome genome (De Gaudenzi et al., 2005) and a few have been validated and characterised. For example, in procyclic forms, the RNA binding protein *Tb*ZFP3 specifically associates with EP1 and GPEET mRNA rather than EP2 and EP3 mRNA (Walrad et al., 2009). It positively regulates expression of EP1 in the procyclic life stage by binding the Loop II and Loop III region on the 3'UTR, potentially competing with an unknown negative regulator (Walrad et al., 2009). Although *Tb*ZFP3 is expressed constitutively throughout the trypanosome life cycle, it only associates with the polysomes during the procyclic life stage (Paterou et al., 2006). Thus, it appears that *Tb*ZFP3 promotes EP1 protein

expression by association with the translational apparatus in a stage specific manner (Walrad et al., 2009).

The EP1 RNA binding protein, *Tb*ZFP3, has a CCCH zinc finger binding domain which is essential for procyclin mRNA association (Walrad et al., 2009). In other eukaryotes, CCCH zinc finger proteins have been shown to regulate post-transcriptional expression by both mRNA degradation and translational control (Kramer et al., 2010). A recent *in silico* analysis of the *T.brucei* genome revealed 48 potential CCCH zinc finger proteins which, with experimental validation, may prove to be further RNA binding proteins (Kramer et al., 2010), the characterisation of which may help further understanding of the mechanisms of post-transcriptional gene expression in trypanosomes.

The RNA binding protein, *Tb*DRBD3, is part of a ribonucleoprotein complex that has been shown to bind to and stabilise a subset of mRNAs, with the majority encoding membrane proteins (Estevez, 2008). This is suggestive of specific RNA binding proteins controlling regulation of functionally related groups of proteins as a post transcriptional regulon (Estevez, 2008). Indeed the PUF9 RNA binding protein has been shown to stabilise mRNAs involved in the cell cycle, specifically during S phase, again indicative of functionally related groups of proteins being regulated by common RNA binding proteins (Archer et al., 2009). Furthermore, genome-wide analysis does demonstrate that, perhaps unsurprisingly, functionally related mRNAs show tightly regulated expression patterns (Queiroz et al., 2009). In the case of both *Tb*DRBD3 and PUF9, association is dependent on specific sequences within the 3'UTR of the mRNA (Archer et al., 2009; Estevez, 2008).

Although the identification and characterisation of 3'UTR recognition motifs and RNA binding proteins is underway, there is still much to be learned about the way in which RNA binding proteins recognise specific mRNAs and regulate post-transcriptional gene regulation, particularly between the different life stages. Further, much of the work so far has focused on stage specific gene regulation between bloodstream and procyclic forms with little known about regulation between slender

and stumpy forms. A large number of genes are differentially expressed between slender and stumpy forms (Jensen et al., 2009; Kabani et al., 2009). Recent whole-genome microarray analysis, comparing slender, stumpy and procyclic forms, revealed two motifs overrepresented in the 3'UTRs of transcripts enriched in the stumpy life stage and one of these motifs showed positional enrichment suggesting that it may be a context dependent regulatory signal (Kabani et al., 2009). It may be that further bioinformatic analysis of databases such as these will identify further potential stage-specific regulatory signals in stage-specific 3'UTRs. However, due to the likely vital role of RNA secondary structure in 3'UTR control (Clayton, 2002), and the difficulties in predicting RNA secondary structure *in silico*, experimental analysis will still be required. Indeed, the recently identified *PAD1* is clearly up-regulated at both the mRNA and protein level in stumpy form cells compared to slender forms (Dean et al., 2009). Experimental analysis of the control of gene expression of *PAD1* may reveal further stumpy-specific regulatory signals which may provide the potential for the identification of stumpy-specific RNA binding proteins (see section 1.12).

1.12 *PAD1* as a tool for analysis of the stumpy life stage

The stumpy life stage plays a crucial role in trypanosome infection dynamics and transmission yet has been difficult to study due to a lack of molecular markers for the life stage (see section 1.9). A new molecular marker for the stumpy life stage has recently been identified. *PAD1* is up-regulated in stumpy forms compared to slender forms at both the mRNA and protein level (Figure 1.13) (Dean et al., 2009). *PAD1* is therefore a molecular marker for the stumpy life stage and for transmission competence (Dean et al., 2009).

PAD1 provides a tool for the analysis of stumpy formation, both *in vivo* and *in vitro* as well as the study of stumpy-specific gene expression. By analysing the regulation of gene expression of *PAD1*, mechanisms of control of repression of expression in slender forms and alleviation of repression in stumpy forms may be identified. Additionally, by coupling reporter genes to the *PAD1* control signals for stumpy

specific gene expression, reporter cell lines for the stumpy life stage can be created. These can then be used to monitor stumpy formation *in vitro* either in response to media containing SIF, or fractions thereof. Alternatively, identification of chemical inducers of stumpy formation may identify downstream targets of the SIF signal. Together these could provide data on the identity of SIF and the stumpy induction pathway. Finally, by monitoring *PAD1* expression relative to parasitaemia throughout chronic *in vivo* infections, data on the dynamics of the slender to stumpy transition could be generated. This data could then be fed into existing models of stumpy formation (Savill and Seed, 2004) and allow refinement of those models. This will provide better understanding of the transmission strategy of trypanosomes and may allow for the potential consequences of therapeutic interventions to be predicted and tested. Moreover, comparing the onset of *PAD1* expression with traditional markers of stumpy formation such as cell cycle arrest and morphology may help to further elucidate the temporal order of events during the transition.

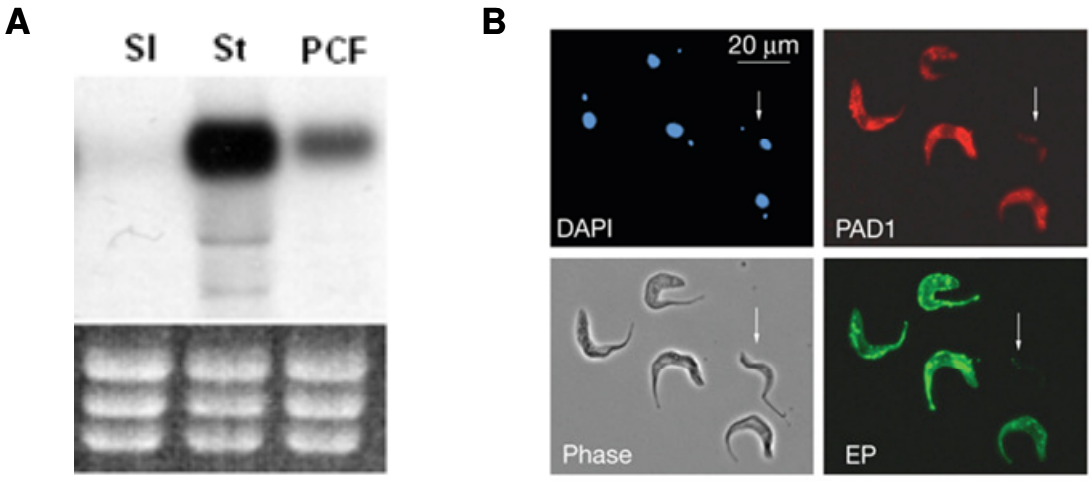


Figure 1.13 *PAD1* is differentially expressed between the slender and stumpy life stages. (A) *PAD1* mRNA is expressed at high levels in stumpy forms compared to slender forms. (B) PAD1 identifies stumpy cells and transmission competence. Cells which are morphologically slender (white arrow) do not express PAD1 on their surface. Cells which do not express PAD1 are not transmission competent, as measured by EP procyclin protein expression after induction of differentiation to the procyclic life stage. Figure modified from (Dean et al., 2009).

1.13 Aims

The aim of this thesis was to study stumpy formation, with the use of *PAD1* as a molecular marker for the purpose of better understanding stumpy-specific gene expression, the stumpy induction pathway and transmissibility in trypanosome infections.

The key aims were as follows:

- Identify the regulatory signals controlling stumpy specific gene expression of *PAD1*.

- Create stumpy-specific reporter cell lines for the analysis of stumpy formation *in vitro.*

- Develop a method to track parasite transmissibility during chronic trypanosome infections.

2 Chapter 2: Materials and Methods

2.1 Trypanosome strains

All monomorphic cell lines used are based on the *Trypanosoma brucei brucei* Lister427 strain. A speculative history of the origin of this strain has been documented (http://tryps.rockefeller.edu/ DocumentsGlobal/lineage_Lister427.pdf) by George Cross at the Rockefeller University

Two pleomorphic trypanosome stains were used in this study. The first is strain *T.b. brucei* AnTat1.1, used for mice infections and harvesting of slender or stumpy cells from mice. The second is *T.b. brucei* AnTat1.1 90:13 (Engstler and Boshart, 2004), a culture adapted pleomorphic cell line used for the creation of the pleomorphic reporter cell line in Chapter 4.

2.2 Trypanosome culture

Bloodstream monomorphic trypanosomes were grown in HMI-9 media (Appendix A) supplemented with a final concentration of 20% foetal calf serum (FCS, Gibco) and 100U/ml penicillin and streptomycin (Gibco) at 37°C in 5% CO_2 (Hirumi and Hirumi, 1989). Cells were grown at a concentration of 1 x 10^5 to 5 x 10^6 cells/ml in 25cm^2 or 75cm^2 filter cap flasks (Greiner Bio-One 690175 and 658 175 respectively). Cells were counted using a Beckman Z2 Coulter particle count and size analyser. All trypanosome work was carried out using sterile techniques.

Cultured pleomorphic trypanosomes were grown in HMI-9 media supplemented with 20% FCS and 100U/ml penicillin and streptomycin at 37°C in 5% CO_2 with the addition of 1.1% methyl cellulose (McCulloch et al., 2004). Presence of methyl cellulose in the media requires that cells be counted using a haemocytometer. However, for the pleomorphic reporter cell line (Chapter 4) more accurate cell counts were required for the assay of CAT reporter gene expression. Thus for experiments involving analysis of CAT expression methyl cellulose was not added to the media and cells were counted using a Beckman Z2 Coulter particle count and size analyser.

Cell lines were stored at -80°C or in liquid nitrogen. For cell lines from culture, 10ml of exponential phase cultures were centrifuged at 800g for 10 minutes in order to pellet the cells. The pellet was resuspended in 500µl of media. To this, 500µl of HMI-9 supplemented with 14% glycerol was added slowly. This was left for 2 minutes before being transferred into a cryovial and stored at -80°C. For cell lines from whole mouse blood, blood was harvested by cardiac puncture (section 2.3). The blood obtained was divided into 100µl aliquots in cryovials. To this, 100µl of HMI-9 supplemented with 14% glycerol was added slowly. This was left for 2 minutes before being transferred into a cryovial and stored at -80°C. After two days at -80°C cell lines could be transferred to a liquid nitrogen storage tank for longer term storage.

2.3 Trypanosome infections in mice

All mice used in this study were strain MF1. All mouse husbandries were carried out by the staff of the March Building, University of Edinburgh. All mouse handling was carried out by Deborah Hall or Prof. Keith Matthews according to the conditions of Home-Office personal and project licenses.

For the harvest of slender forms and for chronic infections, no immunosuppressant was used. For the harvest of stumpy form trypanosomes, mice were injected with 200-250µl of 25mg/ml cyclophosphamide the day before infection to suppress the immune system of the mouse to allow for accumulation of a high parasite density.

To monitor parasitaemia throughout an infection, a tail snip was made and blood collected for a wet smear. Parasitaemia was then estimated from the wet smear by microscopy. For analysis of chronic infections (Chapter 5), 10µl tail blood was collected daily in 10µl capillary tubes (Camlab DMP010) (section 2.12.1) and dry smears were made for microscopy.

Total blood was harvested from mice by cardiac puncture using a Microlance 0.6 x 25mm needle and a 2ml syringe containing 250µl of 2% sodium citrate. Trypanosomes were purified from whole blood through an anion exchange resin column (Whatman 4057200) at pH7.8 in 30ml Buchner flasks.

2.4 Cloning and DNA manipulations

2.4.1 PCR amplification

PCR amplification was carried out in 0.2ml PCR tubes (Axygen 321-10-051) in a Thermo Electron Corporation PCR SPRINT thermal cycler. A typical 50µl reaction is shown below.

PCR:
10µl 5x Flexibuffer
3µl 25mM $MgCl_2$
4µl 2.5mM dNTPs
2µl 25µM Forward primer
2µl 25µM Reverse primer
0.5µl Taq Polymerase (5units/µl)
2µl AnTat1.1 genomic DNA (approx. 50ng/µl)
26.5µl dH_2O

Reaction conditions:
95°C for 5minutes

30x [95°C for 30 seconds, 55°C for 45 seconds, 72°C for 1 minute]

72°C for 4 minutes

2.4.2 Oligonucleotides

All oligonucleotides used in this thesis were produced by Sigma Aldrich and are detailed in Table 2.1.

No.	Name	Sequence
Amplification of the full length PAD1 5' and 3'UTR		
1	PAD1 3'UTR For (BamHI)	TTAGGATCCGCTTAGGGGAGCCAGTGGAGGGC
2	PAD1 3'UTR Rev (XhoI)	TTACTCGAGGGCAATGTTTTGTACAGTACGCTCAG
23	PAD1 3'UTR Rev (BsmAI)	TAATGCTTGAGACGGCAATGTTTTGTACAGTACGCTCAG
20	PAD1 5'UTR For (XhoI)	TTACTCGAGAAACATGGACAGTCAACATCTCCATATG
21	PAD1 5'UTR Rev (HindIII)	TTAAAGCTTGGCAATGTTTTGTACAGTACGCTCAG
Determining the PAD1 3'UTR polyadenylation site		
Along with primer 24 (PAD1 PolyA Round 1) and 1 (PAD1 PolyA Round 2)		
67	ADAPT dT(15)	GGCCACGCGTCGACTAGTACTTTTTTTTTTTTTT
68	AUAP	GGCCACGCGTCGACTAGTAC
Creating the PAD1 3'UTR deletion series		
62	PAD1 3'UTR Δ1-21 (BamHI)	TTAGGATCCGGCGGCTTCCTGACGTTTATTCAAC
63	PAD1 3'UTR Δ1-175 (BamHI)	TTAGGATCCCTTTGTCTATACATATCGCATACCATTAG
28	PAD1 3'UTR Δ1-207 (BamHI)	TTAGGATCCACGGTGCGCCAACTCTGGTAAACG
39	PAD1 3'UTR Δ1-265 (BamHI)	TTAGGATCCATATTGTACGAGCAGTCCATGCATTA
40	PAD1 3'UTR Δ1-354 (BamHI)	TTAGGATCCGAACACACCAACTTATATCGTTAAATTC
29	PAD1 3'UTR Δ1-391 (BamHI)	TTAGGATCCCTCTCGTTTATCGATAAAAGTTCG
41	PAD1 3'UTR Δ1-502 (BamHI)	TTAGGATCCCACTCTTCACCCGTGGAAACACC
30	PAD1 3'UTR Δ1-609 (BamHI)	TTAGGATCCCCTATCTCGTCTTCATATCTTTACTGC
42	PAD1 3'UTR Δ1-776 (BamHI)	TTAGGATCCAAACTTAATGCTTTCATTTCTCCCTTTTTG
Creating internal deletions of the PAD1 3'UTR		
For PAD1 3'UTR Δ386-564, along with primers 1 and 23		
49	1st PAD1 3'UTR Δ386-564 Rev (HindIII)	TAAAAGCTTTGAAGAATTTAACGATATAAGTTGGTGTG
50	2nd PAD1 3'UTR Δ386-564 For (HindIII)	TAAAAGCTTTGCAGACCAACACAGGCACCCCAA
For PAD1 3'UTR Δ354-624, along with primers 1 and 23		
57	1st PAD1 3'UTR Δ354-624 Rev (XhoI)	TAACTCGAGATATAGTGACTGGTGTTGTTGTATCG
58	2nd PAD1 3'UTR Δ354-624 For (XhoI)	TAACTCGAGTATCTTTACTGCTTGTTATAATTTTGTC
Inserting a section of the PAD1 3'UTR into the CAT449 construct		
Along with primer 40		
72	PAD1 3'UTR 624 Rev (BamHI)	TTAGGATCCTGAAGACGAGATAGGTAGCACTACG
Amplification of reporter genes		
For amplification of both GFP and CFP		
3	GFP/CFP For (HindIII)	TTAAAGCTTATGGTGAGCAAGGGCGAGGAGCTG
4	GFP/CFP Rev (BamHI)	TTAGGATCCTTACTTGTACAGCTCGTCCATGCCG
For amplification of GUS		
5	GUS For (HindIII)	TTAAAGCTTATGTTACGTCCTGTAGAAACCCCAAC
7	GUS Rev (BglII)	TTAAGATCTTCATTGTTTGCCTCCCTGCTGCGG
qRT-PCR of PAD1 and *Tb*ZFP3		
24	qRT-PAD1 For	GACCAAAGGAACCTTCTTCCT
25	qRT-PAD1 Rev	CACTGGCTCCCCTAAGCT
17	qRT-ZFP3 For	CAGGGGAAACGCAAAACTAA
19	qRT-ZFP3 Rev	TGTCACCCCAACTGCATTCT

Table 2.1 Oligonucleotides used in this thesis. All primers are written in the 5' → 3' direction.

2.4.3 DNA agarose gel electrophoresis

DNA agarose gels were used to visualise DNA as well as for separating different length DNA fragments and gel purification of those fragments of interest. Most often a 1% agarose gel was used; however, for separation of particularly large fragments of DNA (i.e. above 2kb) a 0.8% agarose gel was used. To prepare a 1% agarose gel, 1g of agarose power is dissolved in 100ml 1x TAE buffer (Appendix A) by heating. Once dissolved the agarose was allowed to cool to approximately 60°C before adding 4µl Safeview (NBS Biologicals). The gel was then poured into a gel tray containing a gel comb. Once the gel was set, the comb is removed and the gel tank is filled with 1x TAE buffer. Samples are loaded into the wells and ran at 100-125V for 25-35 minutes. In all cases a 200-10,000bp DNA ladder (Eurogentic SmartLadder MW-1700-10) was ran alongside the samples for estimation of fragment size.

2.4.4 DNA purification

DNA is purified to remove contaminants such as salts, proteins or primers from previous reactions which may interfere with future reactions. DNA can be purified from solution (i.e. after a PCR reaction) or from an agarose gel when it is desirable to purify single-sized fragments of DNA within a mixed population (i.e. after the removal of an insert from a plasmid via digestion both the plasmid and insert can be gel purified to separate one another). In both cases, DNA was purified using a NucleoSpin Extract II kit (Machery-Nagel) following manufacturer's instructions.

2.4.5 Restriction digests

Restriction digests were carried out at the optimum temperature for the restriction enzyme in the optimum buffer in a 30µl reaction volume according to manufacturers' recommendations. All enzymes and buffers were from either Promega or NEB. Double digests were carried out where an appropriate buffer and reaction temperature were available, otherwise sequential digestions were required. Digests were then purified as in section 2.4.4.

2.4.6 Ligation

Ligation reactions were typically carried out in a 10µl volume at 4°C overnight or at room temperature for 4-5 hours. A control with no digested insert was always included.

Ligation reaction: 1µl 10x Promega T4 DNA Ligase buffer
0.5µl Promega T4 DNA ligase (3U/µl)
1µl Digested vector (approx 200-500ng/µl)
7.5µl Digested insert

2.4.7 Transforming bacteria

For transformation, chemically competent laboratory stocks of XL-1 blue *E.coli* cells are maintained and stored at -80°C. On ice, 5µl of the ligation reaction (section 2.4.6) or 1µl of prepared plasmid is added to 100µl of competent bacteria and left for 30 minutes. The cells are heat shocked at 42°C for 30 seconds in a water bath. To allow the cells to recover before selection, 250µl of L-broth is added to the cells and incubated at 37°C for 45 minutes with shaking. The cells are then plated onto selective LB agar plates (1mg/ml ampicillin) and incubated at 37°C overnight. All plasmids used in this thesis contain an ampicillin resistance gene. Thus, the following day, only cells with the plasmid will have formed colonies. The colonies can then be used to prepare DNA (section 2.4.8) in order to confirm that they have obtained a plasmid and that plasmid contains the appropriate sequence.

2.4.8 Small-scale preparation of plasmid DNA: Mini-prep

For small-scale plasmid DNA preparation a mini-prep is carried out. Mini-prep DNA is used to test for the correct ligation, and often sequence, of an insert into a plasmid. See section 2.4.10 for the method of sequencing.

A discrete colony of interest on an LB agar plate is picked with a sterile pipette tip and used to inoculate 3ml of LB agar. In all cases described in this thesis the colonies were ampicillin resistant and therefore 1mg/ml ampicillin was also added to the culture. The culture is grown overnight at 37°C with shaking. The following day 1.5ml of culture is transferred to a 1.5ml eppendorf tube and centrifuged at 10,000g for 5 minutes at room temperature. The supernatant is removed and 100µl of Solution I is added to the pellet. The pellet is resuspended by rapping the eppendorf along a rack and 200µl of Solution II is added. The tube is mixed by inversion and left to incubate for no more than 5 minutes at room temperature. Then, 150µl of pre-chilled Solution III is added; the tube is again mixed by inversion and incubated on ice for 5 minutes. The sample is then centrifuged at 10,000g for 10 minutes at room temperature. The supernatant is then transferred to a fresh eppendorf containing 900µl of 100% ethanol. The sample is mixed by inversion and incubated at -20°C for 1 hour to allow precipitation of the DNA. The sample is then centrifuged at 10,000g for 10 minutes at room temperature. The supernatant is discarded and 500µl of 70% ethanol is added to the pellet. The sample is centrifuged at 10,000g for 5 minutes at room temperature and the supernatant is removed. The pellet is allowed to air dry for approximately 15 minutes. The pellet is then resuspended in 50µl of dH_2O containing 1µl 10mg/ml RNase. The pellet is allowed to dissolve for 5 minutes before being stored at -20°C.

2.4.9 Large-scale preparation of plasmid DNA: Midi-prep

For large-scale plasmid DNA preparation a midi-prep was used. Midi-prep DNA is of a higher concentration and purity compared to mini-prep DNA and is used for trypanosomes transfections (section 2.5). For midi-preps the QIAGEN® Plasmid Midi Kit (Cat. No. 12145) was used following manufacturer's instructions. The procedure is outlined below.

In a 2000ml autoclaved conical flask, 200µl of liquid culture is added to 200ml of LB broth containing 200µl 1mg/ml ampicillin. The culture is grown at 37°C overnight with shaking. The following day, the culture is centrifuged at 4000g for 30 minutes at 4°C. The supernatant is removed and 4ml of Buffer P1 (containing 100µg/ml RNase) is added to the pellet. The pellet is resuspended and 4ml of Buffer P2 is added. The tube is mixed by inversion and left to incubate for no more than 5 minutes at room temperature. Then, 4ml of pre-chilled Buffer P3 is added; the tube is again mixed by inversion and incubated on ice for 15 minutes. The sample is then centrifuged at 20,000g for 30 minutes at 4°C. The supernatant is then transferred to a fresh tube and spun again at 20,000g for a further 15 minutes at 4°C. Meanwhile, a QIAGEN-tip 100 column is equilibrated with 4ml of Buffer QBT: the solution passes through the column by gravity. The supernatant is applied to the column. The column is then washed twice with 10ml Buffer QC. The DNA is then eluted in 5ml of Buffer QF. To precipitate the DNA, 3.5ml of isopropanol is added to the eluate. The solution is mixed by inversion and centrifuged at 15,000g for 30 minutes at 4°C. The supernatant is removed with care taken to avoid disrupting the pellet. The pellet is washed in 2ml 70% ethanol and centrifuged at 15,000g for 10 minutes at 4°C. The supernatant is removed and the pellet is air-dried for 5 minutes at room temperature. The pellet is then dissolved in 400µl of dH_2O before being stored at -20°C.

2.4.10 DNA sequencing

DNA sequencing was carried out at the School of Biological Sciences Sequencing Service, (The GenePool), University of Edinburgh. Samples were prepared in 6µl volumes with 3.2µM primer and up to 500ng template DNA.

2.4.11 Mapping of the PAD1 3'UTR polyadenylation site

To map the PAD1 3'UTR polyadenylation site in stumpy forms, stumpy AnTat1.1 RNA was used in a reverse-transcriptase (RT) reaction followed by two rounds of nested PCR (Chapter 3).

RT-Reaction: 2µg stumpy AnTat1.1 RNA in 10µl total volume
1µl 50mM ADAPT (67) primer
4µl 5x AMV RT Buffer
1µl 40U/µl RNasin RNase inhibitor (Promega N211A)
2µl 2.5mM dNTPs
1µl 10U/µl AMV Reverse transcriptase (Promega M510A)
1µl 0.1mM 1, 4-dithio-DL-threitol (DTT)

Reaction conditions: 37°C 10 minutes

42°C 1 hour

The cDNA produced was then used in two rounds of nested PCR, firstly with primers AUAP (68) and PolyA Round 1 (24) and then with AUAP (68) and PolyA Round 2 (1), as in section 2.4.1. For details of primers see section 2.4.2, Table 2.1. The products from the second round of PCR were resolved on an agarose gel (section 2.4.3). There was one prominent band which was excised and gel purified (section 2.4.4). The PCR product was then cloned into pGEM-T-Easy (Promega A137A) and transformed into bacteria (section 2.4.7). Five transformant colonies were selected and used to make mini-prep DNA (section 2.4.8). The DNA was then sequenced (section 2.4.10) and the polyadenylation sites determined.

2.5 Trypanosome transfection

2.5.1 Preparing the DNA for transfection by electroporation

A digested DNA plasmid is purified using a NucleoSpin Extract II kit (as described in section 2.4.4) and eluted with 4x 50µl dH_2O to give a total eluate volume of 200µl. To concentrate the DNA by ethanol precipitation, 20µl 3M Na Acetate is added to the purified DNA and mixed by vortex. To this, 500µl of 100% ethanol is added and mixed by vortex. This is then incubated at -80°C for at least 1 hour in order to precipitate the DNA. The mixture is centrifuged at 18,000g for 10 minutes at 4°C and the supernatant removed. The DNA pellet is air-dried for 5 minutes and then resuspended in 5µl 1mM Tris-HCl, pH8, 0.1mM EDTA.

2.5.2 Transfection of monomorphic trypanosomes by electroporation

For transfection of monomorphic bloodstream form cells, 4 x 10^7 cells from a logarithmically growing culture were centrifuged at 800g for 10 minutes in order to pellet the cells. The cells were resuspended in 100µl Mirius Bio Ingenio™ Electroporation solution (MIR 397) and 5ul of DNA is added. Cells were transfected by electroporation in the Amaxa Nucleofector® II (20800714) using program "CD4 T-cells X-001" in Mirius Bio Ingenio™ cuvettes (MIR 50120). Cells were then allowed to recover in 25ml of pre-warmed HMI-9 overnight.

The following day the culture was centrifuged at 800g for 10 minutes to pellet the cells. Cells were resuspended in selective media and transferred to 24 well plates (2ml/well) and serial dilutions were set up (neat, 1/10, 1/100, 1/1000). After 5-7 days, all untransfected cells were dead and remaining live cells were transfectants. From each positive well, the culture is diluted 1/5 into 5ml in 6-well plates and then 1/5 into 10ml in flasks. Transfectant cell lines were then frozen for storage (section 2.2)

2.5.3 Transfection of pleomorphic trypanosomes by electroporation

A mouse was injected intraperitoneally (IP) with *T.b brucei* AnTat1.1 90:30 cells. At 3 days post infection, the population of cells were slender in morphology as determined by microscopy. The mouse blood was harvested by cardiac puncture. Using a sterile 1.5ml eppendorf tube, the blood was centrifuged at 4000g for 10 minutes, separating the blood into plasma, in the upper layer, and red blood cells, in the lower layer. Between the two layers is the buffy coat which is the fraction containing white blood cells, platelets and trypanosomes. The plasma was removed from the upper layer using a pipette. The buffy coat was then taken, with care to avoid red blood cells. The buffy coat was resuspended in 200µl of Mirius Bio Ingenio™ Electroporation solution (MIR 397) and split into two Mirius Bio Ingenio™ cuvettes (MIR 50120). To one cuvette, 5µl of DNA was added, to the other 5µl of dH_2O was added as a negative control. The cells were then electroporated in a Nucleofector® II (20800714) using program "CD4 T-cells X-001". Cells were then allowed to recover in 25ml of pre-warmed HMI-9 containing 1.1% methyl cellulose for 2 hours. After 2 hours, cells were counted using a haemocytometer. If cell concentration was higher than 1 x 10^5 cells/ml then the culture was diluted to this concentration. Cells were transferred to 24 well plates (1ml/well) and serial dilutions were set up (neat, 1/2, 1/4 and 1/8) before recovering overnight.

The following day, 1ml of pre-warmed and pre-equilibrated 2x selective media was added to each well. After 5-7 days, all untransfected cells were dead and remaining live cells were transfectants. From a positive well 1ml of media is taken, centrifuged at 4000g for 10 minutes to pellet the cells. The transfectants are then resuspended in 200µl media and injected IP into a cyclophosphamide-treated mouse (section 2.3). Once the infection has a detectable parasitaemia and cells are still slender in morphology (approximately 3 days post infection) the cells are harvested by cardiac puncture and diluted 1:1 with HMI-9 media plus 14% glycerol. Cells are divided into

200µl aliquots and stored at -80°C for short term storage and in liquid N_2 for long term storage.

2.5.4 Drug concentrations

All transfectants were selected and grown in the presence of the appropriate drug(s). The concentrations of each drug used in this thesis are below.

0.5µg/ml Puromycin (Calbiochem 5404411)
0.5µg/ml Phleomycin (InvivoGen ant-ph-2p)
3µg/ml Hygromycin (InvivoGen ant-hm-1)

2.6 Differentiation to procyclic forms

Bloodstream stage trypanosomes were induced to differentiate to procyclic forms by the addition of 6mM *cis*-aconitate and a drop in temperature from 37°C to 27°C. Percentage differentiation of a population was measured by the percentage of EP positive cells as determined by flow cytometry (section 2.7).

2.7 Flow cytometry

2.7.1 Fixation of cells for flow cytometry

All samples were processed in 5ml polystyrene round bottom tubes (BD Flacon 352052). Approximately 1-5 x 10^6 cells were centrifuged at 800g for 5 minutes at 4°C. The supernatant was removed and cells were washed by filling the FACS tube with cold PBS and centrifuging at 800g for 5 minutes at 4°C. The supernatant was removed and 500µl of fixative (2% formaldehyde, 0.05% glutaraldehyde in PBS) was added. Cells were resuspended in the fix by lightly rapping the tube along a rack and left overnight at 4°C. After fixation, samples being analysed for differentiation to the procyclic form were antibody stained for EP procyclin (section 2.7.2). Samples

being analysed for expression of a fluorescent protein were washed twice in cold PBS to remove the fix and then resuspended in 500µl of PBS.

2.7.2 EP procyclin antibody staining of cells for flow cytometry

Control cells both positive and negative for EP Procyclin were stained concurrently with experimental samples. Cells were washed twice in cold PBS to remove the fix. Cells were then blocked in 200µl of blocking solution for 20 minutes at 4°C to prevent unspecific antibody binding. To remove the blocking solution the FACS tubes were filled with cold PBS and centrifuged at 800g for 5 minutes at 4°C. The supernatant was removed and cells were stained with 200µl of primary antibody for 1 hour at 4°C. Cells were then washed once in cold PBS and 200ul of blocking solution was added and left for 20 minutes at 4°C. The blocking solution was removed as previously and cells were stained in 200µl of secondary antibody for 1 hour at 4°C. Cells were washed twice in cold PBS and resuspended in 500µl of PBS.

Blocking solution:	2% BSA in PBS
Primary antibody:	Anti-EP-Procyclin (VH Bio Ltd CLP001A) diluted 1:500 in blocking solution
Secondary antibody:	Anti-Mouse IgG (whole molecule)-FITC (Sigma F6275) diluted 1:500 in blocking solution.

2.7.3 Analysis by flow cytometry

Stained cells were analysed by flow cytometry using a Beckton Dickinson FACSCalibur machine with CellQuest Pro software. Positive and negative samples were used to determine the machine settings, often with the assistance of Martin Waterfall, University of Edinburgh. In all cases, 1-2 x 10^5 cells/sample were measured. Results were analysed with FlowJo software.

2.8 Immunofluorescent analysis

2.8.1 Preparing a microscope slide

To prepare a slide from an *in vitro* liquid culture of trypanosomes, 1ml of logarithmically-growing cell culture was centrifuged at 4000g for 5 minutes. The supernatant was discarded, and the cells were resuspended in approx 30-40µl media. 15µl of cells were added to each slide and spread carefully across the slide with the length of a 1ml pipette tip. The slides were allowed to air-dry before being fixed and stored in 100% methanol at -20°C until required.

To prepare a slide for immunofluorescence from an *in vivo* infection approximately 5-10µl of tail blood (see section 2.3) is added to a slide. The blood is then spread across the slide with a second, clean, slide. The slides are allowed to air-dry before being fixed and stored in 100% methanol at -20°C until required.

2.8.2 Immunofluorescence

Slides stored in methanol are re-hydrated in 1x PBS for 5 minutes. For chronic infections (Chapter 5) slides were stained for the infecting AnTat1.1 VSG. For this, 100µl primary antibody (Rabbit Anti-AnTat1.1 VSG diluted 1/20,000 in 1x PBS) was added to each slide and placed in a humidity chamber for 45 minutes. Slides were then rinsed in 1x PBS for 3x 5 minutes before adding 100µl Anti-Rabbit IgG (whole molecule)-FITC (Sigma F0382) secondary antibody (diluted 1/50 in 1x PBS). After 45 minutes in a humidity chamber, the slides are again rinsed in 1x PBS for 3x 5 minutes. To visualise the nucleus and kinetoplast, 50µl of 1ug/ml 4',6-diamindino-2-phenylindole (DAPI) was then added to each slide and incubated for 2 minutes in a humidity chamber and washed once in 1x PBS. A cover slip was mounted onto the slide with 40µl of Mowiol (Appendix A) with 1mg/ml p-Phenylenediamine (PDA) and left to set at 4°C overnight in the dark. When staining for AnTat1.1 VSG was not required, slides were stained with DAPI alone. Slides were analysed on a Zeiss Axioskop 2 *plus* using QCapture software for image capture.

2.9 CAT Assays

2.9.1 CAT sample preparation

CAT protein expression was measured using a CAT ELISA kit (Roche 11363727001) according to the manufacturer's instructions. A precise cell count was taken immediately prior to sample preparation (section 2.2). Where appropriate, trypanosomes were growing logarithmically and at a concentration of 1-2 x 10^6 cells/ml. In some cases, due to experimental treatment (i.e. treatment with growth inhibiting drugs), this was not possible but samples were processed in the same manner.

To prepare a sample, 1ml of trypanosome culture was centrifuged in a 1.5ml eppendorf tube at 4000g for 5 minutes. The supernatant was then removed and the cell pellet was washed 3 times in 1ml of cold 1x PBS, on each occasion the sample was centrifuged at 4000g for 5 minutes and the supernatant was removed. After the last wash the PBS was carefully removed before addition of 1ml of 1x Lysis buffer (provided in the CAT ELISA kit). Cells were resuspended in the lysis buffer by rapping the eppendorf tube along a tube rack and incubated for 25 minutes at room temperature. The sample was then centrifuged again at 4000g for 5 minutes to remove cell debris. The supernatant was divided into 2x 500µl into fresh eppendorf tubes and snap frozen in liquid nitrogen. The samples were then stored at -80°C until required.

2.9.2 CAT ELISA

All reagents were included in the CAT ELISA kit and working dilutions were prepared as per the manufacturers instructions. For each ELISA performed, a standard calibration curve ranging from approximately 91ng/ml to 2.8ng/ml (depending on batch) was prepared from a supplied stock of CAT enzyme. This was to ensure that all sample readings were within the linear range of the assay. Any samples which lay outside the linear range were diluted 1:10 or 1:100 as appropriate in sample buffer. In all cases the standard calibration curve had an r^2 value of 0.995 or above.

The ELISA was carried out in the supplied microplate wells which are pre-coated with an Anti-CAT antibody. To each well, 200µl of CAT standards or experimental sample (neat or diluted) were added and incubated at 37°C for 1 hour. Wells were then washed 5 times with 250µl of washing buffer. After removal of the last wash 200µl of Anti-CAT-DIG was added to each well and incubated at 37°C for 1 hour. Wells were washed 5 times with 250µl of washing buffer. After removal of the last wash 200µl of Anti-DIG-POD was added to each well and incubated for 37°C for 1 hour. Wells were then washed 5 times with 250µl of washing buffer. After removal of the last wash 200µl of POD substrate was added to each well and incubated at room temperature for 15-20 minutes to allow colour change (colourless to green) to develop. The absorbance of the samples was measured at 405nm using a BioTek ELx808 Absorbance Microplate Reader with Gen5 data analysis software.

2.10 GUS Assay

GUS enzyme activity was assayed via a colour change reaction which was quantified by an absorbance reading. 4-Methylumbelliferyl-β-D-glucopyranosiduronic acid (MUG) is used as a fluorescent substrate for the β- glucuronidase (GUS) enzyme (as recommended by Prof. I. Roditi, University of Bern) where 50% v/v substrate/lysis solution is added to the trypanosome culture in 96-well plates and absorbance measured after 2 hours of incubation at 37°C, with an excitation wavelength of 355nm and emission wavelength of 460nm on a BioTek Flx800 Fluorescent Microplate Reader with Gen5 data analysis software.

GUS substrate/lysis solution:
1mM MUG
0.82M Tris-HCl pH8
0.6% SDS
0.3mg/ml BSA

2.11 Alamar blue assay

Alamar Blue ® (Serotec) is a chemical which changes colour in response to the redox potential of culture media and therefore can be used to measure cell growth. For the purpose of incorporating a measure of cell growth into the GUS assay described in section 2.10, Alamar Blue® was used. For the assay 10% v/v Alamar Blue® was added to cells that had been treated with drug for 44 hours in 96-well plates. Cells were incubated with the Alamar Blue® for 4 hours and fluorescence was measured at 540nm and 595nm on a BioTek ELx808 Absorbance Microplate Reader with Gen5 data analysis software. Various controls were incuded in the assay, namely media only without Alamar Blue®, media plus Alamar Blue® and untreated cells plus Alamar Blue®. The decrease in growth after drug treatment was then determined by the following calculation:

% Growth = (((Absorbance of treated cells at 450nm)-(Absorbance of treated cells at 595nm) x R_0)/ ((Absorbance of untreated cells at 450nm)-(Absorbance of untreated cells at 595nm) x R_0)) x 100

Where R_0= ((Absorbance of media plus Alamar Blue® at 450nm)-(Absorbance of media without Alamar Blue® at 450nm))/ ((Absorbance of media plus Alamar Blue® at 495nm)-(Absorbance of media without Alamar Blue® at 495nm))

2.12 Quantitative RT-PCR from whole blood

Before beginning bench surfaces and pipettes are cleaned with 70% ethanol and filter tips are used throughout. Care is taken at all points to avoid contamination by keeping samples covered and equipment clean. Nuclease-free sterile water is used throughout.

2.12.1 Sample preparation

From an infected mouse, 10µl of whole blood was collected using a 10µl capillary tube (Camlab DMP010) and immediately transferred into 20µl Nucleic Acid Purification Lysis Solution (ABI 4305895) with 10µl 1x PBS, pre-chilled on ice. The solution was mixed by flicking the tube and was stored at -80°C until required.

For the purposes of production of a standard curve, the concentration of parasites in the blood was required to be known. In this case, slender or stumpy parasites were harvested and purified from a mouse infection (section 2.3). The cells were counted using a Beckman Z2 Coulter particle count and size analyser and a known number of cells were centrifuged at 800g for 10 minutes to pellet the cells. The parasites were then resuspended in a known volume of blood such that the final concentration of parasites per ml blood could be calculated. This blood was divided into 10µl aliquots added to 30µl lysis mix (see above) before being stored at -80°C. These standard curve samples were processed in the same manner as experimental samples.

2.12.2 RNA extraction from whole blood

RNA was extracted using an ABI Prism 6100 Nucleic Acid PrepStation according to manufacturer's instructions. The platform takes 96-well RNA purification trays (ABI 4305673) to which the sample and appropriate washes are applied, and pulls the flow-through into the waste collection by applying a vacuum across the wells. A splashguard (ABI 4311758) was used in order to prevent contamination between wells. All RNA extractions were carried out using the 'RNA Blood-DNA' programme. For this, wells were pre-wet with 40µl of RNA Purification Wash Solution 1 (ABI 4305891) and then 40µl of lysate (section 2.12.1) was added to each well. An 80% vacuum was applied to the wells for 3 minutes. Wells were washed with 650µl of RNA Purification Wash Solution 1 and the wash was removed by applying an 80% vacuum to the wells for 3 minutes. Wells were then washed with 650µl of RNA Purification Wash Solution 2 (ABI 4305890) and the wash was removed by applying an 80% vacuum to the wells for 3 minutes. Taking care to

apply directly to the bottom of the well, 50µl of AbsoluteRNA Wash Solution (ABI 4305545) was added to the wells and incubated for 15 minutes. The AbsoluteRNA Wash Solution contains a DNase which should remove genomic DNA from the sample. Due to the labile nature of DNase this was defrosted on ice with care taken not to destroy the DNase mechanically. Without removing the AbsoluteRNA Wash Solution, 400µl of RNA Purification Wash Solution 2 was added and this was incubated for 5 minutes before removal by applying an 80% vacuum to the wells for 3 minutes. The well was washed twice more with RNA Purification Wash Solution 2, firstly with 650µl and then 400µl, each time removing the wash by applying an 80% vacuum to the wells for 3 minutes. A 90% pre-elution vacuum was applied for 5 minutes to remove any remaining wash solution before moving the purification try from the waste collection position to the sample collection position. To elute the sample, 100µl of Nucleic Acid Purification Elution Solution was added to each well and a 20% vacuum was applied to the wells for 2 minutes. Samples were eluted into a MicroAmp Optical 96-well Reaction Plate (ABI 4306737), sealed with an adhesive film (ABI 4311971) and stored at -80°C until use.

2.12.3 Treatment of RNA with TURBO DNase

Despite the treatment of the RNA with a DNase during the extraction procedure (section 2.12.2), DNA contamination was still present in the RNA. Thus, an additional DNase treatment was added after RNA extraction using an Ambion TURBO DNA-free kit (Applied Biosystems AM1907) according to manufacturers' instructions. A typical 50µl reaction is shown below.

DNase reaction: 5µl 10X TURBO DNase Buffer
44µl RNA (extracted as section 2.12.2)
1µl TURBO DNase

The reaction is mixed gently and incubated at 37°C for 30 minutes before deactivating the DNase with 0.1 volume of DNase Inactivation Reagent (in this case, 5µl). The reaction is mixed well and incubated for 5 minutes with occasional mixing

to disperse the DNase Inactivation Reagent. The sample is centrifuged at 4000g for 1.5 minutes and the RNA is transferred to a fresh tube with care taken not to transfer any DNase Inactivation Reagent which may interfere with downstream reactions. The RNA is stored at -80°C until use and care is taken to avoid repeated freeze thaw cycles.

2.12.4 cDNA production

cDNA is produced using the ABI High-Capacity cDNA Reverse Transcription kit (ABI 4368813) according to manufacturer's instructions. A typical reaction is described below. This kit uses random primers to initiate cDNA production. The cDNA reaction uses a 2x master mix, however, as the RNA is treated with Turbo DNase the RNA must be diluted 1:2 in dH_2O as the Turbo DNase buffer can inhibit the reverse transcriptase reaction at higher concentrations. A typical 20µl reaction is shown below.

RT reaction:	10µl Turbo DNase-treated RNA (diluted 1:2 in dH_2O)
	2µl 10x RT Buffer
	0.8µl 25x dNTP Mix (100mM)
	2µl 10x RT Random Primers
	1µl MultiScribe Reverse Transcriptase
	3.2µl Nuclease-free dH_2O (Promega P1195)
Reaction Conditions:	25°C 10 minutes
	37°C 120 minutes

The reaction is carried out in 0.2ml PCR tubes (Axygen 321-10-051) in a Thermo Electron Corporation PCR SPRINT thermal cycler. Controls which replace the reverse transcriptase with dH_2O are used to ensure that there is not any contamination from DNA in the sample. The cDNA is stored at -20°C until required.

2.12.5 Quantitative RT-PCR

Quantitative RT-PCR (qRT-PCR) was carried out on an ABI StepOnePlus RT-PCR machine to amplify either *PAD1* or *Tb*ZFP3 mRNA. Reactions were carried out in 25µl volumes, as detailed below. A melt curve was added to the end of each qRT-PCR to ensure that there was only one amplification product. For *Tb*ZFP3, a standard curve was for absolute quantification of parasite number. For *PAD1* the ΔΔCT method of data analysis was used for relative quantification, using *Tb*ZFP3 as an internal control (for details see Chapter 5). The data was analysed using the ABI StepOne software version 2.

PAD1 qRT-PCR Reaction:
12.5µl Power SYBR Green PCR Master Mix (ABI 4367659)
0.75µl 10µM Primer 24
0.75µl 10µM Primer 25
4µl dH_2O
7µl cDNA (section 2.12.4) diluted 1/10

ZFP3 qRT-PCR Reaction:
12.5µl Power SYBR Green PCR Master Mix (ABI 4367659)
2.25µl 10µM Primer 17
2.25µl 10µM Primer 19
1µl dH_2O
7µl cDNA (section 2.12.4) diluted 1/10

Reaction Conditions:
95°C 10 minutes

40 x [95°C for15 seconds, 60oC for 1 minute]

+ Melt curve

2.13 Northern blot analysis

Before beginning bench surfaces and pipettes were cleaned with 70% ethanol and filter tips are used throughout. All equipment that to come into contact with the RNA (flasks, gel tank, sandwich boxes etc) was washed with detergent, rinsed in dH_2O, sprayed with 70% ethanol and allowed to air dry before use.

2.13.1 RNA sample preparation

To collect RNA samples, 25ml of logarithmically growing cell culture was centrifuged at 800g for 10 minutes. In some cases, the use of a logarithmically growing culture was not possible (after drug treatment) but the procedure remained the same. The pellet was resuspended in 600µl RLT buffer from a QIAGEN RNeasy Mini Kit (74106) supplemented with 10µl β-mercaptoethanol/ml to lyse the cells. The lysate was then stored at -80°C until processing. Upon thawing, the lysate was homogenised by vortexing at full speed for 1 minute. The RNA was then extracted using the "Purification of Total RNA from Animal Cells Using Spin Technology" protocol from the QIAGEN RNeasy Mini Kit, which uses an RNeasy spin column, according to manufacturers instructions. RNA was eluted into 40µl water. RNA was stored at -80°C and care was taken to avoid repetitive freeze-thaw cycles and was always thawed on ice.

2.13.2 Ribo-probe preparation

The entire CAT ORF was cloned into the pGEM T easy vector (Promega A137A) by Dr Adam Mott, University of Edinburgh. Upon amplification of the CAT ORF from the vector, the amplicon contained T7 and SP6 polymerase binding sites. The amplicon was used to produce digitoxin (DIG) labelled single-stranded RNA that was anti-sense to the CAT mRNA transcript by a Roche DIG RNA labelling kit (11 175 025 910) following manufacturers instructions. The DIG-labelling reaction was carried out in a 0.5ml eppendorf tube. A typical 40µl reaction is shown below.

Ribo-probe reaction: 2µg template
4µl NTP Mix (from kit)
2µl 0.1M DTT
2µl 40U/µl RNasin RNase inhibitor (Promega N211A)
4µl Reaction buffer (from kit)

The reaction mixture was incubated at 37°C for 2 hours before the addition of 4µl DNaseI (from kit). The reaction was then incubated at 37°C for 15 minutes in order to remove any DNA template. The reaction was terminated with the addition of 4µl 200mM EDTA, pH8. To the reaction, 5µl 4M LiCl and 75µl ice cold 100% ethanol was added, mixed well and put at -80°C overnight. The following day the solution was centrifuged at 12,000g for 15 minutes at 4°C to pellet the RNA. The RNA pellet was then washed with 50µl 70% ice cold ethanol and centrifuged again. The supernatant was removed and the pellet was allowed to air dry for 10 minutes before being resuspended in 100µl of water containing 2µl 40µ/µl RNase inhibitor (Promega N211A). The probe was then stored at -80°C.

2.13.3 RNA agarose gel preparation

To create a 1.2% agarose RNA gel in a total final volume of 150ml, 1.8g of agarose was dissolved by heating in 120ml of dH_2O. Once dissolved, 15ml of 10x MOPS, 10.5ml of dH_2O and 4.5ml of 37% formaldehyde is added in a chemical fume hood. The gel is poured into a gel tray containing a 16-well gel comb with care taken to avoid bubbles. This is left to set in a chemical fume hood.

To prepare the RNA for loading onto the gel, an equal quantity of RNA from each sample (0.2-2µg in a volume less than 6µl) is added to 9µl formamide, 3µl of 37% formaldehyde, 2µl of 10x MOPS and 2µl of RNA gel loading buffer. The sample is then heated to 65°C for 5 minutes in order to remove any secondary structure in the RNA. The sample is loaded into the agarose gel and run at 150 Volts in 1x MOPS

running buffer for 90 minutes. Where less than 16 wells are required, the centremost wells are used preferentially. To ensure an equal quantity of RNA has been loaded in each lane, the gel is stained with 20µl of 10mg/ml ethidium bromide in 200ml of 1x MOPS for 15 minutes with shaking. The gel is de-stained by washing at least twice, but often more, in dH_2O for 30 minutes with shaking. The gel is then visualised and photographed using a Fotodyne Incorporated UV transilluminator. This allows visualisation of the ribosomal RNA within a sample, allowing identification of any lanes which have been under or over loaded.

2.13.4 Assembling and dismantling the northern blot

Northern blotting was performed using capillary transfer onto a positively charged nylon membrane. The blot was assembled as follows. A plastic tray was filled up to approximately 2cm of 10x SSC (Appendix A). A sandwich box was placed upside down in the tray. A gel tray the same size as that used to create the agarose gel was then placed upside down on the sandwich box. A piece of chromatography paper (Fisher CHR200) pre-wet in 10x SSC was placed over the gel tray and down into the 10x SSC in the plastic tray to create a wick for buffer transfer. The agarose gel was then placed upside down onto the chromatography paper. Clingfilm was wrapped around the gel and over the plastic tray to prevent evaporation to the environment. A piece of positively charged nylon membrane (Roche 11 417 240 001) pre-wet in 10x SSC was placed on top of the gel and this was rolled flat using a 10ml plastic pipette to remove any bubbles between the gel and the membrane. The membrane was handled only using plastic tweezers throughout. Two pieces of chromatography paper, pre-wet in 2x SSC, are placed on top of the membrane and again rolled flat to remove any bubbles. On top of this, two inches of paper towels are placed, followed by a flat piece of plastic (for even weight distribution) and finally a heavy catalogue to act as the weight. The blot is left to transfer overnight. The next day, the blot is dismantled and the nylon membrane is allowed to dry. The RNA is UV cross linked to the membrane using a Stratalinker (0.120 joules, 254nm). The blot is then stored for up to one week at room temperature until hybridisation.

2.13.5 Hybridisation of the blot

Once wet, care is taken not to let any part of the membrane dry. The membrane is pre-hybridised in 10ml of pre-warmed hybridisation buffer in a hybridisation tube at 68°C for 1 hour in a hybridisation oven (Techne Hybridiser HB-1D) with rotation. To 100µl of hybridisation buffer, 2µl of ribo-probe is added and this is heated at 99°C for 5 minutes. The heated probe in buffer is added to 7ml of hybridisation buffer pre-warmed to 68°C. The pre-hybridisation buffer is removed and the probe in hybridisation buffer is added to the membrane. The membrane is then hybridised overnight at 68°C with rotation.

2.13.6 Detection

After an overnight hybridisation, the blot is washed at 68°C, twice in 100ml 2x SSC with 0.1% SDS for 30 minutes and then once in 100ml 0.5x SSC with 0.1% SDS for 30 minutes. The membrane is removed from the hybridisation tube and all following steps are performed in sandwich boxes. The membrane is rinsed in 100ml of RNA wash buffer for 1 minute at room temperature. The membrane is then blocked in 50ml Maleic Acid buffer with 1% DIG blocking reagent (Roche 11 096 176 001) for 1 hour at room temperature with shaking. After pouring off the blocking solution, the membrane is then incubated in 50ml of Maleic Acid buffer with 1% DIG block and 1.5U of Anti-DIG (Anti-Digoxigenin-AP Fab Fragments Roche 11 093 274 910) for 30 minutes at room temperature with shaking. The membrane is then washed 3 times in RNA wash buffer for 10 minutes before being incubated in 50ml of RNA detection buffer for 2 minutes, both at room temperature.

The membrane is placed in a sleeve of heat sealable polyethylene (known as Bag W) and 10µl of 25mM CDP-Star Chemiluminescence substrate (Roche 11 685 627 001) in 1ml of RNA detection buffer is applied evenly over the surface and left for 2 minutes. The Bag W is closed over the membrane and excess CDP-Star is removed using paper towels to drive the solution out from between the sleeve of Bag W. The sleeve is sealed with a Hulme Martin Impulse heat sealer and incubated at 37°C for 15 minutes.

To visualise the blot by X-ray film, the membrane is placed in an X-ray cassette and taken to a dark room. Once in a dark room a Kodak blue sensitive autoradiography film (GRI) is placed in the cassette on top of the membrane and exposed for 5 minutes initially. After 5 minutes the film is developed in a Compact X2 X-ray processor. Further films are exposed for either a longer or shorter duration depending on the result from the 5 minute exposure. For quantification of northern blots, the membrane, instead of being exposed to film, is visualised by chemilumminescence on a Syngene Gbox with GeneSnap software using no light and no filter settings. mRNA levels are then normalised using the ethidium bromide staining of the ribosomal RNAs on the northern gel as a loading control for each sample. The quantification is carried out using GeneTools software.

2.14 Bioinformatics, statistics and production of figures

All graphs in this thesis were produced using GraphPad Prism 4 software. All statistics were carried out using either GraphPad Prism 4 or Minitab 15. At various points in this thesis a general linear model (GLM) was used to analyse the data. One of the assumptions of a GLM is that the data is normally distributed and this was always tested before proceeding to analysis. If data was not normally distributed, the data was transformed (for example, by logarithmic transformation) before analysis. In all cases, a p-value of less than 0.05 was considered to be significant (i.e. given the null hypothesis, the probability of observing the data is less than 5%).

Immunofluorescent images were manipulated (addition of a scale bar, cropping of image and adjustment of contrast and brightness) using Photoshop software. DNA sequence analysis was carried out online at the Sequence Manipulation Suite (http://www.bioinformatics.org/sms/). Sequence alignments were performed using the Multalin software, publicly available online, using default parameters (http://multalin.toulouse.inra.fr/multalin/multalin.html) (Corpet, 1988). Alignment figures in this thesis were produced in JalView 2.4 (Clamp et al., 2004). Finally, for prediction of mRNA secondary structure the publicly available Sfold software was used (http://sfold.wadsworth.org) (Ding et al., 2005; Ding and Lawrence, 2003).

3 Chapter 3: Identification of regulatory regions that govern the expression of *PAD1*

3.1 Introduction

Gene expression in trypanosomes is controlled almost entirely post-transcriptionally, such that developmental changes in gene expression are controlled largely by mRNA abundance and translation. Genes that are up-regulated in stumpy forms must therefore be repressed during the slender life-stage and must escape translational repression during the stumpy life-stage. So far, regulatory signals controlling expression of a stumpy-specific transcript have not been identified, although bioinformatic analysis has revealed two motifs enriched in the 3'UTR of stumpy-specific transcripts (Kabani et al., 2009). Although linear motifs clearly play an important role in the control of gene expression, RNA molecules are three dimensional molecules with potentially dynamic secondary structures (Clayton, 2002). These structures can form regulatory signals, recognised by RNA binding proteins, and control gene expression (section 1.10). As such, it is unlikely that bioinformatic analysis alone will reveal the mechanisms of stumpy-specific gene regulation.

PAD1 expression is up-regulated at both the mRNA and protein level in stumpy form cells compared to slender forms (Dean et al., 2009). As such, analysis of the control of expression of *PAD1* may reveal mechanisms by which trypanosomes suppress expression of stumpy-specific transcripts during the slender stage of the life-cycle. Additionally, many transcripts are translationally repressed during the stumpy life-stage, and understanding how *PAD1* escapes this translational repression could inform understanding of stumpy-specific gene expression more generally.

As we know that stumpy-specific regulation will require suppression of expression during the slender stage, here monomorphic slender reporter cell lines were used to identify regions in the *PAD1* 3'UTR that prevent expression in proliferative slender stages.

3.2 Sites of polyadenylation in the *PAD1* 3'UTR

The intergenic region between the *PAD1* gene and the downstream *PAD2* gene is 966nt long. In order to identify any regulatory regions in the *PAD1* 3'UTR it was first necessary to know the polyadenylation site. This defines the length of the mRNA and as such provides limits in which to study the 3'UTR.

The polyadenylation sites of almost 6000 genes have recently been identified in slender and procyclic forms (Siegel et al., 2010). For *PAD1*, three polyadenylation sites were found in the procyclic forms, and none in slender forms, likely due to the lack of expression in this life-stage. In procyclic forms the three sites identified were at position 428nt, 572nt and 835nt (Siegel et al., 2010). As these polyadenylation sites were identified in procyclic forms rather than stumpy forms, and there may be stage-specific differences in choice of polyadenylation site (Clement and Koslowsky, 2001; Erondu and Donelson, 1992), it was still necessary to confirm the sites used in stumpy forms.

3.2.1 Identifying the *PAD1* 3'UTR polyadenylation site by BLAST

Initially, to identify the *PAD1* 3'UTR polyadenylation site, a nucleotide BLAST search was carried out. The entire 966nt intergenic region between *PAD1* and *PAD 2* was used to search in the "Non-human, non mouse ESTs" database on NCBI BLAST (http://www.ncbi.nlm.nih.gov/). The most significant alignment was a 273nt sequence (accession number W00184) from a *Trypanosoma brucei rhodesiense* cDNA library. This sequence aligned from position 437 to position 709 in the intergenic region downstream of *PAD1* (Figure 3.1). This suggested that the *PAD1* 3'UTR polyadenylation site is located at position 709. However, as bases 707-710 are all adenine it was not possible to distinguish the exact polyadenylation site from these. As this is only a short fragment of sequence which does not include the *PAD1* ORF and did not contain a stretch of adenine residues indicative of a polyadenylation site, it was still considered necessary to map the polyadenylation site experimentally.

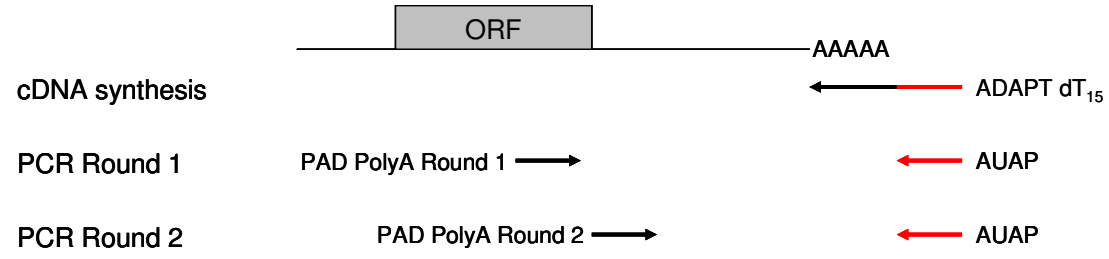
```
EST/1-273       ---------------------------------------------------------------
Pad1-2/1-966    AGCTTAGGGGAGCCAGTGGAGGGCGGCTTCCTGACGTTTATTCAACATTTTGTGTATTTATAT 63

EST/1-273       ---------------------------------------------------------------
Pad1-2/1-966    AATGTTAGCGACGACAGGAGAAAGTAAAATGAACAGATATACTTTGATTCCTTTTTACTATAC 126

EST/1-273       ---------------------------------------------------------------
Pad1-2/1-966    TTTTCCTTTTTTTTTTGCTTCTTTGTCTATACATATCGCATACCATTAGTGAGTTCCTTTATTA 189

EST/1-273       ---------------------------------------------------------------
Pad1-2/1-966    TCGCCTCCGACAATGAAGACGGTGCGCCAACTCTGGTAAACGATAAATAAACGGTTTTAAAAA 252

EST/1-273       ---------------------------------------------------------------
Pad1-2/1-966    ATCATCTTTACACATATTGTACGAGCAGTCCATGCATTATTATTCTTTTTTTTTGAGTTTTCT 315

EST/1-273       ---------------------------------------------------------------
Pad1-2/1-966    CATCAGAAGAGATCGATACAACAACACCAGTCACTATATGAACACACCAACTTATATCGTTAA 378

EST/1-273       ----------------------------------------------------------TTCAG 5
Pad1-2/1-966    ATTCTTCACTTAGCTCTCGTTTATGATCGATAAAAGTTCGAATTCACCTCAAAAAATTTCCAG 441

EST/1-273       AATTAATATTATATTTTCTTTC-ATTCTTTGCTATTTTNAGAGCGATCTCCTTAACTTTCACC 67
Pad1-2/1-966    AATTAATATTATATTTTCTTTCCATTCTTTGCTATTTTCAGAGCGATCTCCTTA-CTCTCACC 503

EST/1-273       CACTCTTCACCCGTGGAAACACCATGTAATCCTTTTCTTCTTGNGGTGGGAGAGAGGAGAGTG 130
Pad1-2/1-966    CACTCTTCACCCGTGGAAACACCATGTAATCCTTTTCTTCTTGCGGTGGGAGAGAGGAGAGTG 566

EST/1-273       CGGAACAANANAGGCACCCCAACACCGTTGATCCGTAGTGCTACCTATCTCGTCTTCATATCT 193
Pad1-2/1-966    CAGACCAACACAGGCACCCCAACACCGTTGATCCGTAGTGCTACCTATCTCGTCTTCATATCT 629

EST/1-273       TTACTGCTTGTTATAATTTTGTCTTTTTTACCTTTTTTTTTTCGNTGCGCTGTCACGGTTTTA 256
Pad1-2/1-966    TTACTGCTTGTTATAATTTTGTCTTTTTTACCTTTTTTTTTCGGCTGCGCTGTCACGGTTTTA 692

EST/1-273       TAAATGTCCATTGCAAA---------------------------------------------- 273
Pad1-2/1-966    TAAATGTCCATTGCAAAATAGTTGCTCTCAAGACACACATCAATAGCATTATATCCGCCCAAT 755

EST/1-273       ---------------------------------------------------------------
Pad1-2/1-966    ACCTCTTCTTTCATTCTCCCAAAACTTAATGCTTTCATTTCTCCCTTTTTGACTTATTGAACC 818

EST/1-273       ---------------------------------------------------------------
Pad1-2/1-966    TTTGAACAATACCACAGACTAACAGCAATAAACAAGCACGCAGTCACGTAAAGGAACCTAACA 881

EST/1-273       ---------------------------------------------------------------
Pad1-2/1-966    TTTTAGGGAAACAAAAATTATCTGGCAGCACACTGGGCGAAACATCAGGAGGAATACAACTGA 944

EST/1-273       ----------------------
Pad1-2/1-966    GCGTACTGTACAAAACATTGCC                                          966
```

Figure 3.1 Alignment between the 966nt *PAD1* to *PAD2* intergenic region and a 273nt sequence from a *Trypanosoma brucei rhodesiense* cDNA library. The intergenic region between *PAD1* and *PAD2* (Pad1-2/1-966) was used to carry out a nucleotide BLAST search. The closest matching sequence (EST/1-273) was a 273nt sequence (accession number W00184) from a *Trypanosoma brucei rhodesiense* cDNA library. Coloured nucleotides represent identity between the two sequences. This alignment suggested that the *PAD1* 3'UTR polyadenylation site was between positions 707-710nt.

3.2.2 Identification of the *PAD1* 3'UTR polyadenylation sites by RT-PCR

To map the *PAD1* polyadenylation site experimentally, RT-PCR was carried out on stumpy form trypanosome RNA using the method and primers as described in Mayho et al., 2006.

A



B

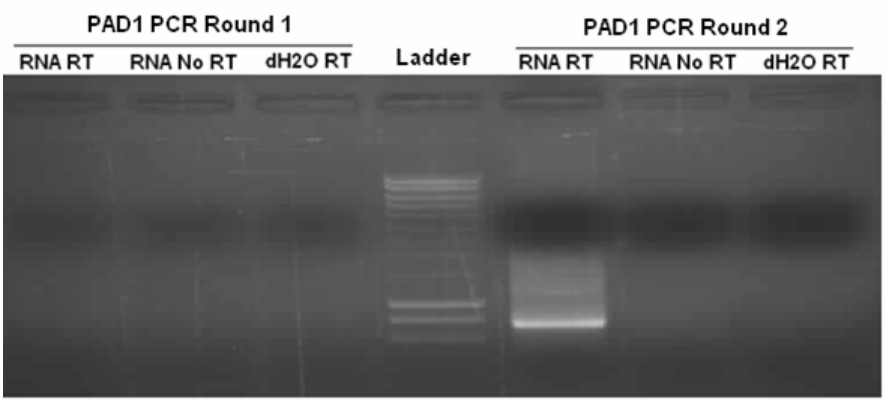


Figure 3.2 Mapping the *PAD1* 3'UTR polyadenylation site by RT-PCR. (A) A schematic diagram of the primers (shown as arrows) used to map the polyadenylation sites. An RT-PCR reaction was carried out using the ADAPT dT_{15} primer to produce cDNA. The ADAPT dT_{15} primer contains a unique sequence as denoted in red. A nested PCR reaction was carried out using the AUAP primer specific to the unique sequence in the ADAPT dT_{15} primer with primers specific for the targets. Diagram not drawn to scale. (B) Round 1 and round 2 PCR products resolved on an agarose gel stained with ethidium bromide and visualised using UV light. The prominent band from the round 2 PCR was excised, purified, cloned into pGEM-T Easy and sequenced. Also shown are the negative controls for each reaction: RNA NoRT = no reverse transcriptase control; dH_2O RT = no template control.

A primer complementary to the polyA tail was used to make cDNA (Figure 3.2). This primer, ADAPT dT_{15}, contains an additional unique sequence at the 5'end. The cDNA was then used as a template for a nested PCR. In the first round PCR, a primer complementary to the additional sequence in the ADAPT dT_{15} primer (AUAP) was coupled with a primer specific to the *PAD1* mRNA (PAD PolyA Round 1). In the second round PCR, the AUAP primer was coupled to a second primer specific to the *PAD1* or mRNA, but further downstream (PAD PolyA Round 2) (Figure 3.2A). The PCR products from both round 1 and round 2 were then resolved on an agarose gel (Figure 3.2). There was one prominent band, although a 'smear' of larger bands was also visible. The prominent band was excised and gel purified. The PCR product was then cloned into pGEM-T Easy and transformed into XL1 *E.coli*. Five colonies were selected, small scale DNA preparations were carried out and the DNA was sequenced.

The *PAD1* 3'UTR polyadenylation site varied between clones (Figure 3.3). Position 707-710 was confirmed as a polyadenylation site, however, a further site at position 722-723 was also identified. From a combination of experimental and bioinformatic analysis it was possible to create a map of the polyadenylation sites (Figure 3.3). Given this map it was then possible to begin the study of the 3'UTRs to identify any regulatory elements they contained.

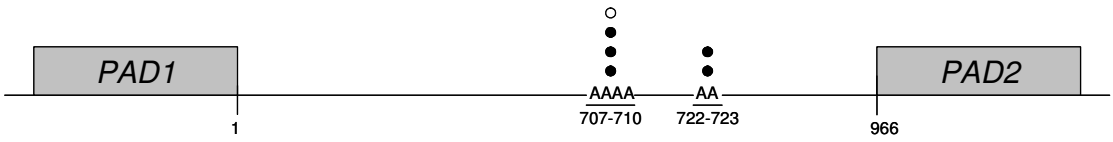


Figure 3.3 A schematic map of the *PAD1* 3'UTR polyadenylation sites in stumpy form *T.brucei*. Sites mapped experimentally by RT-PCR in this analysis are shown by black dots. The site mapped by nucleotide BLAST analysis (section 3.2.1) is shown by a white dot. Genome-wide analysis in procyclic forms identified further polyadenylation sites at positions 428nt, 572nt and 835nt (not shown) (Siegel et al., 2010). Boxes depict ORFs. All distances are shown in nt. Diagram not drawn to scale.

3.3 The *PAD1* 3'UTR does not contain known stumpy-enriched sequence motifs

Recent bioinformatic analyses comparing the 3'UTR of all predicted ORFs from chromosomes 1 and 2 with the predicted 3'UTR of transcripts enriched in stumpy forms, as determined by microarray analysis, have revealed two motifs enriched in the 3'UTR of stumpy-specific transcripts (Kabani et al., 2009). Stumpy motif 1 and stumpy motif 2 (SM1 and SM2) are hexanucleotide sequences (TCTTAC and TTCTTA respectively) with SM1 displaying positional bias, in that it is most often found 150-200nt downstream of the stop codon (Kabani et al., 2009). Upon analysis of the sequence of the *PAD1* 3'UTR neither SM1 nor SM2 was found to be present.

3.4 Creation of reporter constructs for the functional analysis of the *PAD1* 3'UTR

In order to functionally analyse the *PAD1* 3'UTR, a series of reporter cell lines were created. The full length 3'UTR, as well as sequential deletions, were coupled to the chloramphenicol acetyl transferase (CAT) reporter gene and transfected into monomorphic BSF trypanosomes. As *PAD1* is not expressed in slender forms the reporter gene should also not be expressed in monomorphs. If the lack of expression is due to negatively acting elements repressing gene expression in slender forms, then as sections of the 3'UTR are deleted, the negative elements should be deleted. Hence, as negative elements are deleted it would be expected that there would be alleviation of repression of expression causing reporter gene expression to increase.

Reporter constructs were based on the *T.brucei* CAT449 expression vector (Figure 3.4) (Mayho et al., 2006). In this plasmid, the *CAT* gene was coupled to the full length *aldolase* 5'UTR and a truncated *aldolase* 3'UTR providing constitutive CAT protein expression. The construct contains a phleomycin resistance gene under the control of the *actin* 5' and 3'UTRs and was therefore also constitutively expressed. The CAT449 plasmid, when digested with NotI, integrates into the *β-tubulin* locus of the trypanosome genome. Once integrated into the genome, the plasmid is

transcribed by PolII read-through transcription so that the reporter gene is processed by the same polymerase machinery as the endogenous *PAD1* gene.

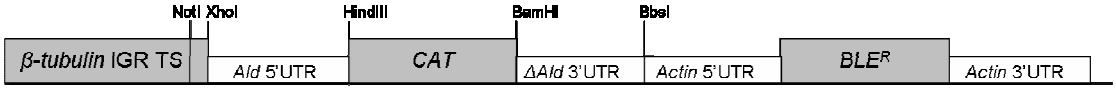


Figure 3.4 Schematic diagram of the CAT449 reporter construct. In the CAT449 expression vector, the *CAT* reporter gene has an *aldolase* 5'UTR (*Ald* 5'UTR) and a truncated *aldolase* 3'UTR (*ΔAld* 3'UTR) providing constitutive expression in *T.brucei*. The construct contains a phleomycin resistance gene (BLE^R) with the *actin* 5' and 3'UTRs providing constitutive expression. The construct contains a *β-tubulin* intergenic region targeting sequence (*β-tubulin* IGR TS); when the plasmid is linearised with NotI the plasmid can integrate into the *β-tubulin* locus. Diagram is not to scale. Figure modified from Mayho et al., 2006.

The truncated *aldolase* 3'UTR can be excised from the CAT449 construct using BamHI and BbsI restriction enzymes which cut at the 5' and 3' of the UTR respectively (Figure 3.4). The BbsI enzyme cuts outside of the recognition sequence and so creates a 5' overhang dependent on the upstream sequence. To insert the *PAD1* 3'UTR into the construct, primers were designed to the 3'UTR sequences such that the 5' primer contained a BamHI site complementary to that in the vector. The *PAD*1 3'UTR contains a BbsI restriction site and thus it was not possible to use a BbsI site to insert the 3'UTR at the 3' end. Therefore, the 3' primer instead contained a BsmAI site which, like BbsI, cuts outside the recognition sequence. Hence, it was necessary to include a short sequence into the 3' primer to create a complementary overhang (Figure 3.5).

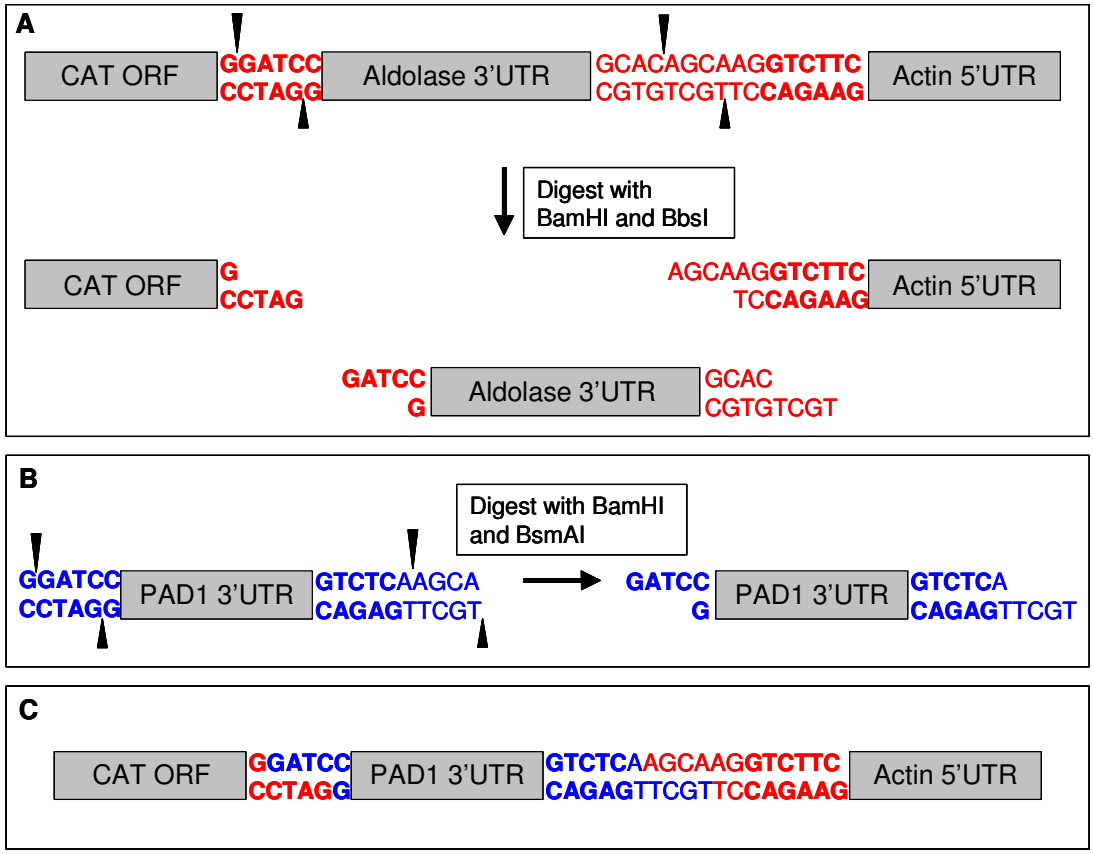


Figure 3.5 The *PAD1* 3'UTR was integrated into the CAT449 construct using BamHI and BsmAI restriction enzymes. (A) The CAT449 vector is digested with BamHI, which recognises and cuts within the sequence GGATCC (shown in bold type), and with BbsI, which recognises the sequence GTCTTC (shown in bold type) and cuts upstream of this sequence. The resulting products are shown. (B) The *PAD1* 3'UTR is amplified from genomic DNA using primers which contain a BamHI site at the 5'end (shown in bold) and a BsmAI recognition site, GTCTC (shown in bold). The BsmAI enzyme cuts downstream of the recognition sequence. The resulting product from a digestion with BamHI and BsmAI is shown. (C) The resulting product of a ligation with the CAT449 vector (digested with BamHI and BbsI) and the *PAD1* 3'UTR (digested with BamHI and BsmAI). Only the regions of interest are shown here. A full diagram of the CAT449 vector is shown in Figure 3.3. Arrowheads indicate cleavage sites for restriction enzymes. Red bases represent the CAT449 vector and blue bases represent the *PAD1* 3'UTR insert.

It was necessary to include additional sequence 3' of the polyadenylation site in order to ensure correct processing of the polyadenylation site (Matthews et al., 1994); therefore, for the full length *PAD1* 3'UTR, the entire 966nt intergenic region between *PAD1* and *PAD2* was inserted into the CAT449 vector. Deletions were

made at intervals along the 3'UTR from the 5' end to allow for correct processing of the polyadenylation sites at the 3' end (Figure 3.6).

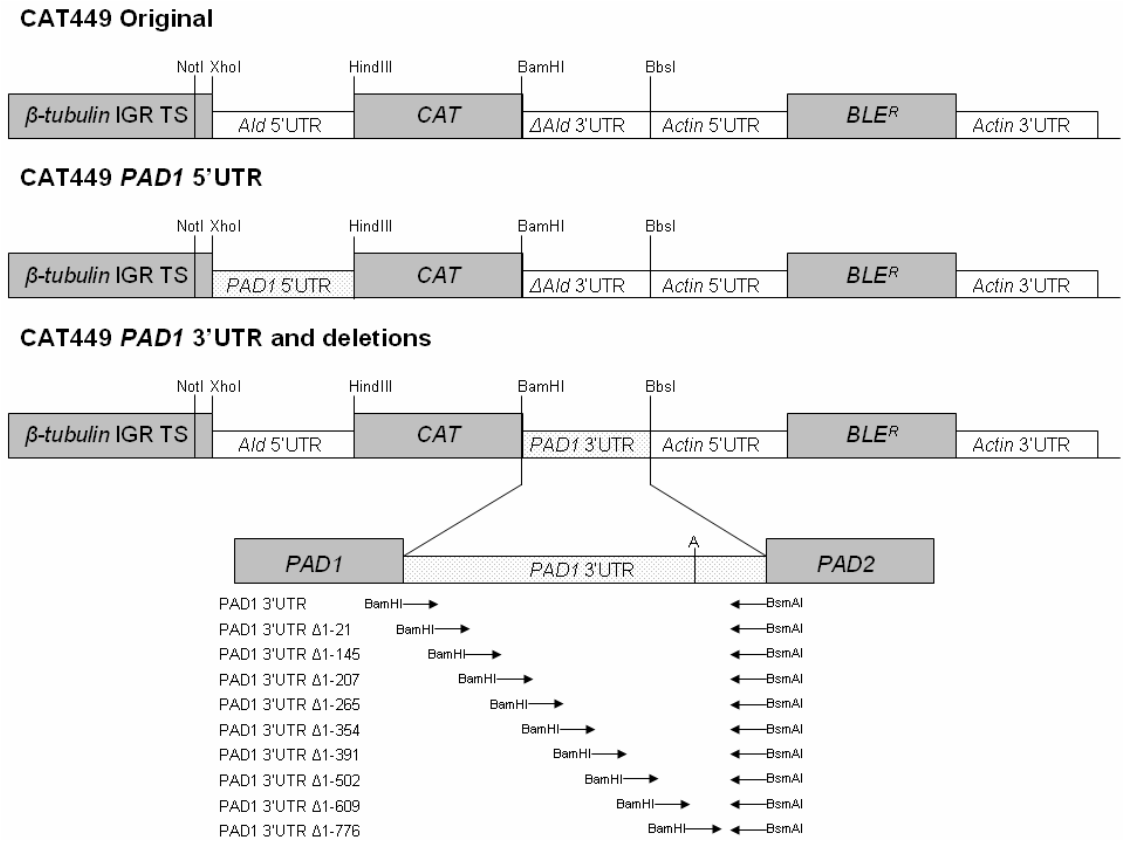


Figure 3.6 A schematic diagram of the CAT449 *PAD1* 3'UTR deletion constructs. To create the CAT449 *PAD1* 5'UTR construct from the CAT449 original construct, the *aldolase* 5'UTR was excised and replaced by the *PAD1* 5'UTR using XhoI and HindIII. To create the CAT449 *PAD1* 3'UTR deletion constructs for the CAT449 original construct the truncated *aldolase* 3'UTR was excised using BamHI and BbsI. This was replaced with either the full length *PAD1* 3'UTR or a truncated form created by amplifying the 3'UTR using a series of primers which progressively deleted regions from the 5' end. All 5' primers contained a BamHI site and the 3' primer contained a BsmAI site with a complementary sequence to that created by the BbsI digestion in the CAT449 construct (Figure 3.5). The 'A' in the *PAD1* 3'UTR depicts the 707-710nt and 722-723nt polyadenylation sites. This diagram is not to scale.

In order to confirm that it is the 3'UTR and not the 5'UTR that is responsible for *PAD1* stumpy-specific expression, the 5'UTR was also integrated into the CAT449 vector. The *aldolase* 5'UTR was excised via XhoI and HindIII restriction sites (Figure 3.4). To introduce the *PAD1* 5'UTR, XhoI and HindIII sites were added to 5' and 3' primers, respectively, designed to amplify the 582nt upstream of the *PAD1* start codon (Figure 3.6). It was expected that if the 3'UTR of *PAD1* was indeed controlling stage-specific gene expression then the 5'UTR would not cause repression of expression in monomorphs.

Once created, constructs were transformed into XL1 *E.coli* and a small scale preparation of each of the plasmids was isolated. The plasmids were sequenced over the regions of interest to ensure the UTR sequences had inserted correctly and then larger-scale midi-prep DNA was prepared. Plasmids were linearised with NotI and transfected into *T.brucei* 427 by electroporation. After 24 hours, transfectants were selected in 2ml selective media in 24-well plates with 0.5µg/ml phleomycin. Addition of the phleomycin caused cell death of non-transformed cells in 2-3 days. After 5-7 days, wells containing proliferating cells were considered as positive transfectants. Two clones were selected for each cell line and grown in 5ml selective media in 6-well plates, then in 10ml selective media in 25cm^2 tissue culture flasks. Cell lines were stored at -80°C and in liquid nitrogen. All cell lines made for the purposes of 3'UTR analysis of the *PAD1* 3'UTR are listed in Table 3.1.

Construct	Description of 3'UTR coupled to CAT reporter gene
CAT449 Original	Truncated *aldolase* 3'UTR
PAD1 5'UTR	Truncated *aldolase* 3'UTR Note: *aldolase* 5'UTR replaced with *PAD1* 5'UTR
Full PAD1 3'UTR	Entire 966nt intergenic region between *PAD1* and *PAD2*
PAD1 3'UTR Δ1-21	Nucleotides 1-21 deleted from 'Full PAD1 3'UTR'
PAD1 3'UTR Δ1-145	Nucleotides 1-145 deleted from 'Full PAD1 3'UTR'
PAD1 3'UTR Δ1-207	Nucleotides 1-207 deleted from 'Full PAD1 3'UTR'
PAD1 3'UTR Δ1-265	Nucleotides 1-265 deleted from 'Full PAD1 3'UTR'
PAD1 3'UTR Δ1-354	Nucleotides 1-354 deleted from 'Full PAD1 3'UTR'
PAD1 3'UTR Δ1-391	Nucleotides 1-391 deleted from 'Full PAD1 3'UTR'
PAD1 3'UTR Δ1-502	Nucleotides 1-502 deleted from 'Full PAD1 3'UTR'
PAD1 3'UTR Δ1-609	Nucleotides 1-609 deleted from 'Full PAD1 3'UTR'
PAD1 3'UTR Δ1-776	Nucleotides 1-776 deleted from 'Full PAD1 3'UTR'
PAD1 3'UTR Δ386-564	Nucleotides 386-564 deleted from 'Full PAD1 3'UTR' and replaced by a HindIII restriction site
PAD1 3'UTR Δ354-624	Nucleotides 354-624 deleted from 'Full PAD1 3'UTR' and replaced by an XhoI restriction site
354-624 PAD1 3'UTR	Nucleotides 354-624 of PAD1 3'UTR inserted upstream of the truncated *aldolase* 3'UTR in 'CAT449 Original'

Table 3.1 All constructs created for *PAD1* 3'UTR functional analysis. All constructs were based on the CAT449 expression vector (Figure 3.4), labelled here as CAT449 Original. For the *PAD1* 3'UTR Δ386-564 and Δ354-624 constructs the deleted region was replaced by a restriction site (section 3.8). For the 354-624 *PAD1* 3'UTR construct the 354-624nt region of the *PAD1* 3'UTR was PCR amplified with primers containing HindIII sites at both the 5' and 3'end. The section was then inserted into the 5'end of the truncated *aldolase* 3'UTR at the HindIII site present between the CAT reporter gene and this UTR (Figure 3.4). All constructs were sequenced to ensure correct insertion. With the exception of the *PAD1* 5'UTR construct, the 5'UTR upstream of the CAT reporter gene was the *aldolase* 5'UTR.

3.5 The *PAD1* 3'UTR controls reporter gene mRNA abundance in slender forms

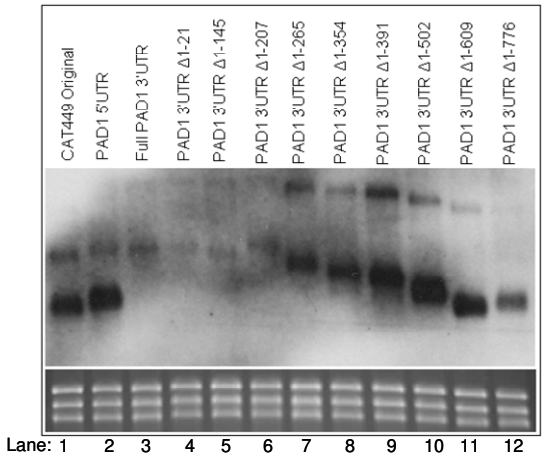
Once the CAT449 *PAD1* 3'UTR cell lines were produced (section 3.4), two independent clones from each cell line were analysed for CAT mRNA expression. Northern blots were prepared for the two clones of each cell line. A representative northern blot analysed by X-ray film is shown in Figure 3.7A and quantification of two northern blots (one from each clone) analysed by luminescence is shown in Figure 3.7B.

First, the data demonstrated that the CAT449 original construct did provide constitutive *CAT* mRNA expression as expected (Figure 3.7). Furthermore, the replacement of the *aldolase* 5'UTR with the *PAD1* 5'UTR caused a 1.6 fold increase in mRNA abundance (Figure 3.7), thus demonstrating that the *PAD1* 5'UTR was not responsible for repression of expression in slender forms.

The replacement of the truncated *aldolase* 3'UTR in the CAT449 original construct with the *PAD1* 3'UTR caused an approximately 3.7-fold decrease in CAT mRNA abundance (Figure 3.7 lane 3). Hence, the *PAD1* 3'UTR was sufficient for repression of mRNA abundance in monomorphic slender forms. This was later confirmed to be the case in pleomorphic reporter cell lines, where a reporter gene coupled to the *PAD1* 3'UTR exhibited low mRNA abundance in slender forms and high mRNA abundance in stumpy forms (Chapter 4).

Progressive deletions into the *PAD1* 3'UTR from the 5' end initially caused very little effect on mRNA abundance, indicating that first 207nt of the *PAD1* 3'UTR was not necessary for repression of expression in slender forms (Figure 3.7 lanes 3-6). Further deletions however, caused gradual alleviation of repression of expression with deletion of 1-502nt of the *PAD1* 3'UTR causing an approximately 8.1-fold increase in reporter mRNA abundance compared to the full length construct (Figure 3.7 lane 10). Hence, the latter section of the *PAD1* 3'UTR appeared to contain a slender form repression element.

A



B

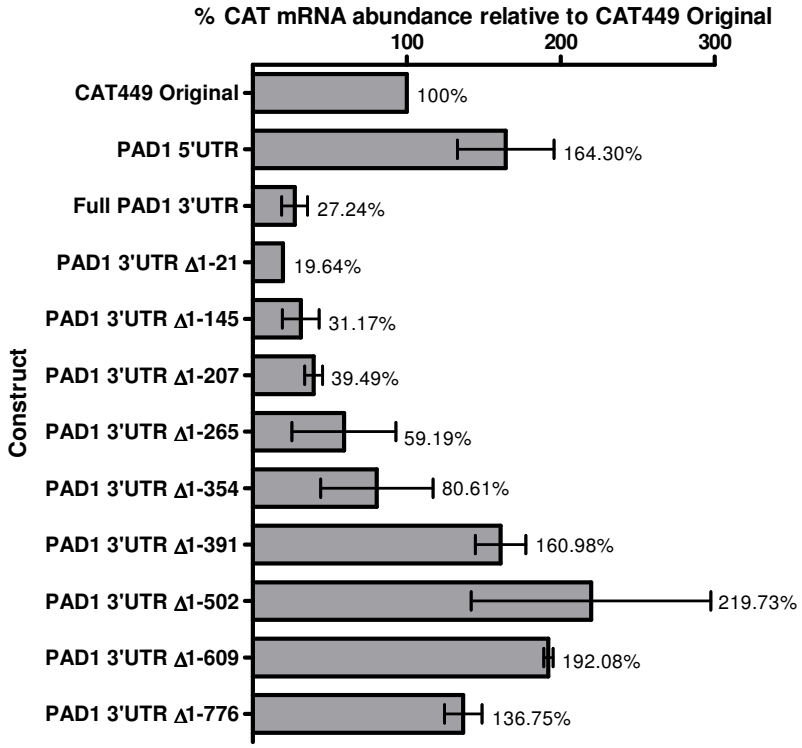


Figure 3.7 CAT mRNA abundance increases with sequential deletion of the *PAD1* 3'UTR. (A) A representative northern blot of the CAT449 *PAD1* 3'UTR deletion series shows that CAT mRNA abundance down-regulated by the *PAD1* 3'UTR compared to the CAT449 original construct. This repression of mRNA expression is alleviated with progressive deletions. Upper bands are processing intermediates, often observed when using the CAT449 construct (Mayho et al., 2006). (B) Quantification of the mRNA abundance in two independent clones for each cell line shows that the full length *PAD1* 3'UTR causes an almost 4-fold decrease in mRNA abundance compared to the CAT449 original construct and this is alleviated with progressive deletions.

The final *PAD1* 3'UTR deletion construct (Δ1-776nt) removed the polyadenylation sites mapped experimentally in section 3.2.2. Indeed, by northern blot analysis, it was observed that the progressive deletions of the full length *PAD1* 3'UTR caused progressive decreases in transcript size until the *PAD1* 3'UTR Δ1-776 construct, where there was a size increase compared to the previous deletion (Figure 3.7A lane 12). It is likely that with the loss of the preferred polyadenylation site, an alternate location is used. This indicated that the polyadenylation sites mapped in stumpy forms (section 3.2.2) are also the polyadenylation sites used in monomorphic slender forms. The sites mapped by Siegel and colleagues in procyclic forms, at positions 428, 572 and 835nt (Siegel et al., 2010) are not major sites of polyadenylation in these cell lines, as the associated changes in transcript size that would be expected are not observed by northern blot analysis (Figure 3.7A).

3.6 The *PAD1* 3'UTR controls reporter gene protein expression in slender forms

The preceding section demonstrated that the *PAD1* 3'UTR caused repression of reporter mRNA abundance in slender forms. To assess regulation at the protein level, the level of CAT protein expression was measured in the CAT449 *PAD1* 3'UTR cell lines. In all cases, two independent clones were analysed for each cell line, with each clone tested in at least two CAT ELISAs and the results were consistent between clones (Figure 3.8).

As expected, the CAT449 original construct provided good levels of CAT protein expression (Figure 3.8). Replacement of the *aldolase* 5'UTR with the *PAD1* 5'UTR caused an approximately 1.6-fold increase in CAT protein expression (Figure 3.8) precisely correlating with the 1.6-fold increase in mRNA abundance observed in section 3.5. Hence, it appears that the role of the *PAD1* 5'UTR in controlling gene expression in this context is probably entirely controlled at the level of mRNA abundance and has no discernable effect on translation efficiency. There is, alternatively, the possibility that, as the levels of expression are compared relative to the CAT449 original construct, that any differences in the relative levels of protein

compared to the mRNA are masked by equivalent difference in the CAT449 construct. It seems that this would be rather unlikely given such a close correlation between relative mRNA and protein expression yet cannot be ruled out. The high levels of protein expression in the *PAD1* 5'UTR cell line reveals, again, that the *PAD1* 5'UTR is not responsible for repression of expression in slender forms.

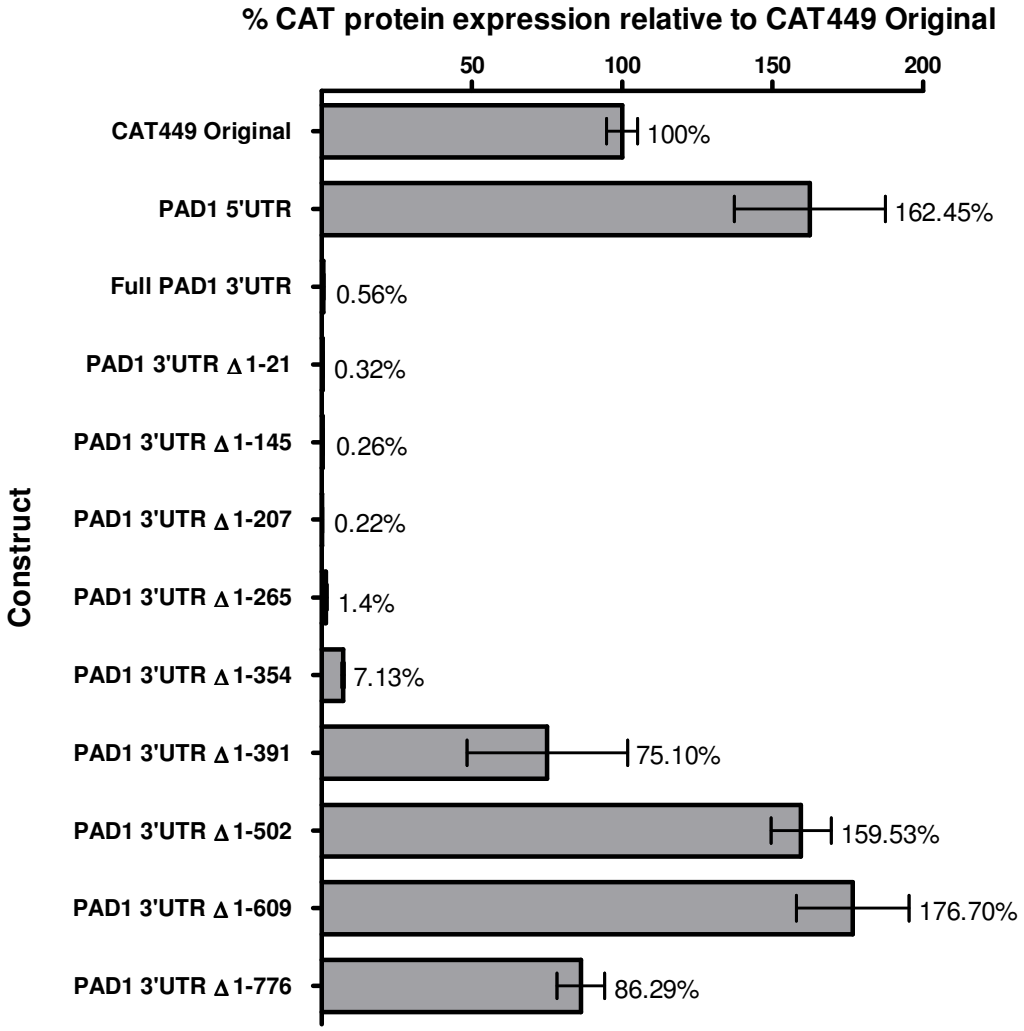


Figure 3.8 CAT protein increases with sequential deletion of the *PAD1* 3'UTR. Analysis of the level of CAT protein expression relative to the CAT449 original construct shows that CAT protein expression is down-regulated by the *PAD1* 3'UTR to a greater degree than that observed at the mRNA level. This repression of protein expression is alleviated with progressive deletions, particularly at the latter end of the molecule. Deletion 1-776nt deletes beyond the mapped polyadenylation sites.

In contrast, CAT protein was almost undetectable when the truncated *aldolase* 3'UTR of the original CAT449 construct was replaced by the full-length *PAD1* 3'UTR (Figure 3.8). There was almost a 4-fold decrease in mRNA expression but an almost 180-fold decrease in protein expression, in the *PAD1* 3'UTR construct compared to the CAT449 construct. This demonstrates that the *PAD1* 3'UTR is repressed at both the level of mRNA abundance (section 3.5) and protein expression.

Deletion of the first 207nt of the *PAD1* 3'UTR had been demonstrated to have no effect on the mRNA levels compared to the full length UTR (section 3.5). Likewise, deletion of these bases caused no discernable effect on reporter protein expression compared to the full length *PAD1* 3'UTR (Figure 3.8). Hence, it appears that the first 207nt is not necessary for repression of mRNA abundance or protein expression in slender forms. Further deletions did cause alleviation of repression of expression, however (Figure 3.8). Deletion of 1-354nt, although showing alleviation of repression of expression compared to the full length transcript, still demonstrated a low level of CAT protein expression compared to the CAT49 original construct. It may be that deletion of sections of the UTR not directly involved in repression of expression cause disruption of the secondary structure of other parts of the molecule (section 3.7). In the case of the *PAD1* 3'UTR Δ1-265 and Δ1-354 constructs, the small rise in CAT protein expression may, speculatively, represent a small disruption in the folding of the downstream region of the UTR which reduces the efficiency of binding to an RNA binding protein, for instance.

Deletion beyond 354nt caused a large rise in CAT protein expression, up to over a 300-fold increase in protein expression in the *PAD1* 3'UTR Δ1-609nt construct compared to the full length UTR, suggesting that there was specific deletion of some element causing repression of protein expression in slender forms (Figure 3.8). Thus, the *PAD1* 3'UTR contains one or more slender form repression elements at the distal end of the 3'UTR which cause repressed mRNA abundance and repressed protein expression. It is also worth noting that the *PAD1* coding sequence is clearly not necessary for stumpy-specific gene expression, although this does not exclude a potential additional level of repression controlled by the protein coding sequence.

3.7 The secondary structure of the *PAD1* 3'UTR

In trypanosomes it is often the 3'UTR of a gene that is responsible for the control of expression via the action of regulatory RNA binding proteins. These proteins recognise and bind to the three-dimensional RNA molecule and can either promote or repress expression (section 1.10). Although linear regulatory motifs have been identified, any interaction will be dependent on the structure of the RNA molecule; even linear motifs can fold into specific loops (Mayho et al., 2006) that may dictate association with the RNA binding protein.

The secondary structure of an RNA molecule can be predicted using algorithms, such as Sfold and Mfold, that are publicly available online (http://sfold.wadsworth.org and http://mfold.bioinfo.rpi.edu/ respectively) (Ding et al., 2005; Ding and Lawrence, 2003; Mathews et al., 1999; Zuker, 2003). Mfold predicts secondary structure based on minimising the free energy of a particular mRNA, such that the molecule folds into the lowest possible energy state (Mathews et al., 1999; Zuker, 2003). In contrast, Sfold predicts a sample of 1000 possible secondary structures, within which, structural features will recur (Ding and Lawrence, 2003). The resulting optimum predicted structure, known as the ensemble centroid, is the structure which is the most similar to all the other predicted structures (Ding et al., 2005). For the analysis described here, the Sfold program was utilised as it was demonstrated to be less prone to error (Ding et al., 2005). Although such algorithms provide models of the potential structure into which an RNA molecule may fold, these are still very much predictions. However, they do provide a framework for directing further experimental analysis.

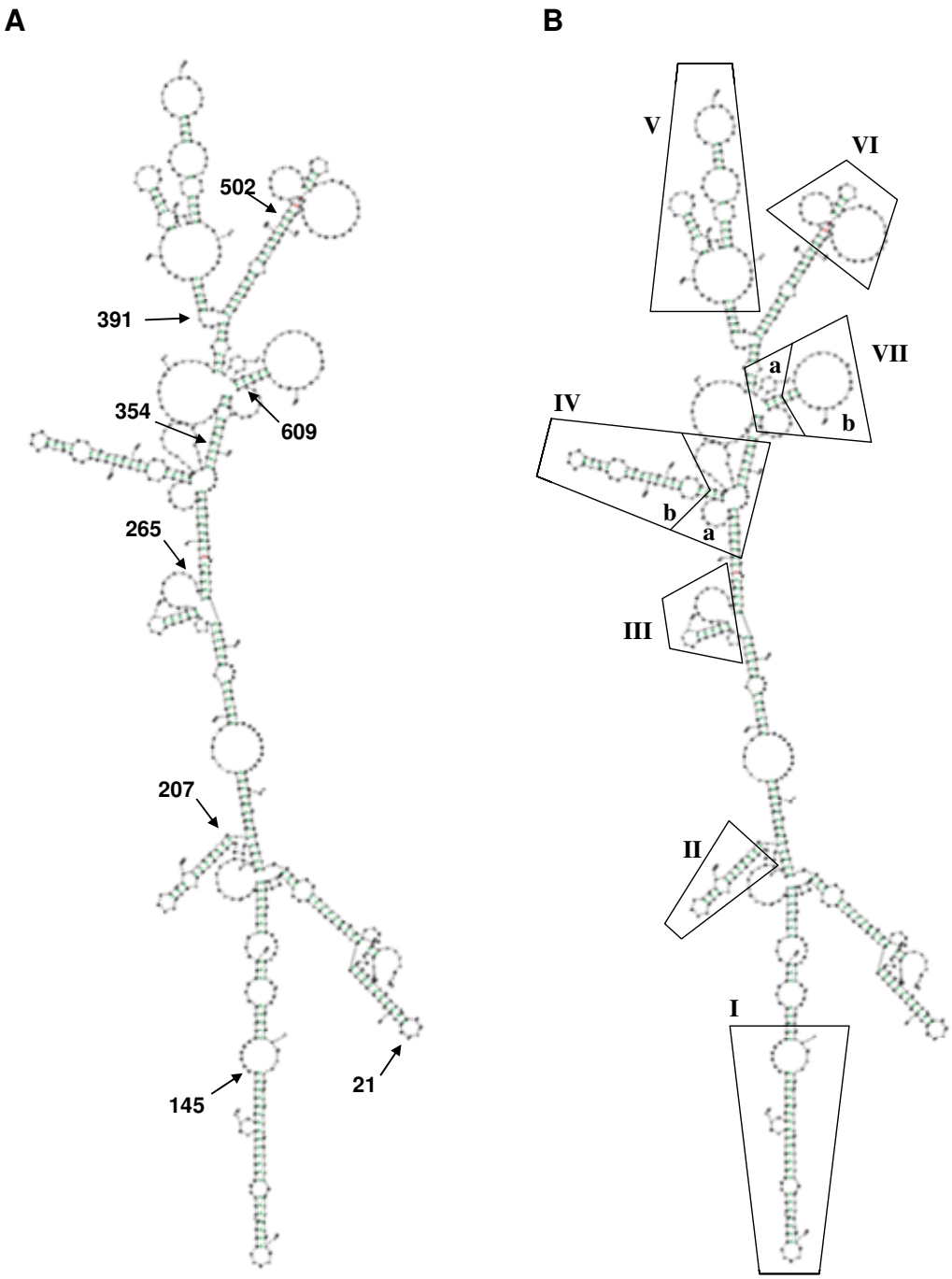


Figure 3.9 Structural features of the *PAD1* 3'UTR as predicted by Sfold. (A) The predicted secondary structure of the full length *PAD1* 3'UTR using position 710nt as the polyadenylation site. The nucleotide positions indicated correlate with the position of the deletions made in the *PAD1* 3'UTR deletion series produced in section 3.4. (B) There are various stem-loops and bulges in the *PAD1* 3'UTR, some of which have been indicated by numbered boxes. This is not an exhaustive classification of the observed structural features and does not relate to known functional features.

Using Sfold analysis (http://sfold.wadsworth.org) (Ding et al., 2005; Ding and Lawrence, 2003), the secondary structure of the *PAD1* 3'UTR was predicted (Figure 3.9A), using position 710nt as the polyadenylation site (see section 3.2). The *PAD1* 3'UTR is fairly long, compared to the median 3'UTR length in trypanosomes of 400nt (Siegel et al., 2010), and has various potential structural features, which may or may not have a functional relevance. Some of these potential features have been highlighted in Figure 3.9B. It appears that the secondary structure clusters into either the proximal (structures I and II) or distal (structures IV-VII) end of the molecule spatially separated by a 'linker' region (containing element III).

In order to determine if any of the potential structural features of the 3'UTR can be implicated in the stumpy-specific expression of *PAD1*, the predicted secondary structure of each of the deletion constructs was produced (Appendix B). As the 1-776nt deletion construct deletes beyond the polyadenylation site it was not considered in this analysis. Table 3.2 indicates both the sequences and potential structures that are disrupted in each deletion. With the aim of identifying potentially important structures in the control of stumpy-specific gene expression, the presence or absence of structures was compared to CAT protein expression in the various cell lines.

Alleviation of repression of CAT protein expression in slender forms was first seen to some degree in deletion 1-354nt and more so with, 1-391nt (Figure 3.8). This correlates with the loss of structures IVb and V (Table 3.2). The most marked increase in CAT protein expression was with deletions 1-502nt and 1-609nt (Figure 3.8) which correlate with the additional loss of structures VI and VIIa (Table 3.2). These structures, IVb-VIIa, are clustered at the distal end of the predicted secondary structure (Figure 3.9). Evidently this region of the molecule plays a role in repression of expression in slender forms. It may not be any one of these structures in particular that are involved in stumpy-specific gene expression, but rather the configuration of two or more in three-dimensional space.

	Structure																	
	I		II		III		IV a		IV b		V		VI		VII a		VII b	
Deletion	Seq	Str	Seq	Str	Seq	Str	Seq	Str	Seq	Str	Seq	Str	Seq	Str	Seq	Str	Seq	Str
Full length																		
Δ1-21																		
Δ1-145																		
Δ1-207																		
Δ1-265																		
Δ1-354																		
Δ1-391																		
Δ1-502																		
Δ1-609																		

Table 3.2 The features predicted to be present or absent in the *PAD1* 3'UTR deletion series. The presence (white squares) or absence (dark grey squares) of the sequence (Seq) and secondary structure (Str) of each of the structural features highlighted in Figure 3.9B, in each of the 3'UTR deletions, are indicated. For four of the deletion constructs structure VIIb was still present but in a somewhat disrupted form (see Appendix B) and this is indicated by light grey squares in the table.

3.8 Internal deletions of the *PAD1* 3'UTR

From the experimental and structural analysis of the *PAD1* 3'UTR deletion series, it appeared that the distal region of the mRNA plays a role in stumpy-specific gene expression. However, in the initial analysis the deletions created were from the 5' end of the molecule, such that a large portion of the 3'UTR had been deleted before alleviation of repression was observed. As such, it was not clear if deletion of this region alone would be adequate for the loss of repression in slender forms. To further investigate this, two constructs containing internal deletions of this region were created.

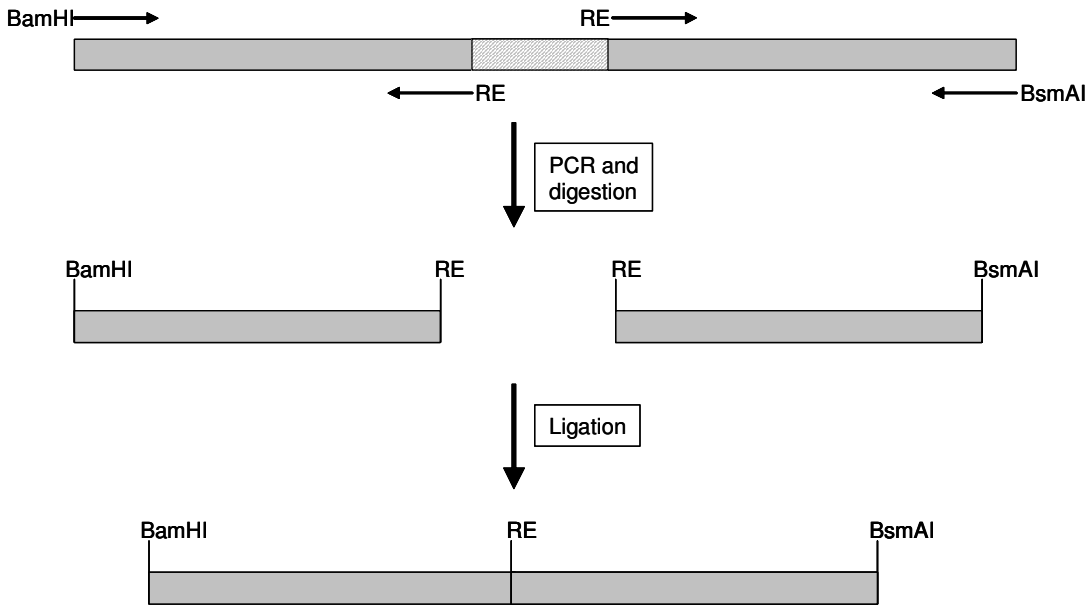


Figure 3.10 Strategy for the production of internal deletion constructs of the *PAD1* 3'UTR. The full length *PAD1* 3'UTR (grey box) is amplified in two halves, without the section to be deleted (hatched box); with primers containing restriction enzyme sites (RE denotes either HindIII or XhoI restriction sites). These sections can then be digested and joined by ligation before being integrated into the CAT449 vector (as section 3.4) using the BamHI and BsmAI restriction sites.

Two deletion constructs were made which delete the 386-564nt and 354-624nt internal sections of the *PAD1* 3'UTR (Table 3.1). This was achieved by amplifying the remaining sections of the 3'UTR independently with the introduction of a restriction site in order that the two sections could then be digested and joined by ligation (Figure 3.10). These inserts were then introduced into the CAT449 expression vector as for the previous inserts (section 3.4). The constructs were designed so as to only minimally disrupt the remainder of the mRNA structure, as predicted by Sfold (Figure 3.11). Indeed, structures I and II at the proximal end of the molecule, as well as structure III in the 'linker' region, remain intact in both internal deletions and structure IV remains in tact in deletion 386-564nt (Figure 3.11).

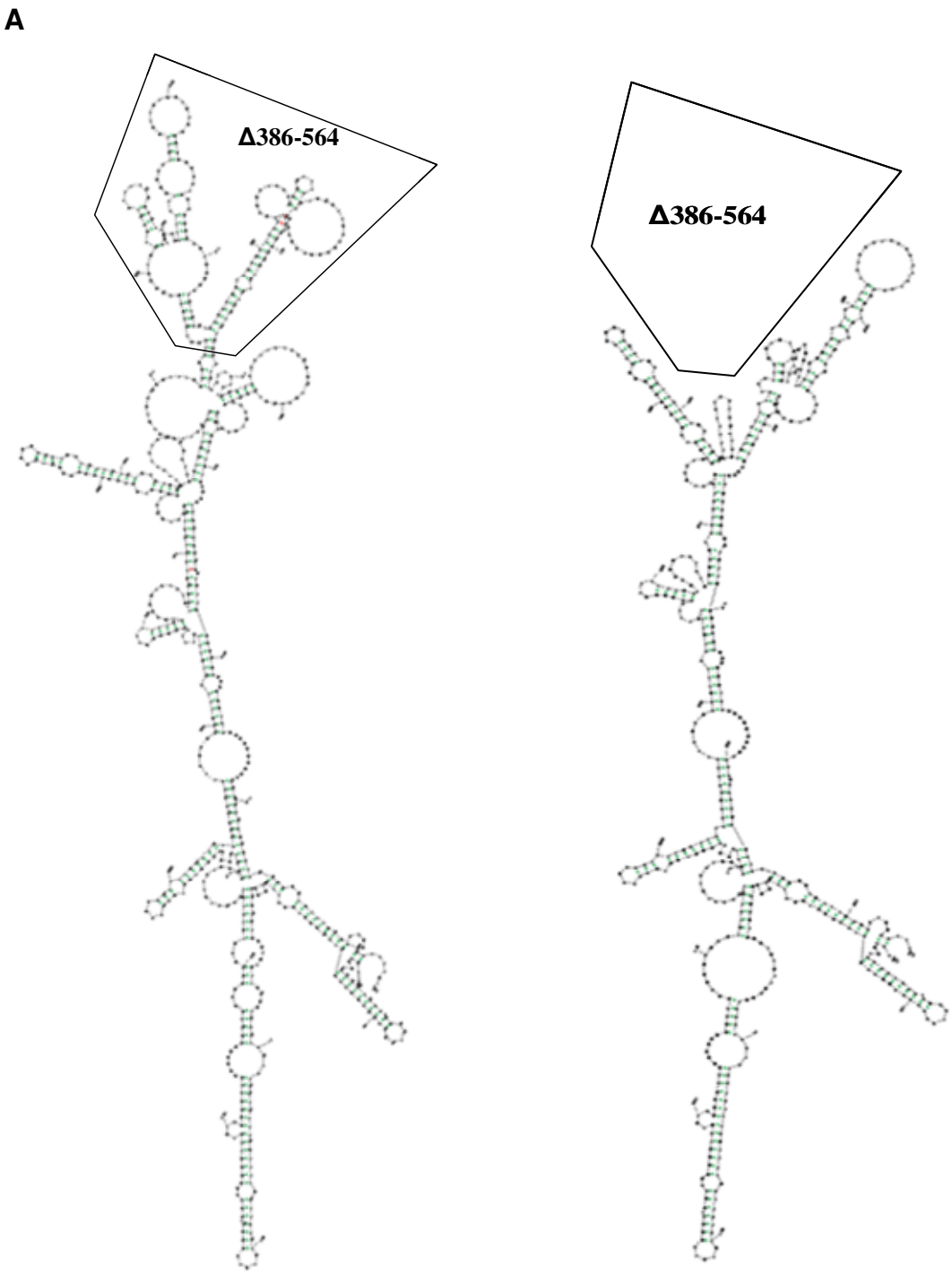


Figure 3.11 continued on the next page

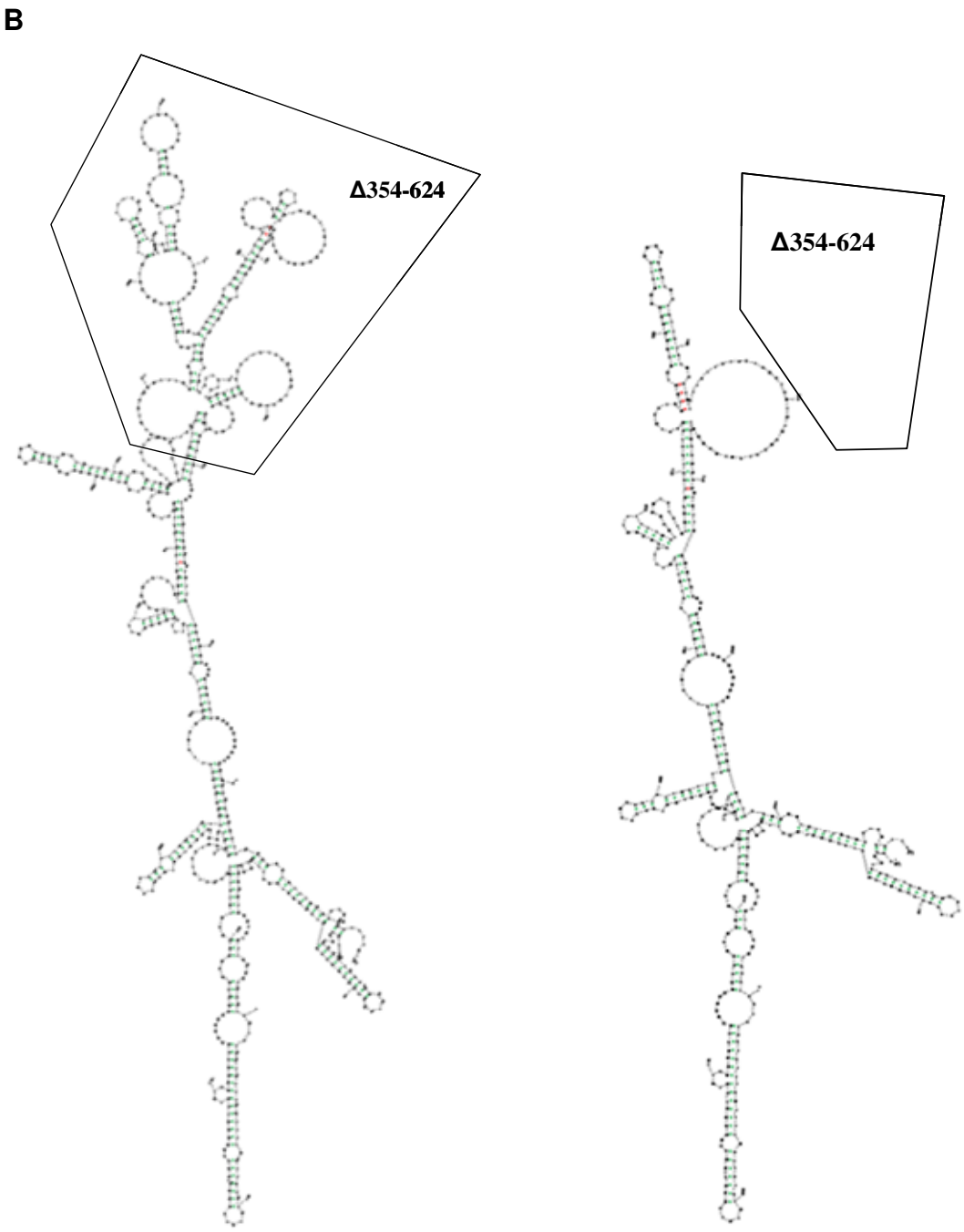


Figure 3.11 Predicted structures of the *PAD1* 3'UTR internal deletions. The predicted secondary structure of (A) the full length *PAD1* 3'UTR compared to the Δ386-564nt internal *PAD1* 3'UTR deletion and (B) the full length *PAD1* 3'UTR compared to the Δ354-624nt internal *PAD1* 3'UTR deletion using position 710nt as the polyadenylation site. Boxes on full length structures indicate the structures that are deleted. Boxes on deletion constructs indicate approximate location of deleted structures. Constructs were designed so as to minimise disruption to the remainder of the structure.

Two clones per construct were created and the level of CAT mRNA and protein was determined (Figure 3.12). The *PAD1* 3'UTR Δ386-564 construct had no discernable effect on the level of CAT mRNA or protein levels compared to the full length *PAD1* 3'UTR, with no alleviation of repression observed. This demonstrated that loss of the 386-564nt region of the *PAD1* 3'UTR was not sufficient for alleviation of repression of expression in monomorphic slender forms and hence this region is not necessary for repression of expression.

On the contrary, the *PAD1* 3'UTR Δ354-624 construct demonstrated an approximately 2-fold increase in CAT mRNA abundance compared to the full length *PAD1* 3'UTR (Figure 3.12A and B) but still remains below the level of mRNA abundance in the CAT449 original construct and therefore does not alleviate repression of mRNA abundance to the same degree as was observed for the deletions made from the 5' end of the molecule (compare with the alleviation of mRNA repression observed in Figure 3.7). Additionally, the *PAD1* 3'UTR Δ354-624 construct caused no difference in CAT protein expression compared to the full length *PAD1* 3'UTR (Figure 3.12C). This suggested that the 354-624nt region of the *PAD1* 3'UTR was not necessary for the repression of expression in monomorphic slender cells. The partial alleviation of repression of mRNA abundance may indicate that part of this region is responsible, at least to an extent, for *PAD1* mRNA abundance. Alternatively, it may be that deletion of this region is not at all sufficient for alleviation of repression, and some more proximal region is acting as a dominant control region at the level of both mRNA abundance and protein expression. The alleviation of repression of mRNA abundance in the *PAD1* 3'UTR Δ354-624 may in this case simply represent some disruption to the remainder of the molecule rather than a direct effect of loss of this region.

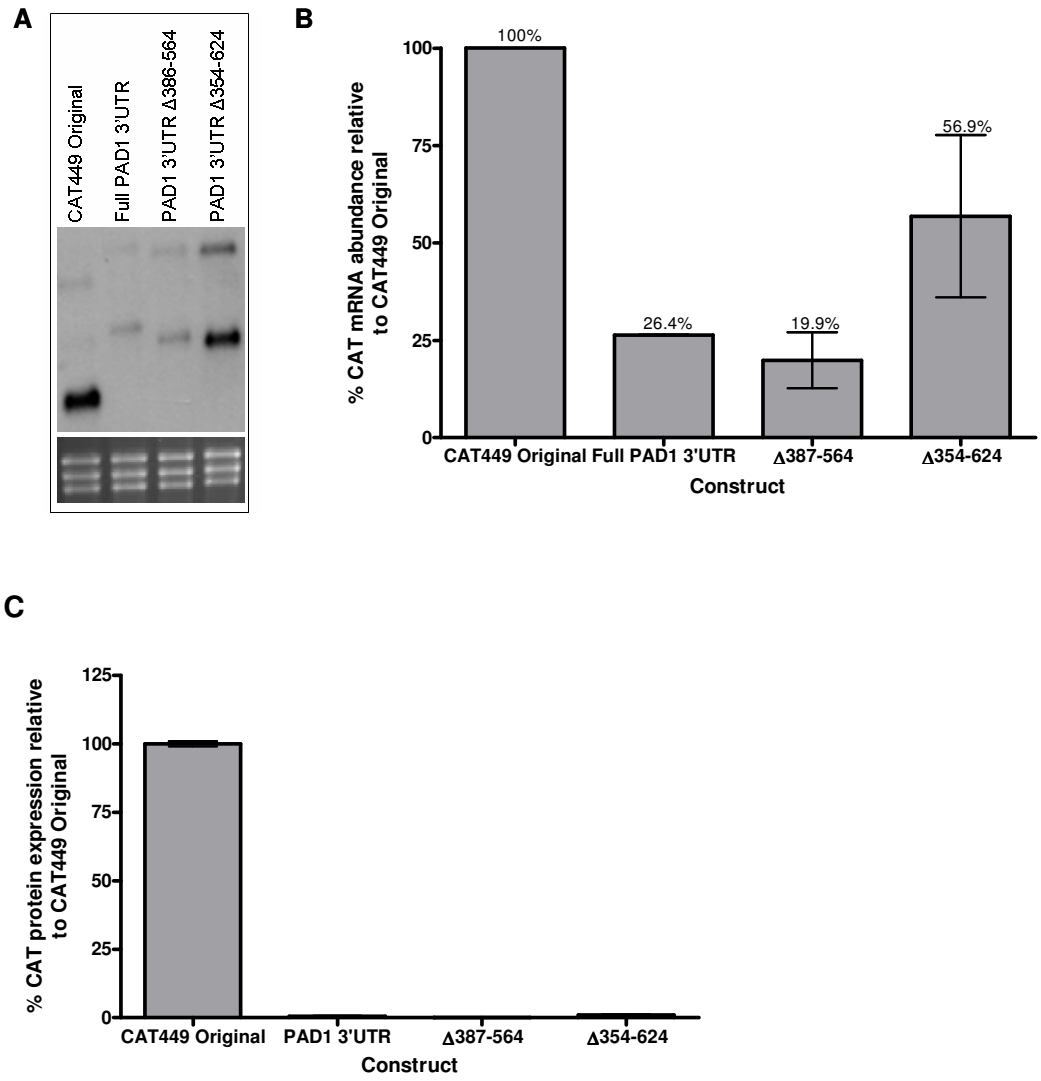


Figure 3.12 Repression of CAT protein expression is not alleviated when internal elements of the *PAD1* 3'UTR are deleted, but there is some effect on CAT mRNA. (A) A representative northern blot of one clone per construct is shown. (B) Quantification of the mRNA abundance in two independent clones for each cell line. As previously observed (Figure 3.7) the full length *PAD1* 3'UTR causes repression of mRNA abundance compared to the CAT449 original construct. The *PAD1* 3'UTR Δ387-564 construct causes no change in the level of mRNA abundance compared to the full length construct demonstrating that the 387-564nt region of the molecule is not necessary for repression of mRNA abundance. Deletion of the 354-624nt region of the *PAD1* 3'UTR did cause some alleviation of repression of mRNA abundance. (C) Neither the *PAD1* 3'UTR Δ387-564 nor the *PAD1* 3'UTR Δ354-624 construct caused alleviation of repression of CAT protein expression in slender forms.

3.9 Section 354-624nt of the *PAD1* 3'UTR can act independently to repress reporter gene expression

It had been demonstrated that the loss of the 354-624nt region of the *PAD1* 3'UTR had not been sufficient to cause complete alleviation of repression of reporter gene mRNA abundance or any alleviation of repression of reporter protein expression (section 3.8). However, in sections 3.5 and 3.6, it was demonstrated that deletion up to position 354 of the *PAD1* 3'UTR had also not been sufficient for complete alleviation of reporter gene expression, with full alleviation of repression not observed until deletions beyond this position (Figure 3.7 and Figure 3.8). This indicated that there were at least two, complementary, elements in the *PAD1* 3'UTR responsible for stumpy-specific gene expression and deletion of either alone was not sufficient for full alleviation of repression of expression. If the 354-624nt region of the *PAD1* 3'UTR did indeed contain one of two (or more) stumpy-specific control regions, then insertion of this region alone into another 3'UTR should cause at least some repression of expression in slender forms.

In order to investigate this, the 354-624nt region was inserted into the truncated *aldolase* 3'UTR of the CAT449 original construct (Table 3.1). In the CAT449 construct, there is a HindIII restriction site between the CAT reporter gene and the truncated *aldolase* 3'UTR (Figure 3.4). The 354-624nt region of the *PAD1* 3'UTR was therefore amplified using primers with the addition of HindIII restriction sites at both the 5' and 3' ends. The amplicon was then digested with HindIII and ligated into the CAT449 construct which had also been digested with HindIII. To ensure that only one copy of the 354-624nt region has inserted into the construct, and that this section was in the correct orientation, the region was sequenced before use.

Two clones containing this construct were created and were analysed for CAT mRNA and protein expression (Figure 3.13). Firstly, by northern blot it can be observed that the insertion of the 354-624nt region of the *PAD1* 3'UTR into the truncated *aldolase* 3'UTR of the CAT449 original construct caused a size-shift of the processed mRNA (Figure 3.13A) indicating that the normal *aldolase* polyadenylation site is utilised. The insertion of this 354-624nt region did cause repression of CAT

mRNA abundance in slender forms compared to the CAT449 original construct where mRNA (Figure 3.13A and B). This repression, was not however, to the same extent as the full length *PAD1* 3'UTR where much greater repression is observed (Figure 3.13A and B). The same is true for CAT protein expression, where the insert causes 3-fold, but not complete, repression of protein expression. This suggests, that either the 354-624nt region is not sufficient for complete repression of gene expression in slender forms or that the region is unable to fold correctly out of context of the remainder of the *PAD1* 3'UTR or in the presence of the truncated *aldolase* 3'UTR. The ability of the element to cause some repression of both mRNA and protein expression, however, suggests that the element is able to fold into a somewhat functional conformation, suggesting incorrect folding may not be completely responsible for the incomplete repression. These data, along with the observation that deletion of this region in the context of the full length *PAD1* 3'UTR was not sufficient to cause alleviation of repression indicates that there is some other element in the 3'UTR controlling stumpy-specific gene expression, yet this region does contribute.

It would be of interest to create a further cell line where an unrelated sequence, of a similar size, is inserted into the truncated *aldolase* 3'UTR in the CAT449 construct in order to ensure that the repression of expression observed here was not a non-specific effect whereby the presence of an additional sequence is interfering with mRNA processing or translation of the truncated *aldolase* 3'UTR.

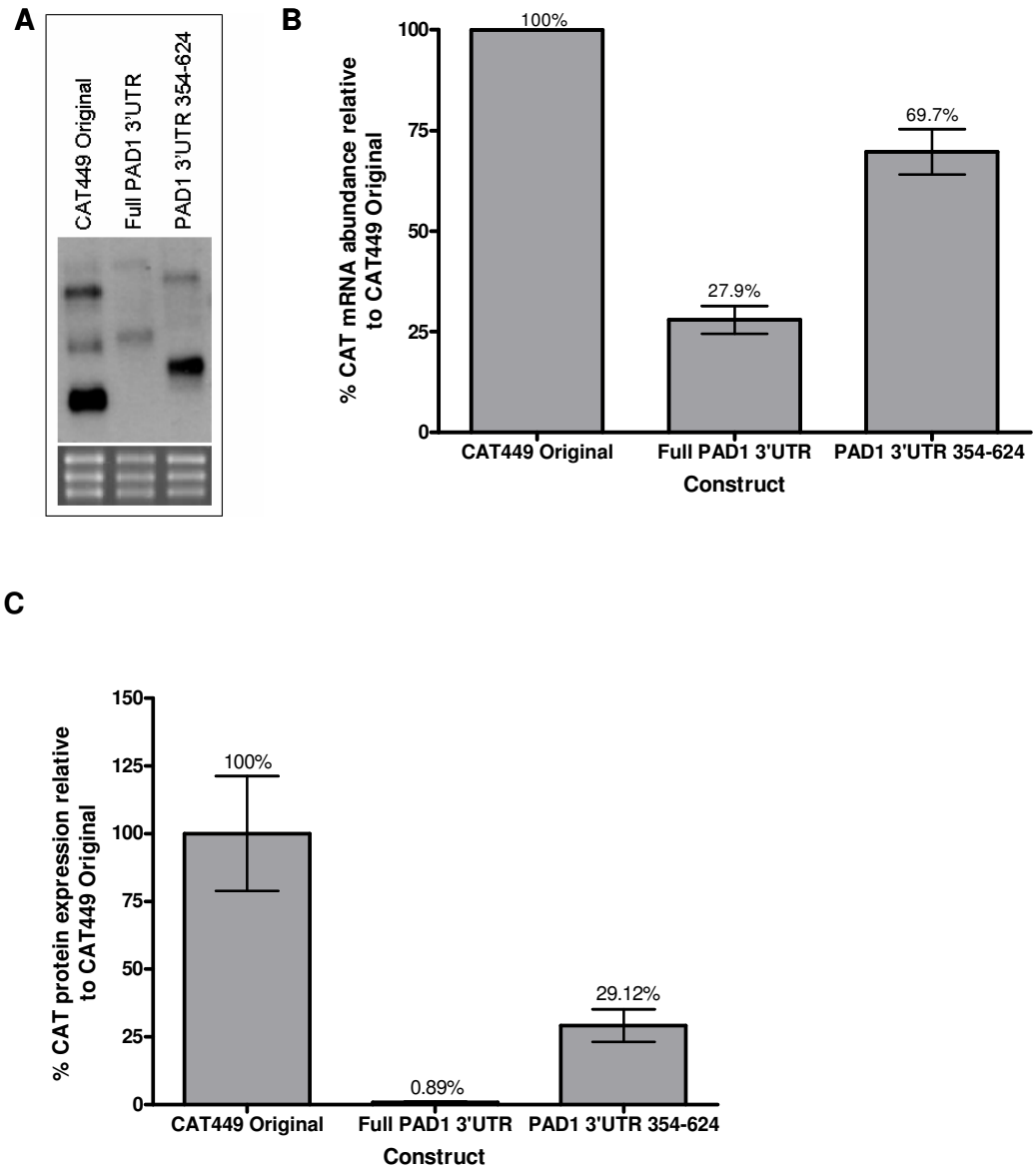


Figure 3.13 Insertion of the 354-624nt region of the *PAD1* 3'UTR into the truncated *aldolase* 3'UTR in CAT449 causes some repression of expression in slender forms. (A) A representative northern blot of one clone per construct is shown. (B) Quantification of the mRNA abundance in two independent clones for each cell line. As previously observed (Figure 3.7) the full length *PAD1* 3'UTR causes repression of mRNA abundance compared to the CAT449 original construct. Insertion of the 354-624nt region into the truncated *aldolase* 3'UTR of the CAT449 construct caused a decrease in the level of mRNA abundance compared to the CAT449 original construct but not to the same extent as the full length *PAD1* 3'UTR demonstrating that the 354-624nt region of the molecule does contribute to repression of mRNA abundance. (C) Insertion of the 354-624nt region into the truncated *aldolase* 3'UTR of the CAT449 construct caused a decrease in CAT protein expression compared to the CAT449 original construct.

3.10 Treatment of the *PAD1* 3'UTR reporter cell lines with a stumpy-inducing compound

It had been established that the *PAD1* 3'UTR causes repression of expression in slender forms. However, during the slender to stumpy transition the repression of expression must be relieved such that the *PAD1* is expressed during the stumpy-life stage. There are numerous possibilities for how this may occur. It may be that the *trans*-acting factors responsible for repression in slender forms are stage-specific and therefore absent in stumpy forms. Otherwise they may associate with different proteins and machinery such that they promote rather than repress expression in the stumpy life stage. Instead, a stumpy-specific RNA-binding protein could potentially out-compete a slender-specific protein for the same region of the 3'UTR. Or alternatively, it may be that *cis*-acting signals other than the slender repression element are responsible for the promotion of expression in stumpy forms. Of course, it could be that a combination of these suggestions act together to provide the developmental regulation of gene expression observed for *PAD1*.

A deletion series of the *PAD1* 3'UTR had been created in monomorphic slender cells (section 3.4). With the large number of transfections required to create the number of cell lines in Table 3.1, it was not practical to create the deletion series in a pleomorphic cell line given the low transfection efficiency of the pleomorphic cell lines available. Therefore in order to examine reporter gene expression in stumpy forms, the monomorphic cell lines were induced to differentiate to 'stumpy-like' forms by the use of the hydrolysable analog of cAMP, 8pCPT-2'-O-Me-cAMP. This drug has been reported to cause 'stumpy-like' formation in monomorphic cell lines (Laxman et al., 2006) and is analysed in Chapter 4 of this thesis. It is worth noting at this point that although 8pCPT-2'-O-Me-cAMP causes some of the changes associated with stumpy formation it does not appear to create 'true' stumpy forms as would be observed in natural infections. This is perhaps unsurprising given that they are derived from culture adapted monomorphic cells which, after serial passage in the laboratory, may have acquired various mutations. The treated cells do however show decreased cell growth and an increased capacity for differentiation to the procyclic life stage (Chapter 4).

The truncated *aldolase* 3'UTR in the CAT449 original construct conveys constitutive CAT protein expression. It was therefore expected that, as cells are induced to differentiate to the stumpy life stage, the CAT reporter gene expression would either remain the same as in slender forms, or if any change is observed, it would be a decrease in protein expression, as there is a global decrease in protein synthesis during the stumpy life-stage (Brecht and Parsons, 1998). Treatment of cells containing the CAT449 original construct with 10µM 8pCPT-2'-O-Me-cAMP did indeed cause a small decrease in CAT protein expression (Figure 3.14). However, the replacement of the *aldolase* 5'UTR with the *PAD1* 5'UTR caused this small decrease in protein expression to be lost. Thus, it may be that the *PAD1* 5'UTR is playing a minor role in stabilising the transcript, or assisting translation, during the stumpy life-stage.

The *PAD1* 3'UTR causes repression of reporter gene expression in slender forms and upon induction to differentiate to stumpy forms would be expected to cause an increase in CAT reporter gene expression. Indeed, treatment of cells containing the full length *PAD1* 3'UTR with 8pCPT-2'-O-Me-cAMP causes alleviation of repression of reporter gene protein expression (Figure 3.14). This indicates that the 8pCPT-2'-O-Me-cAMP is indeed inducing 'stumpy' or 'stumpy-like' formation in monomorphic slender cells and that the *PAD1* 3'UTR is sufficient for stumpy-specific gene expression. Due to very low levels of CAT protein for many of the constructs in the slender forms, it is not possible to accurately measure the fold increase in CAT protein expression in slender forms compared to stumpy forms in all cell lines, but the increase in CAT protein expression upon stumpy induction is approximately 40-fold for the full length construct.

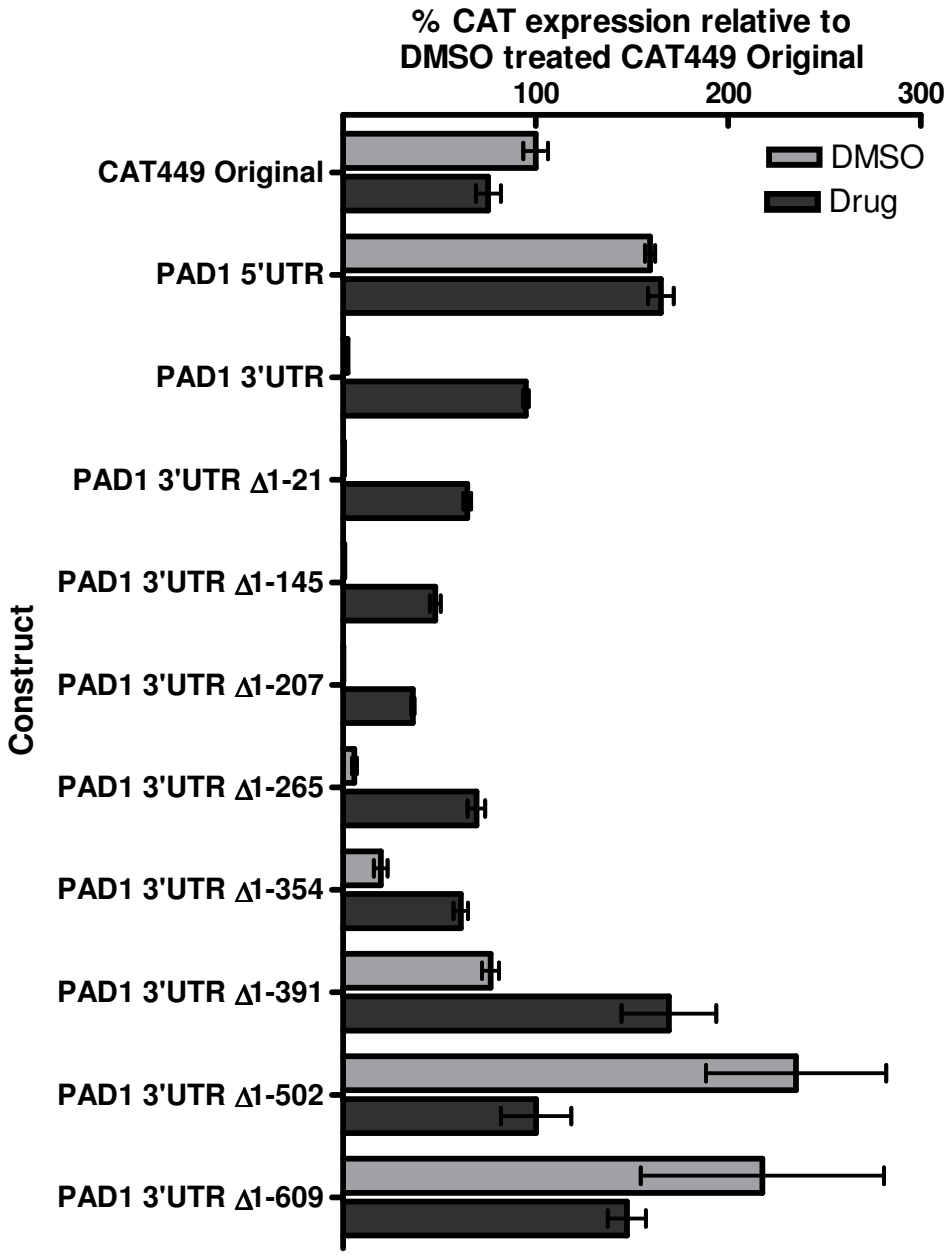


Figure 3.14 8pCPT-2'-O-Me-cAMP causes alleviation of repression of the *PAD1* 3'UTR in monomorphs and this alleviation is lost with progressive deletions of the UTR. Upon chemical induction to the stumpy life-stage, cells containing the CAT449 Original construct show a small decrease in reporter protein expression. Conversely, stumpy induction in cells containing the full length *PAD1* 3'UTR causes alleviation of repression of expression. Thus, the *PAD1* 3'UTR is sufficient for up-regulation of expression during the stumpy life stage. Progressive deletions into the 3'UTR cause loss of responsiveness to stumpy-formation.

Deletion of the even the first 21nt of the *PAD1* 3'UTR reduced the increase in total CAT protein expression upon stumpy induction (Figure 3.14) and was continued to be reduced with progressive deletions. However, it appeared that the fold increase upon stumpy-induction was actually increased with the first three deletions (Δ1-21, Δ1-145 and Δ1-207) due to the very low CAT protein expression in the DMSO-treated cell lines. Hence, it is not possible to ascertain if deletion up to position 207nt caused an increase or decrease in protein expression in response to stumpy-formation. Deletions beyond position 265nt caused a measurable decrease in fold-increase upon stumpy induction; only an 11-fold increase in CAT protein expression was measured in the Δ1-265 cell lines upon stumpy-induction and a 3- and 2-fold increase in the Δ1-354 and Δ1-391 cell lines respectively. Deletion beyond position 502nt actually caused decreased protein expression upon stumpy formation indicating that at this point responsiveness to stumpy induction had been completely lost. These data indicate that the element(s) responsible for up-regulation of protein expression upon stumpy induction is located 5' to position 502nt.

Endogenous *PAD1* is up-regulated at both the level of mRNA abundance and protein expression in stumpy forms compared to slender forms (Dean et al., 2009). Although numerous attempts were made it was not possible to determine the effect of 8pCPT-2'-O-Me-cAMP on mRNA levels. Repeated northern analyses produced inconsistent and non-reproducible results (data not shown). There was not a clear or consistent up-regulation of CAT mRNA levels upon stumpy formation despite the consistent up-regulation of protein expression. Indeed, when monomorphic slender cells were treated with 8pCPT-2'-O-Me-cAMP no up-regulation of *PAD1* mRNA was observed (Dean and Matthews, unpublished data), although PAD1 protein levels were not measured. This may indicate that the stumpy-like cells produced by this drug are not 'true' stumpy forms as would be found in pleomorphic populations (Chapter 4). Due to this, any conclusions made regarding stumpy-specific gene expression after treatment with 8pCPT-2'-O-Me-cAMP would ideally be validated in pleomorphic cell lines. However, given the difficulty in creating transgenic pleomorphic cell lines, drug-induced stumpy formation still provides a useful primary analysis to direct more informed analysis in pleomorphic lines.

3.11 Discussion

3.11.1 The *PAD1* 3'UTR utilises more than one polyadenylation site

From genome-wide analysis, the *PAD1* 3'UTR polyadenylation site used during the procyclic life-stage had been mapped to positions 428, 572 and 835nt (Siegel et al., 2010). For the purposes of the analysis of the *PAD1* 3'UTR in bloodstream forms, the polyadenylation site was determined both experimentally (section 3.2.2) and by database searching (section 3.2.1) and was shown to be at positions 707-710nt and 722-723nt in stumpy forms.

When a deletion series of the *PAD1* 3'UTR, from the 5' to 3' end, was created (section 3.4), it was observed that progressive deletions caused a decrease in transcript length, as determined by northern blot (Figure 3.7). However, the exception to this was a construct which deleted up to position 776nt and therefore beyond the polyadenylation site as determined in this thesis. In this case, the size of the transcript appeared to increase relative to the previous deletion suggesting that at this point an alternative polyadenylation site was being utilised. By contrast, if positions 428 or 572nt were polyadenylation sites in bloodstream forms then it would be expected that deletions beyond these positions would have a similar effect on transcript length, which was not observed. Further, if position 835nt was a polyadenylation site used in bloodstream forms, the deletion up to position 776nt would not be expected to have caused the relative increase in transcript length compared to the previous deletion. These data help to confirm positions 707-710nt and 722-723nt as the polyadenylation sites used in bloodstream forms.

The discrepancy between the polyadenylation sites mapped in this analysis and those mapped by Siegel *et al* may be due the different life stages in which the analysis was carried out as differential use of polyadenylation sites during different life stages has been observed in trypanosomes (Clement and Koslowsky, 2001; Erondu and Donelson, 1992). Given that position 707-710nt or 722-723nt were not identified at all in procyclic forms, this does indicate that for *PAD1* differential polyadenylation

sites may used during different life stages. In order to determine if this is indeed the case, the same experimental analysis carried out here with RNA harvested from stumpy form cells could be carried out on RNA harvested from procyclic form cells.

3.11.2 The *PAD1* 3'UTR is sufficient for stumpy-specific gene expression

As discussed in section 1.10, trypanosomes control their gene expression almost entirely post-transcriptionally (Clayton, 2002). As such, it is often the 3'UTR of a gene that is responsible for gene expression. Although stumpy-specific motifs have been identified by bioinformatic analysis (Kabani et al., 2009), no stumpy-specific gene has been experimentally analysed for the control of gene expression. Here the *PAD1* 3'UTR was coupled to a reporter gene and transfected into monomorphic slender cells. It was hypothesised that if the *PAD1* 3'UTR is sufficient for stumpy specific gene expression, that this would cause repression of expression of the reporter gene in slender cells. Indeed, this was the case (section 3.5 and 3.6), with the *PAD1* 3'UTR causing a 3.7-fold decrease in reporter mRNA abundance and almost 180-fold decrease in reporter protein expression compared to a constitutive control. Due to the very low levels of both mRNA and particularly protein caused by the *PAD1* 3'UTR all measurements were given relative to the constitutively expressing CAT449 original construct which is essentially an arbitrary measurement, but allows comparison between cell lines which would otherwise not be possible.

In this chapter, all analysis of the *PAD1* 3'UTR was carried out in monomorphic cells which do not naturally produce stumpy forms. Chemical induction of stumpy of stumpy-like formation in these monomorphic cells (section 3.10) resulted in alleviation of reporter gene protein expression under the control of the *PAD1* 3'UTR further demonstrating that the 3'UTR is sufficient for stumpy-specific gene expression. However, chemical induction of stumpy formation may not produce fully differentiated cells, as seen in pleomorphic cell lines, and so the sufficiency of the *PAD1* 3'UTR to control stumpy-specific gene expression in both slender and

naturally derived stumpy forms from an *in vivo* infection was later confirmed in pleomorphic reporter cell lines, as will be discussed in Chapter 4.

3.11.3 The *PAD1* 5'UTR does not contribute to repression of gene expression in slender forms

It was established that the *PAD1* 5'UTR does not contribute to repression of gene expression in slender forms, in fact causing an up-regulation of reporter gene expression compared to the control construct during this life-stage (section 3.5 and 3.6). A 1.6-fold increase in both reporter gene mRNA abundance and protein expression compared to the control construct implies that effect of the *PAD1* 5'UTR on protein expression is completely accountable by the increase in mRNA abundance.

When induced to differentiate to the stumpy-life stage by 8pCPT-2'-O-Me-cAMP, the control construct demonstrated a small decrease in reporter gene protein expression. The presence of the 5'UTR however, caused this small decrease to disappear. Thus, the *PAD1* 5'UTR may be playing a minor role in stabilising the transcript, or assisting translation, during the stumpy life-stage. In order to confirm if the *PAD1* 5'UTR does contribute to gene expression during the stumpy life stage, a reporter construct with both the *PAD1* 5' and 3'UTR could be created. If the 5'UTR does contribute to increased gene expression in stumpy forms, it would be expected that this new construct would have greater reporter mRNA abundance and protein expression than with the *PAD1* 3'UTR alone and if it has no effect, then no difference would be observed. Either way, the data here demonstrates that the *PAD1* 5'UTR is not necessary for either down-regulation of gene expression during the slender life stage or for up-regulation of gene expression during the stumpy life stage.

3.11.4 Progressive deletions of the *PAD1* 3'UTR cause alleviation of repression of gene expression in slender forms

Having established that the *PAD1* 3'UTR was indeed sufficient for repression of gene expression during the slender life stage it was hypothesised that deletion of any repressive element in the 3'UTR should cause alleviation of repression in slender forms. Therefore, a deletion series of the *PAD1* 3'UTR was created in monomorphic slender cells, where deletions where made from the 5'end of the UTR (section 3.4). It was observed that progressive deletions into the UTR did cause alleviation of repression of both reporter mRNA abundance and protein expression. Based on the predicted secondary structure of the molecule, the 3'UTR can be divided into various structures and these structures are essentially clustered into the proximal and distal end of the molecule (Figure 3.9 and Table 3.2). Deletion of the proximal end, up to position 265nt, of the molecule had little effect on mRNA abundance and no discernable effect on protein expression. Thus, the proximal end of the molecule is not necessary for repression of expression in slender forms. This did not discount a role for the proximal end of the molecule (as will be discussed in section 3.11.6), but did demonstrate that there was at least one repressive element in the *PAD1* 3'UTR at the distal element of the molecule.

The progressive deletions into the 3'UTR caused alleviation of repression of both mRNA abundance and protein expression, although the repression of mRNA abundance was not enough to account for the repression of protein expression, indicating that *PAD1* is controlled at both levels. However, the pattern of loss of repression of mRNA abundance correlated with the pattern of loss of repression of protein expression suggesting that the same regions were involved in both levels of control.

3.11.5 There is more than one repressive element in the *PAD1* 3'UTR

Having established the proximal end of the *PAD1* 3'UTR was not required for repression of gene expression in slender forms, two internal deletions of the UTR were made, deleting sections 386-564nt and 354-624nt (section 3.8). Having shown that deletion of these regions from the 5'end caused alleviation of repression, it was expected that these internal deletions may have a similar effect. However, the 386-564nt deletion had no effect on reporter gene expression and 354-624nt deletion had only a small effect on mRNA abundance and no effect on protein expression. Thus, having demonstrated that the proximal end of the molecule was not necessary for stumpy-specific expression (section 3.5 and 3.6), it was then shown that the distal end of the transcript was also not necessary (section 3.8). This suggests that there are at least two complementary elements controlling repression of expression in slender forms, each of which can act independently of one another.

Given the hypothesis that there are two elements involved in the control of stumpy specific gene expression in the *PAD1* 3'UTR, the 354-624nt section was then inserted into the 5'end of the truncated *aldolase* 3'UTR in the CAT449 construct (section 3.9). It was expected that if this region was a repressive element in slender forms, that a decrease in reporter gene expression would be observed compared to the original construct. Indeed, a decrease in both reporter mRNA abundance and protein expression was observed. The repression was not complete however, in that there was still greater reporter expression than observed with the full length *PAD1* 3'UTR. Thus, it may be that the structure of this section was disrupted out of the context of the remainder of the *PAD1* 3'UTR or that this section is not sufficient for complete repression of expression in slender forms. Indeed, deletion up to position 354nt from the 5'end (section 3.5 and 3.6) caused some alleviation of repression of expression, correlating with this hypothesis. Regardless of the extent of repression of expression, the insertion of the 354-624nt segment of the *PAD1* 3'UTR into the truncated *aldolase* 3'UTR did cause some repression of expression, demonstrating that the sequence does contain a repressive element. Given that deletion of this sequence alone in the context of the full length *PAD1* 3'UTR did not cause

alleviation of repression, this correlates with the hypothesis that there must be more than one repressive element in the UTR preventing expression in slender forms.

3.11.6 A model for the *PAD1* 3'UTR

Analysis of the *PAD1* 3'UTR has revealed that there are at least two repressive elements controlling repression of expression in slender forms. The data generated in this thesis suggests that one element resides at the proximal end of the molecule and the other at the distal end, correlating with the two predicted clusters of secondary structures (section 3.7), and that these two elements are complementary and can act independently. A schematic model of the mechanism by which the full length *PAD1* 3'UTR causes repression of expression in slender forms, along with the various deletion constructs produced in this thesis is shown in Figure 3.15.

In this model, the *PAD1* 3'UTR has two areas which contain elements responsible for repression of expression in slender forms, one at the proximal end of the molecule, the other at the distal end of the molecule (Figure 3.15A). These elements may act through trans-acting factors such as RNA-binding proteins (depicted in Figure 3.15 by red and blue circles), which may function though a variety of methods (discussed in section 3.11.8)

From the analysis of the *PAD1* 3'UTR deletion series, it is clear that the proximal end of the molecule is not necessary for repression of expression in the presence of the distal element and therefore that the distal end must be sufficient for at least some degree of repression at both the level of mRNA abundance and protein expression. In the model, this correlates to loss of the first element, but maintenance of the second repressive element (Figure 3.15B). In all cases, deletions made from the 5'end of the molecule may disrupt surrounding structures resulting in less effective repression of the element(s) still present. Deletions into the distal end of the molecule, from the 5'end, results in loss of both elements (Figure 3.15C) and therefore complete alleviation of repression.

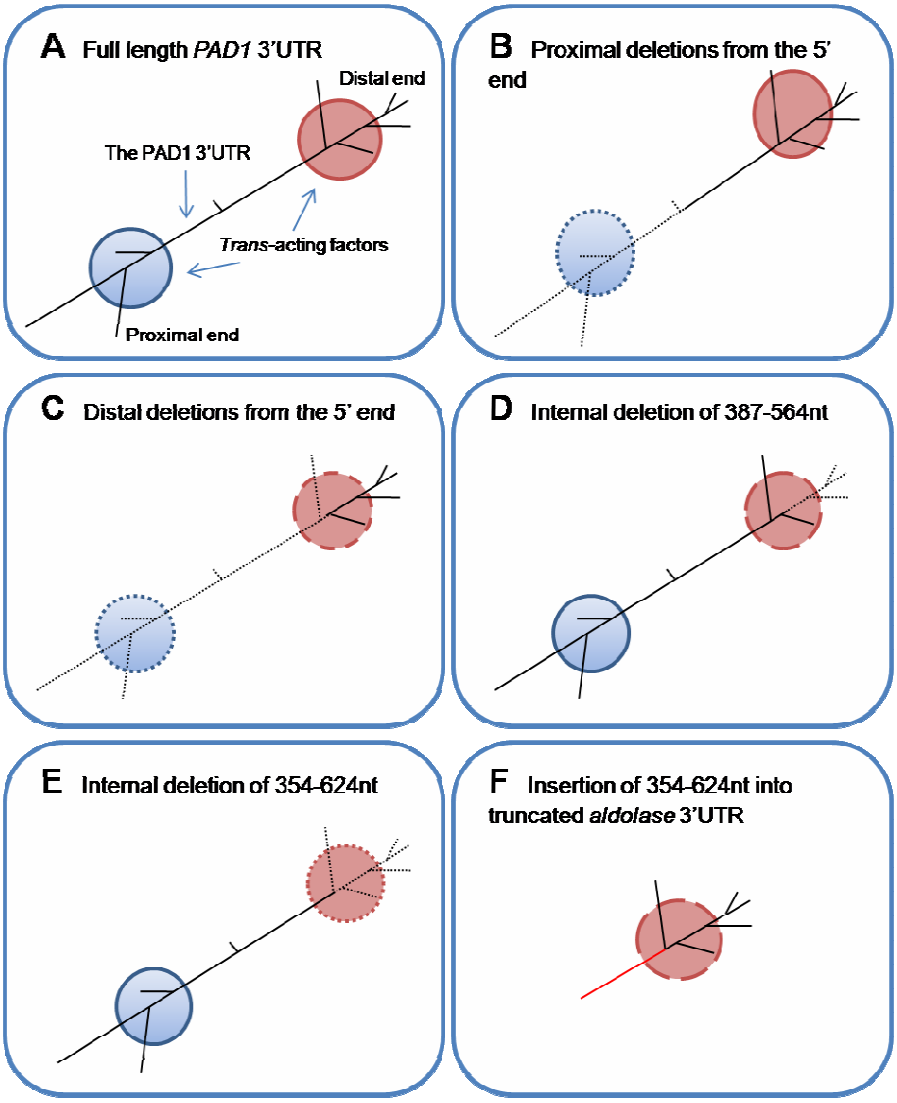


Figure 3.15 A model of the *PAD1* 3'UTR. The structure of the *PAD1* 3'UTR is represented by the black stick diagram and is based on the predicted secondary structure of the 3'UTR as shown in Figure 3.9. Solid black lines represent the presence of the structure; dotted black lines represent the absence of the structure. The circles represent possible RNA-binding proteins or protein complexes (section 3.11.8). The blue circle represents association with the proximal end of the molecule; the red circle represents association with the distal end of the molecule. Solid lines around circles represent an ability to associate with the 3'UTR, dashed lines represent a reduced capacity for association and dotted lines represent an inability to associate with the UTR. In (A) the full length *PAD1* 3'UTR contains two repression elements which can be bound by two sets of trans-acting factors. (B) Deletion of the proximal end of the molecule by deletions from the 5'end results in loss of the proximal element without complete loss of the repression of expression. (C) Deletion of both the proximal and distal elements results in loss of both elements and therefore loss of repression. (D) Internal deletion of the 387-564nt and (E) 354-624nt regions disrupts the distal element without loss of repression due to the remaining proximal element. (F) The 354-624nt section, containing the distal element, into the truncated *aldolase* 3'UTR (red line) causes some repression of expression.

Deletion of the internal 387-564nt region may not be sufficient for complete loss of the distal element (Figure 3.15D) and with the proximal element still functional in isolation there is no alleviation of repression. Internal deletion of the 354-624nt region may cause complete loss of the distal element resulting in a small increase in reporter mRNA abundance observed (Figure 3.12). Alternatively, loss of both regions may result in complete loss of the distal element. In this case, the small increase in mRNA abundance caused by the 354-624nt deletion and not in the 387-546nt deletion could be explained by a disruption to the secondary structure in the remainder of the molecule, slightly reducing the efficiency of the proximal element. Insertion of the 354-624nt segment of the *PAD1* 3'UTR, containing the distal element, into the truncated *aldolase* 3'UTR is sufficient for at least some repression of reporter gene expression. The data is consistent with either a reduced capacity for repression out with the context of the remainder of the *PAD1* 3'UTR (as shown in Figure 3.15F) or with the distal element not being sufficient alone to cause complete repression of expression in slender forms.

The up-regulation of reporter protein expression in response to chemical induction of stumpy formation in the monomorphic cell lines was reduced by deletion beyond position 265 and lost by deletion beyond position 502nt of the *PAD1* 3'UTR. This implies that both the proximal and distal elements of the *PAD1* 3'UTR may play a role in the positive regulation of gene expression in stumpy forms. Furthermore, the effect of 8pCPT-2'-O-Me-cAMP was clear and consistent at the protein level but difficult to characterise at the mRNA level. This implies that the key level of regulation upon stumpy induction is at that of the level of translation efficiency rather than mRNA abundance. It is not clear however to what extent this drug induces stumpy formation in these culture adapted monomorphic cell lines and therefore, for future examination of up-regulation in stumpy forms, analyses would ideally be carried out in pleomorphic cell lines.

3.11.7 Further characterisation of the *PAD1* 3'UTR

Having established that there are at least two repressive elements in the *PAD1* 3'UTR, and having proposed a model which accounts for the data generated in this thesis (section 3.11.6), the UTR could be studied further by creation of additional constructs in order to identify more defined control regions and to test the model. For instance, larger internal deletions could be created that move further towards the 5'end of the molecule. This may help to define a minimal sequence which is still capable of complete repression of expression which could then be used to attempt to identify *trans*-acting factors (section 3.11.8). Then, having defined a minimal sequence, specific deletions could be made that disrupt particular structures which may help to very specifically define a region of importance. Additionally, the proximal end of the molecule could be inserted into the truncated *aldolase* 3'UTR to determine if this conveys complete repression of expression out of context of the remainder of the *PAD1* 3'UTR, and responsiveness to chemical inducers of stumpy formation.

To further characterised the responsiveness of the *PAD1* 3'UTR to stumpy formation, cells containing the *PAD1* 3'UTR Δ386-564, *PAD1* 3'UTR Δ354-624 and 354-624 *PAD1* 3'UTR constructs (Table 3.1) could be treated with 8pCPT-2'-O-Me-cAMP. The most ideal way in which to test if a given element is stumpy-responsive however, would be to create constructs for the insertion into a pleomorphic cell line. Pleomorphic transfections are difficult and time-consuming and could not be used effectively to identify candidate regions, but given preliminary analysis in monomorphic cell lines with chemical induction of stumpy formation, confirmation of results with selected constructs would be possible.

The *PAD1* 3'UTR has been shown here to be responsible for repression of mRNA abundance in slender forms. It is most likely that this is through mRNA stability (rather than regulation of mRNA processing), although this was not demonstrated here. To confirm that it is indeed decreased mRNA stability that causes decreased mRNA abundance an experiment could be carried out whereby transcription is chemically inhibited by the action of Actinomycin-D and mRNA decay is monitored

over time by northern blot in order to determine the estimated half-life of the molecule. A rapid turn-over of mRNA would confirm that low mRNA abundance was due to mRNA instability. The same experiment could be carried out on other constructs from the deletion series to determine if the increase in mRNA abundance is due to increased mRNA stability.

3.11.8 *Trans*-acting factors controlling *PAD1* expression

In trypanosomes, the *cis*-acting elements in the 3'UTRs of differentially regulated mRNAs act through *trans*-acting factors, such as RNA binding proteins which affect mRNA stability and protein translation (section 1.10). In this thesis, only *cis*-acting elements of the *PAD1* 3'UTR have been investigated, but discovery of the RNA-binding proteins or complexes that are responsible for stumpy-specific expression would be of interest given that although there are a large number of potential RNA-binding proteins in the trypanosome genome (De Gaudenzi et al., 2005), relatively few have been characterised (Archer et al., 2009; Estevez, 2008; Paterou et al., 2006), and no stumpy-specific mechanisms of control of gene expression have been reported.

The mechanism by which RNA-binding proteins could be causing repression of expression in slender forms and alleviation of repression in stumpy forms could be occurring by a variety of methods. For example, for each element there may be one RNA-binding protein complex which is repressive, which is present in slender forms but absent in stumpy forms. Alternatively, the complex could have differential associations between slender and stumpy forms such that it is repressive in slender forms but promotes expression in stumpy forms. Further, it could be that two complexes compete for each element, with a repressive complex being dominant in slender forms and a positive complex being dominant in stumpy forms. Indeed, given that there are at least two elements controlling gene expression in the *PAD1* 3'UTR it may be that one acts via one of the above mechanisms and the other acts by another. Identification of proteins which are able to bind the *PAD1* 3'UTR and analysis of the expression profile of these proteins between slender and stumpy forms as well as the

identification of other proteins to which they bind could elucidate the mechanism of stumpy-specific gene expression used by *PAD1*.

In previous analysis, *cis*-acting elements have been used as bait in RNA-affinity purification in order to identify proteins able to bind that element (Estevez, 2008). With further characterisation of the *PAD1* 3'UTR (section 3.11.7) a more defined region of the UTR involved in stumpy specific gene expression could be used in such a method to identify interacting proteins. Alternatively, known RNA-binding proteins could be analysed for their ability to bind the *PAD1* mRNA. For example, *PAD6* was identified as a potential target of *Tb*DRBD3 as depletion of this protein by RNAi causes down-regulation of *PAD6* mRNA (Estevez, 2008) and it has since been shown that this protein can associate with *PAD1* mRNA (Estevez, unpublished data, personal communication). The effect of *Tb*DRBD3 depletion on *PAD1* mRNA abundance could be determined by northern blot or quantitative RT-PCR of *PAD1* mRNA in the *Tb*DRBD3 RNAi cell line. If an effect was observed, this would implicate *Tb*DRBD3 as a regulator of *PAD1* expression.

3.11.9 The *PAD1* coding sequence may still play a role in stumpy-specific expression

In this analysis the *PAD1* 3'UTR was sufficient to convey stumpy-specific gene expression onto the CAT reporter gene. In Chapter 4 the same effect is observed with another reporter gene, GUS, demonstrating that this effect is not sequence specific. However, in analysis of the control of procyclin gene expression it was demonstrated that although the 3'UTR was sufficient for stage-specific regulation in the context of a reporter gene, the coding sequence itself also played a role in preventing expression during the inappropriate life-stage even in the absence of the 3'UTR control elements (Schurch et al., 1997). Here, the *PAD1* coding sequence was not investigated and so it remains a possibility that despite the 3'UTR being sufficient for stage-specific gene expression, the coding sequence could still play a role.

3.11.10 The *PAD1* 3'UTR provides a tool for the creation of reporter cell lines for the stumpy life stage

It has been shown in this chapter that the *PAD1* 3'UTR is sufficient for control of stumpy-specific gene expression of a reporter gene. This provides a valuable tool for the analysis of stumpy formation *in vitro* by the creation of reporter cell lines coupled to the *PAD1* 3'UTR. Previous analyses on stumpy formation have relied on morphological analysis, which is time-consuming and subjective, or analysis of increased capacity for differentiation to the procyclic life-stage or mitochondrial activation (Breidbach et al., 2002; Denninger et al., 2007; Laxman et al., 2006; Reuner et al., 1997; Tyler et al., 1997). The use of reporter cell lines to identify stumpy forms would provide a high-throughput method of analysis for stumpy formation *in vitro,* which would allow for further analysis of chemical inducers of stumpy formation, such as 8pCPT-2'-O-Me-cAMP, as well as potentially providing a method for identification of SIF. The use of the *PAD1* 3'UTR in both pleomorphic and monomorphic reporter cell lines for stumpy formation was therefore investigated and will be discussed in Chapter 4.

4 Chapter 4: Monomorphic and pleomorphic reporter cell lines for the analysis of stumpy formation *in vitro.*

4.1 Introduction

Study of the slender to stumpy transition has been hindered by a lack of molecular marker for the stumpy life stage. The *PAD1* 3'UTR is sufficient for stumpy-specific gene expression, as demonstrated in Chapter 3. Therefore it should be possible to create stumpy-specific reporter cell lines that couple the *PAD1* 3'UTR with reporter genes. This would provide a method to study stumpy formation *in vitro*.

Most laboratory trypanosome strains are monomorphic, that is, they have lost the ability to differentiate into stumpy forms due to serial passage. Although monomorphs do not differentiate in response to cell density this does not exclude their use to study stumpy formation. The inability of monomorphs to differentiate may be due to an inability to detect SIF, for instance, and by targeting downstream effectors of the SIF pathway, stumpy formation could potentially be induced. Therefore, reporter cell lines for the analysis of stumpy formation in monomorphs were created.

Although monomorphic cells are a useful tool for studying the downstream effectors of SIF signalling, a pleomorphic cell line is required for the analysis of the, as yet unidentified, SIF itself. For this purpose, a pleomorphic reporter cell line was therefore created for the analysis of the effects of conditioned media, containing SIF, on pleomorphic cells.

4.2 Production of monomorphic reporter cell lines for analysis of stumpy formation

A series of reporter constructs were created for the purpose of analysing stumpy formation in monomorphs (Table 4.1). All constructs created were based on the plasmid pHD617 (Biebinger et al., 1997) which integrates into the ribosomal spacer region and uses PolI transcription (Figure 4.1). The pHD617 vector originally contained a hygromycin resistance cassette (HygroR). For the purpose of inserting two constructs into one cell line, a second pHD617 vector containing a puromycin

resistance cassette (PuroR) was also used, which had previously been generated in the laboratory (S.Dean, unpublished). Reporter genes were coupled either to the *PAD1* 3'UTR for stumpy-specific expression or to the *Actin* 3'UTR for constitutive expression. Both enzymatic (CAT and GUS) and fluorescent (GFP and CFP) reporters genes were used. Reporter constructs were transfected into trypanosomes either alone, or in combination with another construct with complementary drug resistance (Table 4.1).

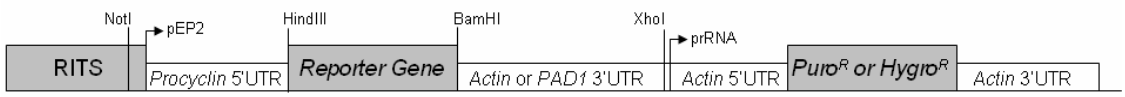


Figure 4.1 The pHD617 vector. The reporter gene (CAT, GUS, CFP or GFP) is controlled by either a constitutive *Actin* 3'UTR or a stumpy-specific *PAD1* 3'UTR. The construct contains either a puromycin or hygromycin resistance gene which is under the control of the *Actin* 3'UTR. After linearisation with NotI, the vector integrates into the ribosomal spacer locus via a ribosomal integration site sequence (RITS). As such, transcription is via PolI and uses the EP2 promoter (pEP2) to drive reporter gene transcription and the rRNA promoter (prRNA) to drive drug resistance gene expression.

Cell Line	Constructs present	Constitutive expression	Stumpy-specific expression
427 GFP-*PAD1* 3'UTR/ CFP-*Actin* 3'UTR	pHD617 GFP-*PAD1* 3'UTR puroR pHD617 CFP-*Actin* 3'UTR hygroR	CFP	GFP
427 GFP-*Actin* 3'UTR	pHD617 GFP-*Actin* 3'UTR hygroR	GFP	None
427 CAT-*PAD1* 3'UTR/ GUS-*Actin* 3'UTR	pHD617 CAT-*PAD1* 3'UTR puroR pHD617 GUS-*Actin* 3'UTR hygroR	GUS	CAT
427 GUS-*PAD1* 3'UTR	pHD617 GUS-*PAD1* 3'UTR puroR	None	GUS

Table 4.1 Cell lines created for the study of stumpy formation in monomorphs. All cell lines arise from transfection of parental 427 cells with one or more variants of the pHD617 construct (Figure 4.1).

4.2.1 Validation of fluorescent cell lines

The 427 GFP-*PAD1* 3'UTR/CFP-*Actin* 3'UTR cell line was expected to express CFP protein in monomorphic cells, but express little, if any, GFP protein (Table 4.1). In order to determine if the cells were indeed constitutively expressing CFP, cells were analysed via flow cytometry. Figure 4.2 shows that 427 GFP-*PAD1* 3'UTR/CFP-*Actin* 3'UTR cells did show a small increase in CFP expression, although most of the cells were still considered negative for CFP expression. This indicates that there was very low constitutive expression of this reporter. Cells were not expected to be positive for GFP expression as cells were monomorphic slender and thus would not express *PAD1*. However, treatment of the 427 GFP-*PAD1* 3'UTR/CFP-*Actin* 3'UTR cell line with stumpy-inducing drugs (as discussed in section 4.3) never resulted in cells positive for GFP expression (data not shown). Hence, this cell line was not considered useful for monitoring stumpy formation.

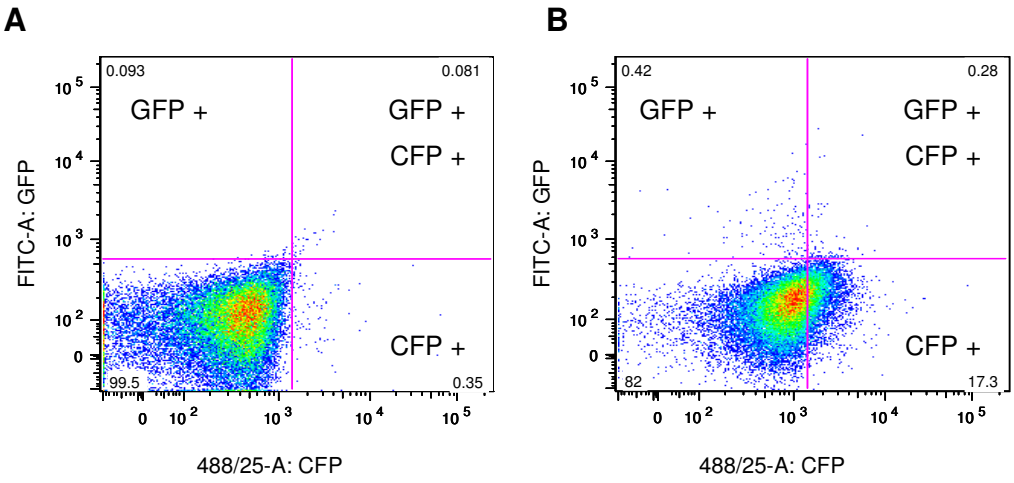


Figure 4.2 427 GFP-*PAD1* 3'UTR/CFP-*Actin* 3'UTR cells have a small increase in CFP signal compared to parental cells. A) Parental 427 cells are negative for CFP and GFP B) GFP-*PAD1* 3'UTR/CFP-*Actin* 3'UTR cells have an increase in CFP signal, yet the shift is largely not out-with the negative range. There is also a small increase in GFP positive cells. This increase in CFP expression is not high enough to clearly distinguish between positive and negative populations. The lower left-hand quadrant in each graph represents cells negative for CFP and GFP expression. The upper left-hand quadrant represents cells positive for GFP and negative for CFP. The lower right-hand quadrant represents cells positive for CFP and negative for GFP. Finally, the upper right-hand quadrant represents cells positive for both GFP and CFP expression. Values within each quadrant represent the number of cells within each quadrant as a percentage of the total number of cells.

Nine independent clones of the 427 GFP-*Actin* 3'UTR cell line have been produced. Of these, none expressed detectable GFP protein, as determined by flow cytometry (data not shown). This suggested that there is a fundamental difficulty with free fluorescent protein expression in these bloodstream cells. As attempts to produce fluorescent reporter cell lines were unsuccessful, enzymatic reporter genes were then tested.

4.2.2 Validation of enzymatic cell line 427 CAT-*PAD1* 3'UTR/GUS-*Actin* 3'UTR

The 427 CAT-*PAD1* 3'UTR/GUS-*Actin* 3'UTR cell line was expected to express GUS protein in monomorphic cells, but express little, if any, CAT protein. To quantify CAT protein levels a Roche CAT ELISA kit was used as described in Chapter 2. As expected CAT protein expression was low in untreated 427 CAT-*PAD1* 3'UTR/GUS-*Actin* 3'UTR cells, but increased in response to stumpy-inducing drugs, as will be demonstrated in section 4.3. This demonstrated the validity of this reporter system in monomorphic cells.

Previously, transgenic trypanosome cell lines have been produced which replaced a procyclin gene with the GUS gene as a marker for differentiation from bloodstream to procyclic forms (Sbicego et al., 1999). GUS activity was then assayed via a colour change reaction which was quantified by an absorbance reading (Sbicego et al., 1999). Here a similar method was used to evaluate the 427 CAT-*PAD1* 3'UTR/GUS-*Actin* 3'UTR cell line where 4-Methylumbelliferyl-β-D-glucopyranosiduronic acid (MUG) is used as a fluorescent substrate for the GUS enzyme (as recommended by Prof. I. Roditi, University of Bern). The substrate was added to the trypanosome culture in 96-well plates and absorbance measured after 2 hours, with an excitation wavelength of 355nm and emission wavelength of 460nm, as detailed in Chapter 2. The 427 CAT-*PAD1* 3'UTR/GUS-*Actin* 3'UTR cell line was shown to express GUS using this method (Figure 4.3), thus confirming successful transfection and GUS expression from the cell line. Supporting this, no GUS activity was detected in parental 427 cells.

Although CAT proved an effective reporter, for high throughput screens the GUS reporter was considered a more cost effective reporter system. Therefore, as described in section 4.4, a cell line coupling the GUS gene with the *PAD1* 3'UTR was also created (Table 4.1).

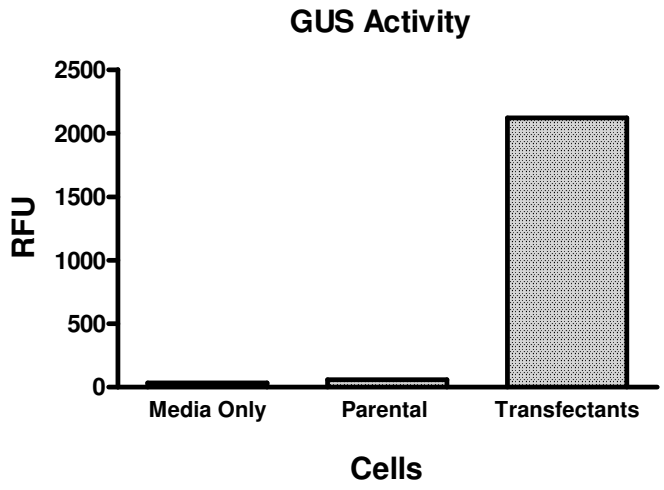


Figure 4.3 427 CAT-*PAD1* 3'UTR/GUS-*Actin* 3'UTR did express GUS protein, whereas parental 427 cells did not. 427 CAT-*PAD1* 3'UTR/GUS-*Actin* 3'UTR (transfectants), 427 parental cells and media only were tested for GUS activity using the 4-Methylumbelliferyl-β-D-glucopyranosiduronic acid fluorescent substrate. The solution was excited at 355nm and emission read at 460nm after 2 hours incubation at 37°C. The media was tested to ensure the coloured pH indicator in the media did not affect the colour-based assay, which it did not.

4.3 Previously studied inducers of stumpy formation in monomorphs

Monomorphic slender cells, by very definition, do not differentiate into stumpy forms in response to high cell density. However, it appears that monomorphic cells have not lost the capacity to differentiate *per se,* as various drugs have been reported in the literature to cause stumpy or stumpy-like formation (section 1.7.3). In order to demonstrate that the 427 CAT-*PAD1* 3'UTR/GUS-*Actin* 3'UTR cell line could be used to detect stumpy formation *in vitro*, these cells were treated with three of the reported stumpy-inducing drugs: 8pCPT-cAMP, 8pCPT-2'O-Me-cAMP and Troglitazone (Breidbach et al., 2002; Denninger et al., 2007; Laxman et al., 2006;

Vassella et al., 1997). In each case the concentration of drug used was based on effective concentrations in previous studies. After drug treatment, samples were collected at 0 hours, 24 hours and 48 hours in order to measure cell proliferation, CAT reporter gene expression and GUS reporter gene expression. After 48 hours of drug treatment, the cells were induced to differentiate to the procyclic life stage and differentiation capacity was measured.

4.3.1 8pCPT-cAMP causes monomorphic slender cells to differentiate to stumpy-like cells

Despite years of study, the cAMP signalling pathway in trypanosomes is not yet understood. However, cAMP signalling has been implicated in playing a role in trypanosome proliferation and differentiation (Breidbach et al., 2002; Mancini and Patton, 1981; Rolin et al., 1993; Vassella et al., 1997). For instance, cell permeable cAMP has been shown to mimic SIF in pleomorphic cells (Vassella et al., 1997) as well as in monomorphic forms (Breidbach et al., 2002). Furthermore, treatment with low concentrations (1-2μM) of the phospodiesterase inhibitor, etazolate, also mimicked SIF activity. It was therefore hypothesised that SIF acts via the cAMP signalling pathway. However, high concentrations of cAMP are required, with previous studies using 250μM in pleomorphs and 1mM in monomorphs. Here the monomorphic reporter cell line was treated with a lower concentration of 100μM 8pCPT-cAMP, freshly dissolved in dH_2O after it was shown in preliminary experiments that this concentration was sufficient to cause up-regulation of reporter gene expression (data not shown).

After 24 hours drug treatment there was a clear decrease in cell proliferation compared to control cells as would be expected if proliferative slender cells are differentiating to cell-cycle arrested stumpy forms (Figure 4.4A). There was also an approximately 24-fold increase in stumpy-specific reporter gene expression after 24 hours, rising to an over 36-fold increase after 48 hours (Figure 4.4B). Contrasting with this, there was small decrease in constitutive reporter gene expression after 48 hours of drug treatment (Figure 4.4C), which may be expected given that stumpy

forms are repressed in translation (Brecht and Parsons, 1998). To further confirm that cells had differentiated to stumpy-like forms, cells treated with drug for 48 hours were then induced to differentiate to the procyclic life stage by the addition of cis-aconitate and a decrease in temperature. After only 6 hours, approximately 75% of drug-treated cells had differentiated to the procyclic life stage compared to less than 2% of the control cells (Figure 4.4D), as assessed by the expression of EP procyclin. This indicated that the drug treated cells had an increased capacity for differentiation to the procyclic life stage, as would be expected of a stumpy population.

Despite the various indications that cells had differentiated to the stumpy life stage, cells treated with drug did not develop a truly stumpy morphology (Figure 4.4E) unlike previous reports (Breidbach et al., 2002). In this study, 10 times less drug was used than that required to induce stumpy-like morphology in monomorphs previously (Breidbach et al., 2002). It may be that at the lower concentration the drug was causing some, but not all, of the molecular changes that are required for full differentiation to the stumpy forms. Alternatively, it may be that there is a requirement for a further differentiation trigger, or that the monomorphic cells are not capable of morphological differentiation. This does, however, indicate that morphological change is not the initial step in the process of differentiation, as shown previously (Tyler et al., 1997).

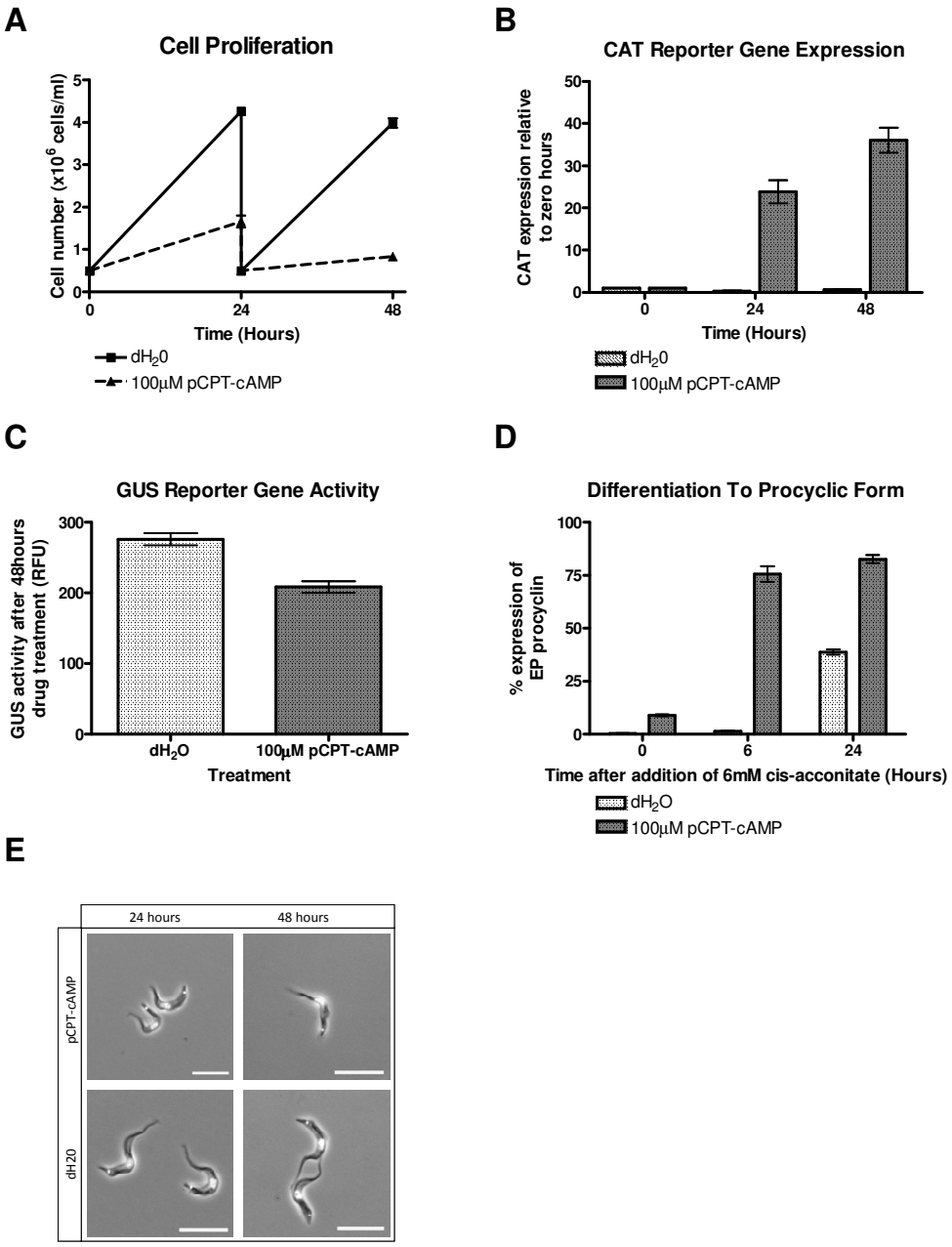


Figure 4.4 100µM 8pCPT-cAMP causes monomorphic slender cells to differentiate to stumpy-like forms. Treatment of 427 CAT-*PAD1* 3'UTR/GUS-*Actin* 3'UTR monomorphic slender cells with 100µM 8pCPT-cAMP causes (A) a decrease in cell proliferation, (B) an increase in stumpy-specific CAT reporter gene expression, normalised to cell number and (C) a decrease in constitutive GUS reporter gene expression, normalised to cell number. (D) After 48 hours of 8pCPT-cAMP treatment, cells were induced to differentiate to the procyclic life stage with the addition of 6mM cis-aconitate and a drop in temperature from 37°C to 27°C. 8pCPT-cAMP treated cells had higher EP procyclin expression compared to controls. (E) After 24 and 48 hours of 8pCPT-cAMP treatment, the cell morphology had not changed in comparison to dH_20-treated controls and therefore the cells had not turned stumpy in morphology. Error bars represent s.e.m, n = 3.

4.3.2 8pCPT-2'O-Me-cAMP causes monomorphic slender cells to differentiate to stumpy-like cells

The cAMP signalling pathway is not understood in kinetoplastids and may differ significantly from other eukaryotes (section 1.7.3). Although the cAMP pathway has been implicated as the mechanism by which SIF acts (Vassella et al., 1997), recent data has suggested that it is actually the hydrolysis products of cAMP that are inducing stumpy formation (Laxman et al., 2006). Therefore, the effect of 10µM 8pCPT-2'O-Me-cAMP freshly dissolved in DMSO on the monomorphic reporter cell line was also analysed (Figure 4.5). Control cells were treated with DMSO alone.

After 24 hours drug treatment there was again a decrease in cell proliferation compared to control cells (Figure 4.5A). There was also an over 18-fold increase in CAT reporter expression after 24 hours, rising to an over 41-fold increase after 48 hours (Figure 4.5B). There was a small decrease in constitutive reporter gene expression after 48 hours of drug treatment (Figure 4.5C). Again, to further confirm that cells had differentiated to stumpy-like forms, cells treated with drug for 48 hours were then induced to differentiate to the procyclic life stage by the addition of *cis*-acconitate and a decrease in temperature. After 6 hours, approximately 83% of drug-treated cells had differentiated to the procyclic life stage compared to less than 2% of the control cells as monitored by EP procyclin expression (Figure 4.5D). This indicates that the 8pCPT-2'O-Me-cAMP treated cells had an increased capacity for differentiation to the procyclic life stage, as would be expected of a stumpy population.

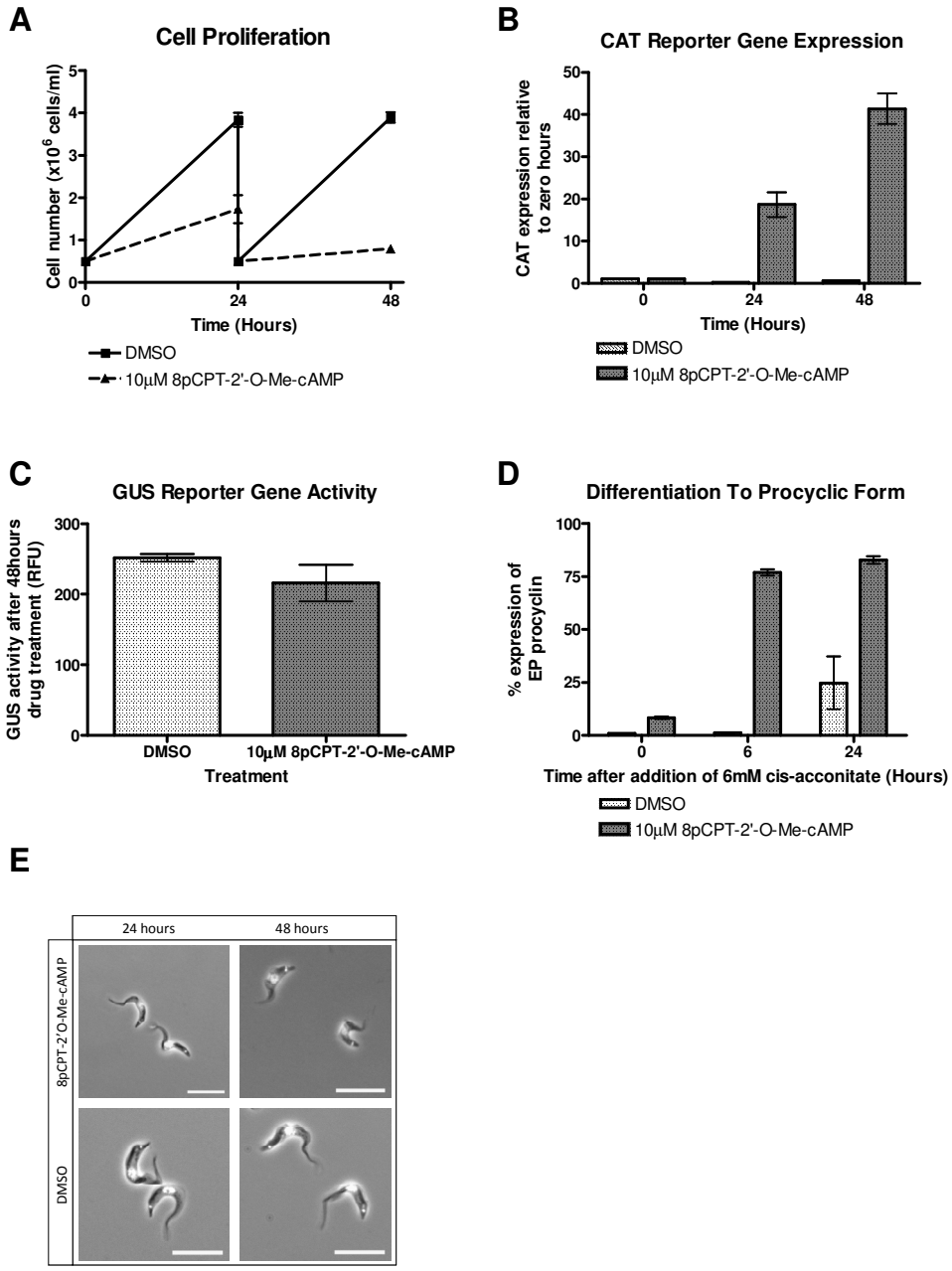


Figure 4.5 10µM 8pCPT-2'-O-Me-cAMP causes monomorphic slender cells to differentiate to stumpy-like forms. Treatment of 427 CAT-*PAD1* 3'UTR/GUS-*Actin* 3'UTR monomorphic slender cells with 10µM pCPT-2-O-Me-cAMP caused (A) a decrease in cell proliferation, (B) an increase in stumpy-specific CAT reporter gene expression normalised to cell number and (C) no change in constitutive GUS reporter gene expression, normalised to cell number. (D) After 48 hours of pCPT-2-O-Me-cAMP treatment, cells were induced to differentiate to the procyclic life stage with the addition of 6mM cis-acconitate and a drop in temperature from 37°C to 27°C. 8pCPT-2-O-Me-cAMP treated cells had higher EP procyclin expression compared to controls. (E) After 24 of 8pCPT-cAMP treatments, the cell morphology had not changed in comparison to DMSO-treated cells, however, after 48 hours the cells looked 'stumpy-like'. Error bars represent s.e.m, n = 3.

Unlike the cells treated with 8pCPT-cAMP (Figure 4.4E), those treated with 8pCPT-2'O-Me-cAMP did have an altered morphology after 48 hours with cells beginning to appear 'stumpy-like' (Figure 4.5E). However, monomorphic slender forms can look more morphologically intermediate than true slender forms from a pleomorphic population (compare dH_2O-treated monomorphic control cells in Figure 4.4 with pleomorphic slender cells from day 4 of an *in vivo* infection in Figure 5.9). This makes analysis of morphological changes very subjective. However, as a morphological change was observed with 8pCPT-2'O-Me-cAMP treatment and not with 8pCPT-cAMP treatment, along with the lower dosage required, it may be that the 8pCPT-2'O-Me-cAMP could elicit more of the molecular changes that are required for full differentiation to the stumpy-life stage.

4.3.3 Troglitazone does not cause monomorphic slender cells to differentiate to stumpy-like cells

Having analysed the effect of cell permeable cAMP and a hydrolysable cAMP analogue, the effect of a distinct drug was tested. Troglitazone is a thiazolidinedione and has been reported to induce stumpy formation in monomorphic trypanosomes at a concentration of 5µM (Denninger et al., 2007). Therefore, the effect of 5µM Troglitazone freshly dissolved in DMSO on the monomorphic reporter cell line was analysed (Figure 4.6). Control cells were treated with DMSO alone.

After 48 hours of treatment with Troglitazone there was no effect on cell proliferation compared to control cells (Figure 4.6A). There was no detectable increase in stumpy-specific reporter gene expression (Figure 4.6B) and no change in constitutive reporter gene expression (Figure 4.6C) after the 48 hours. Thus, it appeared that Troglitazone had not caused stumpy formation in these cells.

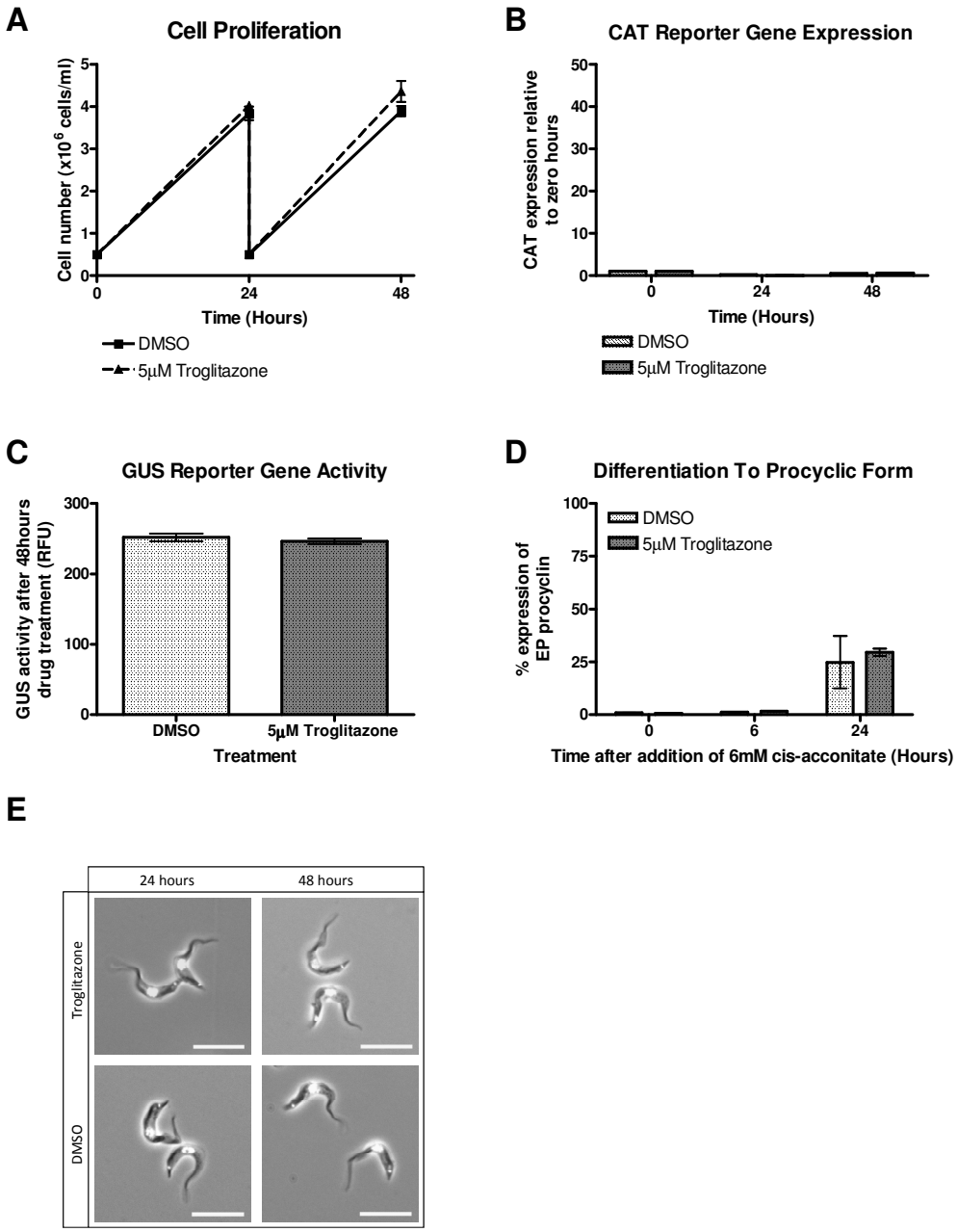


Figure 4.6 5μM Troglitazone does not cause monomorphic slender cells to differentiate to stumpy-like forms. Treatment of 427 CAT-*PAD1* 3'UTR/GUS-*Actin* 3'UTR monomorphic slender cells with 5μM Troglitazone caused no effect on (A) cell proliferation, (B) stumpy-specific CAT reporter gene expression, normalised to cell number and (C) constitutive GUS reporter gene expression, normalised to cell number. (D) After 48 hours of troglitazone treatment, cells were induced to differentiate to the procyclic life stage with the addition of 6mM cis-acconitate and a drop in temperature from 37°C to 27°C. Troglitazone treatment had no effect on EP procyclin expression compared to controls. (E) After 24 and 48 hours of 8pCPT-cAMP treatments, the cell morphology had not changed in comparison to DMSO-treated controls and did not therefore turn stumpy in morphology. Error bars represent s.e.m, n = 3.

To further confirm that cells had not differentiated to stumpy-like forms, cells treated with drug for 48 hours were then induced to differentiate to the procyclic life stage by the addition of *cis*-acconitate and a decrease in temperature. Even after 24 hours, cells had not differentiated to the procyclic life stage any more efficiently than compared to control cells, as determined by EP procyclin expression (Figure 4.6D). This indicates that the Troglitazone treated cells had not differentiated to the stumpy life stage. Morphological analysis also revealed no change in Troglitazone treated cells compared to control cells (Figure 4.6E).

The discrepancy between the results shown here and previously published data (Denninger et al., 2007) may be due to differences between the monomorphic strains used. The cells in each laboratory will have very different passage histories and therefore there are likely to be numerous differences between strains. This does however indicate that Troglitazone may not be a biologically significant inducer of stumpy formation, potentially unlike cAMP and its hydrolysis products.

4.4 High-throughout screening for pharmaceutical inducers of stumpy formation

Treatment of the 427 CAT-*PAD1* 3'UTR/GUS-*Actin* 3'UTR cell line with stumpy-inducing drugs caused an increase in CAT reporter gene expression. This ability to monitor stumpy induction provided the opportunity to develop a high throughput assay for compounds that accelerate stumpy formation at abnormally low parasite density. This reporter line could provide a simple assay for enhanced stumpy formation accessible to screening with small molecule or natural product libraries. If administrable compounds were identified that are able to force the irreversible differentiation of slender forms to stumpy forms at low levels of parasitaemia then this could have potential therapeutic consequences. For example, by inducing stumpy-formation at low parasitaemia the potential for parasite transmission may be reduced (Figure 4.7). In order to examine this further, and establish a high-throughput method of analysis, a collaboration was established with the Drug Discovery Unit at the University of Dundee.

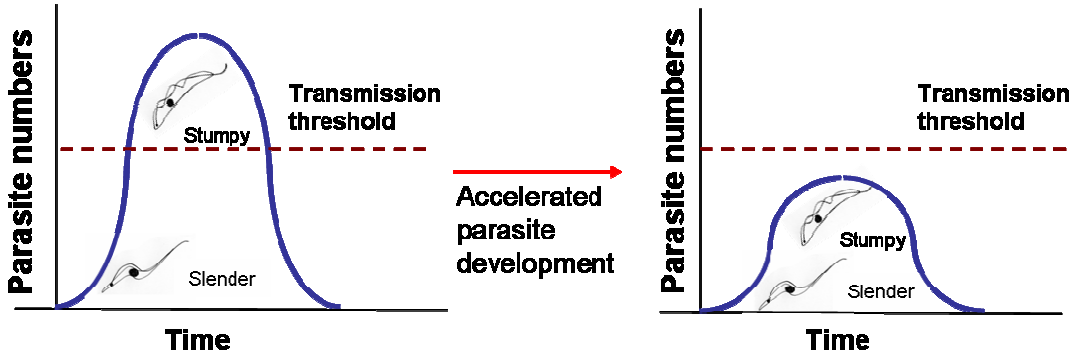


Figure 4.7 Accelerated parasite development as a route to limiting transmission and parasitaemia. In the normal course of infection slender cells generate stumpy forms once parasite numbers are sufficient to ensure transmission if taken up during a tsetse blood meal. By accelerating stumpy formation at a lower parasitaemia, the parasite load in the host would be reduced, potentially reducing pathogenicity. Furthermore, the consequent limitation in parasite numbers could reduce the potential for parasite transmission during a tsetse blood-meal. Figure from (MacGregor and Matthews, 2010).

4.4.1 The 427 GUS-*PAD1* 3'UTR reporter cell line can act as a marker cell line for stumpy formation

Although the detection of the CAT reporter expression was straightforward, involving a simple ELISA protocol, this could not be incorporated into a high throughput screen as it is expensive, labour-intensive and time-consuming. Therefore, the CAT reporter gene was replaced with the GUS reporter gene to produce the 427 GUS-*PAD1* 3'UTR reporter cell line (Table 4.1). Detection of the GUS enzyme involves incubating drug treated cells with lysis buffer containing a substrate of the enzyme and detection by the resulting fluorescence. Thus, this fast, cheap assay could be utilised in high-throughput analysis.

Having replaced the reporter gene, it was necessary to confirm that GUS activity did increase in stumpy forms as expected. Therefore, 427 GUS-*PAD1* 3'UTR cells were treated with 10µM 8pCPT-2'-O-Me-cAMP, which had caused an increase in stumpy specific CAT reporter gene expression in 427 CAT-*PAD1* 3'UTR/GUS-*Actin* 3'UTR cells (section 4.3.2). When GUS activity was normalised to cell number there was a clear increase in GUS activity after treatment with 10µM 8pCPT-2'-O-Me-cAMP (Figure 4.8) and therefore the 427 GUS-*PAD1* 3'UTR reporter cell line could act as a

marker cell line for stumpy formation. This increase was, however, detected without adjusting the relative fluorescence units reading for cell number (Figure 4.8). This was because the decreased cell proliferation caused by the drug generates fewer cells with greater GUS gene expression in drug treated cells compared to the large number of cells with low GUS gene expression in untreated cells. Hence, cell number adjustment was appropriate to measure the increase in reporter activity.

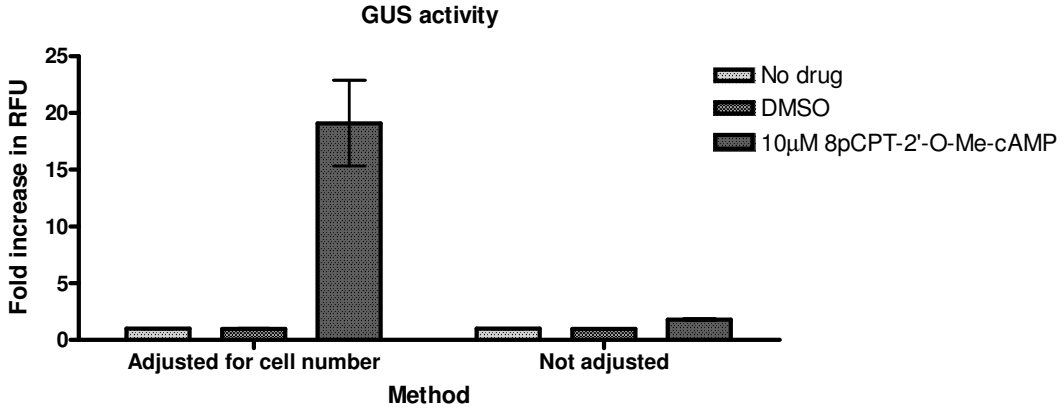


Figure 4.8 427 GUS-*PAD1* 3'UTR cell line responds to 10µM 8pCPT-2'-O-Me-cAMP. Treatment of 427 GUS-*PAD1* 3'UTR monomorphic slender cells with 10µM pCPT-2-O-Me-cAMP caused an increase in GUS activity which was detectable when the data are adjusted for cell number but was not detectable when not adjusted for cell number. Error bars represent s.e.m, n = 3.

4.4.2 Modification of the assay for high-throughput analysis

Because it appeared that monitoring total cell number was important to evaluate the GUS activity results, use of the Alamar Blue® assay was investigated to incorporate some measure of cell number into a high-throughput analysis (Raz et al., 1997; Sykes and Avery, 2009), although it is important to note that the Alamar Blue® assay measures metabolic activity rather than absolute cell number. Additionally, the initial protocol involved: culturing cells in a volume of 10ml in 25cm^2 flasks at a starting concentration of 1 x 10^5 cells/ml; incubating with drug for 48 hours; measuring cell number; and then carrying out the GUS assay. However, for high-throughput analysis the entire method required to be carried out in 96-well plates. To this end, cells were seeded at 1 x 10^5, 5 x 10^5 and 1 x 10^6 cells/ml in 100µl volumes in 96-well plates. Cells were treated with or without 100µM 8pCPT-cAMP which induces

stumpy formation in monomorphic slender cells (section 4.3.1) and incubated for 48 hours. At 44 hours, 10% v/v Alamar Blue® was, or was not, added to the cells and incubated for 4 hours. At 48 hours (4 hours post addition of Alamar Blue®) fluorescence was measured at 540nm and 595nm for the Alamar Blue® assay. Immediately subsequent to the alamar fluorescent reading, 100µl of MUG substrate was added to all cells and incubated for 2 hours at 37°C. All combinations of treatments were performed in triplicate.

After 48 hours in culture, all cells seeded at 1 x 10^6 cells/ml had become very dense and died (data not shown). In cells seeded at 1 x 10^5 cells/ml or 5 x 10^5 cells/ml, however, the presence of Alamar Blue® had no effect on the measurement of GUS activity with or without drug (Figure 4.9A). Therefore, it was clear that the Alamar Blue® assay could be used in combination with the GUS assay without interference, although the GUS assay was required to be the latter of the two assays performed, as the MUG substrate lyses the cells.

At a starting concentration of 1 x 10^5 cells/ml, a 1.9-fold increase in GUS activity was observed upon incubation with 100µM 8pCPT-cAMP compared to untreated cells (Figure 4.9B). As determined by the Alamar Blue® assay, the growth of cells treated with the drug was 55.2% that of untreated cells. Although this cannot provide an accurate measure of total cell number it can provide an estimate of population growth. More specifically, we can estimate that for every 100 hundred cells in the untreated populations, there are only 55.2 cells in the drug treated populations. When the difference in estimated cell number is taken into account, a 3.1-fold increase in GUS activity could be observed (Figure 4.9B).

At a starting concentration of 5 x 10^5 cells/ml, a 4.1-fold increase in GUS activity was observed upon incubation with 100µM 8pCPT-cAMP compared to untreated cells (Figure 4.9C). As determined by the Alamar Blue® assay, the growth of cells treated with the drug was 71.6% that of untreated cells. When the difference in estimated cell number was taken into account, a 7.6-fold increase in GUS activity in response to drug treatment could be observed (Figure 4.9C).

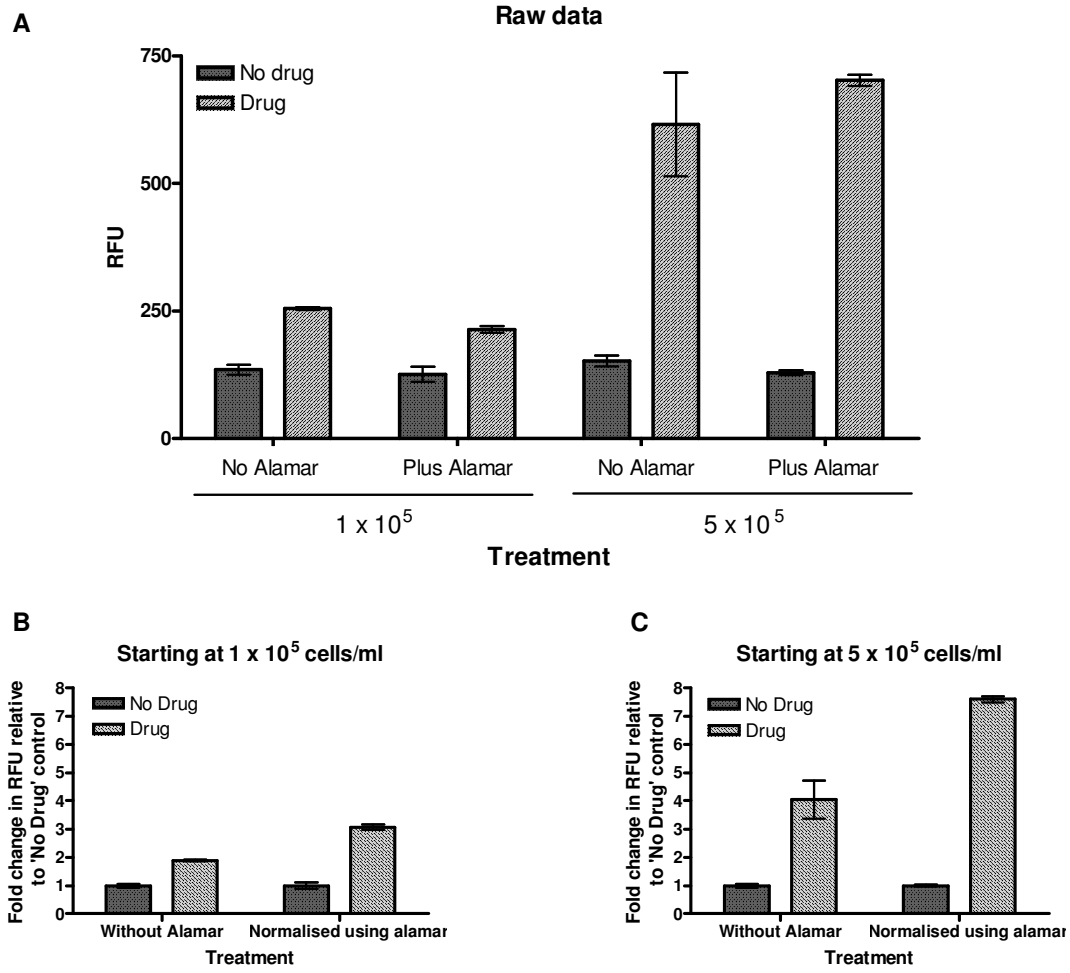


Figure 4.9 The Alamar Blue® assay can be used to provide an estimate of cell number for the GUS assay. Cells were seeded at 1 x 10^5 or 5 x 10^5 cells/ml and incubated with 100µm 8pCPT-cAMP for 48 hours. At 48 hours GUS activity was measured. (A) GUS activity was unaffected by the presence/absence of Alamar Blue® (added at 44 hours) with or without drug, in cells seeded at 1 x 10^5 cells/ml or 5 x 10^5 cells/ml. Therefore the Alamar Blue® assay could be used in combination with the GUS assay without interference. (B) At a starting concentration of 1 x 10^5 cells/ml, a 1.9-fold increase in GUS activity was observed upon incubation with the drug compared to untreated cells. When the difference in estimated cell number, using the Alamar Blue® assay, was taken into account, a 3.1-fold increase could be observed. (C) At a starting concentration of 5 x 10^5 cells/ml, a 4.1-fold increase in GUS activity was observed upon incubation with the drug compared to untreated cells. A 7.6-fold increase in GUS activity in response to drug treatment was observed when the difference in estimated cell number is taken into account. Error bars represent s.e.m, n = 3.

The aim of the inclusion of the Alamar Blue® assay was to better observe a clear distinction between cells treated with drugs that are inducing stumpy formation and untreated cells. The Alamar Blue® assay does provide this clearer distinction (Figure 4.9B and C) without interfering with the GUS assay. Additionally, it could easily be incorporated into the 96-well plate format as required for high-throughput screening.

Cells seeded at a concentration of 5 x 10^5 cells/ml provided a better assay than those seeded at 1 x 10^5 cells/ml. It may be that cells seeded at 1 x 10^5 cells/ml have not reached a high enough cell number in the 96-well plates for positive GUS expression to clearly be distinguished from background GUS activity. Although a starting concentration of 1 x 10^6 cells/ml was also tested, all untreated cells had grown dense and stopped growing within the 48 hour incubation period. Therefore in the 96-well format a starting concentration of 5 x 10^5 cells/ml was considered the most appropriate.

Taking the above observations into account, the 96-well format assay outlined in Figure 4.10A was provided to the Drug Discovery Unit at the University of Dundee. The protocol was then subject to further assay development at the Drug Discovery Unit in order that the compound screening could be carried out in 384-well plates (section 4.4.3).

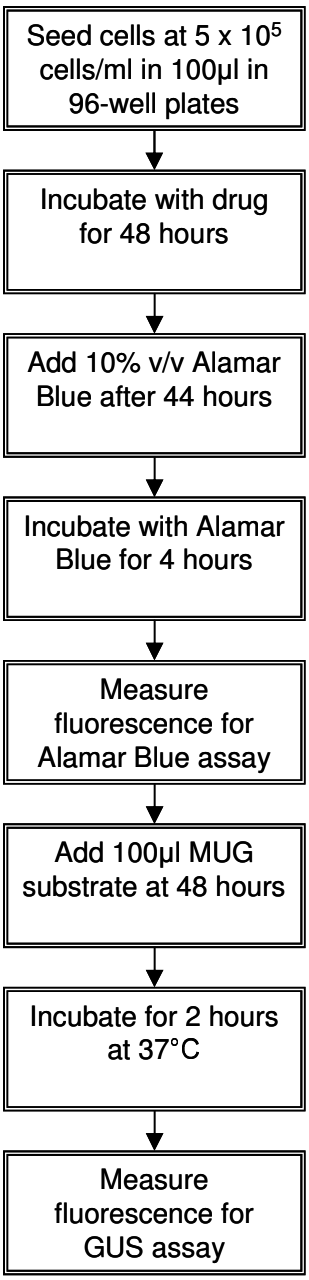


Figure 4.10 Flow diagram of the 96-well format assay. The 96-well format assay provided to the Drug Discovery Unit operated by seeding cells at a density of 5 x 10^5 cells/ml in 100µl in 96-well plates. Cells are then incubated at 37°C with the drug of interest for 48 hours. At 44 hours post addition of the drug, 10% v/v Alamar Blue® is added to the cells and incubated for 4 hours at 37°C. Fluorescence is read at 540nm and 595nm for the Alamar Blue® assay and difference in cell growth between untreated and drug treated cells is calculated (Chapter 2). At 48 hours post addition of the drug, 100µl MUG substrate is added and incubated at 37°C. After 2 hours, GUS activity is quantified by exciting wells at 355nm and detecting emission at 460nm.

4.4.3 Kinase inhibitors will be screened for ability to induce stumpy formation

Having established a method for the high-throughput screening of compounds able to induce stumpy formation, it remained to select an appropriate group of compounds to test. It was decided that a library of 6708 kinase inhibitors would be an appropriate set of compounds. These were chosen for two key reasons: firstly, kinases play an important role in many cell processes and there is evidence that they are involved in the stumpy induction pathway (Domenicali Pfister et al., 2006; Vassella et al., 2001b) (section 1.7.3); secondly, kinases are thought to be potential drug targets in trypanosomes (Doerig, 2004; Naula et al., 2005; Parsons et al., 2005) and therefore they have therapeutic potential.

At the time of production of this thesis, assay development and compound screening were still underway at the Drug Discovery Unit at the University of Dundee. The screen will include a primary single point screen at 16.6μM, repeats and potency testing of positive results. Any compounds identified from the screen as inducers of stumpy formation will be subject to further analysis, including monitoring the ability for compounds to cause an increased capacity for differentiation to the procyclic life stage (section 4.6.7).

4.5 Production of a pleomorphic reporter cell line for analysis of stumpy formation

Monomorphic reporter cell lines for the analysis of stumpy formation *in vitro* have successfully been produced. However, for the analysis of SIF, the use of a pleomorphic reporter cell line was required. Serial passage of bloodstream trypanosomes in culture selects for monomorphism: cells which respond to SIF and differentiate to the stumpy-life stage die and are lost from the population, whereas any slender forms which do not respond to SIF will continue to proliferate and take over the culture (see section 4.6.5). As such, pleomorphic cell lines are less frequently used in laboratory studies due to the requirement to maintain very low cell

density and the difficulty in successfully culturing and genetically manipulating these cells.

However, over recent years, advancements have been made in developing methods to culture pleomorphic cells (McCulloch et al., 2004; Vassella and Boshart, 1996). In particular, the addition of methyl cellulose to the media (McCulloch et al., 2004) aids pleomorphic cell growth. Additionally, a pleomorphic cell line, AnTat1.1 90:13, which has been culture adapted and transfected to enable tetracycline-inducible gene expression or RNAi, is available (Engstler and Boshart, 2004).

The original pHD617 CAT-*PAD1* 3'UTR construct (Table 4.1) contained a tetracycline-repressor binding site upstream of the CAT reporter gene. If transfected into the AnTat1.1. 90:13 cell line, which expresses the tetracycline repressor protein, this construct would require addition of tetracycline for expression of the construct. The tetracycline-repressor binding site was therefore removed from the construct before transfection into AnTat1.1 90:13 cells. After numerous attempts, a transgenic pleomorphic reporter cell line, AnTat1.1 90:13 CAT-*PAD1* 3'UTR, was successfully created for the analysis of the effects of conditioned media, containing SIF, *in vitro*.

4.5.1 Validation of the AnTat1.1 90:13 CAT-*PAD1* 3'UTR reporter cell line in vivo

In order to determine if the AnTat1.1 90:13 CAT-*PAD1* 3'UTR reporter cell line exhibited the expected pattern of CAT mRNA and protein expression, two mice were infected with the cell line. Trypanosomes were harvested from one mouse on day three post infection when the population was predominantly slender. Trypanosomes were harvested from the second mouse on day 6 post infection when the population was predominantly stumpy. Slender form AnTat1.1 90:13 CAT-PAD1 3'UTR cells express very little CAT protein or mRNA (Figure 4.11). Stumpy forms, however, express a large amount of CAT protein and mRNA compared to the slender forms (Figure 4.11). Therefore the AnTat1.1 90:13 CAT-*PAD1* 3'UTR cell line can act as a reporter cell line for the stumpy life-stage. Additionally, the appropriate expression

of the reporter gene under control of the *PAD1* 3'UTR confirms the conclusions of Chapter 3, that the *PAD1* 3'UTR is responsible for the stumpy-specific gene expression of *PAD1*.

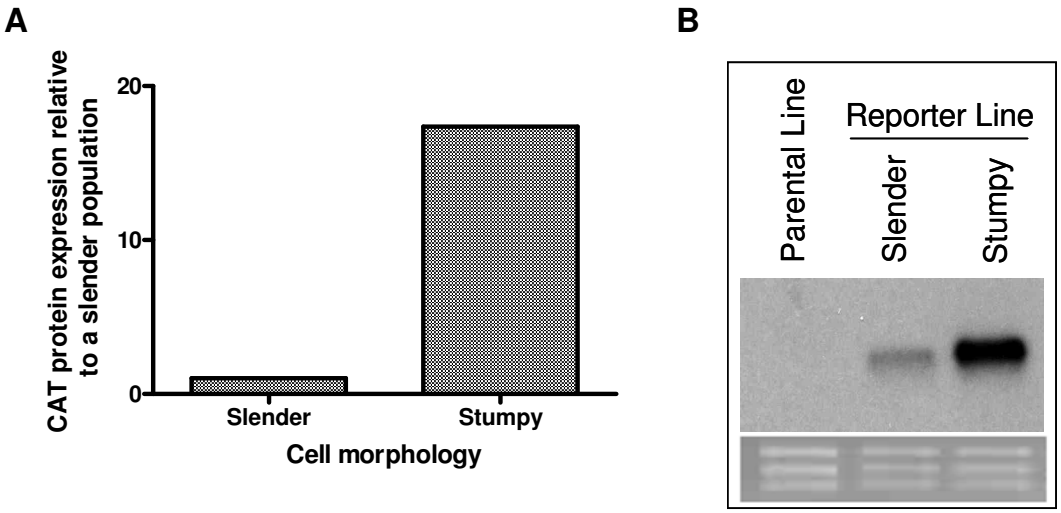


Figure 4.11 The AnTat1.1 90:13 CAT-*PAD1* 3'UTR expresses the CAT reporter gene in stumpy forms. Slender and stumpy populations of the AnTat1.1 90:13 CAT-*PAD1* 3'UTR cell line were isolated from mice infections. (A) The slender population expresses little CAT protein whereas the stumpy population expresses comparatively high levels of CAT protein (B) The slender population expresses little CAT mRNA whereas the stumpy forms express comparatively high levels of CAT mRNA. Parental cells express no CAT mRNA, as expected.

4.5.2 Culturing the reporter cell line *in vitro*

Having established that the pleomorphic reporter cell line expressed the reporter gene appropriately *in vivo,* the cell line was then analysed *in vitro*. To grow the pleomorphic cell line *in vitro* for the purpose of transfection, media containing methyl cellulose had been used. However, the inclusion of methyl cellulose in the media prevents counting of the cells by the use of a coulter counter and instead requires the use of a haemocytometer. As the CAT reporter assay requires that an accurate cell count is taken for analysis, it was decided that the use of methyl cellulose in the media was not compatible with this assay and was therefore not included.

As this cell line had previously been cultured in the presence of methyl cellulose, it was initially thought that the trypanosomes may need to adapt to media without methyl cellulose. For that reason, cells were grown for up to one week without methyl cellulose in order to attempt to further culture-adapt the cells before the start of any experiment. However, as cell density had to be maintained at very low densities to prevent stumpy formation, it was difficult to obtain enough cells to start an experiment. For example, for an experiment to compare only two factors, each in triplicate, in 10ml flasks at a starting concentration of 1 x 10^5 cells/ml, a total of 6 x 10^6 cells are required. From a culture of cells maintained at no greater than 5 x 10^4 cells/ml to prevent stumpy formation, a minimum of 120ml of culture is required. Additionally, cells did not reliably grow at such low densities and as such, would often die. The cell death was not a result of growth in culture *per se,* as cells could be grown reliably in culture at a higher density (for example, see Figure 4.12B).

Despite being unable to maintain a slender population *in vitro*, it was discovered that slender forms of the pleomorphic reporter cell line harvested from a mouse infection could be grown in culture without the use of methyl cellulose, at least over a 72 hour time frame (for example, see Figure 4.12B) without any further prior adaptation to culture. Therefore, all cells used for experimentation were obtained from mice three days post infection when the population is predominantly slender. Cells were purified from the blood and incubated in pre-warmed media without methyl cellulose.

4.5.3 Characterisation of the pleomorphic reporter cell line *in vitro*

To establish if the AnTat1.1 90:13 CAT-*PAD1* 3'UTR reporter cell line was expressing the reporter gene appropriately *in vitro,* slender cells were harvested from a mouse and seeded at 10^5 and 10^6 cells/ml in triplicate. Cell density and CAT protein expression was measured every 24 hours for 72 hours (Figure 4.12). Cells seeded at 10^5 cells/ml showed an increase of the CAT protein expression over time (Figure 4.12A) as the population density increased slowly (Figure 4.12B). In contrast, cells seeded at 10^6 cells/ml did not grow well in culture (Figure 4.12C). The

lack of proliferation was not due to an inability of these cells to survive in culture as they were from the same starting populations as the cells seeded at 10^5 cells/ml (Figure 4.12B). These cells did, however, show increased reporter gene expression over the first 48 hours, reaching expression levels almost as high as an *in vivo*-derived stumpy population (Figure 4.12A) indicating that the population, although not growing well, was able to differentiate to stumpy forms. Indeed the lack of growth may be due to rapid cell cycle arrest of the population. At 72 hours, the cells seeded at 10^6 cells/ml had reduced CAT expression compared to 48 hours, which may represent senescence of the stumpy population. Additionally, the decrease in CAT protein expression in the population between 48 and 72 hours suggested that the protein is being turned over and that the increase in protein expression over time in culture in both cells seeded at 10^5 and 10^6 cells/ml is an increase in CAT protein production and not simply accumulation of the protein.

The reduced increase in CAT expression in cells seeded at 10^5 cells/ml compared to cells seeded at 10^6 cells/ml, demonstrates that the AnTat1.1 90:13 CAT-*PAD1* 3'UTR reporter cell line is responding to cell density (Figure 4.12). That is, those cells at a lower density have lower reporter gene expression. This would be expected from a stumpy-specific marker as cells differentiate in a density dependent manner (Vassella et al., 1997). These results demonstrate that the pleomorphic reporter cell line can report on stumpy formation *in vitro*.

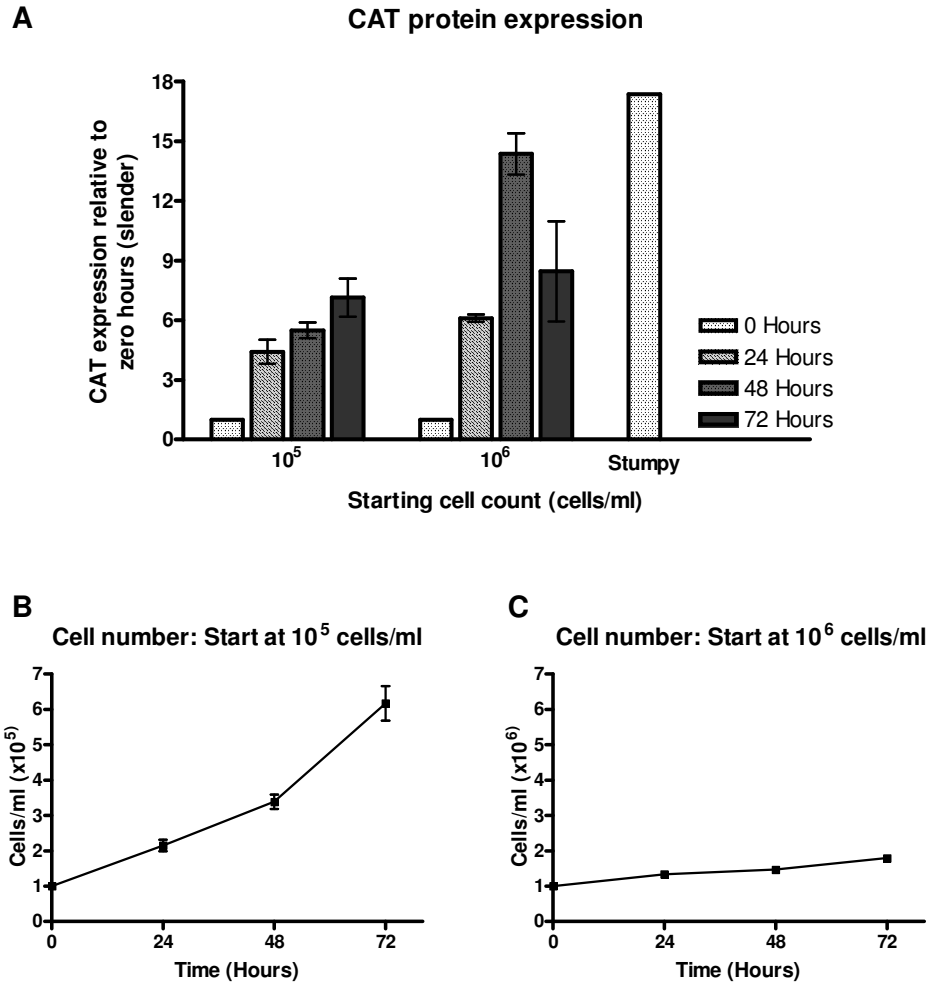


Figure 4.12 The AnTat1.1 90:13 CAT-*PAD1* 3'UTR responds to cell density. Slender AnTat1.1 90:13 CAT-*PAD1* 3'UTR cells were harvested from a mouse infection and transferred into culture medium. Cells were seeded at a density of 10^5 or 10^6 cells/ml in triplicate. (A) CAT protein expression was monitored every 24 hours. The cell line showed an increase in CAT reporter expression over time (normalised to cell number) in culture in a density dependent manner. The CAT reporter gene expression of a stumpy population from a mouse infection is shown for comparison. (B) Cells seeded at 10^5 cells/ml grew slowly over the 72 hour time period. (C) Cells seeded at 10^6 cells/ml did not grow well over the 72 hour time period. As these cells were from the same starting population as those seeded at 10^5 cells/ml, the lack of growth is not due to a general inability to grow in culture. Error bars represent s.e.m, n = 3.

4.5.4 The pleomorphic reporter cell line does not respond to monomorphic or pleomorphic conditioned media *in vitro*

The AnTat1.1 90:13 CAT-*PAD1* 3'UTR reporter cell line provides a method to study stumpy formation in pleomorphic cells *in vitro* and hence may be used to further characterise the putative density-signalling factor, SIF. Although the identity of SIF is unknown, there are numerous characteristics of this factor that have been determined (section 1.7.3). SIF is produced by the parasite and is sufficient to induce the slender to stumpy transition without the requirement for host processes (Reuner et al., 1997; Vassella et al., 1997). The factor, which may be a combination of chemical inducers, is small enough to pass through a filter with a 500Da cut-off and therefore is thought to potentially be a small metabolite. The factor is stable *in vitro* and can be maintained at 37°C for up to 27 days and can be heated to 90°C for 15 minutes without loss of activity (Vassella et al., 1997), but has been proposed through mathematical modelling to have a half-life of 1.4 hours *in vivo* (Savill and Seed, 2004). Further, it has been demonstrated that both pleomorphic and monomorphic conditioned media is able to induce stumpy formation in pleomorphic cells with similar kinetics (Vassella et al., 1997). Thus, it appears that monomorphic cells produce SIF at the same rate as pleomorphic cells, supporting the proposal that SIF may be a metabolite.

Given that conditioned media contains SIF, it would be expected that treatment of the AnTat1.1 90:13 CAT-*PAD1* 3'UTR reporter cell line with conditioned media from both monomorphic and pleomorphic reporter cell lines would cause the cells to differentiate to the stumpy form and that the CAT reporter expression would consequently increase. If this were the case, the cell line would provide a method to track SIF activity. It follows that different conditioned media could be analysed by both ability to induce stumpy formation and composition by mass spectrometry: by comparing the results, it may help to identify SIF.

To prepare monomorphic conditioned media, 427 cells were seeded at 1 x 10^5 cells/ml and incubated for 72 hours without passage such that cells reached a density

of 7.8 x 10^6 cells/ml. The culture was then centrifuged to pellet the cells and the supernatant was collected. The conditioned media was then filter sterilised through an 20µm filter, which should not cause loss of the SIF, given that it can pass through a filter with a 500Da cut-off (Vassella et al., 1997). In case there was any effect of keeping media at 37°C with 5% CO_2, 'untreated' media without cells was also kept as a control in the incubator and was also filter sterilised before use.

As pleomorphic cells did not grow well at high density (Figure 4.12C), to prepare pleomorphic conditioned media, AnTat1.1 90:13 cells were grown in culture, maintained at a density less than 1 x 10^6 cells/ml. Cells were passaged daily and the remaining culture media was collected and pooled. As such, the density of the pleomorphic culture was not particularly high and SIF was not allowed to accumulate over time. Despite this, there should still be a small amount of SIF compared to fresh media which will not contain SIF. The pleomorphic conditioned media was also filter sterilised through a 20µm filter.

To treat AnTat1.1 90:13 CAT-*PAD1* 3'UTR cells with conditioned media, slender cells were harvested from a mouse infection and transferred to pre-warmed culture media at a density of 1 x 10^5 cells/ml. Cells were incubated in media composed of 50% fresh media and 50% monomorphic or pleomorphic conditioned media, or untreated media, which had been incubated at 37°C but contained no trypanosomes. Cells were then incubated for 48 hours, with cell density and CAT reporter expression measured at 10, 24 and 48 hours (Figure 4.13).

Cells incubated in conditioned media grew less well than those incubated in untreated media (Figure 4.13A). Disappointingly however, cells treated with either pleomorphic or monomorphic conditioned media showed decreased CAT reporter gene expression compared to those maintained in untreated media after 48 hours (Figure 4.13B). Hence, under these conditions, the AnTat1.1 90:13 CAT-*PAD1* 3'UTR reporter cell line did not respond to either pleomorphic or monomorphic conditioned media.

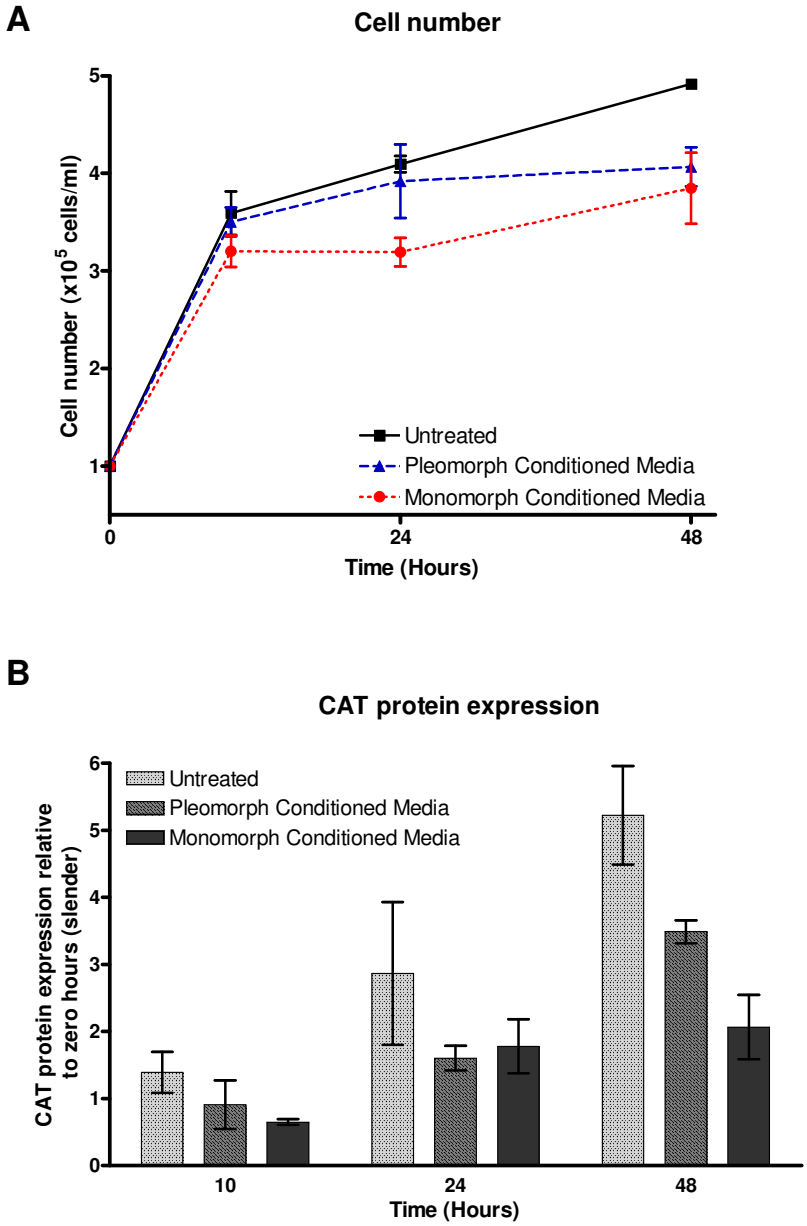


Figure 4.13 The AnTat1.1 90:13 CAT-*PAD1* 3'UTR does not respond to pleomorphic or monomorphic conditioned media. Slender AnTat1.1 90:13 CAT-*PAD1* 3'UTR cells were harvested from a mouse infection and transferred into culture medium composed of 50% fresh media and 50% monomorphic or pleomorphic conditioned media or untreated media. (A) Cells grown in conditioned media grew less well than those grown in untreated media after the first 10 hours. (B) Cells grown in untreated media showed the greatest increase in CAT reporter gene expression, normalised to cell number, over time in culture. Error bars represent s.e.m, n = 3.

Although the conditioned media was diluted 50:50 with fresh media, it is likely that the conditioned media had reduced nutritional capacity compared to untreated media. It would not be expected that this would be a limiting factor for growth given the low cell density of the culture (starting at 1 x 10^5 cells/ml) and the addition of fresh media. However, this may possibly account for the decrease in cell proliferation observed (Figure 4.13).

The difference in cell density between the three groups of cells may account for the observed differences in CAT expression as the AnTat1.1 90:13 CAT-*PAD1* 3'UTR cell line responds to population density (section 4.5.3). Biologically, it is likely that it is not cell density at a given time point that is relevant but rather the total number of cells producing SIF over the duration of the experiment. For example, imagine two populations, A and B, starting at 1 x 10^5 cells/ml. Population A grows to 5 x 10^5 cells/ml in 24 hours and then grows no further in the following 24 hours whereas population B does not grow for the first 24 hours, but in the following 24 hours grows to 5 x 10^5 cells/ml. Both populations will have a final density at 48 hours of 5 x 10^5 cells/ml, but population A will have had 24 hours at maximum population density to accumulate SIF, compared to population B which will have a lower accumulation of SIF. Of course, this may be complicated further by the rate of turnover of SIF, and if this is dependent on cell density, which is unknown. Additionally, it is not known if cells which are committed to stumpy formation produce SIF and as such inequalities in SIF production relative to the growth phase are unknown, but could also complicate analysis.

Regardless of the reason for the decrease in CAT protein in cells treated with conditioned media compared to untreated media, the most notable outcome of this experiment was that there was no increase in CAT expression in response to conditioned media under these conditions. As previously mentioned, one of the possible explanations for the lack of response of the reporter cell line to conditioned media is that the media was no longer optimised for cell growth. It is unlikely that at such low cell density the media, which is composed of 50% fresh media, would have been depleted in nutrients but perhaps the alteration in pH of the media by the

trypanosomes had a detrimental effect. Trypanosomes, and indeed all cell types, must regulate their internal pH in order to maintain the correct pH for intracellular processes. As a result, culture media is prepared at a pH of 7.5 but growth of the trypanosomes in the media causes a decrease in the pH, making the conditioned media more acidic. Although trypanosomes are able to regulate their internal pH it has been shown that an external pH of 5.5 can induce bloodstream to procyclic differentiation in the absence of CCA (Rolin et al., 1998) and slender forms are particularly sensitive to pH stress compared to stumpy forms (Nolan et al., 2000). The pH of the conditioned media was not as low as 5.5 (actual pH unknown) but if it were causing differentiation to the procyclic form it would be expected that actually this would also cause an increase in CAT reporter gene expression. Despite this, it is still possible that more acidic pH of the conditioned media affected the differentiation processes or growth of the cells. Indeed, Vassella *et al*., prepared conditioned media by readjusting the pH to 7.5 and supplementing the cells with additional FCS. Thus, this was examined by treating the AnTat1.1 90:13 CAT-*PAD1* 3'UTR cells with modified conditioned media (Vassella et al., 1997). To achieve this, monomorphic conditioned media was prepared by seeding 427 cells at a density of 5 x 10^5 cells/ml and incubating them for 48 hours without passage such that cells reached a density of 8.7 x 10^6 cells/ml. The culture was centrifuged to pellet the cells and the supernatant was collected. To modify the conditioned media, the pH was readjusted to 7.5 by the addition of NaOH and 20% FCS was added. The medium was then filter sterilised through a 20µm filter. Unmodified media and untreated media were also included as controls.

To treat AnTat1.1 90:13 CAT-*PAD1* 3'UTR cells with modified conditioned media, slender cells were harvested from a mouse infection and transferred to pre-warmed culture media at a density of 1 x 10^5 cells/ml. Cells were incubated in media composed of 50% fresh media and 50% monomorphic conditioned media or untreated media which had either been unmodified, or modified as detailed above. Cells were then incubated for 24 hours, with cell density and CAT reporter gene expression measured at 10 and 24 hours (Figure 4.14). Each treatment was performed in duplicate.

For cells grown in untreated media, there was an initial increase in cell proliferation in modified media compared to unmodified media, which then reversed after 24 hours (Figure 4.14A). This may have represented the cell line proliferating at a greater rate initially, and then responding to the increased cell density and starting to exit the cell cycle as they differentiated to stumpy forms. Indeed this correlates with the increase in CAT expression observed in cells incubated in modified media compared to unmodified media after 24 hours (Figure 4.14B). Cells grown in conditioned media, both modified and unmodified, had decreased proliferation compared to untreated media (Figure 4.14A). Modified conditioned media caused a greater decrease in cell proliferation than unmodified conditioned media, suggestive that the modified conditioned media was perhaps causing cell cycle arrest and stumpy formation. However, there was no increase in CAT reporter expression in cells treated with either modified or unmodified conditioned media compared to those treated with control media (Figure 4.14B). Although this analysis only considered effects up to 24 hours, further analysis with modified media, up to 72 hours (see section 4.5.5) confirmed this observation. Thus, under these conditions the reporter cell line did not respond to monomorphic conditioned media. Further, this was not due to altered pH of the culture media or concentration of FCS in the media.

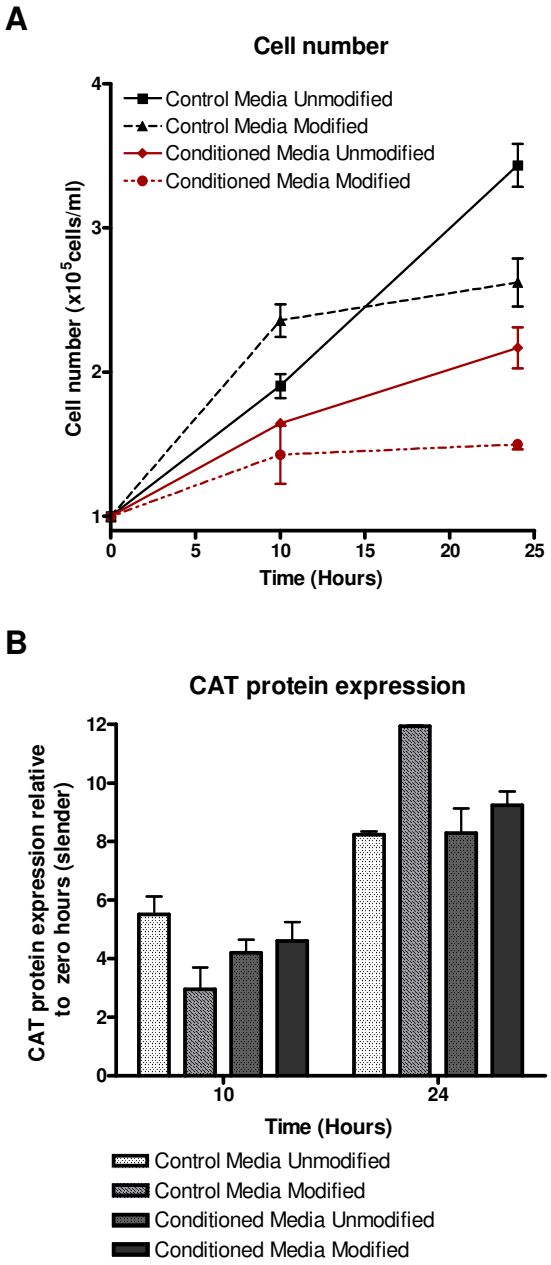


Figure 4.14 Adjusting the pH and addition of FCS to monomorphic conditioned media does not cause a response in AnTat1.1 90:13 CAT-*PAD1* 3'UTR cells. Slender AnTat1.1 90:13 CAT-*PAD1* 3'UTR cells were harvested from a mouse infection and transferred into culture medium composed of 50% fresh media and 50% monomorphic conditioned media or untreated media which had either been unmodified, or modified by the addition of 20% FCS and adjusted to pH 7.5. (A) Cells grown in conditioned media grew less well than those grown in untreated media. After 24 hours, cells grown in modified media grew less well than those in unmodified media (B) All cells increased CAT reporter expression over time in culture, normalised to cell number. However, after 24 hours, there was no difference between cells grown in modified conditioned media compared to unmodified media. Error bars represent s.e.m, n = 2.

4.5.5 Maintaining low cell density can prevent increased reporter gene expression in the pleomorphic reporter cell line

The AnTat1.1 90:13 CAT-*PAD1* 3'UTR cell line had not responded to conditioned media, with or without modification of that media. Nonetheless, the population of cells will themselves be producing SIF in culture. As the half-life of SIF is unknown, although it has been shown to be stable over a 27 day period, it may be that the conditioned media used had lost some activity. As such, it was possible that the SIF accumulated in the population will have a greater effect than SIF from conditioned media and, given the difference in cell proliferation between cells grown in conditioned media versus control media, this may have confounded the analysis.

If the AnTat1.1 90:13 CAT-*PAD1* 3'UTR cell line could be maintained at low cell density this should prevent the increase in CAT reporter gene expression over time in culture. Moreover, the addition of conditioned media to cells maintained at low cell density should allow the observation of an effect of conditioned media, as there should be only a low amount of the SIF produced by the population.

To test these hypotheses, monomorphic conditioned media was prepared by seeding 427 cells at 1 x 10^5 cells/ml and incubating them for 72 hours without passage such that cells reached a density of 6.7 x 10^6 cells/ml. The culture was centrifuged to pellet the cells and the supernatant was collected. Despite the modification of the conditioned media not causing an increase in CAT reporter expression (section 4.5.4), the pH was again readjusted to 7.5 and 20% FCS was added. The media was then filter sterilised through a 20µm filter.

To maintain AnTat1.1 90:13 CAT-*PAD1* 3'UTR cells at low density with and without conditioned media, slender cells were harvested from a mouse infection and transferred to pre-warmed culture media at a density of 1 x 10^5 cells/ml. Cells were then incubated in media composed of 50% fresh media and 50% monomorphic conditioned media or untreated media. Cells were then incubated for 72 hours, with cell density being measured every 12 hours after the first 24 hours. Every 12 hours the cells were diluted back to the starting concentration with 50% fresh media and

50% monomorphic conditioned media or untreated media. A control group of cells were not split during the experiment. In each case, CAT reporter expression was measured every 24 hours (Figure 4.15) and all treatments were performed in triplicate.

Figure 4.15 shows that it was possible to maintain low CAT expression in the AnTat1.1 90:13 CAT-*PAD1* 3'UTR cell line by maintaining the cells at a low density. This confirms that the reporter cell line responds to cell density as shown previously in section 4.5.3. At the 24 hour time point all cells had been treated equally (as no samples were diluted until this time). Therefore, the differences in CAT expression observed at 24 hours did not reflect an effect of treatment and the cause of these differences in unknown.

Figure 4.15A shows that whilst there may have been a small effect of monomorphic conditioned media on cell growth at 48 and 72 hours, no effect on CAT reporter gene expression was seen compared to those cells diluted with untreated media (Figure 4.15). This indicates that the lack of observed effect of monomorphic conditioned media on the AnTat1.1 90:13 CAT-*PAD1* 3'UTR cell line was not due to cell density, via the action of the population's own SIF, confounding the effect.

With more time it would be possible to explore the numerous possible reasons that the cell line does not respond to conditioned media, as will be discussed in section 4.6.3. Despite the lack of response to conditioned media under these conditions, the AnTat1.1 90:13 CAT-*PAD1* 3'UTR pleomorphic reporter cell line did however display appropriate reporter gene expression *in vivo* (section 4.5.1) and did respond to population density *in vitro* (section 4.5.3) Hence, it still provides an invaluable tool for the analysis of stumpy formation and potentially for the characterisation of SIF in the future.

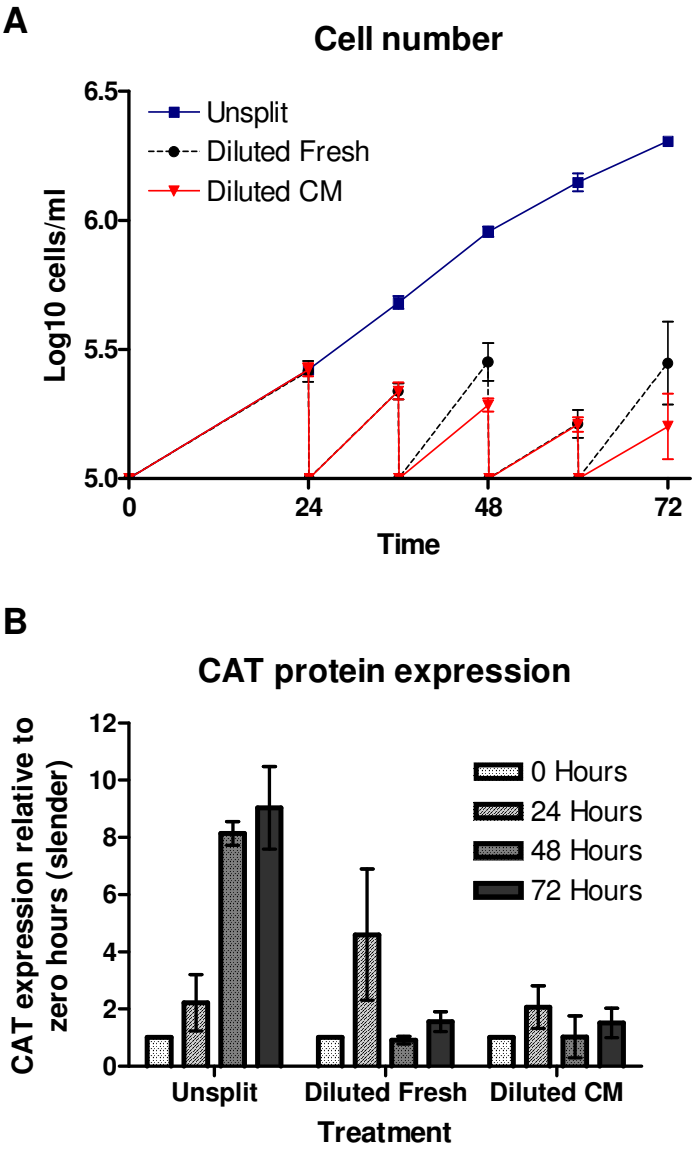


Figure 4.15 CAT reporter expression could be kept low by maintaining the AnTat1.1 90:13 CAT-*PAD1* 3'UTR cells at a low cell density. Slender AnTat1.1 90:13 CAT-*PAD1* 3'UTR cells were harvested from a mouse infection and transferred into culture medium composed of 50% fresh media and 50% monomorphic conditioned media or untreated media. Cells were diluted to a starting concentration of 1 x 10^5 cells/ml every 12 hours. A control group of cells were not passaged. (A) Cells grown in conditioned media as well as those than grown in untreated media. (B) Cells which had not been diluted showed increased CAT reporter expression, normalised to cell number, over time in culture. Cells diluted into fresh or conditioned media both showed reduced CAT reporter gene expression compared to undiluted cells. There was no difference between CAT reporter gene expression in cells diluted with fresh or conditioned media. CM = conditioned media. Error bars represent s.e.m, n = 3.

4.6 Discussion

4.6.1 Monomorphic and pleomorphic reporter cell lines have been produced for analysis of stumpy formation *in vitro.*

The identity of the parasite-derived density sensing molecule, SIF, remains unknown, despite being proposed over 10 years ago (Vassella et al., 1997). Additionally, little is known about the signalling pathway through which SIF acts. This is largely due to a lack of efficient tools for the analysis of stumpy formation both *in vivo* and *in vitro*. The discovery of the stumpy-specific surface protein, PAD1 (Dean et al., 2009), provided a molecular marker for the stumpy life stage. It was demonstrated in Chapter 3 that the 3'UTR of the *PAD1* gene is sufficient for stumpy-specific gene expression and, as such, this could be utilised for the production of reporter cell lines for the stumpy life stage. Here, both monomorphic and pleomorphic reporter cell lines have been created which allow the analysis of stumpy formation *in vitro*. The monomorphic cell lines were used for rapid and objective analysis of chemical inducers of stumpy formation, which may aid in the characterisation of the stumpy induction pathway and allow for identification of compounds with potential therapeutic uses. The pleomorphic cell line has the potential for the characterisation and, potentially, the identification of SIF. So far, however, this reporter system has not demonstrated stumpy inducing activity in conditioned media.

4.6.2 The AnTat1.1 90:13 CAT-*PAD1* 3'UTR cell line responds to cell density *in vivo* and *in vitro*

It was confirmed that in the AnTat1.1 90:13 CAT-*PAD1* 3'UTR pleomorphic reporter cell line, both the CAT mRNA and protein are developmentally regulated *in vivo* (section 4.5.1). Hence, the cell line is pleomorphic, as expected, and the *PAD1* 3'UTR is confirmed as sufficient for stumpy-specific gene expression.

In culture, the reporter cell line responded to high cell density (section 4.5.3) and reporter expression could be maintained at a low level by maintaining a low population density (section 4.5.5). Thus, this cell line appears to be a promising tool

for the analysis of stumpy formation (see section 4.6.4). Despite the appropriate expression of the cell line in response to the population's own cell density, the cell line has, so far, failed to respond to conditioned media however.

4.6.3 The AnTat1.1 90:13 CAT-*PAD1* 3'UTR cell line does not respond to conditioned media

SIF is the density-sensing signal produced by the parasite, which induces the slender to stumpy transition. As the parasite produces SIF it is released into the culture media such that conditioned media from a dense parasite population contains SIF activity and can be used to induce stumpy formation at low cell density (Vassella et al., 1997). Therefore, the AnTat1.1 90:13 CAT-*PAD1* 3'UTR cell line was treated with conditioned media and reporter gene expression was monitored (section 4.5.4 and 4.5.5). It was expected that the addition of conditioned media would result in an increase in reporter expression compared to control media due to the presence of SIF. However, despite numerous attempts, not all of which are shown here, this was never observed. As we have no positive control, interpretation of these negative results becomes difficult. Indeed it had been expected that perhaps conditioned media would act as a positive control for further experimentation given that it is proposed to contain SIF.

There are numerous possible reasons for the lack of response of the cell line to conditioned media. The two broad possibilities are that (a) the conditioned media does not support stumpy formation or (b) that the cell line is unable to respond to SIF. Firstly, the possibility that the conditioned media does not support stumpy formation shall be considered.

It could be that there is no SIF in the conditioned media. For the pleomorphic conditioned media, a low density of cells were used and SIF was not allowed to accumulate over time. Although, it may be expected that even a small amount of SIF would result in an increase in reporter expression compared to control media it may be that a minimum threshold is required to induce stumpy formation and therefore

this cannot be ruled out as a possible explanation for the lack of response to pleomorphic conditioned media. For the monomorphic conditioned media, cell density was high (always above 6.5 x 10^6 cells/ml) and SIF was allowed to accumulate for at least 48 hours. These conditions should definitely cause the conditioned media to contain SIF, unless the monomorphic cell line was not producing SIF at all. This is highly unlikely. A monomorphic cell line, related to the one used here, was shown to produce SIF at approximately the same rate as two different pleomorphic cell lines (Vassella et al., 1997), which would concur with the hypothesis that SIF is a small metabolite. Additionally, it is unlikely that monomorphic cell lines are monomorphic due to an inability to produce SIF; rather they would be predicted to be unable to perceive the signal (see section 4.6.5). Although unlikely, however, it cannot be excluded that our monomorphic cells do not produce SIF. Hence, it would be appropriate to repeat the attempts at inducing stumpy formation via the action of pleomorphic conditioned media which had been allowed to accumulate SIF to a greater extent than carried out here. Alternatively, the cell line used by Vassella et al., could be evaluated.

Another explanation is that the conditioned medium may have lost SIF activity. This would be unexpected given that Vassella et al., were able to store the conditioned media for up to 27 days, heat it to 90°C, and filter through a 500Da cut-off without loss of SIF activity (Vassella et al., 1997). Indeed our, conditioned medium was stored for less than one week and was only filtered through a 20µM filter to remove any potentially remaining trypanosomes. To establish if loss of SIF activity during storage is indeed a problem, however, the pleomorphic reporter cell line could potentially be grown in a transwell system along with a high density of pleomorphic or monomorphic cells. This would result in the cell line being exposed directly to fresh conditioned media without the requirement for processing or storage of that media.

Given that all conditioned media used by Vassella et al., was ultra-filtered, it could be speculated that there is an inhibitory factor preventing stumpy formation in the conditioned media produced here, but removed by the ultra-filtration procedure

(Vassella et al., 1997). This seems unlikely, as it would require complex dynamics between a positive and negative stumpy induction factor, for which there is no evidence. It also seems unlikely that the conditioned media is having some toxic effect on the cells, which then prevents stumpy formation, as cell growth is still observed (Figure 4.13, Figure 4.14 and Figure 4.15). Also, those cells which do grow would presumably still be capable of responding to the high concentration of SIF expected to exist in the monomorphic conditioned media.

Perhaps it would be appropriate to attempt to induce stumpy formation via the action of stumpy inducing chemicals such as cell permeable cAMP ((Vassella et al., 1997) and see section 4.3). Should the cell line respond to chemical induction this would indicate that indeed the conditioned media does lack stumpy inducing activity and that the cell line is capable of responding to triggers of stumpy formation under these conditions. However, if the cAMP was unable to induce stumpy formation, the results would imply that the cell line is unable to respond to stumpy inducing signals under these conditions.

It is not the case that our reporter line is unable to differentiate to stumpy forms at all, as it has been shown that it responds to its own cell density both *in vivo* and *in vitro*. Therefore, the possibility that the cell line is simply unable to differentiate can be excluded. Nonetheless, it may be that under the conditions used in this study the cell line is unable to respond to SIF. A fundamental difference in the conditions used is that here cells are grown in liquid culture, whereas Vassella *et al*., use semi-solid agarose plates for cell culture. It was decided not to use semi-solid agarose plates at the start of this study due to the tendency for cells to form localised clusters of cells which may create regions of high local density on an otherwise low density plate (Reuner et al., 1997). This has two implications; the first being that it may create high local concentrations of SIF which would confound analysis of SIF. Alternatively, however, it may provide cells with a high degree of cell-to-cell contact in a local area. Perhaps, to induce stumpy formation two factors are required: the diffusible stumpy induction factor and cell-to-cell contact. It was concluded in the previous analysis that cell-to-cell contact was not required because stumpy formation

was achieved at low cell density (Vassella et al., 1997). However, if the local cell density in a cluster of cells on a semi-solid agarose plate was sufficient to fulfil this requirement, but the local SIF concentration is not, then this would result in the same positive response to conditioned media as if SIF is required alone. If this were the case, it may be that the AnTat1.1 90:13 CAT-*PAD1* 3'UTR cell line does not respond to SIF due to a lack of cell-to-cell contact in a homogenous liquid culture. To test if this is the case, it would be appropriate to replicate the conditions used by Vassella et al., to grow the reporter cell line on semi-solid agarose plates and determine if there is any observed effect of conditioned media. If there was, this would confirm that the conditioned media contains SIF, but also opens up the possibility that SIF is not the only factor required.

Indeed, there is the potential for many further experiments to ascertain the reason for the lack of response of the pleomorphic reporter cell line to conditioned media which will then allow use of the cell line for the characterisation of SIF. Also, understanding the reason why the cell line does not respond might in itself help to further characterise the stumpy induction pathway.

4.6.4 The AnTat1.1 90:13 CAT-*PAD1* 3'UTR cell line could provide a method for the identification of SIF

Although the AnTat1.1 90:13 CAT-*PAD1* 3'UTR cell line has not responded to conditioned media, this problem may be resolved with further analysis, as has been discussed (section 4.6.3). If this can be resolved, the cell line provides a valuable tool for the characterisation and perhaps identification of SIF. For instance, the cell line could be used to track SIF activity in conditioned media. The media could potentially be fractionated and SIF activity tracked to specific fractions. Those fractions with SIF activity could then be subject to mass spectrometry which may identify the chemical identity of SIF. Alternatively, different conditioned media produced and treated in different ways, such as different durations of SIF accumulation or filtration of the media could be compared for both SIF activity and composition by mass spectrometry, again potentially identifying SIF.

Considering the identity of SIF has remained elusive for so long, there is a huge requirement for a method to track stumpy formation in order to identify this factor. The developed cell line will provide a key reagent for this, albeit one requiring further optimisation.

4.6.5 The 427 CAT-*PAD1* 3'UTR cell line further demonstrates that monomorphic cells are not incapable of stumpy-formation *per se*

It was confirmed in the 427 CAT-*PAD1* 3'UTR monomorphic reporter cell line, that both CAT mRNA and protein increase in response to stumpy-inducing compounds. This confirms that the cell line is a suitable tool for the analysis of stumpy formation in monomorphs and that monomorphs are capable of, at least partially, undergoing some steps in the progression to the stumpy form.

Monomorphic cells are essentially artefacts of cell culture procedures. In a natural infection, only stumpy cells are capable of transmission. As such, the capacity for perception of population density and differentiation to the stumpy-life stage is maintained by cyclical transmission. Conversely, in the laboratory there is a selective advantage for any mutation that prevents an individual cell from perceiving, or responding to, increased population density, as stumpy cells are exited from the cell cycle and, in culture, will die. The resulting monomorphic cells are most commonly studied in the laboratory because they proliferate well and show low levels of antigenic variation. The latter phenomenon, in particular, has greatly aided the molecular understanding of VSG switching mechanisms.

In a whole population, a mutation in a single cell that prevents the production of SIF would not prevent the differentiation of that individual cell given that the rest of the population would continue to make SIF. As such, it is unlikely that monomorphic cells arose from a mutation in the production of SIF. It is not impossible, however, that after some initial mutation, monomorphs subsequently lost the capacity to produce SIF following a further mutation. This is highly improbable given that SIF is

hypothesised to be a small metabolite (Vassella et al., 1997) and monomorphic conditioned media has been shown to be capable of inducing stumpy formation as well as pleomorphic conditioned media (Vassella et al., 1997), but until the identity of SIF is exposed, it is not possible to be certain. Moreover, different passage histories of even the same monomorphic lines in different laboratories may generate different phenotypes through independent mutations after the populations diverged.

It remains, therefore, that monomorphs are either unable to perceive or unable to respond to SIF. As monomorphic cells are able to generate stumpy or stumpy-like cells in response to analogues of cAMP ((Laxman et al., 2006; Vassella et al., 1997) and section 4.3), they are not incapable of differentiating to stumpy forms *per se*. Nonetheless, these compounds do not appear to induce full differentiation to the stumpy life stage (see section 4.6.6 for further discussion) and one of the possible explanations for this is that monomorphic cells have a reduced capacity for transmission of the SIF signal.

It seems that the most likely initial mutation in monomorphic cells is in the reception of SIF. If SIF is transmitted through a single receptor then a single mutation could prevent stumpy formation. In this case, induction of stumpy formation by compounds such as cAMP analogues, would act by by-passing the reception mutation by inducing downstream effectors of the stumpy induction pathway.

As detailed above, monomorphs generated at different times in different laboratories could also have different initial mutations. Moreover, it is certain that monomorphic cell lines will have acquired numerous additional mutations during long term cell culture that may or may not impact on the stumpy induction pathway and that will vary between laboratories. Differences between the cell lines in different laboratories may account for differences in the effects of drugs on different monomorphic cell lines. For instance, it was observed that Troglitazone was able to induce stumpy-like formation in monomorphic cells, yet this was not replicated here (section 4.3.3).

A large-scale analysis of the differences between monomorphic and pleomorphic cells, either by genome sequencing or by microarray analysis, may identify mutations in the stumpy induction pathway or potentially reveal the SIF receptor. However, it is likely that many mutations irrelevant to the stumpy induction pathway between cell lines would also be observed. Perhaps the comparison between a newly created monomorphic cell line and its parental pleomorphic line would reduce the number of differences, making for less complicated analysis. Alternatively, the comparison of numerous pleomorphic lines with numerous monomorphic lines may identify recurring mutations between the two groups which may pinpoint genes of interest in this pathway.

4.6.6 cAMP signalling in the stumpy induction pathway

cAMP signalling has been implicated in the differentiation of trypanosomes since it was discovered that there are changes in the intracellular cAMP concentration during the slender to stumpy differentiation (Mancini and Patton, 1981) and during the bloodstream to procyclic differentiation (Rolin et al., 1993). When it was first recognized that the slender to stumpy transition was induced by the unidentified SIF, it was shown that cell permeable cAMP was able to mimic the actions of SIF in monomorphic and pleomorphic cell lines (Breidbach et al., 2002; Vassella et al., 1997). It therefore seemed likely that cAMP is playing some role in the stumpy induction pathway. However, little is known about the mechanisms by which cAMP signalling act in kinetoplastids (section 1.7.3). Generally, in eukaryotes, cAMP transmits a signal via protein kinase A which phosphorylates transcription factors and causes changes in gene expression. This is clearly unlikely to be the mechanism that cAMP works in trypanosomes given the broad lack of regulation at the transcriptional level (section 1.10). Although PKA-like proteins have been identified, these often lack the cAMP-binding domain essential for cAMP-dependence (Seebeck et al., 2001). Proteins predicted to bind cAMP have been identified (Laxman and Beavo, 2007; Seebeck et al., 2001), but as yet, no protein has been shown to be an effector of cAMP in trypanosomes. When it was demonstrated that the hydrolysis products of cAMP were more potent at inducing stumpy or stumpy-like formation in

monomorphic cells, this implied that a cAMP-type signalling pathway was unlikely to be acting during induction of stumpy formation (Laxman et al., 2006).

Here, both 8pCPT-cAMP and 8pCPT-2'-O-Me-cAMP were used to validate reporter cell lines which act as molecular markers for the stumpy life stage. Both drugs caused increased stumpy-specific reporter gene expression as well as decreased cell proliferation and increased capacity for differentiation to the procyclic life-stage, all characteristics of stumpy cells. However, 8pCPT-cAMP did not cause any morphological change and although 8pCPT-2'-O-Me-cAMP did cause some morphological change, these cells did not look like true stumpy forms from a pleomorphic population (compare Figure 4.5E with Figure 5.9 and 5.13). As the drugs did not induce a full morphological change, this suggests that these drugs are not causing complete differentiation of slender to stumpy cells, but rather they are becoming 'stumpy-like'. It is possible that this is due to a decreased capacity for culture-adapted monomorphic cells to differentiate to the stumpy life-stage, even after by-passing a perception mutation (see section 4.6.5). Alternatively, it may be that SIF has more than one downstream effector pathway and that the cAMP analogues are only targeting one of these pathways. The analysis of additional compounds able to induce stumpy formation may indicate if this is a likely explanation. A final possibility is that the addition of cAMP has other unknown effects on the cells which confound analysis.

4.6.7 The 427 GUS-*PAD1* 3'UTR cell line provides a method for the identification of pharmaceutical inducers of stumpy formation

The identification of a molecular marker that distinguishes the non-proliferative stumpy form provides the ability to develop high throughput screens for compounds that accelerate stumpy formation at abnormally low parasite density. The 427 GUS-*PAD1* 3'UTR cell line provides a fast, simple assay for enhanced stumpy formation that can be performed in 384-well plates.

Each of the compounds identified by the 427 GUS-*PAD1* 3'UTR cell line will be validated initially by testing their action on the 427 CAT-*PAD1* 3'UTR/GUS-*Actin* 3'UTR cell line. This will determine if the response seen in a high-throughput screen can be mimicked on a smaller-scale in the laboratory. The use of the constitutive reporter in the 427 CAT-*PAD1* 3'UTR/GUS-*Actin* 3'UTR cell line will ensure that there is not a global up-regulation of protein in response to the drug. Additionally, the testing for the capacity for differentiation to the procyclic life stage will ensure that the compounds cause the cells to become functionally stumpy. Having established that these compounds are able to cause stumpy formation in monomorphs, it will then remain to test the compounds on pleomorphic cells, potentially on the AnTat1.1 90:13 CAT-*PAD1* 3'UTR cell line to further confirm that they are inducing stumpy formation.

The identification of these compounds has two potential rewards: their use as therapeutic compounds for the acceleration of parasite development; and potentially in the elucidation of the stumpy induction pathway.

4.6.8 Further elucidation of the SIF pathway

The compounds used in the screen for ability to induce stumpy formation were from a library of kinase inhibitors. There are approximately 176 putative protein kinases in the *T.brucei* genome (Parsons et al., 2005) and they likely play a role in the cell cycle and differentiation events (Naula et al., 2005). Given that kinases play an important role in many cell processes, it is likely that they also play a role in stumpy formation. Indeed, knock-out of two different kinases, ZFK and *Tb*MAPK5, have both been shown to cause increased stumpy formation, suggesting a role in the stumpy induction pathway (Domenicali Pfister et al., 2006; Vassella et al., 2001b).

By identifying kinase inhibitors that cause stumpy formation we may be able to identify kinases that act in the stumpy induction pathway. The targets of the kinase inhibitors could potentially be identified by a technique created by Cellzome which allow identification of the proteins to which a kinase inhibitor is binding; using

Kinobeads™ to select for those proteins and mass spectrometry for their identification (Bantscheff et al., 2007). This would then identify candidate proteins involved in stumpy formation that could then be studied further. This may provide key insights into the production of stumpy forms and the signalling pathways involved.

4.6.9 Accelerated parasite development as a therapeutic tool

If compounds are identified that are able to force the differentiation of slender forms to stumpy forms at low parasitaemia then this may have potential therapeutic consequences. Firstly, as trypanosome infections last from months to years, each infected individual can act as a reservoir for the disease for a considerable period of time. The degree of periodicity in their transmission capacity is poorly understood but potentially vital in a context where tsetse bites are rare or show seasonality (Baylis, 1997). During the chronic stage of a *T.brucei* infection in cattle, the parasitaemia fluctuates such that there are periods when the parasitaemia is sufficiently low to prevent the infection of tsetse flies feeding on the host (Van den Bossche et al., 2005). Therefore, differentiation of the population to the cell-cycle arrested stumpy form specifically at these points of very low parasitaemia could cause decreased transmissibility.

Although accelerated parasite development has the potential to reduce host parasitaemia it also has the potential to increase the transmission potential of the population through increasing the proportion of stumpy forms. The extent of this risk depends on several factors: the density of parasites in the bloodstream, the number of parasites required to infect a tsetse fly and the efficiency of accelerated development. Assuming a typical parasitaemia in cattle of 10^5 parasites/ml (Van den Bossche et al., 2005) and an infective dose to tsetse under optimal conditions of 1 trypanosome (Maudlin and Welburn, 1989), then a reduction in parasite numbers of at least three orders of magnitude would be required to minimise the probability of parasite transmission in a tsetse blood meal. However, in the field chronic infections of cattle can exhibit refractory periods when there are insufficient parasites for tsetse

transmission, suggesting that the parasites may be on the threshold of transmissibility, particularly when compounded by seasonal effects. This might make transmission-blocking feasible.

Moreover, if stumpy formation could be driven to 100% efficiency then the irreversible division arrest of these forms would ensure complete parasite elimination through cell senescence within only a few days, limiting the risks of enhanced transmission and providing a curative regime. Clearly, without better understanding of the transmission biology of trypanosomes in the field the prospects for a transmission blocking approach are difficult to assess and a more thorough understanding of the within-host dynamics would be required to predict the outcome of such a treatment. This does, however, offer great potential for disrupting normal life-cycle progression and thereby restricting both pathogenicity and transmissibility of the parasite.

It is satisfying, that the identification of the stumpy marker gene, *PAD1* offers the ability to both screen for new drugs to accelerate stumpy formation, as described in this chapter, and to refine the transmission models required to optimise their potential deployment in the field, as will be described in Chapter 5.

5 Chapter 5: Monitoring within-host dynamics of stumpy formation *in vivo*

5.1 Introduction

Trypanosome infections are characterised by periodic fluctuations in parasitaemia (Ross and Thomson, 1910). These fluctuations are caused by a combination of the slender to stumpy transition and antigenic variation (section 1.7). However, within-host dynamics are not as well characterised due to the technical difficulty of quantifying slender and stumpy forms by microscopy, especially given the presence of cells with an intermediate morphology, which are in the transition process.

As trypanosome infections last from months to years, each infected individual can act as a reservoir for the disease for a considerable period of time. However, the degree of periodicity in their transmission capacity is poorly understood but potentially vital in a context where tsetse bites are rare or show seasonality (Baylis, 1997). Furthermore, through mathematic modelling it has been predicted that the production of stumpy forms is a key contributor to the ordered expression of distinct antigenic variants that sustain chronic trypanosome infections (Lythgoe et al., 2007). This, combined with the likelihood of co-infection with different trypanosome strains (Balmer and Caccone, 2008), generates the potential for complex infection dynamics with important possible implications for disease progression and transmission in endemic areas. Quantitative models for the dynamics of slender and stumpy parasites and fluctuating parasitaemia could allow the consequences of potential therapeutic strategies, such as those discussed in Chapter 4, to be predicted.

PAD1 is a marker for transmission competence in individual cells (Dean et al., 2009). Monitoring *PAD1* expression provides a route to analyse the proportions of slender and stumpy forms during infections and to understand how the dynamics of these change throughout the course of a classical chronic infection. Here, the use of *PAD1* as a marker for the transmission stage during *in vivo* infection is investigated using quantitative RT-PCR (qRT-PCR).

5.2 Establishing chronic trypanosome infections in mice

In order to monitor within-host infection dynamics, it was first necessary to determine if it was possible to establish chronic infections in the mouse model system and, further, if characteristic fluctuations of the parasitaemia could be observed. As chronic infections had not previously been established in this laboratory there was the potential that either the trypanosome infection would overwhelm and kill the mouse, or that the mouse would be able to control infection after the first peak of parasitaemia. To investigate this, three male MF1 mice were infected with approximately 500 AnTat1.1 pleomorphic trypanosomes and the parasitaemia was monitored by microscopy over a 32 day period (Figure 5.1) using the rapid matching method for the quantification of parasitaemia (Herbert and Lumsden, 1976).

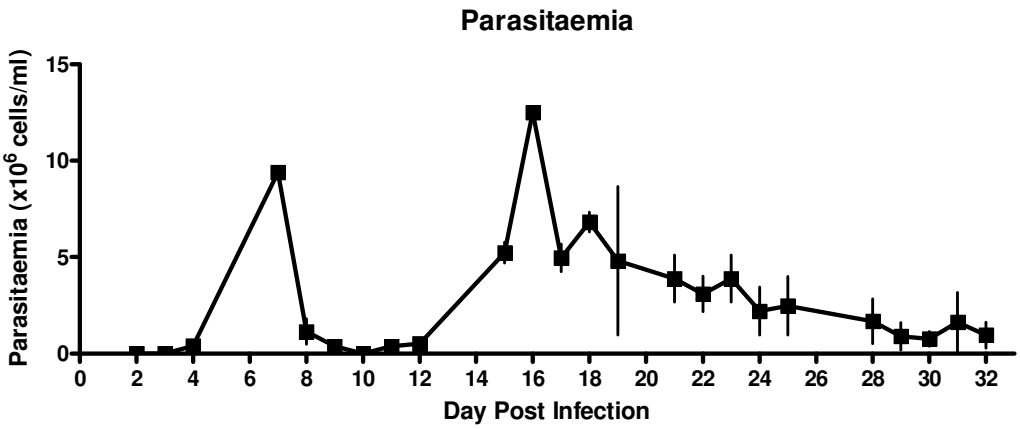


Figure 5.1 Chronic trypanosome infections in MF1 mice. Three MF1 mice were injected I.P. with 500 AnTat1.1 pleomorphic trypanosomes at day 0. Parasitaemia was monitored by microscopy over a 32-day period. Each "■" represents a point at which the parasitaemia was measured. Error bars represent standard error of the mean.

In all three mice a trypanosome infection was established, with the parasitaemia peaking twice at a density of approximately 10^7 parasites/ml blood, as determined by microscopy. This demonstrated that not only is it possible to establish a chronic infection in this model system, but that fluctuations in parasitaemia could be observed, as in natural infections.

5.3 RNA extraction, DNase treatment and cDNA production of trypanosome mRNA from whole-blood samples

In order to track disease progression it was necessary to examine constitutive transcript levels in order to monitor parasitaemia (section 5.4) and relative stumpy-specific transcript levels to monitor transmissibility (section 5.5) throughout an infection. Hence, it was necessary to establish a method of collection and processing of whole-blood samples for qRT-PCR. This had already been established for the analysis of gamete production during rodent malaria infection (Wargo et al., 2006) and so a similar protocol was utilised here.

In order to acquire blood samples from mice daily, 10µl capillary tubes were used to draw blood from a small tail snip. The 10µl samples were transferred immediately into 30µl of lysis mix (Chapter 2). RNA was extracted from whole blood samples using an ABI Prism 6100 Nucleic Acid PrepStation (Chapter 2) (Wargo et al., 2006).

Although a DNase treatment step was included in the RNA extraction protocol, it was discovered that this did not remove all contaminating DNA from the samples which caused a high degree of background signal in "no reverse transcriptase" (NoRT) controls (Figure 5.2). Hence, it was necessary to add a further DNase treatment step into the protocol after RNA extraction, but before cDNA production, using the Ambion TURBO DNase kit (Chapter 2). This reduced the DNA contamination of the samples (Figure 5.2) and was therefore used in the treatment of all samples.

The Ambion Turbo DNase treatment was carried out in a buffer which, in high concentrations, can inhibit the reverse transcriptase reaction and this was observed to be the case here (data not shown). Therefore the DNase-treated RNA samples were diluted $^{1}/_{10}$ before cDNA production. Reverse transcription was carried out using the ABI High-Capacity cDNA Reverse Transcription kit (Chapter 2). The cDNA produced was diluted $^{1}/_{10}$ before the qRT-PCR reaction so as to prevent cDNA-inhibition. Unfortunately, the necessity to dilute the sample at various steps along the pathway, combined with the small starting volume of 10µl of mouse blood, resulted

in a limitation of detection of target transcripts (section 5.4 and section 5.5), reducing the overall sensitivity of the assay.

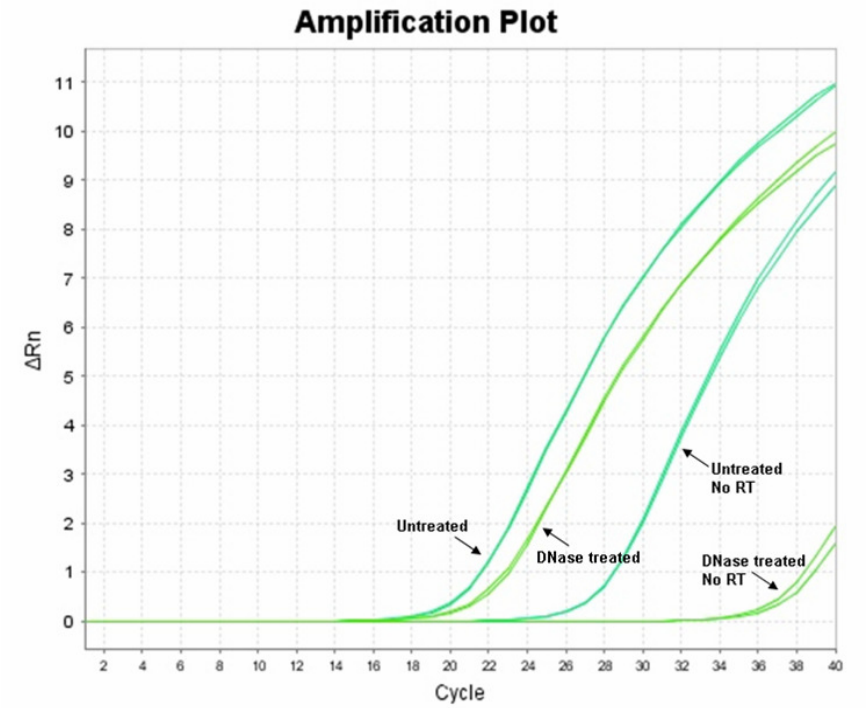


Figure 5.2 TURBO DNase treatment removes contaminating DNA from the reaction. Shown in dark green are samples without the additional Turbo DNase treatment after RNA extraction. Although a positive signal was detected in the untreated samples, there was a large signal in the 'NoRT' control samples. Shown in light green are the Turbo DNase treated samples. There was a small decrease in the signal of the DNase treated samples compared to the untreated samples, due to loss of the contaminating DNA and because the sample was diluted into a buffer. Importantly, there was a large decrease in the signal in the DNase treated 'NoRT' control samples compared to the untreated 'No RT' control samples demonstrating that the Turbo DNase treatment was effective. Each sample was run in duplicate. Rn is the fluorescent signal of the sample relative to the ROX passive reference dye. ΔRn is the Rn minus the baseline, where the baseline is the background fluorescent signal in all samples.

5.4 Determining parasitaemia of chronic trypanosome infections in mice

In order to measure parasitaemia using qRT-PCR it was necessary to identify a target transcript that is constitutively expressed in bloodstream trypanosomes. *Tb*ZFP3 was selected as an endogenous control as it has been shown to be constitutively expressed

throughout the slender, stumpy and procyclic stages of the trypanosome lifecycle (Paterou et al., 2006). Thus, it was expected that *Tb*ZFP3 expression levels would be the same in all samples for any given concentration regardless of cell morphology.

5.4.1 Production of slender and stumpy standards curves

In order to measure parasitaemia using *Tb*ZFP3 as a target, it was necessary to create a standard curve from a sample of known parasitaemia. To do this, slender and stumpy trypanosomes were harvested from mice infections, purified and counted. A known number of parasites were then used to inoculate a known volume of fresh whole blood. This blood was then divided into 10µl aliquots and added to 30µl lysis mix, as is the procedure for experimental samples (section 5.3). These control samples were then processed with experimental samples and used to create the standard curve.

5.4.2 *Tb*ZFP3 is an appropriate target for the quantification of parasitaemia in chronic trypanosome infections in mice

To test that *Tb*ZFP3 expression levels are indeed equal in slender and stumpy trypanosome RNA samples, three slender and three stumpy standard curves were produced by processing these samples simultaneously (section 5.4.1). A semi-log regression line was calculated for the slender samples and the stumpy samples (Figure 5.3A). Neither the slope nor the intercept of the regression lines of the slender and stumpy samples were significantly different ($F_{1,25}$= 0.048, p= 0.8276 and $F_{1,26}$= 0.004, p= 0.9493, respectively) and therefore the data could be combined (Figure 5.3B).

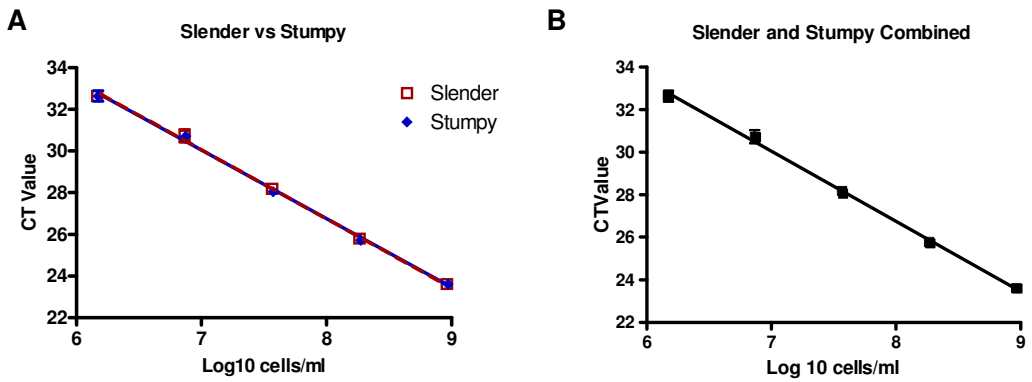


Figure 5.3 *Tb*ZFP3 is constitutively expressed between slender and stumpy forms. (A) Three slender form and three stumpy form standard curves were used to calculate a semi-log regression line for each cell type (y= 53.07 – 3.289x, r^2= 0.9926 and y= 53.24 – 3.312x, r^2= 0.9949, respectively). The slope and intercept of the slender and stumpy semi-log regression lines are not significantly different from one another ($F_{1,25}$= 0.048, p= 0.8276 and $F_{1,26}$= 0.004, p= 0.9493, respectively). (B) Combining the slender and stumpy data produces a semi-log regression line of y= 53.15 – 3.299x, r^2 = 0.9937 which can be used to calculate estimated PCR efficiency of 100.92%. Error bars represent s.e.m.

Importantly, this demonstrates that *Tb*ZFP3 is indeed constitutively expressed between slender and stumpy forms. This has been tested over an approximately 1000-fold range from approximately $1.5x10^6$ to $9x10^8$ trypanosomes/ml mouse blood. Additionally, given that the six samples were processed and standard curves produced simultaneously, this demonstrates that there is very little error in the procedure. Furthermore, the efficiency of the PCR can be estimated using the following calculation:

PCR Efficiency = $(10^{-1/slope}-1) \times 100$

From the combined slender and stumpy *Tb*ZFP3 data (Figure 5.3B), the estimated PCR efficiency is 100.92% indicating excellent PCR performance. Overall therefore, *Tb*ZFP3 can be used as a target in qRT-PCR to measure parasitaemia. *Tb*ZFP3 is also an appropriate endogenous control for the ΔΔCT method of calculation of relative *PAD1* expression, which shall be discussed in section 5.5.2.

5.5 Measuring stumpy formation in chronic trypanosome infections in mice

As *PAD1* mRNA is differentially expressed in slender and stumpy forms, relative *PAD1* expression could potentially be used as a stumpy-specific target transcript in qRT-PCR during an *in vivo* infection. For this the ΔΔCT method of relative quantification of *PAD1* expression was used: the ΔΔCT method uses an endogenous control (in this case, *Tb*ZFP3) in order to calculate the expression of *PAD1* normalised to the total number of cells in the population and relative to a reference sample (such as a given day during an *in vivo* infection).

5.5.1 Design of *PAD1*-specific RT-PCR primers

PAD1 is a member of an eight gene family and so care had to be taken to ensure that *PAD1* alone was detected. Although other possible primer pairs were tested (data not shown), only one target location was identified that allowed specific amplification of *PAD1* and not the other *PAD* gene family members without contaminating primer dimers. These primers spanned both the protein coding region of the gene and the 3'UTR (Figure 5.4).

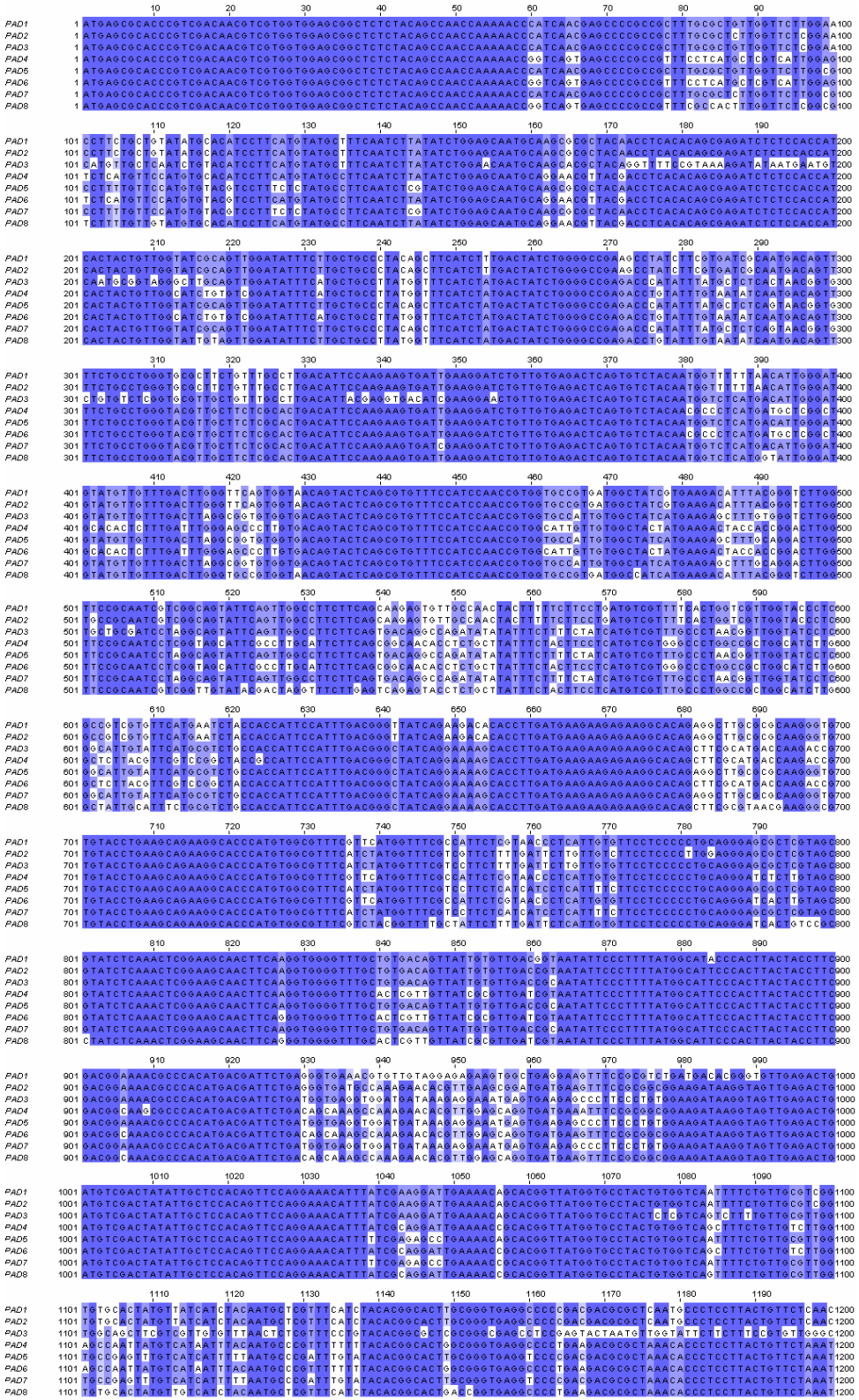


Figure 5.4 continued on the following page

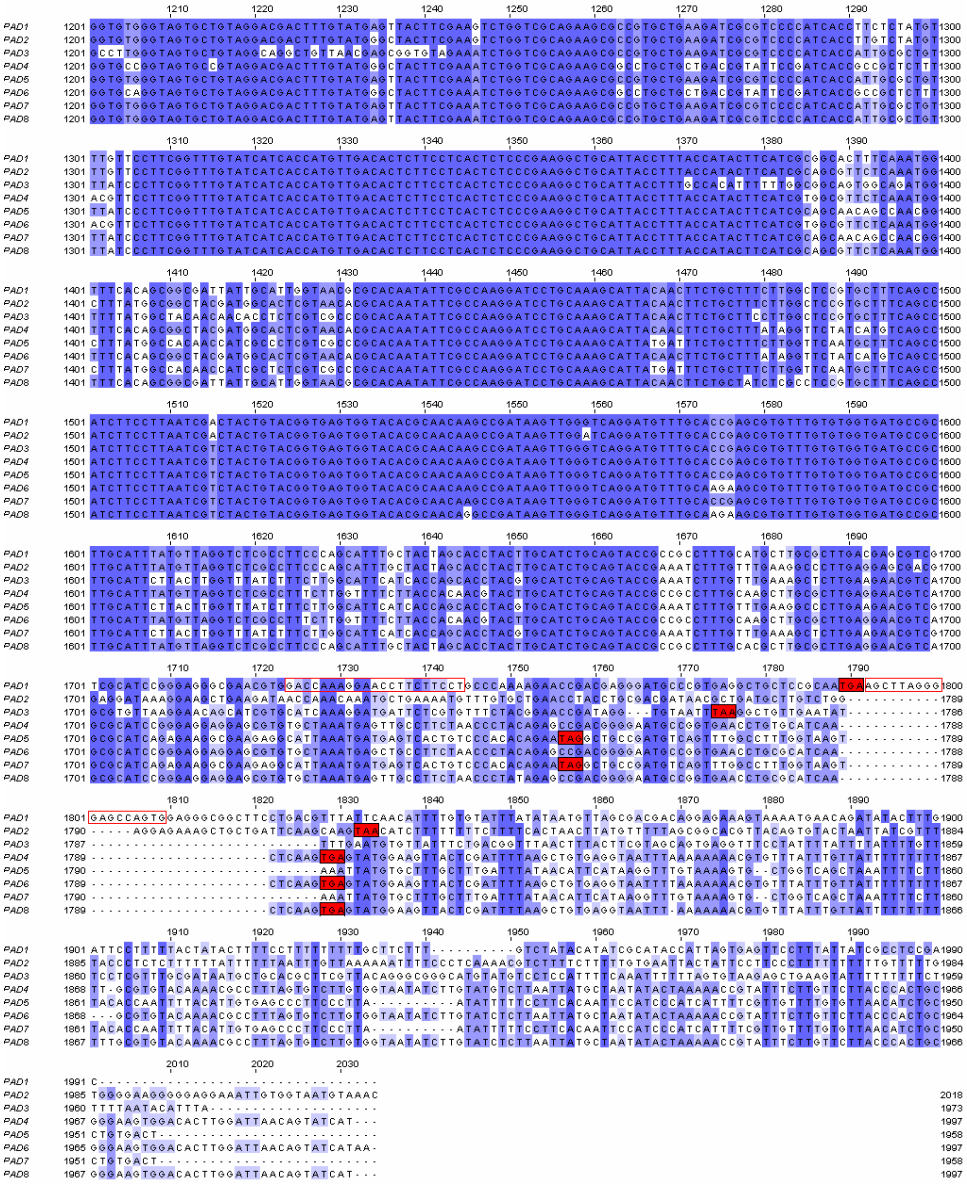


Figure 5.4 Alignment of *PAD* array coding sequence and 3'UTRs showing the location of the *PAD1* specific primers. As the polyadenylation site of most of the *PAD* genes is unknown only the first 200bp of the 3'UTR of each gene is shown. Filled red boxes indicate the stop codons. Unfilled red boxes indicate the locations of the sequences targeted by the *PAD1* specific primers at position 1724-1744bp and 1792-1809bp of the alignment. 5'UTRs of the genes are not shown.

5.5.2 Validation test for use of *PAD1* in the ΔΔCT method calculation

For the ΔΔCT method calculation to be valid, it was necessary to show that both the endogenous control, *Tb*ZFP3, and the target of interest, *PAD1*, had approximately equal amplification efficiencies. This was to ensure that for any dilution of a sample, within the range of the assay, the calculation would give consistent results and therefore that it was appropriate to compare samples of different concentrations.

If the amplification efficiencies of the control and target gene are approximately equal then the ΔCT, i.e. ($CT_{PAD1} - CT_{ZFP3}$) should not vary with dilution of a given sample. To test this, a standard curve for both *PAD1* and *Tb*ZFP3 was produced from the same stumpy trypanosome RNA sample (section 5.4.1) and the ΔCT was calculated over a range of dilutions from approximately $7.3x10^6$ to $9x10^8$ trypanosomes/ml mouse blood (Figure 5.5). The slope of the semi-log regression line in this test was 0.023. A slope of less than ±0.1 is considered to be acceptable for use of the ΔΔCT method. Thus, this experiment was a suitable validation that the PCR efficiencies of *PAD1* and *Tb*ZFP3 were approximately equal and do meet the requirements for use of the ΔΔCT method.

Although a further dilution to approximately $1.5x10^6$ /ml mouse blood was included, this was below the limit of detection of *PAD1* in this assay. Therefore when *PAD1* is not detectable at low parasitaemia this may not necessarily reflect a lack of *PAD1* expression. Although attempts were made to increase the sensitivity of the assay, the limitations of the small initial sample size (10µl) and the various dilutions required (section 5.3) prevented further improvements to the assay.

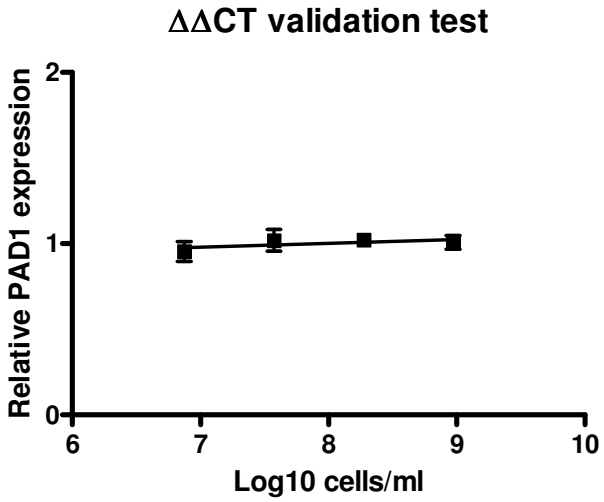


Figure 5.5 Validation test for the ΔΔCT method calculation. Stumpy trypanosome RNA extracted from whole blood was used to produce a standard curve for both *PAD1* and *Tb*ZFP3 ranging from 7.3 x 10^6 to 9 x 10^8 cells/ml using 1/5 serial dilutions with three replications per dilution. The ΔCT ($CT_{PAD1} - CT_{ZFP3}$) was calculated for each dilution and plotted against the log of the input cell concentration. A semi-log regression line was calculated for the data. As the slope of the semi-log regression line was 0.023, this demonstrated that it was appropriate to use the ΔΔCT method calculation. Error bars represent s.e.m.

5.6 Analysis of chronic trypanosome infections in mice using qRT-PCR

Having developed a qRT-PCR approach that could be used to monitor stumpy formation and parasitaemia during an infection, six male MF1 mice were infected I.P. with approximately 1000 AnTat1.1 pleomorphic trypanosomes from a donor mouse. Every day, for 30 days, samples were taken for qRT-PCR (section 5.3) and for measurement of parasitaemia by microscopy using the rapid matching method (Herbert and Lumsden, 1976). Further, from day 4-9 microscope slides were prepared in order to monitor morphology, VSG expression and cell cycle status such that the temporal order of events during the first slender to stumpy transition could be ascertained. The experimental procedure is outlined in Figure 5.6.

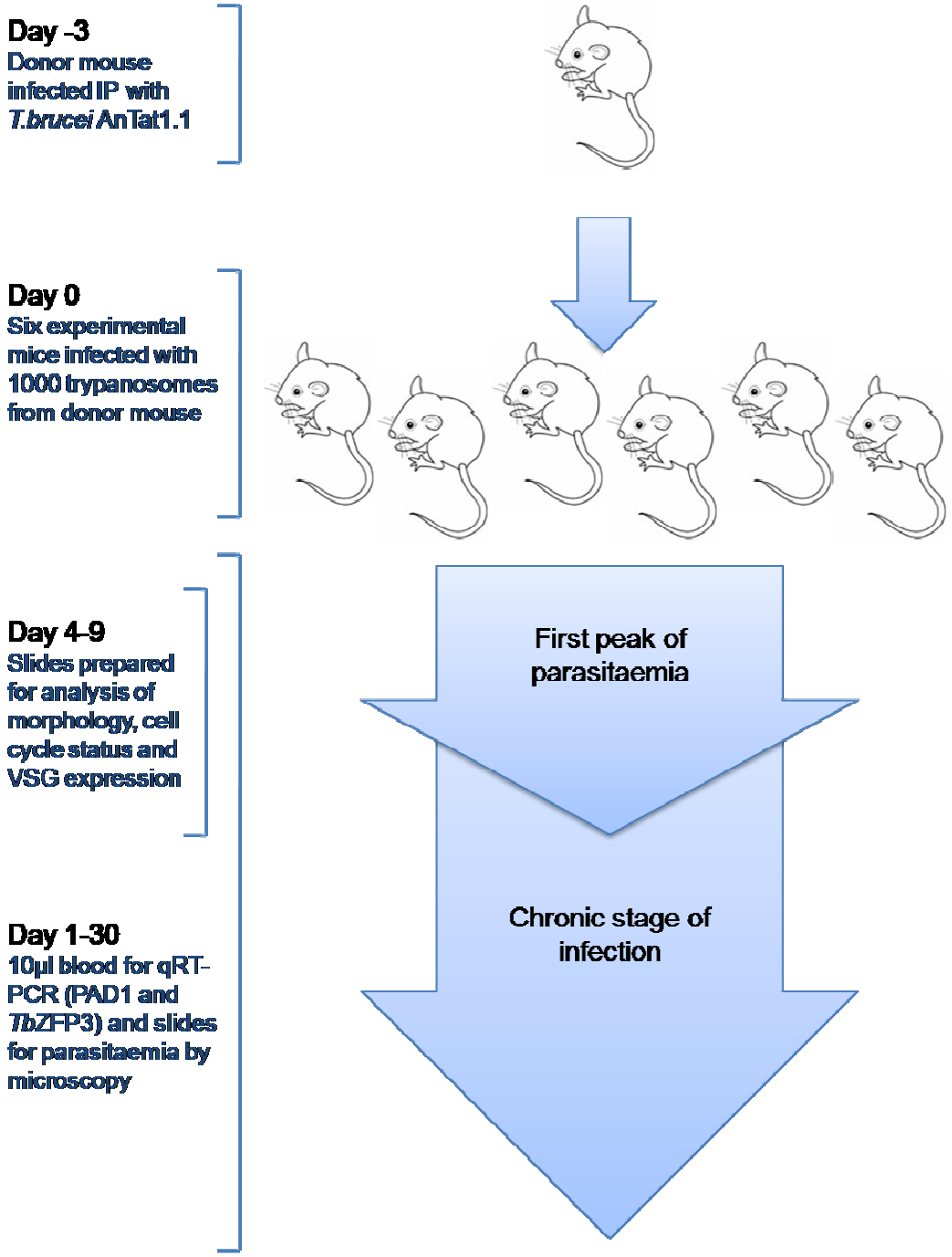


Figure 5.6 The experimental procedure for analysis of within-host dynamics during chronic mouse infections. A donor mouse was infected with an unknown quantity of AnTat1.1 pleomorphic trypanosomes on day -3. On day 0, the donor mouse had a detectable parasitaemia composed of slender cells. Whole blood was collected from the donor mouse by cardiac puncture and trypanosomes were purified (Chapter 2) and counted. Approximately 1000 trypanosomes were then injected I.P. into each of 6 experimental male MF1 mice. Infection was allowed to progress for 30 days post infection (dpi). Every day samples were collected for qRT-PCR and for determining parasitaemia by microscopy. On days 4-9 microscope slides were prepared for the analysis of cell cycle status, morphology and VSG expression.

5.6.1 Parasitaemia fluctuates throughout the infection

Every day throughout the infection samples were collected for qRT-PCR to determine parasitaemia, as well as relative *PAD1* expression (section 5.4 and 5.5). Additionally, microscope slides were prepared for the purpose of analysis of parasitaemia by microscopy. By both microscopy and qRT-PCR, it was observed that in all mice, infection followed a similar pattern of progression during the first 12 days post infection (dpi), where the parasitaemia rose from below detection on 2dpi to an initial peak of parasitaemia on 5-7dpi. In all cases, the parasitaemia then fell dramatically over 9-12dpi before fluctuating, with individual variability, for the remainder of the infection (Figure 5.7).

By comparing the estimated parasitaemia as determined by microscopy with the calculated parasitaemia as determined by qRT-PCR, there was found to be a good correlation with the patterns of parasitaemia in most cases (for example, see mouse 2, 23-30dpi), but a less good correlation between exact cell numbers, particularly during the first peak of parasitaemia. This indicates that the estimation of parasitaemia by microscopy in this case is not particularly accurate, especially at very high parasitaemia. The microscopy method used requires that a wet film of blood is analysed and the approximate number of parasites per field is translated into parasites/ml (Herbert and Lumsden, 1976). It may be that at very high density, parasitaemia was being underestimated. In very occasional cases, the microscopy data and the qRT-PCR data appear to differ in the pattern of change in parasitaemia. For example, in mouse 3, 20-25dpi a large fall and then rise in parasitaemia was observed through qRT-PCR which was not detected by microscopy. Given that the qRT-PCR data shows a gradual change over time this indicates that this is not likely an error in the qRT-PCR data, although it is possible, and is perhaps more likely due to human error. Although there are a few discrepancies with the qRT-PCR and the microscopy method for analysis of the parasitaemia, the qRT-PCR method, using *Tb*ZFP3 as a constitutively expressed transcript, provides a valuable technique for the accurate determination of parasitaemia throughout the course of an infection. Expansion of this method for use in other organisms, such as cattle, could provide a method for analysis of chronic infections in the field (section 5.8).

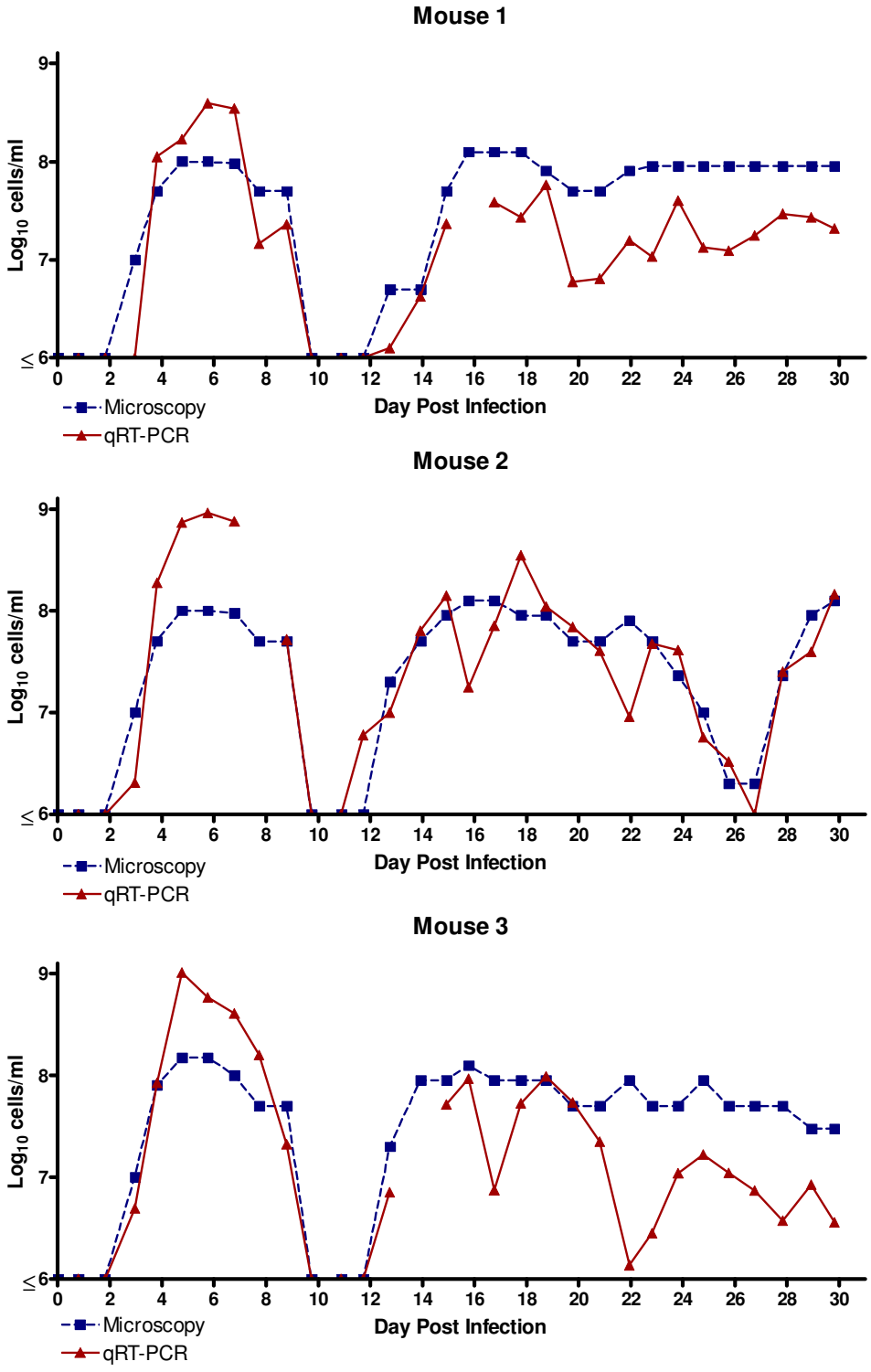


Figure 5.7 continued on the following page

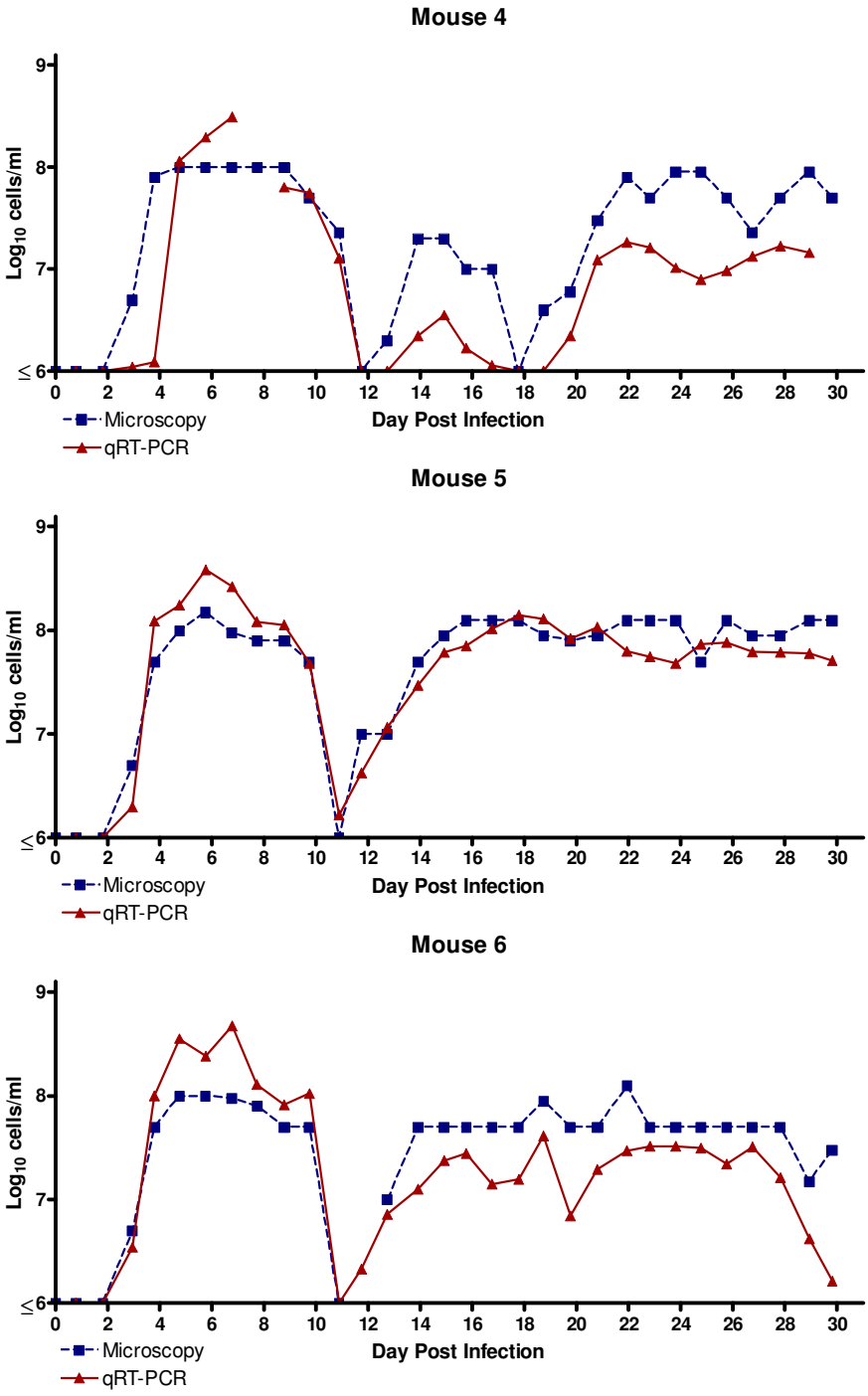


Figure 5.7 Comparison of parasitaemia as determined by microscopy and qRT-PCR. Throughout the infection parasitaemia was monitored daily by both microscopy and qRT-PCR. After an initial peak and fall in parasitaemia (measured in cells per ml mouse blood) the parasitaemia fluctuates throughout infection. Generally, the patterns of parasitaemia are equivalent as determined by both experimental methods, although differences are observed in absolute parasitaemia, likely due to errors by microscopy or qRT-PCR. Gaps in the data represent missing data points.

As already mentioned, the parasitaemia in all six mice fluctuates throughout the 30 day infection. During the chronic stage (day 12 onwards) the parasitaemia and fluctuations in parasitaemia vary greatly between the six mice. For instance, in mice 1-3 and 5-6 parasitaemia recovers well between 12-16dpi from the dramatic fall after the first peak. In mouse 4, however, the parasitaemia does not recover well and rises to a peak of only 3.5 x 10^6 cells/ml on 15dpi before falling again. Interestingly, the parasitaemia then does increase to a maximum parasitaemia of 1.8 x 10^7 cells/ml, which is similar to levels of parasitaemia observed in other mice.

5.6.2 The within-host dynamics of parasitaemia and stumpy formation during chronic infection

As well as measuring parasitaemia, the relative expression of *PAD1* was monitored throughout infection (Figure 5.8). The expression of *PAD1* is displayed relative to the parasitaemia to account for differences in cell number and in all cases is normalised to mouse 5, 4dpi which has been given an arbitrary value of 1. The first observation to note is that the *PAD1* expression of the population did fluctuate throughout infection, corresponding to fluctuations in parasitaemia (Figure 5.8). Initially the *PAD1* expression was very low, or below detection, at 3dpi when the parasitaemia was first detected in most cases, correlating with the day 3 population being composed of predominantly slender forms, but rose between 4-6dpi as the parasitaemia continued to rise. The delay between a rise in parasitaemia and a rise in *PAD1* expression correlates with the concept that the slender cells proliferate, and as the density increases, produce SIF, which in turn causes stumpy formation. Although there was individual variation between the mouse infections, generally the parasitaemia began to fall at 7-10dpi whereas the relative *PAD1* expression remained high. This is indicative of the loss of proliferative cells by differentiation and immune killing, with the cells that are persisting at this stage being differentiated forms and therefore expressing *PAD1* mRNA. It is indeed known that stumpy forms are more resistant to immune killing than slender forms and persist therefore for longer in the face of the developing host immune response (Engstler et al., 2007).

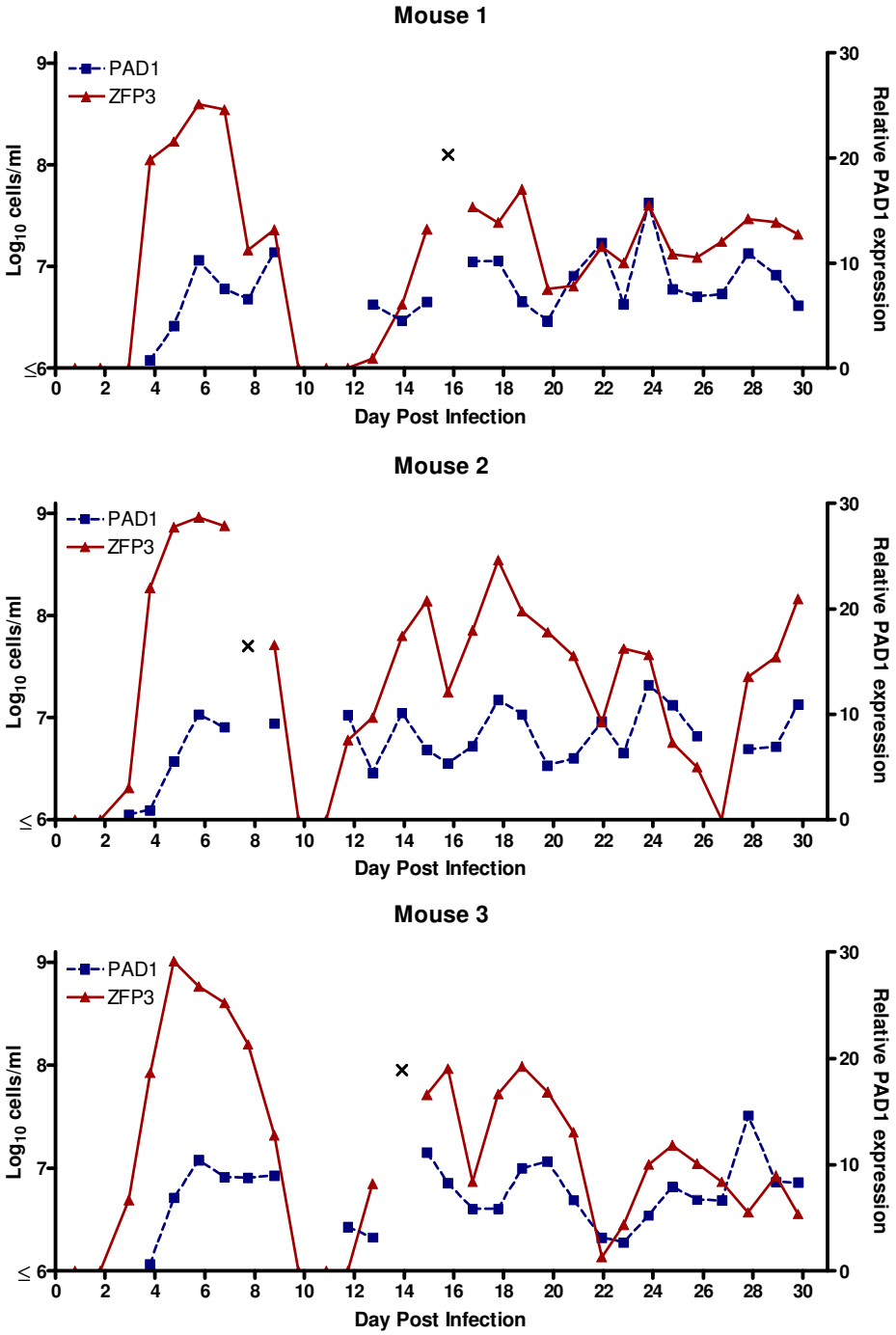


Figure 5.8 continued on the following page

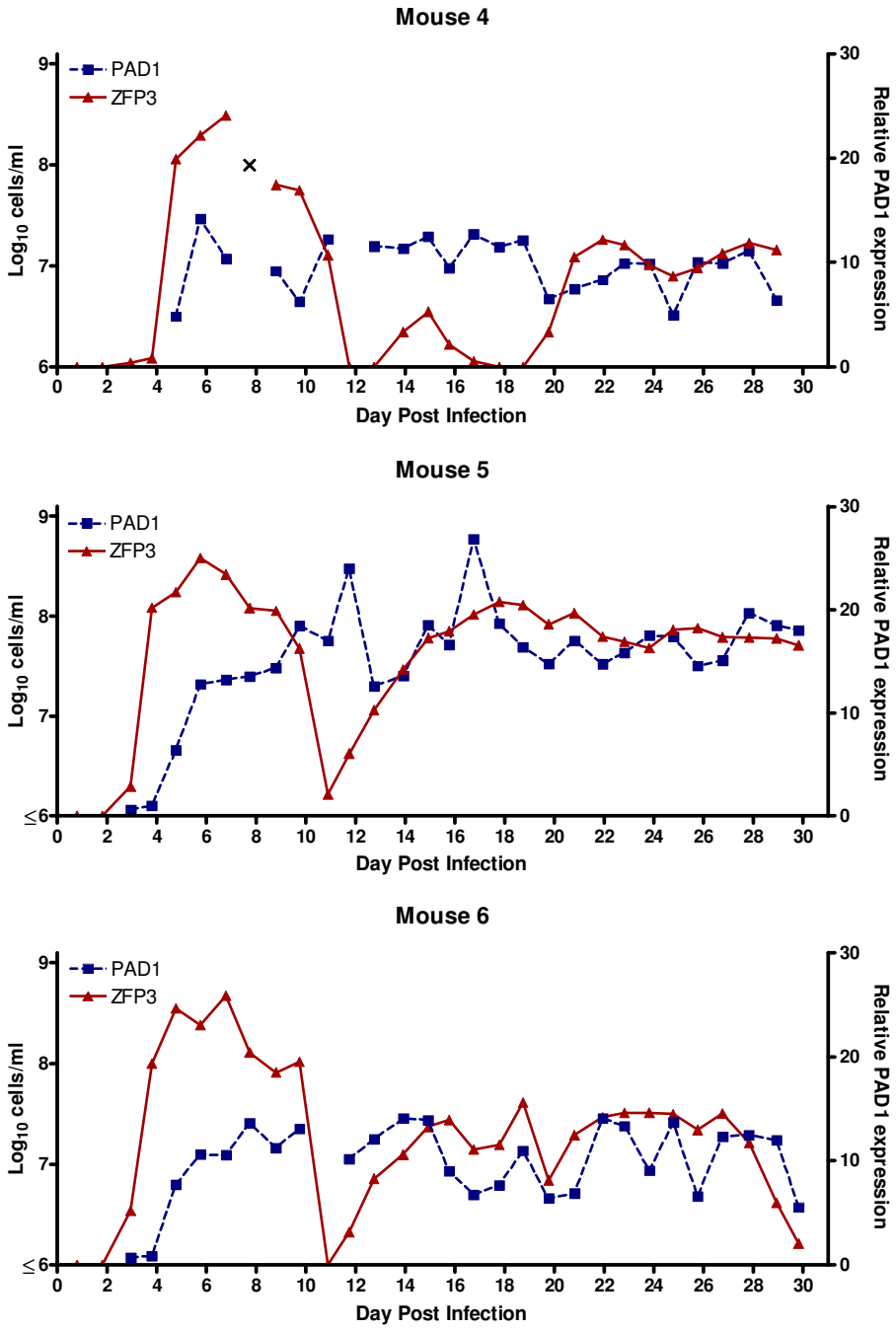


Figure 5.8 Parasitaemia and stumpy formation in chronic infections over 30 days. Parasitaemia, as determined by qRT-PCR for TbZFP3, is denoted by the red line, using the left y-axis. Where qRT-PCR data points are missing, a cross (X) represents the parasitaemia as measured by microscopy for that day. *PAD1* expression, denoted by a blue line using the right y-axis, is given normalised to the parasitaemia and relative to the expression in mouse 5, 4dpi, which has been given an arbitrary value of 1. Relative PAD1 expression is a measure of the total PAD1 expression of a population of cells (normalised to cell number) and does not represent a measure of the number of differentiated cells.

Throughout the chronic stage of infection the level of *PAD1* expression fluctuated somewhat with fluctuating parasitaemia but generally remained high. This demonstrates that at all times during the chronic stage of infection the parasitaemia contained transmission competent differentiated forms. Indeed, even at low parasitaemia, such as observed in mouse 4, 13-19dpi (Figure 5.8), the relative *PAD1* expression remained high, indicating that the few cells that did make up the parasitaemia were stumpy forms. This is logical given that if the cells were predominantly slender forms it would be expected that the parasitaemia would rise and not remain at this low level. Indeed, when parasitaemia did begin to rise on 19dpi (Figure 5.8), this was accompanied by a drop in the relative *PAD1* expression, due to the increasing proportion of slender forms.

The maintenance of a high proportion of stumpy forms throughout infection, even during periods of low parasitaemia, could have important implications on infection transmissibility (section 5.8.5). Additionally, during an infection trypanosomes periodically switch their surface coat in order to evade the host immune system. Only slender forms can functionally switch their VSG coat and so if the population contains a relatively low number of slender forms, then this could prolong antigen chronicity by reducing the number of antigenic variants used in a given time frame compared to a population with a higher proportion of slender forms. Therefore the maintenance of low slender density throughout infection could contribute to low antigen variant usage as well high transmission capacity.

5.6.3 The temporal order of events during the slender to stumpy transition (within a population)

Only during the first wave of parasitaemia in a new infection can the temporal order of events during stumpy formation be monitored at a population level *in vivo.* Therefore, during this first peak of parasitaemia, microscope slides were prepared for the analysis of morphology and cell cycle status.

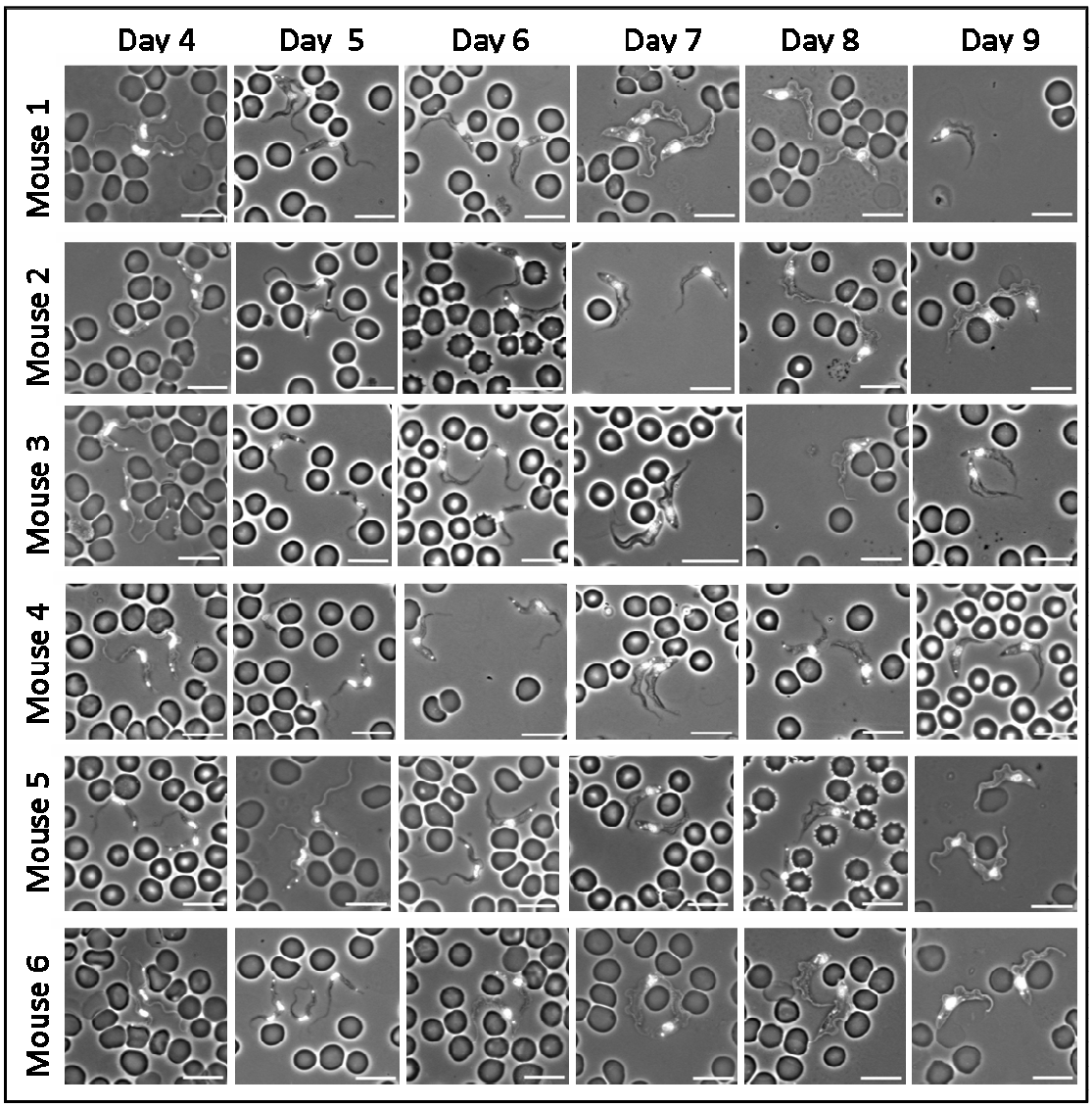


Figure 5.9 Representative microscope images of the morphology of cells at 4-9dpi in 6 mice. For all six mice, at 4-9dpi, slides were prepared from blood from a tail snip. Slides were fixed and stored in 100% methanol at -20ºC. Slides were stained with DAPI to visualise the nucleus and kinetoplast to aid in the identification of cell type. Dividing cells were always classified as slender forms. Scale bars represent 10µm.

The analysis of cell type by morphology can be somewhat subjective and is rather time-consuming. Morphological change is, however, the main way by which stumpy forms have been distinguished previously (Breidbach et al., 2002; Denninger et al., 2007; Laxman et al., 2006; Reuner et al., 1997; Savill and Seed, 2004; Tyler et al., 1997). By comparing morphological cell cycle data with the qRT-PCR data, this allows the traditional methods to be compared to the protocol developed here. Additionally, morphological change, cell cycle exit and altered gene expression may

not all coincide temporally during the slender to stumpy transition and comparing the data will allow for dissection of the order of events.

Firstly, 250 cells per mouse per day (4-9dpi) were scored as slender, intermediate or stumpy forms. Slender forms are longer and thinner than stumpy forms with a larger distance between the nucleus and the kinetoplast (section 1.6, figure 1.6). Additionally, all dividing cells were classified as slender which presumes that cell cycle arrest occurs before or concurrently with morphological change (section 5.7.6). Stumpy forms are wider with a shorter distance between the nucleus and the kinetoplast (section 1.6, figure 1.6). Those cells which could not definitively be classified as either slender or stumpy were classified as intermediate forms. Due to the subjective nature of the analysis it is possible that some early intermediate cells could be scored as morphologically slender or that late intermediates could be scored as morphologically stumpy. Figure 5.9 shows a panel of representative images over the six day period in the six mice and Figure 5.10 shows the quantitative analysis. In all mice, the trypanosome population at 4-5dpi was predominantly slender in morphology. At 6dpi there was a peak of intermediate forms and the first fully stumpy forms became visible. At 7-9dpi the majority of cells were stumpy in morphology, with very few remaining slender forms. The combined data from all six mice can be seen in Figure 5.12.

Next, 250 cells per mouse per day (4-9dpi) were scored for their position in the cell cycle (section 1.5). Cells were classified as 1K1N, 2K1N or 2K2N according to the number of nuclei and kinetoplasts present in the cell, with 2K1N and 2K2N cells representing proliferating cells and 1K1N cells representing cells in G1 or G0 (or early S phase) of the cell cycle. It was observed that at 4dpi approximately 25% of the population was 2K1N or 2K2N, representing a typical proliferative population. This correlated with the predominantly slender morphology observed at this point in infection (Figure 5.10). After 4dpi, the proportion of 2K1N and 2K2N cells in the population decreased until 7dpi where almost the entire population was 1K1N, correlating with the predominantly stumpy morphology of the cells by 7dpi. The combined data from all six mice can be seen in Figure 5.12.

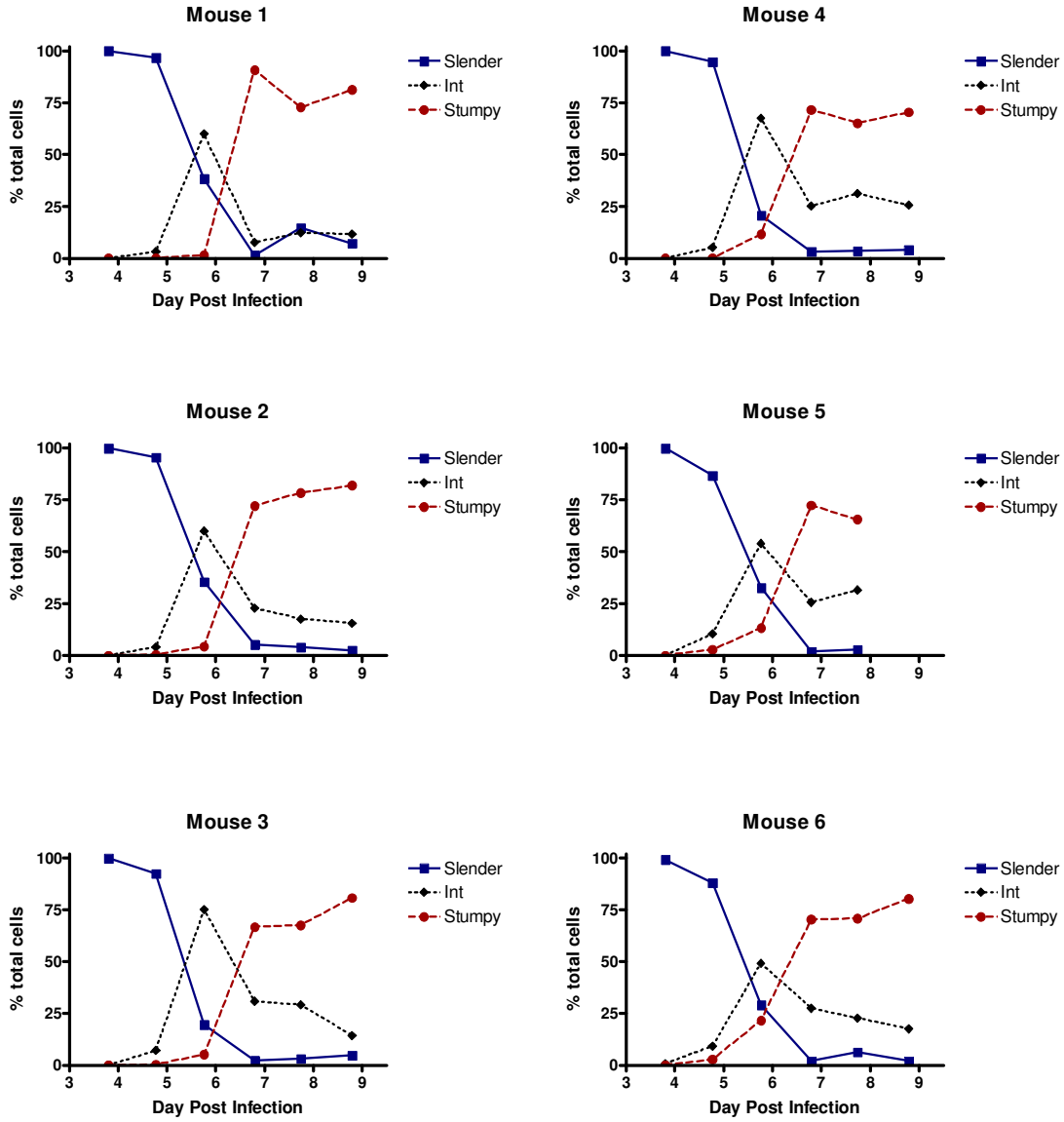


Figure 5.10 Cell morphology during the first peak of parasitaemia *in vivo*. 250 cells per mouse per day (4-9dpi) were scored for morphology by microscopy (see Figure 5.9 for examples of cell types). All mice showed a similar pattern of morphological change. In all cases, slender forms predominated on 4-5dpi followed by a peak of intermediate forms on 6dpi. Stumpy forms began to appear on 6dpi and predominated in the infection at 7-9dpi. The combined data from all six mice can be seen in Figure 5.12.

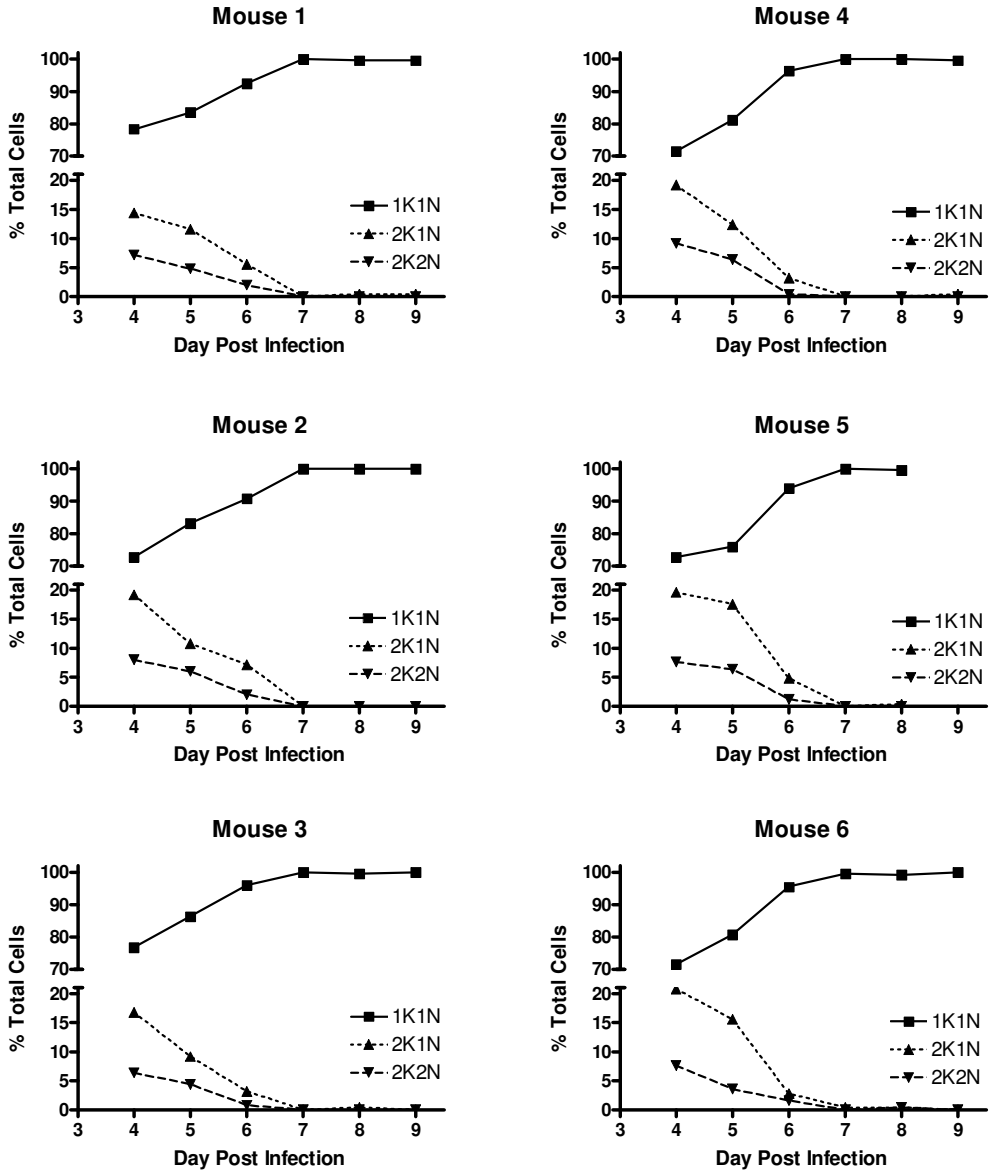


Figure 5.11 Population cell cycle status during the first peak of parasitaemia *in vivo*. 250 cells per mouse per day (4-9dpi) were scored for cell cycle status by microscopy (see Figure 5.9 for examples of 1K1N, 2K1N and 2K2N cells). All mice showed a similar pattern of cell cycle change as infection progressed. At 4dpi approximately 25% of the cells in the population were either 2K1N or 2K2N. From 4dpi to 7 dpi the number of dividing cells dropped and the number of 1K1N cells increased such that by 7-9dpi there were very few dividing cells in the population. This decrease in proliferating cells correlated with the loss of slender cells from the population (Figure 5.10). The combined data from all six mice can be seen in Figure 5.12.

By combining the data from all six mice for both morphology and cell cycle status, and comparing this to the qRT-PCR data generated for this time period it was possible to observe the pattern of changes that were occurring in the population during the first peak of infection (Figure 5.12). After infection, the parasitaemia was below the limit of detection up to 2dpi, with the first detectable parasitaemia on 3dpi. On 4 and 5dpi, the parasitaemia rose as dividing slender forms predominated in the infection (99.9% ±0.1 s.e.m and 92.4% ±1.7 s.e.m respectively). As the first intermediate forms began to appear between 4 and 5 dpi, fewer replicating cells were present (with the number of 1K1N cells rising from 74% ±1.2% s.e.m to 81.9% ±1.4% s.e.m) and the *PAD1* mRNA levels rose. As *PAD1* mRNA began to rise before the presence of a large number of intermediate and stumpy forms it may be that *PAD1* mRNA expression is a very early marker for differentiation and mRNA expression rises before the onset of detectable morphological change (section 5.7.5), although PAD1 protein is not detected until after morphological change (Dean et al., 2009). On 6dpi there was a peak of morphologically intermediate forms, comprising 61% (±3.8% s.e.m) of the population, along with the peak of *PAD1* mRNA. Thus, *PAD1* was expressed in morphologically intermediate forms to a similar degree as in stumpy forms confirming that it is an early marker for stumpy formation. Stumpy forms began to emerge at 6dpi (9.6% ±3.0% s.e.m). Then, as stumpy forms predominated in the infection at 7-9dpi (7dpi, 74% ±3.5%; 8dpi, 70.1 ±2.1%; 9dpi, 79% ±2.2% s.e.m), there were correspondingly few dividing cells (with 99.9% ± 0.1% s.e.m at 7dpi being 1K1N) and the *PAD1* mRNA levels plateaued. Parasitaemia then began to fall between 7 and 8dpi, likely due to a combination of both stumpy formation and immune clearance of the infecting antigen type (see section 5.7). Indeed, these events during the first wave of parasitaemia correspond to the expression of only one VSG variant (section 5.6.4).

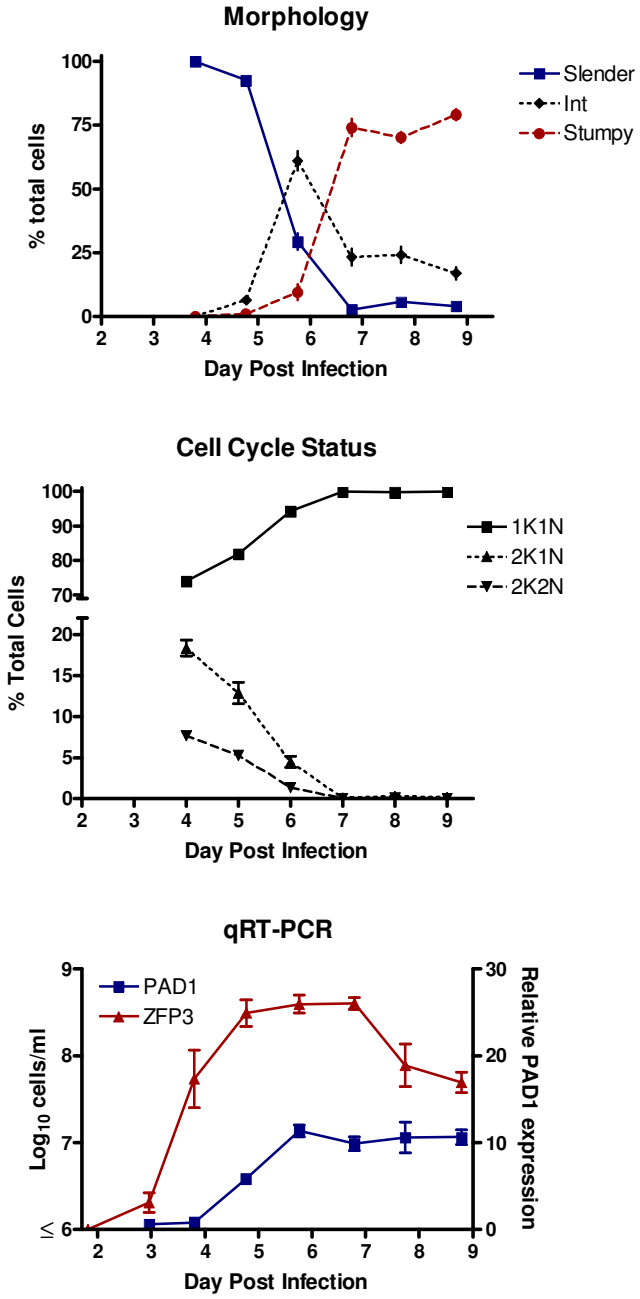


Figure 5.12 Comparison of morphology, cell cycle status, parasitaemia and *PAD1* expression for all six mice. Comparing both the microscopy and qRT-PCR data for all six mice the temporal order of events during the first stage of parasitaemia could be deduced. On 4 and 5dpi, the parasitaemia rose and dividing slender forms predominated in the infection. On 6dpi there was a peak of morphologically intermediate forms correlating with a peak of *PAD1* expression as *PAD1* demonstrating that *PAD1* is an early marker for differentiation. Stumpy forms predominated in the infection at 7-9dpi and there were correspondingly few dividing cells. The *PAD1* mRNA levels plateaued at 6-9dpi demonstrating that *PAD1* is expressed equally in intermediate and stumpy forms. Error bars represent s.e.m.

Given the consistency of data between individual mice, this temporal progression of stumpy formation of the population seems to be well regulated and this analysis provides a temporal profile of the slender to stumpy transition early in infection. Furthermore, the sensible correlation between cell morphology, cell cycle status, parasitaemia and *PAD1* mRNA expression demonstrates that the qRT-PCR method developed here for the analysis of stumpy formation was a suitable way to monitor within-host dynamics. This method is less subjective than microscopy as well as less time-consuming, allowing large numbers of samples to be processed simultaneously. There are numerous future applications of this method, as will be discussed in section 5.8.

5.6.4 The first peak of parasitaemia correlates to the expression of one antigen variant

The pleomorphic strain AnTat1.1 used in this analysis is named after the infecting antigen variant that is expressed on its surface. Using an antibody to the AnTat1.1 VSG, microscopy slides were analysed for loss of AnTat1.1 VSG expression, which would show that antigen switching had occurred (Figure 5.13).

Although cells negative for AnTat1.1 staining were detected as early as 7dpi, even by 9dpi more than 99% of the population remained positive. Due to the very low numbers of negative cells present, accurate estimations of the proportion of switched cells were not obtained. Interestingly, of the very few cells that were identified as being negative for AnTat1.1, both slender and stumpy morphology was observed (Figure 5.13, 8dpi and 7dpi respectively). Thus, either stumpy forms are capable of antigen variation (although this would not be a functional switch given that stumpy forms are cell cycle arrested and could therefore not populate the next wave of parasitaemia) or slender cells which have already switched antigen coat are not refractory to stumpy formation.

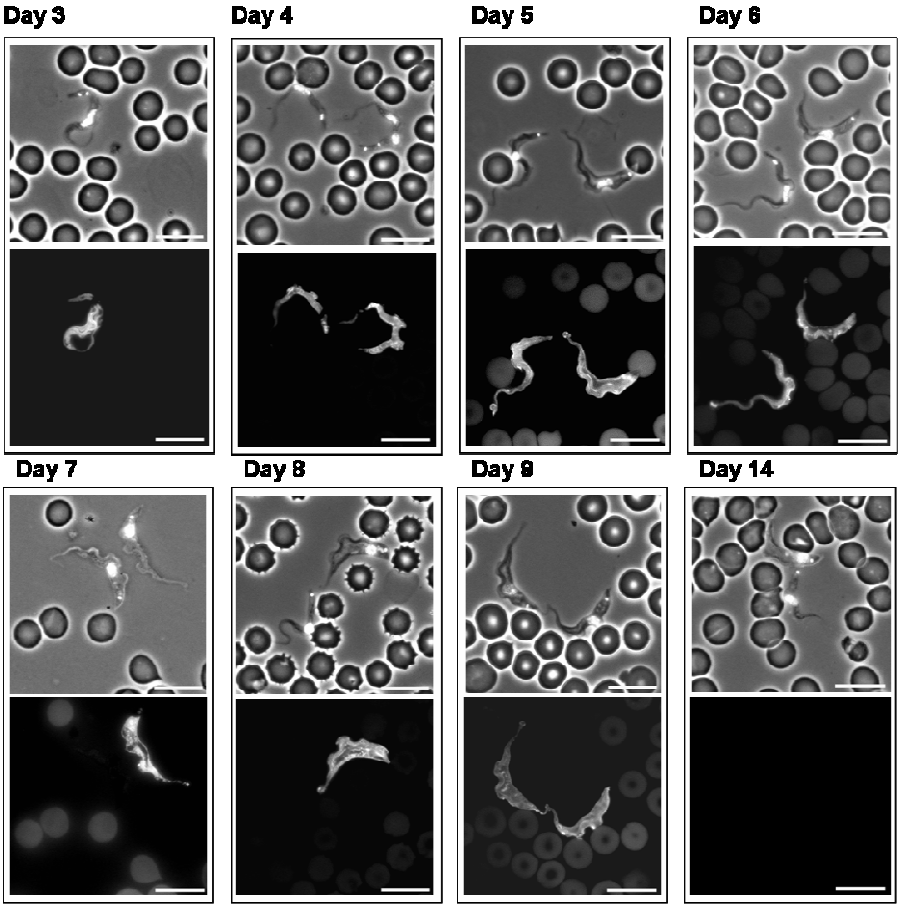


Figure 5.13 Loss of AnTat1.1 expression during infection. Microscope slides were stained for AnTat1.1 VSG expression. Shown are examples from the six mouse infections. From 3-6dpi no cells were observed that were negative for AnTat1.1 expression. From 7-9dpi the occasional cell negative for AnTat1.1 was observed, but this remained less than 1% of the population. Both slender and stumpy form cells could be identified as AnTat1.1 negative (for example, an AnTat1.1 negative stumpy form is shown on day 7 and a slender form on day 8). On 14dpi all cells were negative for AnTat1.1. Upper panels for each day are phase contrast plus DAPI staining. Lower panels for each day are stained for AnTat1.1 expression using a rabbit Anti-AnTat1.1 VSG primary antibody and an Anti-Rabbit-FITC secondary antibody. Scale bars represent 10μm.

5.7 Mathematical modelling of the slender to stumpy transition during infection

The slender to stumpy transition is density-dependent and controlled by a parasite-derived factor, SIF (Vassella et al., 1997). Most of the mathematical models produced to describe the slender to stumpy transition have not incorporated SIF-induced differentiation into the model, but rather use slender parasite density to

determine stumpy formation (Black et al., 1982; Lythgoe et al., 2007; Seed and Black, 1997; Seed and Black, 1999; Tyler et al., 2001). One model has been produced that considered SIF-induced differentiation during the slender to stumpy transition (Savill and Seed, 2004). This model was shown to be a good fit to existing data on parasitaemia over the first seven days of infection in immunocompromised mice and the resulting parameter estimates correlated well with previous studies (Savill and Seed, 2004). However, as with most models of stumpy formation produced thus far, this model was limited by the dependence on subjective morphological analysis of cell type and was considered in the absence of the immune system. Rather than the use of morphology to monitor stumpy formation, Tyler *et al* used increased expression of mitochondrial activity and did incorporate the host immune system into their model demonstrating that both the host immune system and density-dependent differentiation is responsible for limitation of parasitaemia, but that differentiation alone can account for fluctuations in parasitaemia throughout infection (Tyler et al., 2001). This model did, however, only quantify stumpy formation during the first six days of infection and did not consider SIF-induced differentiation.

The production of accurate mathematical models is limited by the data to which the model is fitted. In section 5.6, data on both parasitaemia and stumpy formation had been generated for six mice, which were not immunocompromised, every day for 30 days during a trypanosome infection, providing a large amount of non-subjective quantitative data to which a mathematical model could be fitted. Thus, this data was fitted to the pre-existing Savill and Seed, 2004, model which considers SIF-induced differentiation (Savill and Seed, 2004), but now incorporating a simple model of the host immune system. All mathematical modelling was carried out by Dr Nick Savill, University of Edinburgh. All figures produced from the data generated by the mathematical modelling (Figure 5.15 to Figure 5.22) were produced by Dr Savill. The data that was incorporated into the model was the measured parasitaemia and the total *PAD1* expression of the population, as measured by the raw C_T values from the qRT-PCR analysis, in the six mice over 30 days, generated in section 5.6.

5.7.1 The mathematical model

A detailed description of the mathematical model and the equations used, written by Dr Savill, is shown in Appendix C. The model comprises a set of differential and partial differential equations describing: (1) the concentration of non-committed slender cells of a particular antigen variant; (2) the age distribution of cells after the point at which they commit to differentiate (i.e. committed slender cells, intermediate cells and stumpy cells); (3) the dynamics of SIF concentration and (4) the immune clearance rate of each antigen variant. The model was fitted to the parasitaemia and *PAD1* expression data for each mouse and finds the most likely values for the parameters for each mouse. Many of the parameters are unknown, but parameter priors were incorporated into the model, where the expected values and distribution of a parameter, based on prior knowledge, was specified. Details of all the parameter priors are found in Appendix C, but some are also detailed below. When a parameter prior is described, the distribution of that parameter is specified i.e. if a parameter has a prior of $N_T(\mu, \sigma)$, that states that the prior has a normal distribution truncated at zero, with a mean of μ and a standard deviation of σ.

There are four distinct cell types in the model: non-committed, replicating slender cells, committed replicating slender cells, non-replicating intermediate cells and non-replicating stumpy cells (Figure 5.14). The concentration of non-committed slender cells of a particular variant is determined by the slender cell replication rate, the rate of switching to and from that variant, the immune clearance rate, and the rate of SIF-induced differentiation, all of which are model parameters. The total number of non-committed slender cells is the sum of all slender cells of all variants.

Non-committed slender cells of each variant are induced to commit to differentiation in response to SIF. After becoming committed to differentiation, the slender cell will undergo further replication as a committed slender cell before cell cycle arrest. It has indeed been shown that commitment to differentiation occurs before cell cycle arrest and morphological change (Tyler et al., 1997). The cell will then become intermediate before becoming fully stumpy. The lifespan of each cell type through differentiation are parameters of the model and the age distribution of committed

cells at any given time point is described by a partial differential equation. Committed replicating slender cells and intermediate cells of each variant are lost, through time, by differentiation to the next cell type as well as by immune clearance (Figure 5.14). Stumpy forms are lost, through time, by stumpy cell death as well as by immune clearance. The total concentration of cells is the sum of all cell types of all variants at any given point in time.

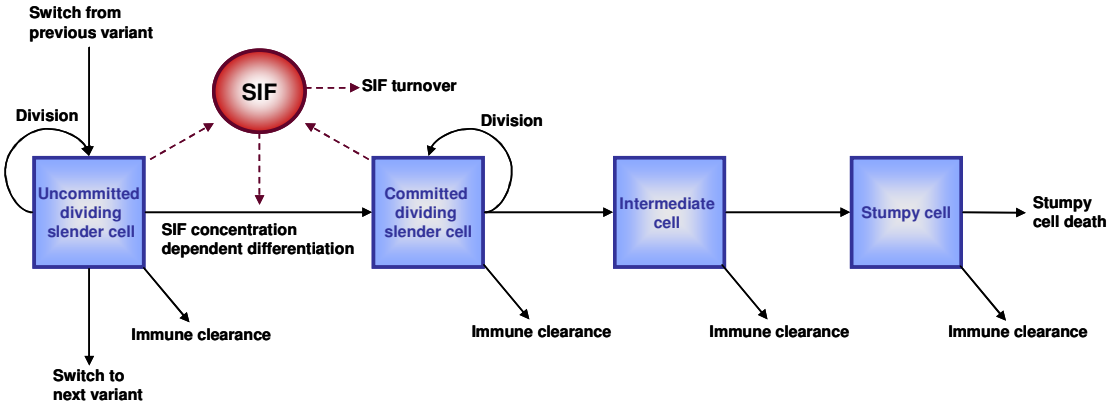


Figure 5.14 A model of stumpy formation *in vivo*. A schematic representation of the mathematical model used to describe stumpy formation during infection. Details of the model are discussed in the text.

SIF is produced by non-committed slender cells, as well as by committed slender cells up to a certain age (section 5.7.7). Intermediate and stumpy forms do not produce SIF. SIF is lost by turnover or clearance in the host. SIF induces non committed slender cells to differentiate at a rate proportional to SIF concentration. This model, incorporating SIF-induced stumpy formation, was shown to better account for the quantitative data than the use of parasite density induced differentiation (section 5.7.8).

PAD1 begins to be expressed as cells become committed to differentiation and is not expressed in non-committed slender cells. There are only two phases of change of *PAD1* expression through differentiation. Firstly, *PAD1* expression rises linearly until a certain age through differentiation, then *PAD1* expression may remain constant or rise or fall linearly until stumpy cell death. Based on the data collected in

section 5.6, the relative *PAD1* expression per stumpy cell is approximately 10^{-7} (see Appendix C) and therefore a parameter prior of $N_T(10^{-7}, 10^{-7})$ was set for the *PAD1* expression after the initial rise and at stumpy cell death. The change in *PAD1* expression through differentiation is discussed in section 5.7.5.

At time zero (i.e. at the point of infection) the infection is composed of one variant, expressing the infecting VSG (in this case AnTat1.1). Later, slender cells switch between variants sequentially (i.e. one after another) with differential switching rates, such that at any given time the infection is composed of one or more variants. Biologically, there is a degree of ordered expression of VSGs, but rather than being caused by sequential switching, it is determined by differential probabilities of activation of each variant (Morrison et al., 2005). Indeed, differential activation rates combined with density-dependent stumpy formation was demonstrated to be sufficient for ordered antigenic variation by mathematical modelling (Lythgoe et al., 2007). However, with the exception of analysis of the loss of the infecting variant (section 5.6), no data was collected regarding antigenic variation during these infections. As such, no data is available to model antigen switching and therefore associated parameters must be estimated. When producing a model, one of the aims is to explain the data in as few parameters as possible and when modelling a complicated biological system with numerous parameters, this will, at times, mean simplifying the system for the purposes of the model. Sequential antigenic variation is a simple model which reduces the number of parameters that must be estimated but still results in the expression of multiple variants at the same time (section 5.7.4). As such, the use of this model can be justified by parsimony. Furthermore, the number of variants expressed during the 30 day period is unknown. Therefore the most parsimonious situation would be for only one variant to be expressed. Thus, the parameter prior for the number of variants is set to $1 + N_T(0,5)$ and hence the minimum number of variants in the infection is 1. Clearly more than one variant will be required; otherwise the immune system would have been able to clear the infection. Indeed the model requires more variants in order to explain the data (section 5.7.4), but this prior dictates that when the parameters are fitted to the model, the fewest number of variants possible should be used.

Finally, as the data were collected in immunocompetent mice, the host immune system was considered. The mammalian immune system is complex and no data, such as the presence of antibodies against specific VSGs, was collected here. The model therefore incorporates a simple immune response against each variant which is activated by the replicating slender cells of that variant, as well as by a self-enhancing rate. The rate of immune clearance of each type of committed cell is proportional, but different, to the rate of immune clearance of slender cells of that variant, and can be different from one another. This is to account for the differential efficiency of the host immune system to clear slender and stumpy forms (Engstler et al., 2007). There is a time delay between the initiation of infection and the activation of the immune response against the infecting variant, but the immune response against further variants is not delayed. Although this may not represent the complexity of the host immune system, the lack of data regarding the immune response limits the production of a detailed model.

At time 0, i.e. the initiation of infection, all cells are assumed to be non-committed slender cells of the same infecting variant. The starting concentration of SIF is zero and the starting immune response to all variants is zero. Mice were injected with approximately 10^3 cells and therefore the parameter prior for the $\log_{10}$ starting concentration of cells was set as N(3, 0.5), presuming the volume of blood per mouse was approximately 1ml (section 5.7.7).

5.7.2 The adequacy of the fit of the model to the quantitative data

In order to fit the quantitative data of each mouse to the model, the parameters were varied until the 'best fit' model of each mouse is found. This model then provided a set of estimated values for each parameter which best explained the biological data (section 5.7.7). To assess how well the model fitted the data, Dr Savill compared the position of the mean of the standardised residuals at each time point to the 95% predictive interval of the mean (Figure 5.15). If the model is accurate, it would be expected that the mean at each time point would lie within this predictive interval 95% of the time and any mean that lies out with this interval suggests that the model

does not suitably explain that time point. Further, if there are outliers, where a standardised residual is more than about 3 standard deviations from zero, then this also indicates a poor fit of the model to the data. The model generated here had a very good fit to the quantitative data, with no outliers for either parasitaemia or *PAD1* expression (Figure 5.15). There was, however, an underestimation of parasitaemia on day 4, indicated by the mean of the standardised residuals being greater than the upper limit of the 95% predictive interval. This suggested that the model was underestimating the rate of growth of slender cells in the initial wave of parasitaemia.

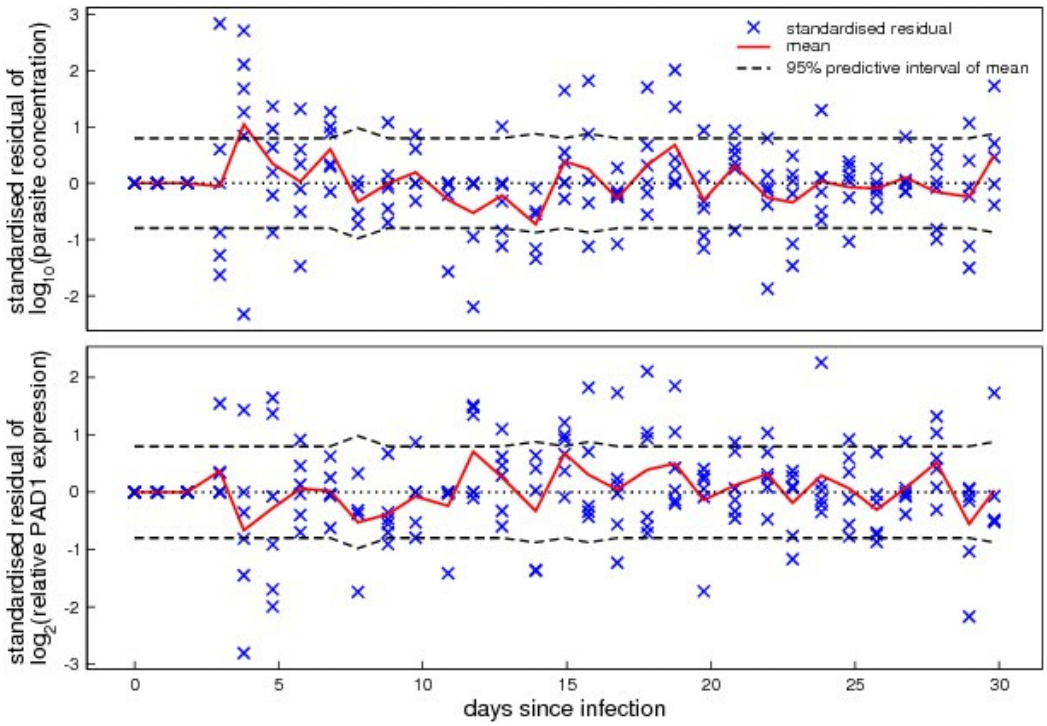


Figure 5.15 The adequacy of the fit of the mathematical model to the quantitative data. To test the adequacy of fit of the model to the data, the standardised residuals of each data point from each mouse (blue crosses) for both parasitaemia (upper panel) and *PAD1* expression (lower panel) was compared to the 95% predictive interval of the model (dashed lines). It would be expected that the mean of the data would be within that interval 95% of the time, if the model were correct. The mean of the standardised residuals from the six mice is shown as a solid red line. This mean lies within the 95% predictive interval of the mean in all cases, except for the parasite concentration on day 4. Standardised residuals which are greater than 3 standard deviations from 0 are considered to be outliers. There are therefore no outliers for this model. With the exception of day 4 parasitaemia the model is a very good fit to the data.

5.7.3 The predicted infection dynamics

The model fits well to quantitative data of the six mice, as determined by analysis of the standardised residuals from all six mice (section 5.7.2). Confirming this, the predicted and measured parasitaemia and *PAD1* expression in individual mice were well correlated throughout infection in most cases (Figure 5.16). The underestimation of parasitaemia on day 4 can be observed in Figure 5.16B, where the measured parasitaemia lies above the 95% predicted interval of the model solutions in all mice except mouse 4 where there is a delay in the increase in parasitaemia compared to the other mice.

In all cases, the model predicted that during the fall in parasitaemia after the first peak (Figure 5.16B) there was a second smaller peak, or a delay in the rate of decrease in parasitaemia. This could indeed be observed by the data alone in mouse 1, 4, 5 and 6, where the parasitaemia began to fall around 6-7dpi and then that rate of decrease slowed around 8-9dpi in most mice. For mouse 2, it was predicted that the second peak was of the same approximate magnitude as the first. However, for mouse 2 there was no qRT-PCR data available for day 8. Thus, it may be that this lack of data affects the fitting of the model such that the model inaccurately predicted a higher parasitaemia at this point.

From the model, predictions could be made regarding the dynamics of components of the model which were not experimentally tested, such as the dynamics of SIF concentration, the proportions of different cell types throughout infection, dynamics of antigenic variation (section 5.7.4) and *PAD1* expression per cell throughout differentiation (section 5.7.5).

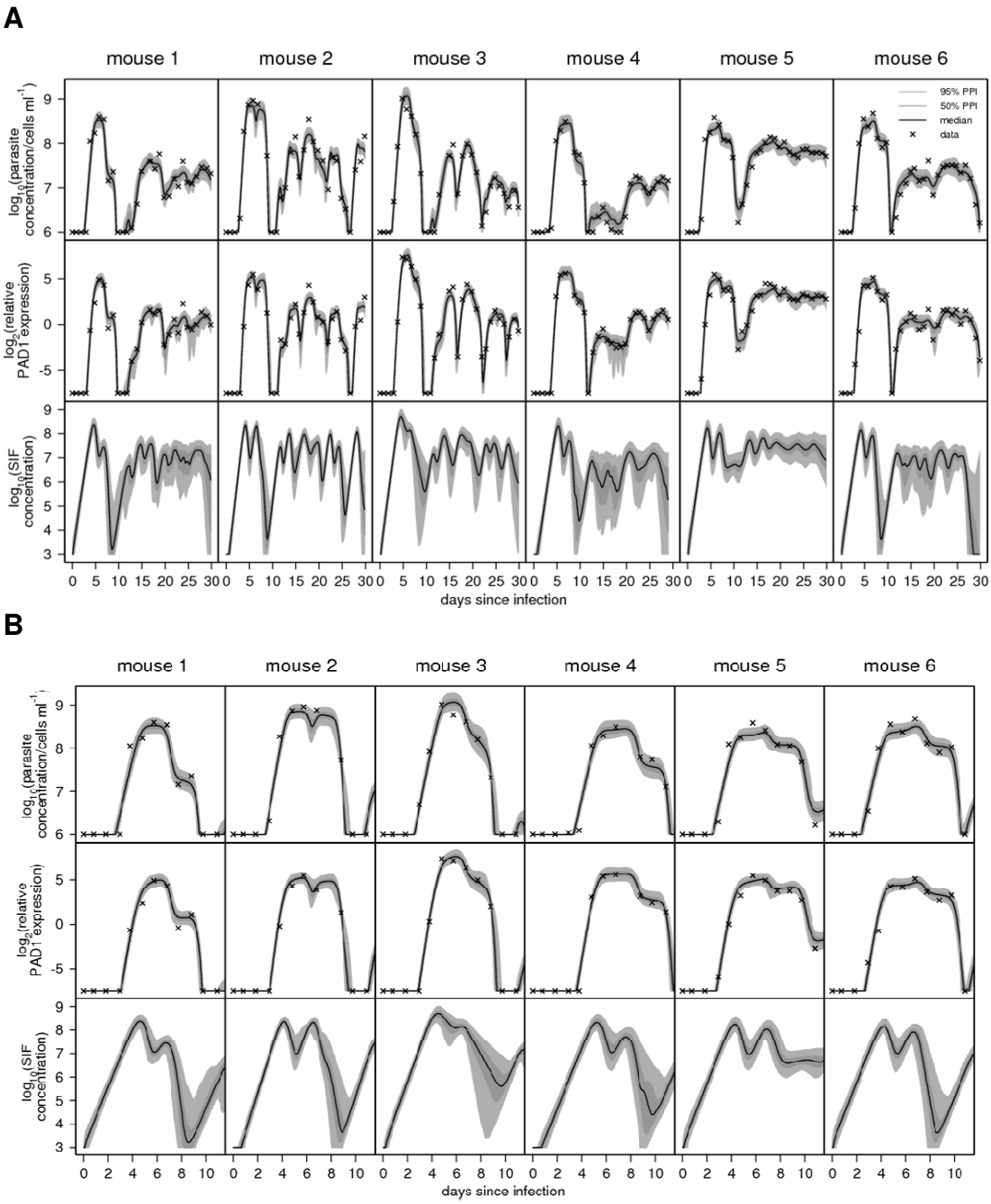


Figure 5.16 The fit of the mathematical model to the quantitative data for each mouse. The measured and predicted infection dynamics over (A) the full 30 days and (B) the first 11 days of infection. The measured parasitaemia (crosses) compared to the predicted parasitaemia (lines) according to the model is shown in upper panels. The measured $\log_2$ relative *PAD1* expression (crosses) compared to the predicted $\log_2$ relative *PAD1* expression (lines) is shown in centre panels. Note, that total *PAD1* expression of the population, relative to mouse 5 day 4 is fitted to the model, and plotted here, rather than the *PAD1* expression normalised to parasitaemia, as plotted in section 5.6. The predicted dynamics of SIF concentration is shown in lower panels. In all cases the 50% and 95% posterior predictive intervals are shown in dark and light grey, respectively.

Firstly, the model states that SIF induces stumpy formation at a rate proportional to its concentration. As the SIF concentration was unable to be measured due the lack of knowledge of its identity, there was no data regarding SIF that could be incorporated into the model, and so the predictive intervals are larger than those for parasitaemia and *PAD1* expression (Figure 5.16). During the first peak of parasitaemia, SIF was predicted to rise with parasitaemia. As the parasitaemia began to fall after the first peak, the concentration of SIF also began to fall; reducing the rate of differentiation. This may be the cause of the delay in rate of parasitaemia loss around day 8-9. As well as the differentiation to stumpy forms, the immune response will also contribute to changes in parasitaemia but comparing the parasitaemia during the first peak to the predicted immune response against the first variant demonstrates that the clearance rate of the first variant rises after the initial drop in parasitaemia but before the second large fall after day 8-9 (section 5.7.4, Figure 5.18C). Hence, the initial drop in parasitaemia was predicted to be caused by SIF-induced differentiation and not by the immune response.

The data here indicates that during the first peak of parasitaemia the SIF concentration rises (Figure 5.16B). The parasitaemia is then limited by SIF-induced differentiation to the stumpy form. As the parasitaemia begins to fall, so does the SIF concentration (Figure 5.16B), resulting in a reduced rate of differentiation. Thus, more of the remaining slender forms in the population will continue to proliferate, reducing the rate of loss of parasitaemia. This would predict a small rise in the number of slender forms (section 5.8.7) after the initial decrease in parasitaemia. Then as SIF concentration rises once again, and an immune response has been established against the first variant (section 5.7.4), the rate of loss of parasitaemia increases causing the low parasitaemia observed around day 10.

Additionally, throughout the chronic stage of the infection SIF concentration was predicted to fluctuate, as would be expected from the fluctuating parasitaemia and *PAD1* expression and fluctuations in cell types (discussed below). After the initial peak, the parasitaemia and *PAD1* expression in mouse 5 remained high throughout infection, this correlates with a high concentration of SIF throughout the infection.

The dynamics of the proportion of the different cell types could also be predicted throughout the course of infection (Figure 5.17). The model predicted that there were fluctuations in proportions of cell type, as would be expected, throughout infection. There were clear differences between different mice, where fluctuations were more pronounced in mouse 2 and 3 than in mouse 5 and 6, for example. After the initial fall in parasitaemia and proportion of slender cells there was a peak of slender forms in all mice as the parasitaemia is re-established. However, overall the proportion of slender cells tended to remain low during the chronic stage of infection, with tight peaks where this proportion did increase. The proportion of differentiated cells, however, tended to remain high although there were periodic sharp decreases followed by recovery of the differentiated population to a high proportion. Hence, during infection there was maintenance of a high level of transmissible forms, with no prolonged period of a low proportion of transmissible forms, implying that the population was transmission competent throughout infection.

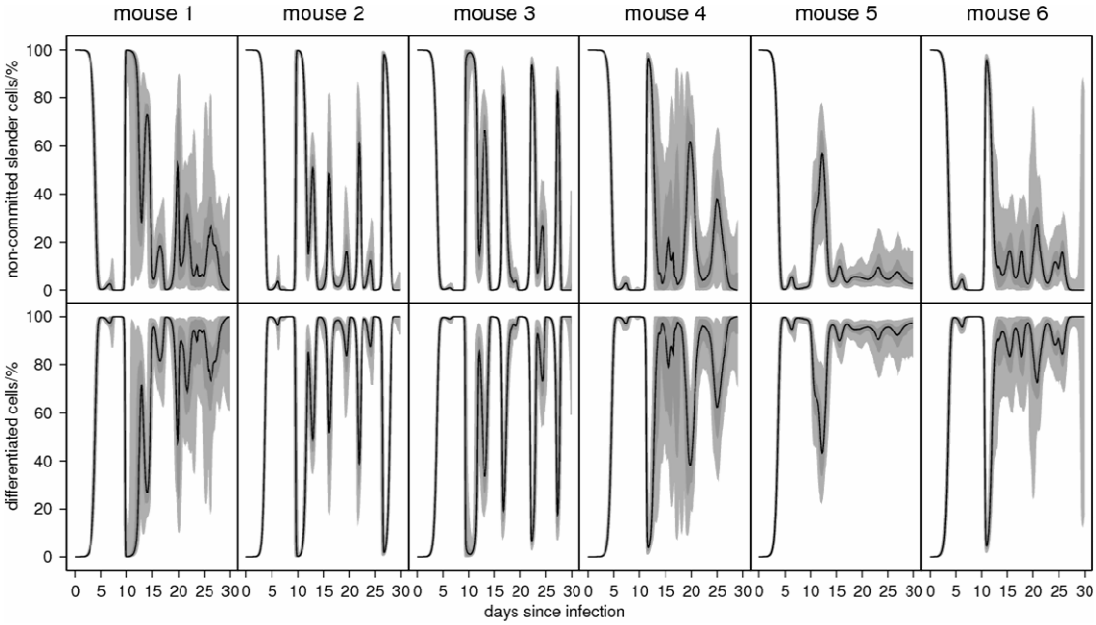


Figure 5.17 The predicted proportion of cell types throughout infection. The proportion cell types fluctuate during infection, but the proportion of non committed slender forms tends to be quite low throughout infection (upper panel). The proportion of differentiated cells tends to remain quite high throughout infection (lower panel).

5.7.4 The predicted dynamics of antigenic variation throughout infection

Trypanosomes have a large repertoire of VSGs which they can utilise to evade the host immune response throughout infection. As discussed in section 5.7.1, a simple model of antigen switching was used here, where switching between antigenic variants was sequential. Further, the parameter prior was set so that the model used the least number of variants possible to explain the data. The predicted dynamics of antigenic variation from the model are shown in Figure 5.18A. In all cases, the first peak of parasitaemia was predicted to be composed of the infecting variant. This correlates with the experimental observation in section 5.6.4 that the majority of cells in that first peak of infection were expressing the infecting AnTat1.1 VSG. However, the model does seem to underestimate the switch rate from the first to second variant. According to the model only approximately 1 in 10^6 cells should have switched during the first peak of parasitaemia (Figure 5.18A) and it would therefore be unlikely to observe as many cells not expressing the AnTat1.1 VSG when counting only 250 cells per infection per day (section 5.6.4). After this first peak, multiple variants were expressed and simultaneous expression was observed.

The median predicted number of variants used by each mouse ranged from 8 to 12 variants over the course of the 30 day infection with a mean of 10.5 ± 0.7, (section 5.7.7, Figure 5.21). The actual number of variants used is not known. It was hypothesised in section 5.6.2, that the maintenance of a low number of slender cells during the chronic stage of infection may limit the number of variants used and therefore prolong antigen chronicity. Indeed, for mouse 5 and 6, where the parasitaemia, *PAD1* expression and predicted proportion of differentiated cells (Figure 5.17) remained mostly high throughout infection only 8 variants were predicted to be used. For mouse 2, which showed the largest fluctuations in predicted SIF concentration (Figure 5.18) and more slender forms throughout infection than mice 5 and 6 (Figure 5.17), approximately 12 variants were predicted to be used. This correlates with the hypothesis that a low number of slender cells and high number of transmissible forms throughout an infection reduces antigen usage.

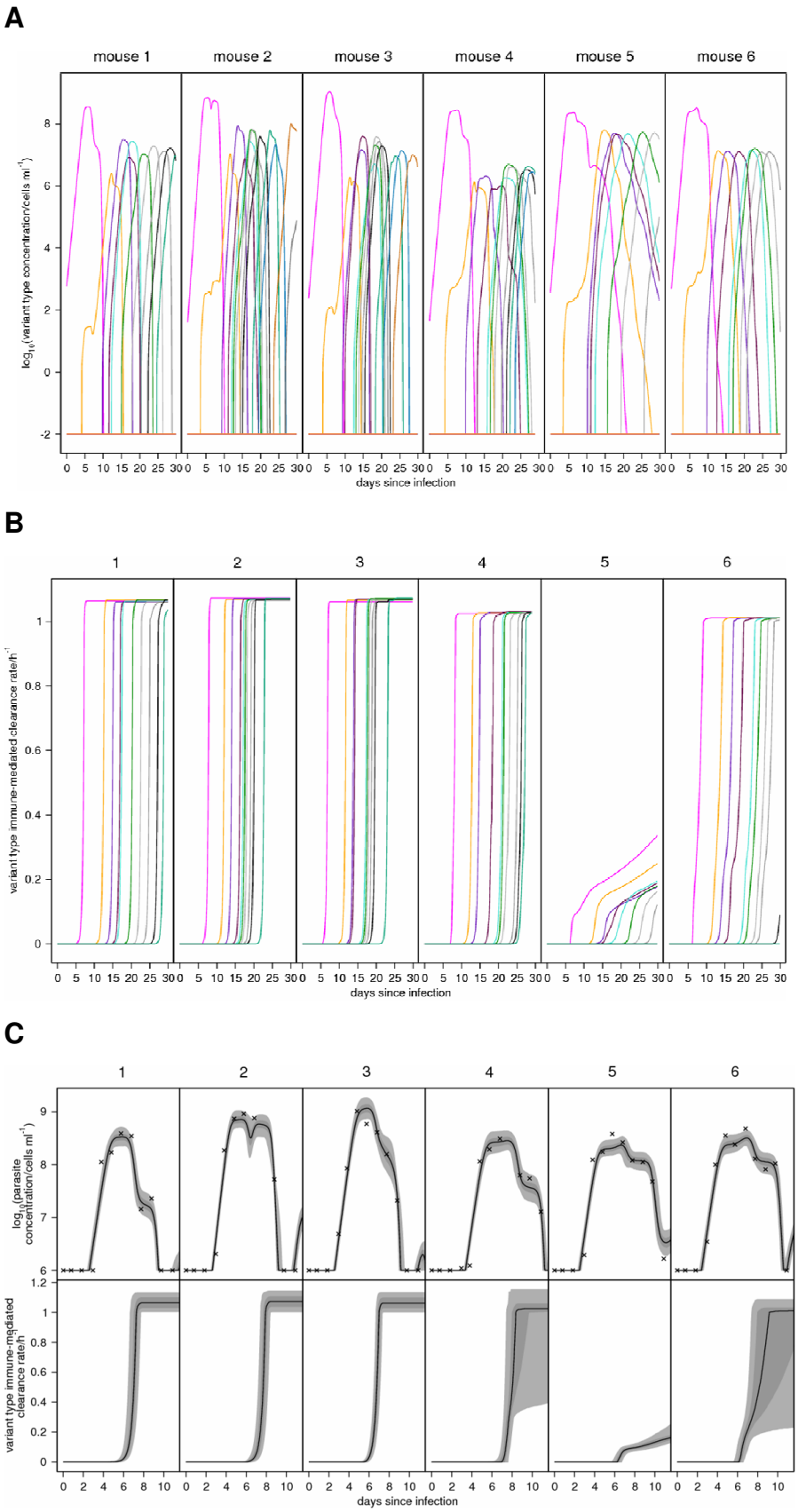


Figure 5.18 continued on the following page

Figure 5.18 The dynamics of antigenic variation and immune clearance as predicted by the model. (A) The predicted dynamics of different antigen variant types arising through infection. The first wave of parasitaemia (approx days 4-11) is composed primarily of the infecting variant. Later, antigen switching does occur, resulting in the expression of multiple variants at once. (B) The predicted immune mediated clearance rates of each variant expressed. As defined by the model, there is a delay in the onset of the immune response to the first variant but not to subsequent variants. Mouse 5 has a reduced immune response compared to the other mice. (C) The predicted immune mediated clearance rate of the first variant compared to the predicted and measured dynamics of parasitaemia. Immune clearance is not predicted to rise until after the initial fall in parasitaemia demonstrating that the model predicts this fall was due to SIF-induced differentiation rather than immune clearance (see section 5.7.3).

There is a variant-specific immune response that arises in response to the presence of each variant (Figure 5.18B). The model specifies that there is a delay in the response to the infecting variant but that, for the purposes of the production of a simple model, the response to later variants is instantaneous. The estimated mean duration before activation if the initial immune response was 142.2 ± 6.9 hours, or approximately 6 days, (section 5.7.7, Figure 5.21) which is not hugely dissimilar to a previous estimation of 122 hours (or approximately 5 days) from infections in a different strain of mouse (Tyler et al., 2001). In all cases, except mouse 5, the immune-mediated clearance of each variant, after initial activation, reached a maximum rate of clearance rapidly. For mouse 5, it was predicted that there was a decreased capacity for immune clearance (Figure 5.18B), which may reflect the health of the mouse. This predicted reduced immune response contributes, in the model, to the maintained high parasitaemia throughout infection in mouse 5 (Figure 5.16).

By comparing the parasitaemia during the first peak of infection to the onset of the immune response, it can be observed that the initial drop in parasitaemia occurs before the clearance rate of the first variant reaches it maximum rate (Figure 5.18C). This correlates with the initial fall in parasitaemia being caused by stumpy formation in response to SIF (section 5.7.3).

5.7.5 The predicted dynamics of *PAD1* mRNA expression per cell during differentiation

In the model, cells begin to express *PAD1* mRNA when they commit to differentiation and *PAD1* expression rises for a certain time, that time being a parameter of the model. Then *PAD1* expression can remain constant or rise or fall linearly from that time until stumpy cell death. Thus, in this model, committed slender cells were able to express *PAD1*. Indeed the model predicted that *PAD1* mRNA levels rose in the cell for approximately 14-15 hours (mean from six mice = 14.4 hours ± 1.5 s.e.m) (Figure 5.19) and that the lifespan of a committed slender cell, before becoming the cell cycle arrested intermediate form, is also approximately 14-15 hours (mean from six mice = 14.6 hours ± 0.7 s.e.m) (section 5.7.7). Thus, the model predicted that there was a rise in *PAD1* mRNA expression as a cell moved from being non-committed slender cell to a committed slender cell and that this rise continued until cells became intermediate. After the initial rise in *PAD1* mRNA expression there was no evidence in the data for any further change in expression until stumpy cell death.

The model appears to suggest a large amount of variability between mice; however, there are large predictive intervals for all the mice, which for the most part overlap (Figure 5.19). It is expected that there will be measurement errors in both the values obtained for parasitaemia and *PAD1* expression. To estimate the relative *PAD1* expression per cell, both parasitaemia and *PAD1* expression are taken into account and as such the measurement errors from both data sets will be reflected in the resulting predictions.

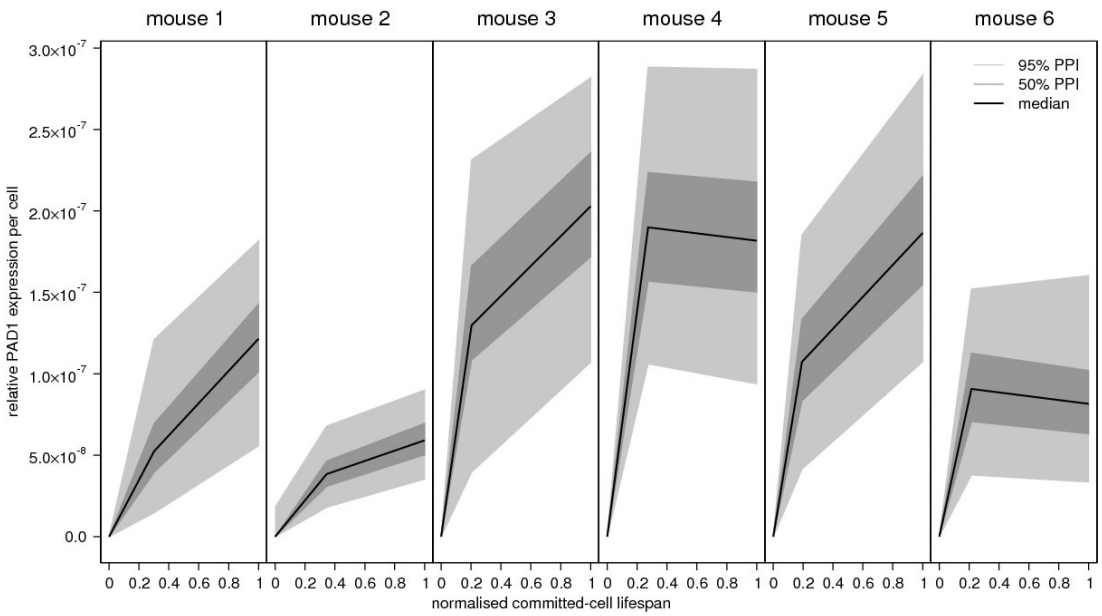


Figure 5.19 Predicted *PAD1* expression of differentiating cells. The predicted dynamics of the rise in *PAD1* mRNA expression during differentiation in individual cells is shown. The 50% and 95% posterior predictive intervals are shown in dark and light grey, respectively.

A model where committed slender cells were not able to produce *PAD1* mRNA was also generated, but this produced a less good fit of the model to the data (data not shown). This implies that cells that are differentiating, but are still slender in morphology, express at least a low level of *PAD1* mRNA. At the protein level, PAD1 is not expressed in cells which are morphologically slender and, as determined by northern blot, *ex vivo* populations which are predominantly slender have a very low level of *PAD1* mRNA expression compared to predominantly stumpy populations (Dean et al., 2009). In Chapter 3 it was shown that *PAD1* expression through the action of the 3'UTR is regulated at both the level of mRNA abundance and protein expression. It may be that this initial rise in *PAD1* mRNA in cells which are in the very early stages of differentiation, but are still morphologically slender, is not reflected at the protein level and only upon transition to the intermediate life stage is the protein expressed. Indeed, nuclear encoded mitochondrial genes, that are up-regulated in stumpy forms compared to slender forms, are up-regulated at the mRNA level before morphological change, but not at the protein level (Tyler et al., 1997). Thus, it may be common amongst stumpy-specific genes that the mRNA is up-regulated in response to commitment, but that these gene remains translationally repressed until cell cycle arrest and morphological change.

5.7.6 Quantitative morphological data could not be incorporated into the model

In section 5.6.3 data was collected on the morphology of the cells during the first peak of parasitaemia. It was thought that perhaps this data could be incorporated into the model as had been done for parasitaemia and *PAD1* expression. However, the predicted morphology, as determined by the model, did not fit well to the observed morphology, as determined by microscopy, even when intermediate and stumpy forms were combined (Figure 5.20). Firstly, the model predicts that the proportion of slender forms should fall, and the number of intermediate and stumpy form should rise, more rapidly than was observed by microscopy. Secondly, in section 5.7.3, it was shown that as parasitaemia falls, after the first peak, the model predicts that there is a second peak, or decrease in the rate of fall of parasitaemia (Figure 5.16). It was proposed that this was due to a small rise in the number of slender cells as the SIF concentration fell. Indeed, the model predicts that that there is a small rise in the proportion of non-committed and committed slender forms in the population after the initial fall. The measured data for mouse 1 did seem to correlate with a small rise in slender forms after the initial fall but there did not seem to be evidence for this in the other five mice (Figure 5.20).

The morphological data may not fit the model predictions for a number of reasons. Firstly, the model states that there are four distinct cell types (non-committed slender, committed slender, intermediate and stumpy). The committed slender form is proliferative and produces SIF, but has begun the process of differentiation and produces *PAD1* mRNA (section 5.7.5). Intermediate forms are those cells that have exited the cell cycle and are fully expressing *PAD1* mRNA but are not yet fully stumpy. The transition is a complex differentiation event that comprises changes in gene expression, metabolism, morphology and cell cycle status. Rather than the existence of 4 distinct groups of cells, in reality the differentiation process is continuous and changes in cell cycle status, morphology and gene expression will very likely not correlate exactly.

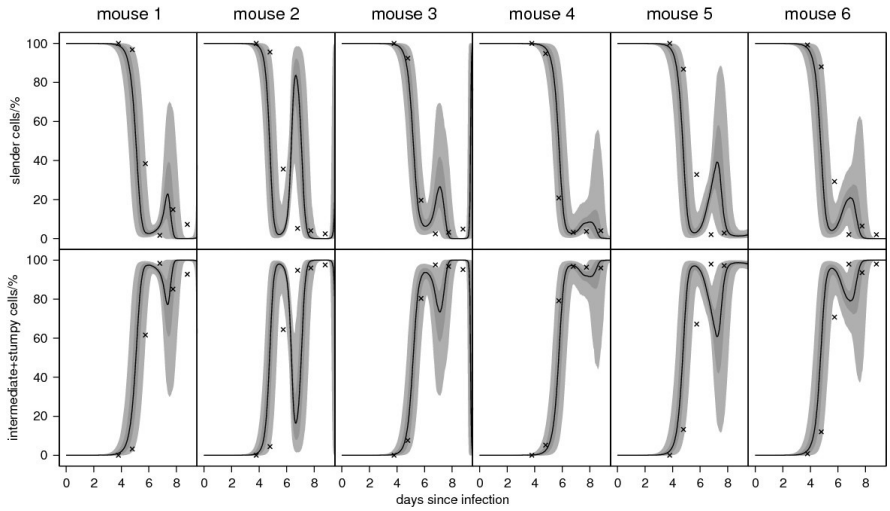


Figure 5.20 Comparison of the predicted and measured morphological data during the first peak of parasitaemia. The measured morphology (crosses) compared to the predicted morphology (lines) according to the model is shown. The 50% and 95% posterior predictive intervals are shown in dark and light grey, respectively. As non-committed and committed slender cells cannot be distinguished by microscopy the upper panel represents all slender cells. Due to the subjective nature of determining morphology by microscopy, the intermediate and stumpy forms were grouped to reduce the potential error arising from this subjectivity. A small rise in the proportion of slender forms in the population was predicted to occur in all mice after the initial fall.

Furthermore, the assessment of morphology by microscopy is subjective: this has been one of the difficulties in studying stumpy formation. It may be that very early intermediate forms could be mistakenly scored as slender in morphology when very little change has occurred so far. This would account for the delay in the loss of slender forms in the measured data compared to the predicted data. From the morphology and cell cycle data collected in section 5.6.3, on day 4, where the population was scored as predominantly slender, 74.0% ± 1.2 s.e.m of the slender population in the six mice were 1K1N. On day 6, when the population was predominantly scored as intermediate 80.5% ± 1.5 s.e.m of the slender population were 1K1N. This implies that either the cell cycle is not in equilibrium, which certainly may be the case given the complex infection dynamics, or presuming the cell cycle is in equilibrium, that some of the cells which were scored as slender were actually cell cycle arrested. These may represent very early intermediate forms and this may account for the differences in the predicted and measured change in morphology through time. This also suggests that cells that are intermediate in

morphology are indeed cell cycle arrested and it is not incorrect to score proliferative cells as slender in morphology (as in section 5.6.3). It is worth noting that this does not account for the expression of *PAD1* mRNA in committed slender forms (as discussed in section 5.7.5) as this morphological data was not incorporated into the model given the poor fit.

5.7.7 Parameter estimation

The model contains numerous parameters, which, when the model is produced, are either unknown or can be estimated based on prior knowledge (Appendix C). When the model was fitted to the data for each mouse these parameters were varied, until a set of values for each parameter were created that best explained the data. Thus, the model can be used to estimate the values of unknown parameters for which data was not collected (Figure 5.21 and Table 5.1). In cases where the estimated parameter distribution includes zero, there is not sufficient information in the data to provide information regarding this parameter.

Firstly, the duration of the slender cell cycle was estimated, by the model, to be approximately 5 hours (Figure 5.21 and Table 5.1). Replication rates vary between strains (Turner et al., 1995) and estimations have varied from as low as 1.7 hours to as much as 9.6 hours, with most studies converging on a cell cycle duration of approximately 4 to 6 hours (Savill and Seed, 2004; Seed and Black, 1997; Seed and Black, 1999; Turner et al., 1995; Vassella et al., 1997). As the estimations of the mean cell cycle duration from the six mice in this case range from 4.0 – 5.5 hours, this correlates well with these other, similar, studies.

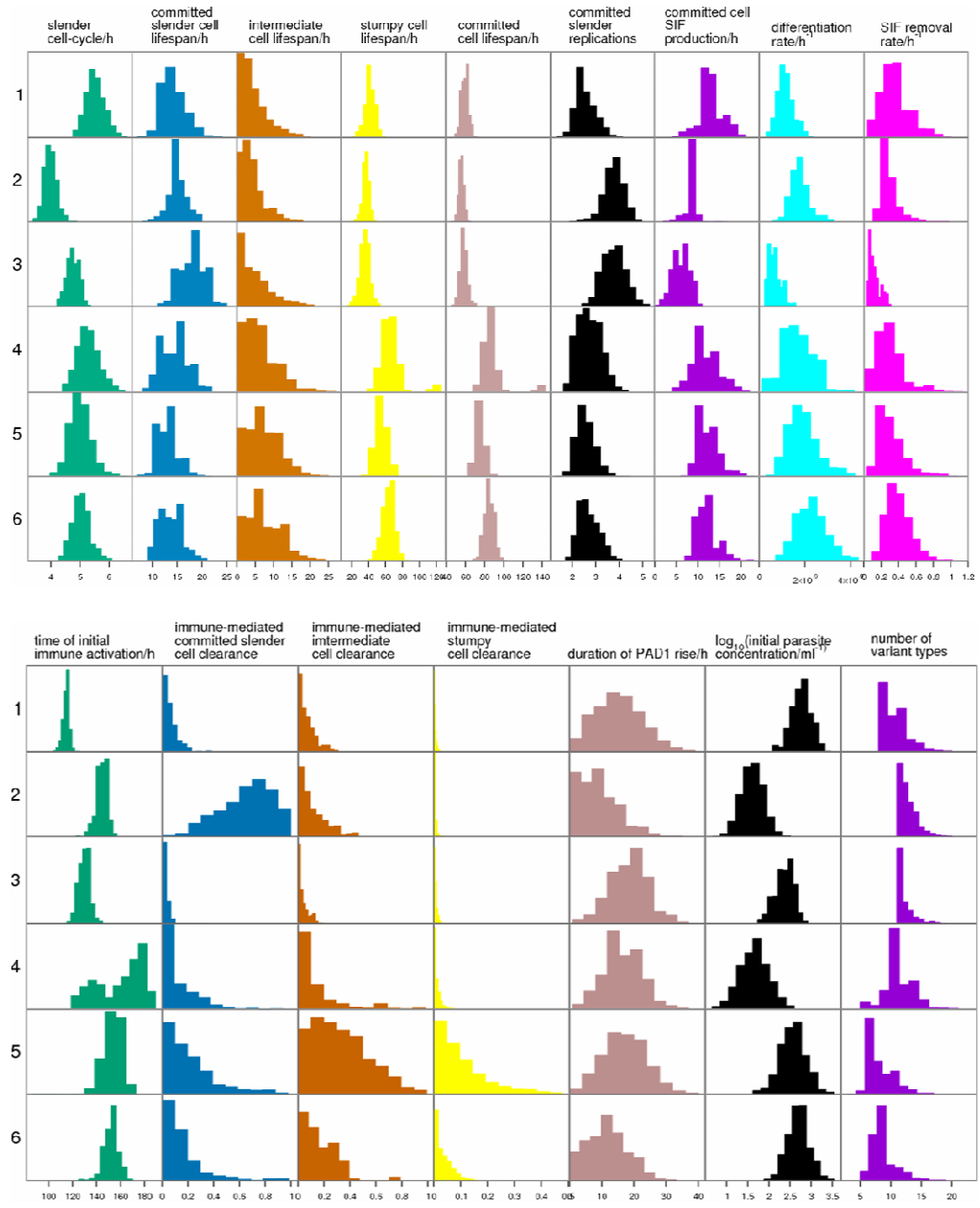


Figure 5.21 Predicted parameter distributions as determined by the model. When the model is fitted to the data for each mouse the model parameters are varied, until a set of values for each mouse for each parameter are created which best explain that data. Each graph represents a probability distribution and as such, the area under each curve = 1. Parameters can be interdependent. In cases where the estimated parameter distribution includes zero, there is not sufficient information in the data to estimate the values of that parameter. Immune-mediated clearance rate has a minimum value of 0 and maximum value of 1. Details of each parameter are discussed in the text.

Parameter	Mean (± s.e.m)
Slender cell cycle	4.9 ± 0.2 hours
Committed slender cell lifespan	14.6 ± 0.7 hours
Intermediate cell lifespan	5.6 ± 0.6 hours
Stumpy cell lifespan	49.9 ± 5.7 hours
Total committed cell lifespan	70.0 ± 5.7 hours
Committed slender replications	3.0 ± 0.2
Committed cell SIF production	10.7 ± 1.2 hours
Differentiation rate	1.6 E-09 ± 2.3 E-10 /hour
SIF removal rate	0.30 ± 0.04 /hour
Time of initial immune activation	142.5 ± 6.9 hours
Duration of *PAD1* rise	14.4 ± 1.5 hours
log10 initial parasite concentration	2.3 ± 0.2 /ml
Number of variant types	10.5 ± 0.7

Table 5.1 Mean parameter estimates from all six mice. The mean parameter values from all six mice, as calculated from the mean parameter values of individual mice.

The lifespan of a cell from the point of commitment to differentiate until stumpy cell death was estimated to be approximately 70 hours (Figure 5.21 and Table 5.1). This duration encompasses the cumulative life spans of the committed slender, intermediate and stumpy forms. The committed slender forms are replicating and are therefore clearly distinguished from intermediate forms. These committed slender cells were predicted to have a lifespan of approximately 14.6 hours, correlating with 3 further replications after commitment to differentiation (Figure 5.21 and Table 5.1). After commitment to differentiate, the level of *PAD1* mRNA per cell rises for approximately 14.4 hours, correlating with the lifespan of a committed cell (as discussed in section 5.7.5). Furthermore, upon commitment to differentiate, cells continued to release SIF for approximately 10 hours (Figure 5.21 and Table 5.1). Thus, the parameter estimations suggest that a committed slender cell receives the signal to differentiate then begins to express *PAD1* mRNA while continuing to proliferate for 3 further cell divisions and to produce SIF during the first two of those divisions, before undergoing cell cycle arrest and therefore transition into the intermediate life-stage. Based on morphological data (discussed in section 5.7.6), it

would not be expected for any morphological change to begin before cell cycle arrest.

The intermediate and stumpy forms are both cell cycle arrested and are therefore only distinguished, in this model, by morphology, for which no data has been incorporated. Therefore, although the model predicted that the stumpy lifespan is approximately 50 hours and the intermediate lifespan is 5 hours (Figure 5.21 and Table 5.1), this may not be accurate. The predicted distribution for the intermediate lifespan includes zero, which, as mentioned above, indicates that there was not enough information in the data to determine this parameter. Without additional data to input into the model, the lifespan of these two life-stages cannot be accurately disentangled. However, the combined predicted lifespan of approximately 55 hours does provide an estimate for the duration from cell cycle arrest to stumpy cell death.

From analysis of the parameter distributions of individual mice (Figure 5.21), it appears that the total lifespan of a committed cell was lower in mouse 1-3 than in mouse 4-6, where the mean from the three mice in each group is 37.8 ± 1.9 hours and 61.8 ± 3.5 hours respectively. It is not clear why there should be such a clear difference in stumpy lifespan between two groups of the experimental infections. However, mice 1-3 were housed in the same cage and mice 4-5 were housed in another (although all mice originated from the same litter and were infected in a random order and were distributed between the two cages at random). It could be speculated that some unknown difference between the conditions in the two cages had some effect on the within host dynamics. This is, however, merely speculation and the reason for this difference remains unknown at this time.

The model states that SIF is lost from the system at an estimated rate of 30% per hour (Figure 5.21 and Table 5.1). Initially, a model was produced where loss of SIF was dependent on background loss as well as at a slender concentration dependant rate, representing the internalisation of SIF by slender cells. The loss of SIF by internalisation did not however have any discernable effect on the model and so this function was removed from the final model described here. This is not to say that SIF

is not internalised by slender cells, but that this is a minor contributor to the overall removal of SIF from the system. It may be that SIF is cleared from the blood of the host by host-mechanisms or that SIF is able to enter other cell types and as such, the density of slender cells has little effect on the total loss. Without knowing the identity of SIF it is not possible to critically assess the model predictions.

The parameters associated with antigenic variation, and the host immune response, were discussed previously in section 5.7.4. Briefly, the model predicts that the time taken for the activation of the immune response against the infecting variant is 142.2 ± 6.9 hours. The number of variants used during this 30 day infection was predicted to be 10.5 ± 0.7 with variability between infections (Figure 5.21 and Table 5.1).

When the model was produced, the immune response against each class of differentiated cells was set to be proportional, but different, to the response against non-committed slender cells. With one exception, the parameter estimates for the immune response for all three classes of differentiated cell type included zero (Figure 5.21). Thus, given this model, the data does not contain evidence for immune clearance of differentiated cells. This does not indicate that immune clearance does not occur, but rather that it is not a significant contributor to the loss of each cell type. Thus, for committed slender forms and intermediate forms, loss of these cell types is primarily by differentiation to the next cell type. For stumpy forms, loss is predicted to be primarily by non-immune mediated cell death (i.e. cell senescence).

However, for mouse 2, there was a predicted immune response against the committed slender forms (Figure 5.21). It is not clear why this is the case but perhaps there is a fine balance between clearance via differentiation and clearance by the immune response within the host. Unlike the other five infections, the mouse 2 infection was predicted to show a lesser reduction in parasitaemia in response to cell density, with the parasitaemia only falling dramatically after the onset of the immune response (section 5.7.3). Thus, where loss of committed slender cells by immune clearance plays only a small role in the other mice, it may be a significant contributor during the first wave of infection in mouse 2. However, as previously mentioned,

there is no qRT-PCR data for day 8 of the mouse 2 infection and microscopy data suggest that the predicted parasitaemia on this day may not be accurate (section 5.7.3). Hence, the rise in immune mediated clearance of committed slender cells in mouse 2 may be an artefact of this error in the model.

Finally, the model had a parameter prior of the $\log_{10}$ initial concentration of cells in the blood set to N(3, 0.5) as approximately 1000 trypanosomes were injected I.P. into each mouse, and the volume of blood per mouse was estimated to be approximately 1ml. The model predicts that, on average, only 190 cells/ml were infected per mouse (Figure 5.21 and Table 5.1), with large variability between mice, ranging from as little as 41 to as high as 590 cells/ml. Indeed, the actual volume of blood in the mouse will not be 1ml which will cause the infecting concentration to vary from the predicted value. Also, as the cells are infected I.P. many of the cells may not successfully enter and establish infection in the bloodstream.

5.7.8 Induction of stumpy formation is better explained by SIF than parasite density

Having established a mathematical model which was a good fit to the quantitative biological data, this model could then be used to test hypotheses regarding the within-host dynamics. For instance, various models previously have used density-dependent differentiation to account for stumpy formation rather than SIF-induced differentiation (Lythgoe et al., 2007; Turner et al., 1995; Tyler et al., 2001). Thus, Dr Savill removed SIF-induced differentiation from the model and replaced it by density-induced differentiation (see Appendix C). The parameters were then optimised for this new model. The adequacy of the model to explain the data was assessed by comparing the mean standardised residuals of each time point from each mouse to its 95% expected interval (Figure 5.22A), as had been carried out for the SIF-induced differentiation model (section 5.7.2).

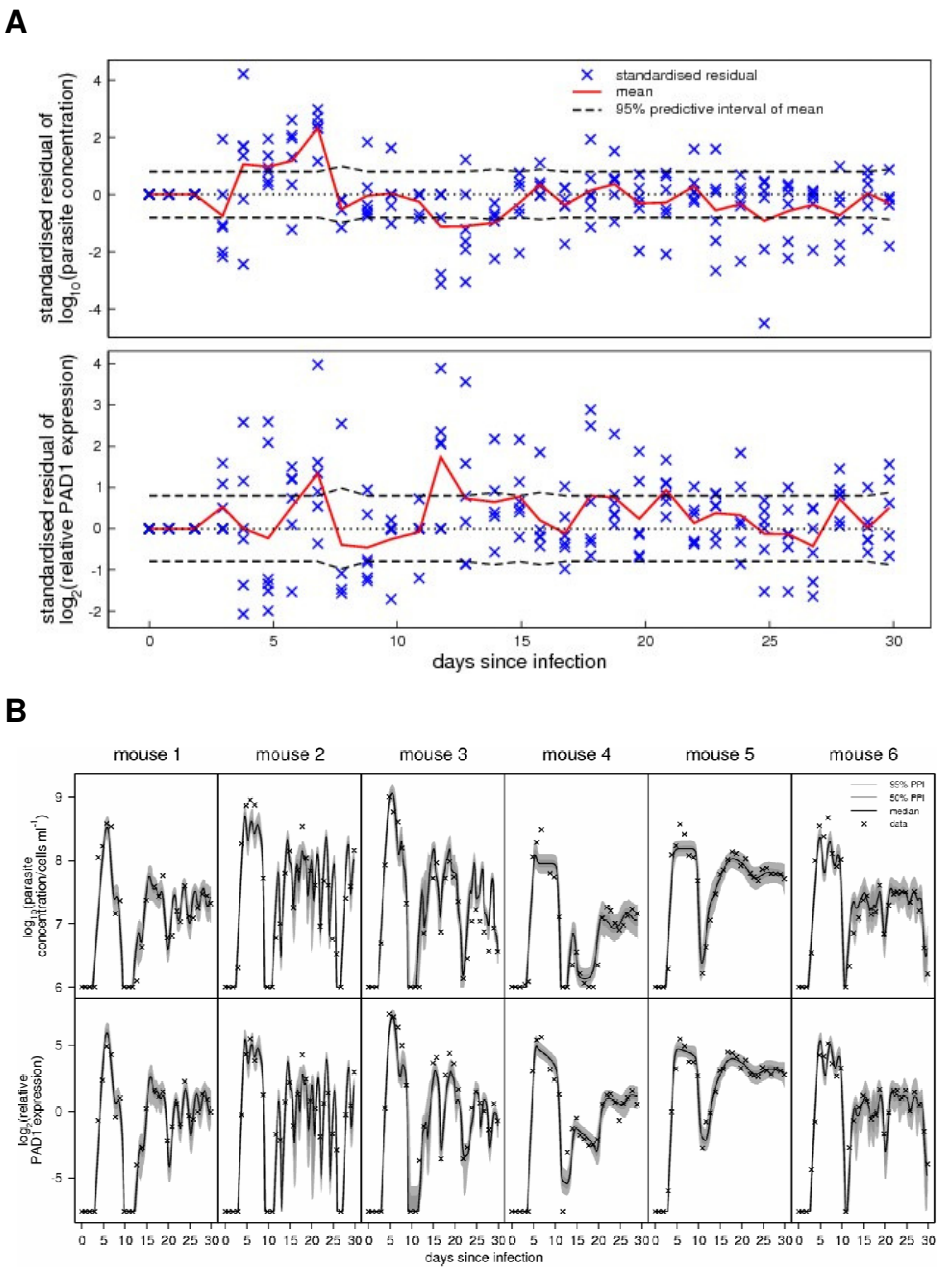


Figure 5.22 Density-dependent differentiation does not provide a good fit to the data. (A) The standardised residuals of each data point from each mouse (blue crosses) for both parasitaemia (upper panel) and *PAD1* expression (lower panel). The mean of the standardised residuals from the six mice is shown as a solid red line. This mean lies out with the expected 95% interval on 8 days for the parasitaemia and 3 days for *PAD1* expression and there are numerous outliers in this model. (B) The measured and predicted infection dynamics for parasitaemia (crosses and lines, respectively) according to the model is shown in upper panels. The measured $\log_2$ *PAD1* expression (crosses) compared to the predicted $\log_2$ *PAD1* expression (lines) is shown in lower panels. In all cases the 50% and 95% posterior predictive intervals are shown in dark and light grey, respectively.

As can be seen in Figure 5.22A the mean of each time point lies outwith the 95% interval and there are numerous outliers, where the standardised residual is more than about 3 standard deviations from zero, at numerous days for both parasitaemia and *PAD1* expression, suggesting a poor fit of the model to the data. Furthermore, analysis of the fit of the model to the data in individual mice (Figure 5.22B), reveals numerous data points lying outwith the model predictions. Comparing the likelihood of each of the two models, reveals that the SIF-induced differentiation model is 10^{68} times more likely than the density-dependent differentiation model (Appendix C). As such, the data provides huge support for SIF-induced differentiation *in vivo*.

5.8 Discussion

5.8.1 A method for the quantitative analysis of parasitaemia and stumpy formation *in vivo* has been established

During a chronic trypanosome infection, the level of parasitaemia fluctuates but is maintained below a lethal threshold. These fluctuations and the maintenance of a sub-lethal parasitaemia are controlled by both the slender to stumpy transition and antigenic variation. Furthermore, given that the stumpy forms are the transmissible forms, stumpy formation is important for transmission and fluctuations in stumpy formation could impact on transmission capacity, especially given that tsetse bites can be rare (Baylis, 1997). However, the within-host dynamics of the slender to stumpy transition have not been well characterised previously, largely due to the difficulty in quantifying stumpy formation. Studies have been carried out using morphology to distinguish between cell types, but this can be subjective and is time-consuming. Furthermore, morphology may not be an accurate way to distinguish differentiating forms (section 5.8.7).

Here, a qRT-PCR method was established whereby RNA was extracted from a starting volume of 10µl of whole blood from an infected mouse and used to analyse parasitaemia and relative *PAD1* expression. Parasitaemia was measured by monitoring *Tb*ZFP3 transcript expression and comparing to a standard curve (section 5.4). *Tb*ZFP3 was known to be a constructively expressed transcript (Paterou et al.,

2006), and it was confirmed in section 5.4.2 that transcript expression is indeed equal in slender and stumpy forms. The limit of detection of parasitaemia was 10^6 cells/ml, due to the limited starting volume and the dilution steps required along the processing pathway which was satisfactory for mouse infections where the parasitaemia is predominantly higher than 10^6 cells/ml but would need improved for analysis of infection in other hosts, such as cattle, which would be relevant in the field (section 5.8.9).

Stumpy formation was monitored via the expression of *PAD1* mRNA which is up-regulated in stumpy forms compared to slender forms and is a functionally relevant marker: PAD1 protein expression correlates with capacity to differentiate to the procyclic life stage (Dean et al., 2009). Expression was normalised to parasitaemia and the changes in *PAD1* expression over time therefore represent a change in the *PAD1* expression of the population.

This method has various future uses, as will be discussed in section 5.8.10. Here, the utilisation of this method allowed parasitaemia and *PAD1* expression to be monitored during the infection of six mice with *T.b. brucei* AnTat1.1 over the course of a 30 day period. As well as analysis of infection dynamics by qRT-PCR, traditional analyses of morphology and cell cycle status were also performed for the first peak of parasitaemia over 4-9dpi. The combined analyses allowed the temporal order of events during early infection to be dissected (section 5.8.3).

5.8.2 Quantitative data has been incorporated into a mathematical model for stumpy formation

Having generated a large amount of quantitative data in section 5.6, this data was incorporated into an existing model of stumpy formation (section 5.7). Mathematical models are useful tools for the analysis of complex within host dynamics: they allow unknown parameters to be estimated and once generated, allow the impact of changes, such as introduction of a drug or presence of a competing strain, to be predicted (section 5.8.10). The model is described in detail in section 5.7.1.

The model was fitted to the data that had been generated in section 5.6 for both parasitaemia and *PAD1* expression. The raw data input into the model was the calculated parasitaemia, as was used in section 5.6, and the total *PAD1* expression of the population, before normalisation to parasitaemia. By analysing parasitaemia and *PAD1* expression every day for 30 days during infection in six mice, this produced over 350 data points to which the model could be fitted. The model provided a good fit the experimental data on all days except day 4, where the predicted parasitaemia was lower than the actual parasitaemia in five of the mice (section 5.7.2). This implies that there is some factor missing or incorrect in the model during this first peak. An underestimation of parasitaemia at this early time point may imply that the duration of the slender cell cycle was overestimated at the early stage of infection. Perhaps the rate of proliferation is faster during the first days of infection compared to later in infection. Overall, however, this model does provide a very good fit to the data and as such could be used to estimate associated parameters and make predictions about the system.

5.8.3 The order of changes in the trypanosome population early in infection

By comparing traditional morphological and cell cycle analysis with qRT-PCR analysis for parasitaemia and *PAD1* expression the temporal order of events during differentiation from slender to stumpy forms *in vivo*, at the population level, were monitored. This can only be analysed during the first peak of parasitaemia where the population is originally composed of almost entirely slender cells so the process is somewhat synchronised and also no SIF has yet accumulated. The differentiation process is not simply a morphological switch but rather a series of events that occur in response to a parasite-derived signalling factor, SIF, including commitment to differentiate in response to SIF, appropriate changes in gene expression, cell cycle status, metabolism and morphology.

A summary of the temporal order of events in the population during the first peak of parasitaemia in the six *T.b brucei* mouse infections, described in detail in section 5.6,

is shown in Figure 5.23. The changes that occurred during the first peak of parasitaemia, including parasitaemia, morphology, cell cycle status and *PAD1* expression was similar in all six infections suggesting that these are a highly co-ordinated series of events.

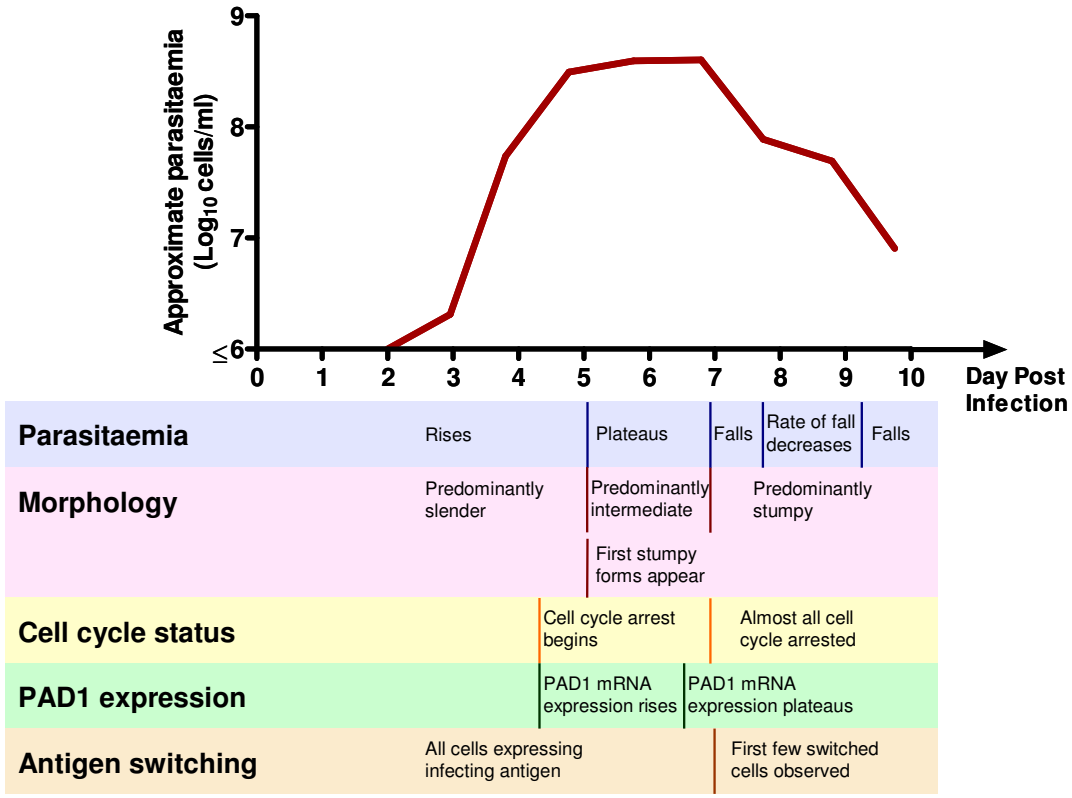


Figure 5.23 The order of events of the population during the first peak of parasitaemia. During the first wave of parasitaemia, the parasitaemia rises as slender forms predominate in the infection. As cell cycle arrest begins, the parasitaemia plateaus, the population becomes predominantly intermediate and the first stumpy forms appear. *PAD1* expression rises with the appearance of differentiating cells and then plateaus for the remainder of the first peak of the infection. The population becomes predominantly stumpy in morphology, almost all of the population is cell cycle arrested and the parasitaemia begins to fall as the first few cells which have switched their VSG coat begin to appear.

5.8.4 The order of events during differentiation of an individual cell

By combining the empirical data and modelling data generated here, the temporal order of events during stumpy formation in an individual cell can be mapped (Figure 5.24) and the timing of these events can be predicted. Tyler *et al.,* established that commitment and division occurred before morphological change, and that nuclear-encoded mitochondrial genes that were up-regulated in stumpy forms compared to slender forms showed up-regulation of the mRNA, but not protein, before morphological change (Tyler et al., 1997). From the analyses in this chapter, it also appears that for *PAD1*, a non-mitochondrial stumpy specific gene, there is also up-regulation of mRNA expression prior to morphological change. The PAD1 protein is not up-regulated until after morphological change (Dean et al., 2009). Further, the model generated in section 5.7 further confirms that commitment to differentiation precedes cell cycle arrest and predicts that a committed slender cell undergoes approximately 3 further divisions after the point of commitment such that the committed slender lifespan if approximately 14-15 hours. The model also predicts that the committed slender cell will continue to produce SIF for approximately 10 hours, or two cell divisions, after the point of commitment (section 5.7.7). The differentiating cell will then undergo cell cycle arrest prior to the detectable change in morphology, although the process of changing morphology may begin at the same point in time. These cell cycle arrested forms are termed intermediates and they do express at least some stumpy-specific proteins. After morphological change is complete the cell is considered fully stumpy. From the point of commitment to stumpy cell death is estimated to be approximately 70 hours, with the time since cell cycle arrest estimated at approximately 55 hours.

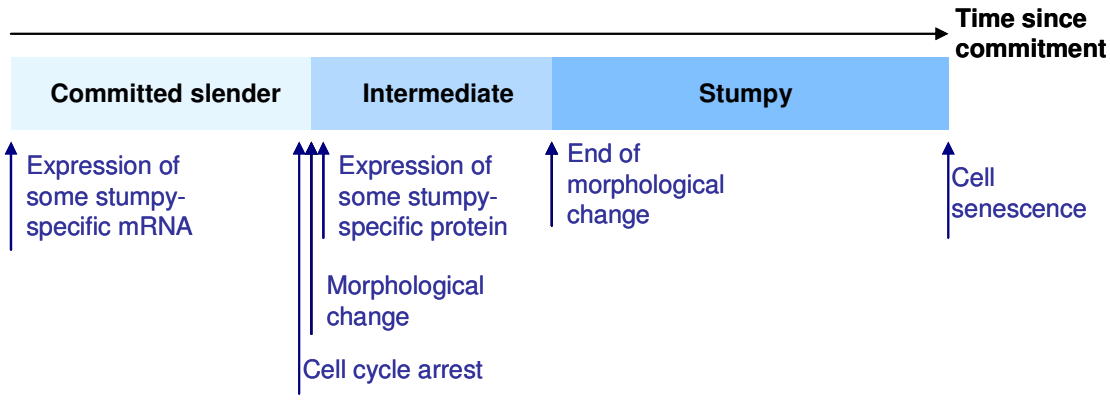


Figure 5.24 The order of events in an individual cell during stumpy formation. After the point of commitment the committed slender cell begins to express some stumpy-specific mRNA but will undergo further divisions before cell cycle arrest. After cell cycle arrest there is a detectable change in cell morphology such that the cell is described as intermediate and at least some stumpy-specific proteins become expressed. At the end of morphological change the cell is fully stumpy. The estimated time from the point of commitment to stumpy cell death is approximately 70 hours, with cell cycle arrest at approximately 14-15 hours.

5.8.5 The proportion of transmission stages is high throughout infection

At the initial stage of infection *PAD1* mRNA levels were very low and rose with the first peak of infection. This correlates with a rise in the number of cells in the population differentiating to the transmissible stumpy form. When parasitaemia fell after the first peak of infection, relative *PAD1* mRNA expression of the population remained high and although it did fluctuate with parasitaemia throughout infection, it did not fall below detection at any point after the first peak. Indeed, even when parasitaemia was low the *PAD1* mRNA expression relative to the number of cells in the population remained high. This indicates that even at low parasitaemia, the number of differentiating or differentiated form remained high. Further, by mathematical modelling, it was predicted that the proportion of non-committed slender forms was generally low (section 5.7.3). Although increases in these forms were observed, these were generally short-lived. This low proportion of non-committed slender forms and high proportion of differentiated forms has two key implications in infection dynamics: firstly, the proportion of transmission competent cells remains high throughout infection and secondly, as only slender cells can

functionally switch the VSG coat this may reduce the number of variants used in a given period, thus prolonging antigen variant chronicity.

Antigenic variation has been shown to display a degree of order or hierarchy that is caused by a combination of differential activation rates and the slender to stumpy transition, with multiple variants being expressed simultaneously (Lythgoe et al., 2007; Morrison et al., 2005). Only slender cells are able to functionally switch to the next variant, as stumpy forms are cell cycle arrested and so will not be able to populate the next wave of parasitaemia. Therefore, if there are fewer slender cells in the population, there will be fewer cells able to switch to a new variant and as such fewer variants will be used in a given time frame. Thus, maintaining a low proportion of slender cells could prolong antigen chronicity by extending the useful life of the VSG repertoire.

Trypanosome infections can last for months to years and as such, reservoirs, such as infected humans and livestock, will play a crucial role in the epidemiology of both human and cattle disease. Although a single trypanosome is capable of causing a tsetse infection (Maudlin and Welburn, 1989), it has been shown that cattle infections can have intermittent periods where they are not infective to the tsetse vector (Van den Bossche et al., 2005). The transmission potential of an infection could be impacted by both total parasitaemia and the proportion of transmissible stages in the population. For instance, if the parasitaemia is so low that no trypanosomes are taken into a tsetse blood meal then the infection is not transmissible at that point; similarly, if there are very few transmissible forms such that only slender forms are taken into a blood meal then this would also reduce transmission potential. When parasitaemia is high, however, the proportion of transmissible forms in the population does not seem to affect tsetse infectivity (Janelle et al., 2009), indicating that at high parasitaemia, such as in mouse models, there is less of a requirement for a high proportion of transmissible forms to be tsetse infective. Here, the proportion of stumpy cells in the population, as determined by microscopy, reached over 75% in the first peak of parasitaemia and then the number of differentiated cells was predicted by mathematical modelling to remain high

throughout the later stage of infection. Other studies have also observed close to 100% stumpy formation in the first week of *T.b brucei* infection (Balmer et al., 2009; Tyler et al., 2001; Tyler et al., 1997; Van Den Abbeele et al., 1999). Indeed, even at low parasitaemia, the proportion of differentiated cells remained high (section 5.6). This implies that these infections remain transmission competent throughout infection.

In contrast, a recent analysis of morphology in *T.b gambiense* mouse infections did not observe a high proportion of stumpy forms (Janelle et al., 2009). The parasitaemia rose to over 10^8 cells/ml in the first wave if infection and then, in most cases, fell as would be expected. The mice used were immunocompromised and therefore this fall in parasitaemia was only due to stumpy formation and not immune clearance, yet in all cases intermediate and stumpy forms represented less than 30% of the total population. Firstly, it may be that there is a considerable difference in the rate of stumpy formation between different sub-species. It is indeed known that growth rates vary between strains (Turner et al., 1995) which may impact on stumpy formation and different strains induce different host responses, which will impact on virulence and disease progression (Morrison et al., 2010). It seems very likely therefore that different sub-species and different strains will have differences within host dynamics and transmissibility. Alternatively, it may be that during stumpy formation in *T.b. gambiense,* morphological change occurs later in the differentiation pathway than in *T.b. brucei*, such that early intermediate cells are scored as slender forms for a longer period or the committed slender cell lifespan is longer, meaning that more morphologically slender forms are present but that many of these are already committed to differentiation. It would be interesting to use the qRT-PCR method established in this thesis to monitor stumpy formation with different strains and sub-species of trypanosome to address these differences (section 5.8.10).

5.8.6 The mathematical model supports SIF induced differentiation *in vivo*

Stumpy formation has been shown *in vitro* to be caused by the parasite-derived SIF (Vassella et al., 1997) which essentially reports on the population density. However, due to the lack of identification of SIF it has not been possible to characterise the role of SIF *in vivo*. Therefore, most mathematical models incorporating stumpy formation have used a density-dependent model whereby the rate of differentiation is proportional to the total number of cells in the population (Lythgoe et al., 2007; Turner et al., 1995; Tyler et al., 2001). Although SIF-induced differentiation will be greatly affected by population density, it will provide different dynamics than density alone. For example, in the density-induced model, as parasitaemia falls stumpy formation will also fall: on the contrary, in the SIF-induced model, SIF will remain in the system after parasitaemia falls and delay the fall in rate of stumpy formation. Here the model had been produced to include SIF-induced stumpy formation. To test if this provided a better or worse fit to the data, SIF-induced differentiation was replaced with density-induced differentiation and the adequacy of each model to explain the data was compared. The SIF-induced model produced a greatly better fit to the data than the density-dependant model, supporting SIF-induced differentiation *in vivo.* Without the identity of SIF, it is not possible to measure the concentration or rate of loss of SIF in the system and thus it is not possible to test the predictions made here. Given future characterisation of SIF, the model could be refined with empirical evidence regarding SIF dynamics.

5.8.7 Morphology is not an accurate tool for the analysis of stumpy formation

Differentiation between life stages involves a complex series of events in order to alter the phenotype of a cell to adapt to changing environmental conditions. In the case of the slender to stumpy differentiation the proliferative cell will receive the SIF signal, transmit that signal through the cell and respond by altering gene expression, cell cycle status, metabolism and, most obviously, morphology. Indeed, morphology is the way in which the different life-stages throughout the trypanosome life cycle

were identified and how the slender and stumpy stages were named (Vickerman, 1965a). The morphological change during stumpy formation is continuous, and as such, intermediate forms can be observed. Functionally these intermediate forms represent cells that are in the process of differentiation and may have already undergone many of the other changes associated with stumpy formation.

There are various problems with the use of morphology as a way to monitor stumpy formation. Firstly this method is subjective; the presence of morphologically intermediate forms complicates analysis by making it difficult to distinguish between cell types. Furthermore, as discussed in section 5.8.4, commitment to differentiation has been shown to occur before final cell division and morphological change and the mRNA levels of stumpy specific genes rise before cell cycle arrest (Tyler et al., 1997), which was supported by the prediction in this thesis that committed slender cells express *PAD1* mRNA before cell cycle arrest. It was also demonstrated here that very early cell cycle arrested intermediate forms are likely being scored as slender in morphology, suggesting that cell cycle arrest also precedes detectable morphological change. Hence, it appears that morphological change is actually a very late marker in the process of differentiation and as such it may be easy to underestimate the number of differentiating forms.

The use of *PAD1* as a molecular marker for stumpy forms, as used here *in vivo* by qRT-PCR and *in vitro* in Chapter 4, removes the subjective nature of the analysis, and provides an early marker for differentiation. Indeed, *PAD1* mRNA expression was shown to increase in the absence of full morphological change in response to stumpy-inducing drugs in Chapter 4: since *PAD1* expression is an earlier marker for differentiation than morphology, it appears these drugs were able to initiate the early steps of differentiation but not full stumpy formation (as discussed in Chapter 4).

5.8.8 Possible improvements to the mathematical model

When mathematical models are produced, they aim to explain the available data as well as possible in as simple a way as possible. Clearly, for complex biological systems simple mathematical models will not be able encompass all the factors that are influencing that system; especially when experimental data regarding many of the factors are unknown. As discussed in section 5.8.2, a mathematical model for stumpy formation was fitted to the data that had been generated in section 5.6. These infections were carried out in immunocompetent mice and as such the host immune system and trypanosome antigenic variation were required to be taken into account in the model. However, no data regarding the host immune response or antigenic variation throughout the infection was collected. Thus, a simple model of the immune system and the antigen switching was incorporated in order to reduce the number of unknown parameters.

Firstly, in the model of the immune system, there was a delay in the response to of the immune system to the infecting variant, but not to later variants. It may improve the model to incorporate a time delay between the responses to each variant to account for the time taken to raise an immune response to each variant. Indeed, in two experimental cattle infections, the VSG-specific transcripts were, in some cases, detected over a week before the immune response specific to that variant could be detected (Morrison et al., 2005).

To incorporate antigenic variation a simple sequential switching model was used whereby the first variant could only switch to the second variant which could only switch to the third and so on, with the rate of switching between two variants different for each pair. This simple model still resulted in the prediction of multiple variants expressed simultaneously and requires less unknown parameters to be incorporated into the model. However, the model did predict a slower switch rate between the first and the second variant than was observed (section 5.7.4). To improve the model regarding antigenic variation, it may be useful to incorporate a model similar to that produced by Lythgoe *et al.,* whereby the rate of switching away from any one variant is constant but that it may switch to any other variant in the

system, with differential probabilities of activation of each individual variant (Lythgoe et al., 2007). This model may be more biologically correct and produces a hierarchical model of expression where variants with a high activation rate are expressed earlier in infection and those with a lower activation rate are expressed later in infection. However, inclusion of such a model would require a much greater number of unknown parameters to be included in the model since no data regarding antigen switching was collected here.

Finally, the error in the measured parasitaemia and *PAD1* expression was assumed, based on previous studies of malaria infection in mice (Mideo et al., 2008; Miller et al., 2010). However, if the errors had been empirically calculated for the assay developed here, this would have reduced the number of assumptions in the model and may have allowed the model to be more accurately fitted to the data. This could have been determined by collecting two or more blood samples from a mouse at the same time point and processing the samples simultaneously. Given that processing three slender and three stumpy standard curve samples simultaneously resulted in very consistent results (section 5.4), the errors would be expected to be low and therefore similar or less than the assumed errors, although this analysis did not account for error during blood sampling.

5.8.9 Possible improvements to the qRT-PCR method

The qRT-PCR method established here provides a way in which to quantify stumpy formation and parasitaemia *in vivo* using only a 10µl sample. However the limit of detection of parasitaemia was 10^6 cells/ml blood due to the small starting volume and the number of dilutions that were required throughout the protocol in order to DNase treat the sample and prevent inhibition during cDNA production and PCR. To improve the method it may be useful to attempt to collect a greater volume of blood per day or to try different reagents throughout processing to determine if fewer dilutions could be carried out. In the mouse infections, the parasitaemia was high enough during the chronic stage of infection that the detection was not limited. However, the current detection limit is too high to monitor cattle infections where the

parasitaemia can range from above 10^6 cells/ml to less than 10^2 cells/ml (Desquesnes and Davila, 2002; Morrison et al., 2005; Rurangirwa et al., 1986). Much larger samples can be collected from cattle; millilitre volumes can be collected every day compared to micro litre volumes from mice. Therefore, with some time developing a new protocol, it should be possible to extend this method to allow for analysis of experimental cattle infections as well as monitoring natural cattle infections in the field (section 5.8.11).

Here, *PAD1* was used to monitor stumpy formation throughout infection. This qRT-PCR method could also be used to monitor up-regulation of mRNA expression of mitochondrial markers of stumpy formation in order to further elucidate the order of events during stumpy formation. Additionally, recent whole genome microarray analysis has identified that a large number of genes are differentially expressed between slender and stumpy forms (Jensen et al., 2009; Kabani et al., 2009). By using qRT-PCR to monitor the rise in expression of genes, other than *PAD1*, that are up-regulated in stumpy formation, during the first wave of infection, it may be possible to identify groups of genes that are up-regulated simultaneously, implying similar mechanisms of gene regulation and common RNA-binding complexes, or determine which genes are early markers of differentiation and which genes are late markers.

5.8.10 Future applications of the qRT-PCR method and the mathematical model in laboratory infections

Firstly, it would be potentially interesting to use the qRT-PCR method established here to analyse the within-host dynamics in chronic infections with different strains of trypanosome in order to establish the variability in stumpy formation between strains. It would perhaps also be useful to analyse infection in different mouse strains, given that different strains show different susceptibility to trypanosomiasis (Antoine-Moussiaux et al., 2008) in order to determine the degree to which the genotype of the host can impact on transmissibility and parasitaemia.

This method could also be used to monitor stumpy formation under different conditions in order to determine if the parasite is able to alter the rate of stumpy formation in response to these conditions. It is not known if trypanosomes are able to alter stumpy formation in response to changes, other than parasite density, in the environment such as the presence of competing strains or availability of host resources. The dynamics between proliferation and stumpy formation may be able to be altered by a change in the rate of replication or the rate of production of, or response to, SIF. Furthermore, there may be changes in the rate of loss of SIF in the host dependent on host factors. To determine if indeed changes in differentiation do occur, this method could potentially be used to analyse the effect of host health on stumpy formation. It has been observed that host glucose levels can fall drastically in infected mice just before host death (Balmer et al., 2009). It may be expected that if the host is very near to death, and the parasites could perhaps detect this in some way, then they would maximise transmission potential and therefore increase stumpy formation. Stumpy formation could be monitored late in infections prior to host death to determine if this is indeed the case. If an increase in stumpy formation was observed, this would indicate that the trypanosome can monitor the health of the host in some way, thus detecting changes in their environment, and respond accordingly.

At least 8.8% of trypanosome infections in the field are thought to contain more than one parasite strain, which is likely to affect within-host dynamics and as such, have an effect on the parasitaemia, virulence and the transmission capacity of the infection (Balmer and Caccone, 2008; Balmer et al., 2009; Turner et al., 1996). Mixed infection can reduce overall parasite virulence compared to single infections (Balmer et al., 2009), with parasite strains suppressing the growth of each other during infection (Balmer et al., 2009; Turner et al., 1996). It is also not known if, and how, the parasite detects the presence of a second strain within the host. The effect of competition on one strain appears to be dependent on the density of the other (Balmer et al., 2009) but the mechanism by which this suppression occurs is unknown.

As discussed above, it is not known if trypanosomes can alter their response to SIF, but this could be a key factor in competition. It has been predicted that the parasites may reduce stumpy formation in the presence of a competitor in order to increase the number of proliferative slender forms in the population and to therefore out-compete the other strain (Pollitt et al., in review). Alternatively, it may be that the strain with the higher growth rate is able to increase to a greater density than the other before the total population density reaches a higher density before stumpy formation begins. In either case, if the parasites cannot distinguish between the SIF produced by each strain, then perhaps this is involed in the density-dependent suppression of one strain on another (Balmer et al., 2009) and it may be that the SIF produced by the dominating strain causes stumpy formation in the other strain at low parasitaemia, thus preventing an increase in parasitaemia of that second strain. Now that a method for quantify stumpy formation during infection has been established it may be possible to modify this in order to determine if stumpy formation varies between strains and if this is altered by the presence of a competing strain, thus demonstrating if trypanosomes are able to detect and alter stumpy formation in response to presence of a second strain. To do this, a means in which to distinguish between the two strains and between stumpy formation of the two strains would be necessary.

For similar analysis regarding gametocyte production in malaria infections, Wargo et al., used strain-specific qRT-PCR primers for the analysis expression of a gametocyte specific gene (Wargo et al., 2007). However, for *PAD1*, which is a member of an 8-gene family, only a very limited region of the gene allowed for specific amplification. Thus, it may not be possible to design strain-specific *PAD1* primers, although this has not been investigated. As an alternative, transgenic pleomorphic cell lines could be created which couple the *PAD1* 3'UTR which controls stumpy-specific gene expression (Chapter 3 and 4), with unique sequences, such as a reporter genes, in order that different strains could be identified. Alternatively, Balmer and Tostado have created different fluorescently labelled strains of *T.b. brucei* and *T.b. rhodesiense* such that they can be distinguished in co-infections (Balmer and Tostado, 2006). They established a method whereby two strains could be infected into the same host and the proportion of each strain could be

monitored by FACS from only 5µl of mouse blood (Balmer and Tostado, 2006). Importantly, the use of FACS allows two populations to be separated: it may be possible to therefore develop this method for the purposes of monitoring differences in stumpy formation in a mixed infection by separating the two strains and then carrying out qRT-PCR on the two populations.

A further application of both the model and the qRT-PCR method would be to predict and investigate the impact of stumpy-inducing compounds, on within-host dynamics. In Chapter 4 a drug screen was carried out to identify compounds which could potentially induce stumpy-formation at low parasitaemia. Although these compounds will need validated and their effect measured *in vitro,* the qRT-PCR method could be used to validate the ability of the drugs to induce stumpy formation and the model could be used to predict the effect of these compounds on the infection dynamics. Inducing stumpy formation may cause decreased parasitaemia and, if able to cause complete stumpy formation would cause eradication of the parasite. Alternatively, the induction of stumpy formation may temporarily cause parasitaemia, and therefore SIF concentration, to fall, but then due to a lack of SIF in the system following treatment, the parasitaemia may surge and the effect would therefore be short-lived. The model could potentially predict such problems and be used to predict the most appropriate treatment regime. The qRT-PCR method could then be used to monitor the effects *in vivo* on within-host dynamics and to test the predictions made by the model.

Mouse models are a valuable tool to analyse many aspects of trypanosome infection, including differences between strains and the effect of competition during infection (Antoine-Moussiaux et al., 2008; Balmer et al., 2009; Morrison et al., 2010; Turner et al., 1996). However, it would also be interesting to analyse natural trypanosome infection dynamics in cattle in the field (section 5.8.11).

5.8.11 Future applications of the qRT-PCR method and the mathematical model in the field

As discussed in section 5.8.9, it should be possible to modify the qRT-PCR method established here for use in analyses of cattle infections. If this were achieved it would then be possible to monitor stumpy formation in natural infections in the field. Samples could be collected daily from cattle in affected regions and transferred directly into RNA lysis solution (see Chapter 2). Samples could then be frozen and stored before being transferred back to the laboratory for analysis. Then, both the changes in parasitaemia and stumpy formation during natural chronic infections could be tracked over time. This would provide information regarding the dynamics of long term chronic infections, potentially identifying if there are large fluctuations in stumpy formation or if the proportions of stumpy forms remains similar throughout infection. This data could then be incorporated into the mathematical model to determine: (1) if the model is suitable for natural cattle infections, or if it needs additional factors to be considered and, (2) what parameter estimates are calculated from these natural infections. Thus, some of the similarities and differences between experimental mice infections and natural cattle infections could be ascertained.

Given the potentially complex within-host dynamics of trypanosome infections, stumpy formation could play an important role in disease progression and transmission in endemic areas. Overall, there is a plethora of exciting experiments, analysing stumpy formation in both laboratory and field infections, that could be carried out following the establishment of this qRT-PCR method, which could further our understanding of trypanosome infections in the field.

6 Chapter 6: Summary and future directions

The African trypanosome differentiates between different life-stages in order to adapt to and prepare for the different environmental conditions it must encounter throughout its life cycle. As such, the parasites must be able to detect and respond to their external environment to cause changes in gene expression which in turn results in changes in metabolism, proliferation status and morphology. Such changes in gene expression must occur largely through mechanisms of post-transcriptional regulation, rather than by the effects of transcription factors as seen in most eukaryotes.

One such life-stage transition is the density-dependent differentiation of proliferative slender forms to the cell-cycle arrested, transmissible stumpy forms in the bloodstream of the mammalian host. This differentiation is triggered by a parasite-derived signalling factor used to report on population density rather than large scale changes in environmental temperature, pH or nutrients. Stumpy formation plays a crucial role in trypanosome infection dynamics, prolonging host survival, and therefore duration of infection, ensuring transmissibility between hosts and contributing to ordered antigenic variation. Despite the importance of stumpy forms, fairly little is known about the mechanisms controlling density signalling, stumpy formation, the mechanisms of control of stumpy-specific gene expression and the dynamics of stumpy formation *in vivo*. This has largely been due to a lack of molecular markers to distinguish stumpy forms, with traditional methods relying on morphological analysis, which is subjective, time-consuming and complicated by the presence of morphologically intermediate forms that arise during differentiation. By comparison, much is known about the molecular events causing bloodstream to procyclic differentiation and differential gene expression between these two forms has studied more extensively. This is largely due to the ability to clearly and efficiently distinguish bloodstream from procyclic forms molecularly by their surface protein coats, as well as the ability to chemically induce the differentiation process in the laboratory.

Recently, the first stumpy-specific surface protein PAD1 was identified. *PAD1* is up-regulated at both the level of mRNA abundance and protein expression during the stumpy life stage. Further, *PAD1* expression has been shown to correlate with

competence to differentiate to the procyclic life-stage, providing the first functional marker for stumpy forms and transmission competence. The aims of this thesis were to exploit the stumpy-specific expression profile of *PAD1* in order to further understanding of control of stumpy-specific gene expression and of stumpy formation both *in vitro* and *in vivo.*

Firstly, the *PAD1* 3'UTR was shown to be responsible for repression of expression during the slender life-stage and for up-regulation of gene expression upon differentiation to the stumpy life-stage in both monomorphic and pleomorphic cell lines. The *PAD1* 3'UTR is the first stumpy-specific 3'UTR that has been experimentally analysed; it is able to repress expression in monomorphic slender forms with regulation exerted at both the level of mRNA abundance and translation. Further, the creation of a series of reporter cell lines in monomorphic cell lines, where section of the *PAD1* 3'UTR were deleted and analysed for reporter gene expression, revealed that there are at least two repressive elements which act to prevent inappropriate gene expression in slender forms, again controlling both mRNA abundance and protein expression.

There is much still to be learned about differential control of gene expression in trypanosomes between life-stages; few RNA-binding proteins have been characterised and the mechanisms by which gene expression is altered by environmental cues is not well understood. During the slender to stumpy transition there are large changes in gene expression, including global translational repression, in response to a single environmental cue, reporting on parasite density. Additional analysis of the *PAD1* 3'UTR, identifying more defined repressive elements or motifs, and identification of interacting RNA-binding proteins would further contribute to the understanding of repression of expression during the slender life-stage and up-regulation and escape of global translational repression during the stumpy life stage.

The changes in gene expression, as well as cell morphology, metabolism and proliferation status that occur during the slender to stumpy transition are the result of a response to density of the population within the mammalian bloodstream. This

response is reportedly induced by the parasite-derived SIF, yet despite being characterised in 1997 as a small soluble molecule, or group of molecules, released into the extra cellular environment, the identity of SIF is not yet known. Moreover, no components of the stumpy induction pathway have been described. Despite the ability for cell permeable cAMP to mimic SIF, implicating cAMP signalling in this pathway, the hydrolysis products of cAMP analogs have since been shown to have a more potent effect, arguing against a classical cAMP signalling pathway. Identification of SIF would provide a much enhanced capacity for analysis of stumpy formation. Much of the analysis of bloodstream to procyclic differentiation has been facilitated by the ability to induce differentiation by high concentrations of CCA. This would provide a biologically relevant method of inducing stumpy formation and aid in the identification of the downstream pathway and study molecular changes in the cell throughout the transition. Further, knowing the identity of SIF may aid understanding into ability, or lack of ability, for trypanosomes to vary their response to SIF depending on conditions within the host. Finally, drugs which mimic its activity could potentially be informatively designed in order to drive stumpy formation with prospective therapeutic effects.

Hence, next in this thesis, both pleomorphic and monomorphic reporter cell lines that can report on stumpy formation *in vitro* were created for the analysis of conditioned media and of chemical inducers of stumpy formation. In the pleomorphic cell line, the reporter gene was down-regulated in slender forms and up-regulated in stumpy forms harvested from *in vivo* infections and responded to cell density *in vitro* indicative of appropriate reporter gene expression for analysis of stumpy formation. Despite the promising potential of this cell line, no response was observed upon treatment with conditioned media, expected to contain the elusive SIF. In the absence of a positive control, the explanation for this is not yet clear: it may be that the conditioned media generated here does not contain SIF suggesting some component has been lost or destroyed during processing or that some other requirement for stumpy formation, as well as SIF, is not fulfilled under the conditions used here. Further experimentation with this cell line, including recreating the conditions used by Vassella *et al.,* should ascertain the reason for the lack of response. If the

pleomorphic cell line could be successfully used to track SIF activity in conditioned media, the identity of this unknown factor could be discovered by comparative mass spectrometry. The activity of candidate compounds identified by mass spectrometry could then be analysed out of the context of conditioned media for validation. This pleomorphic cell line therefore has excellent potential and given more time could be a very useful resource.

Given the lack of advancement in identifying SIF in over 13 years, since its initial characterisation, a second approach to understanding the molecular events that occur during stumpy formation was also taken. Monomorphic reporter cell lines were created and used to analyse previously reported chemical inducers of stumpy formation. If monomorphic cell lines are simply unable to receive the SIF signal then perhaps molecules which induce stumpy formation in monomorphic cells are acting downstream in the stumpy induction pathway. Identification of pathway components would provide a starting point for the characterisation of the remainder of the pathway, potentially including the identity of SIF and the mechanism by which the SIF pathway leads to global changes in gene expression and the formation of stumpy forms.

The analysis of monomorphic reporter cell lines treated with chemical inducers of stumpy formation confirmed that monomorphs are not incapable of stumpy formation *per se,* although, how similar these "stumpy-like" cells are to naturally-derived stumpy forms is not clear. Reporter gene expression in these cell lines correlated with cell cycle arrest and an increased capacity for differentiation to the procyclic life-stage, demonstrating that where an up-regulation of reporter gene expression was observed, other indicators of stumpy formation were also seen. Hence, these cell lines can be used for the analysis and identification of stumpy-inducing compounds *in vitro* in laboratory adapted strains. This allowed for the establishment of a high-throughput assay for stumpy formation in monomorphic cell lines. This assay was used to screen 6000 kinase inhibitors for their stumpy-inducing potential with two objectives: firstly, compounds able to induce stumpy formation at low parasitaemia may have therapeutic potential and, secondly, these compounds

may be acting on components of the stumpy induction pathway and could therefore lead to their identification. Upon receipt of the results from the drug screen, the candidate compounds identified with need to be experimentally validated further. Any kinase inhibitors which can be confirmed as causing stumpy-formation can then be used to search for the molecule(s) which it inhibits. This drug screen therefore has huge potential to identify components of the SIF pathway. As previously mentioned, identification of one component involved in stumpy formation would greatly increase the capacity to study the remainder of the pathway by searching for interacting proteins both up and down-stream.

As well as a lack of understanding of the molecular events causing stumpy formation in individual cells, there is also fairly little known regarding the temporal order of events during stumpy formation *in vivo* and the effects that stumpy formation has on infection dynamics and transmission. Therefore, next, a quantitative method to analyse parasitaemia and stumpy formation *in vivo* was developed here. Stumpy formation was monitored early in mouse infections by traditional morphological and cell cycle analysis, as well as by qRT-PCR for *PAD1* expression. The resulting analysis allowed the definition of the co-ordinated order of events in the generation of stumpy forms in the population early in infection.

Further, *PAD1* expression was monitored throughout chronic mouse infections revealing dominance of transmissible forms, which has important consequences on infection dynamics. Firstly, there is no prolonged periods of a low proportion of differentiated forms suggesting that the infection is transmission competent throughout infection, which may be crucial when tsetse bites are rare. Secondly, as only slender cells are functionally competent for antigenic variation, a low proportion of slender cells throughout infection will result in a reduced use of the antigen repertoire than if there were a high proportion of slender forms, potentially prolonging antigen chronicity.

Dr Nick Savill, University of Edinburgh, then incorporated the quantitative data into a pre-existing model for stumpy formation, allowing for estimation of numerous

parameters relating to infection dynamics *in vivo*. Importantly, this model incorporated SIF-induced stumpy formation and it was demonstrated that an equivalent model incorporating density-induced differentiation did not provide as good a fit to the quantitative data, supporting a quorum-sensing mechanism for differentiation *in vivo* rather than absolute cell density. Combining data from both the experimental analysis and the mathematical model, it was possible to define the temporal order of events that occurred during differentiation of an individual cell, with a cell undergoing further cell divisions following commitment to differentiation and both cell cycle arrest and up-regulation of mRNA expression of stumpy-specific proteins preceding morphological change.

Both the qRT-PCR technique and the improved mathematical model provide valuable tools for future analysis of within host dynamics both in the laboratory and the field. These tools can be used for the analysis of potential differences in stumpy formation between strains, the effect of environmental factors, such as resources and competition, on stumpy formation and the within-host dynamics of chronic infections in the field. The model can also be used to predict the effects of stumpy-inducing compounds, or other therapeutic compounds, on the infection dynamics *in vivo* and may direct informed experimental analysis.

Stumpy forms play an important role in the life-cycle of the trypanosome; influencing transmissibility, prolonged host survival and antigenic variation but have been under studied due to the lack of tools to distinguish the life-stage. Overall, the methods created, and the data then generated in this thesis, provide a valuable step forward towards fully understanding the mechanisms and order of events underlying stumpy formation, control of stumpy-specific gene expression and the role of stumpy formation in infection dynamics of *Trypanosoma brucei.*

Bibliography

Alexandre, S., P. Paindavoine, J. Hanocq-Quertier, F. Paturiaux-Hanocq, P. Tebabi, and E. Pays. 1996. Families of adenylate cyclase genes in Trypanosoma brucei. *Mol Biochem Parasitol*. 77:173-82.

Allsopp, R. 2001. Options for vector control against trypanosomiasis in Africa. *Trends Parasitol*. 17:15-9.

Ansede, J.H., M. Anbazhagan, R. Brun, J.D. Easterbrook, J.E. Hall, and D.W. Boykin. 2004. O-alkoxyamidine prodrugs of furamidine: in vitro transport and microsomal metabolism as indicators of in vivo efficacy in a mouse model of Trypanosoma brucei rhodesiense infection. *J Med Chem*. 47:4335-8.

Antoine-Moussiaux, N., S. Magez, and D. Desmecht. 2008. Contributions of experimental mouse models to the understanding of African trypanosomiasis. *Trends Parasitol*. 24:411-8.

Archer, S.K., V.D. Luu, R.A. de Queiroz, S. Brems, and C. Clayton. 2009. Trypanosoma brucei PUF9 regulates mRNAs for proteins involved in replicative processes over the cell cycle. *PLoS Pathog*. 5:e1000565.

Balber, A.E. 1972. Trypanosoma brucei: fluxes of the morphological variants in intact and X-irradiated mice. *Exp Parasitol*. 31:307-19.

Balmer, O., and A. Caccone. 2008. Multiple-strain infections of Trypanosoma brucei across Africa. *Acta Trop*. 107:275-9.

Balmer, O., S.C. Stearns, A. Schotzau, and R. Brun. 2009. Intraspecific competition between co-infecting parasite strains enhances host survival in African trypanosomes. *Ecology*. 90:3367-78.

Balmer, O., and C. Tostado. 2006. New fluorescence markers to distinguish co-infecting Trypanosoma brucei strains in experimental multiple infections. *Acta Trop*. 97:94-101.

Bangs, J.D., P.F. Crain, T. Hashizume, J.A. McCloskey, and J.C. Boothroyd. 1992. Mass spectrometry of mRNA cap 4 from trypanosomatids reveals two novel nucleosides. *J Biol Chem*. 267:9805-15.

Bantscheff, M., D. Eberhard, Y. Abraham, S. Bastuck, M. Boesche, S. Hobson, T. Mathieson, J. Perrin, M. Raida, C. Rau, V. Reader, G. Sweetman, A. Bauer, T. Bouwmeester, C. Hopf, U. Kruse, G. Neubauer, N. Ramsden, J. Rick, B. Kuster, and G. Drewes. 2007. Quantitative chemical proteomics reveals mechanisms of action of clinical ABL kinase inhibitors. *Nat Biotechnol*. 25:1035-44.

Barrett, M.P. 1999. The fall and rise of sleeping sickness. *Lancet*. 353:1113-4.

Barrett, M.P. 2006. The rise and fall of sleeping sickness. *Lancet*. 367:1377-8.

Barrett, M.P., D.W. Boykin, R. Brun, and R.R. Tidwell. 2007. Human African trypanosomiasis: pharmacological re-engagement with a neglected disease. *Br J Pharmacol*. 152:1155-71.

Barry, J.D., L. Marcello, L.J. Morrison, A.F. Read, K. Lythgoe, N. Jones, M. Carrington, G. Blandin, U. Bohme, E. Caler, C. Hertz-Fowler, H. Renauld, N. El-Sayed, and M. Berriman. 2005. What the genome sequence is revealing about trypanosome antigenic variation. *Biochem Soc Trans*. 33:986-9.

Bass, K.E., and C.C. Wang. 1991. The in vitro differentiation of pleomorphic Trypanosoma brucei from bloodstream into procyclic form requires neither intermediary nor short-stumpy stage. *Mol Biochem Parasitol*. 44:261-70.

Bastin, P., T. Sherwin, and K. Gull. 1998. Paraflagellar rod is vital for trypanosome motility. *Nature*. 391:548.

Baylis, M. 1997. The daily feeding rate of tsetse (Diptera: Glossinidae) on cattle at Galana Ranch, Kenya and comparison with trypanosomiasis incidence. *Acta Trop*. 65:81-96.

Benne, R. 1994. RNA editing in trypanosomes. *Eur J Biochem*. 221:9-23.

Benne, R., J. Van den Burg, J.P. Brakenhoff, P. Sloof, J.H. Van Boom, and M.C. Tromp. 1986. Major transcript of the frameshifted coxII gene from trypanosome mitochondria contains four nucleotides that are not encoded in the DNA. *Cell*. 46:819-26.

Berriman, M., E. Ghedin, C. Hertz-Fowler, G. Blandin, H. Renauld, D.C. Bartholomeu, N.J. Lennard, E. Caler, N.E. Hamlin, B. Haas, U. Bohme, L. Hannick, M.A. Aslett, J. Shallom, L. Marcello, L. Hou, B. Wickstead, U.C. Alsmark, C. Arrowsmith, R.J. Atkin, A.J. Barron, F. Bringaud, K. Brooks, M. Carrington, I. Cherevach, T.J. Chillingworth, C. Churcher, L.N. Clark, C.H. Corton, A. Cronin, R.M. Davies, J. Doggett, A. Djikeng, T. Feldblyum, M.C. Field, A. Fraser, I. Goodhead, Z. Hance, D. Harper, B.R. Harris, H. Hauser, J. Hostetler, A. Ivens, K. Jagels, D. Johnson, J. Johnson, K. Jones, A.X. Kerhornou, H. Koo, N. Larke, S. Landfear, C. Larkin, V. Leech, A. Line, A. Lord, A. Macleod, P.J. Mooney, S. Moule, D.M. Martin, G.W. Morgan, K. Mungall, H. Norbertczak, D. Ormond, G. Pai, C.S. Peacock, J. Peterson, M.A. Quail, E. Rabbinowitsch, M.A. Rajandream, C. Reitter, S.L. Salzberg, M. Sanders, S. Schobel, S. Sharp, M. Simmonds, A.J. Simpson, L. Tallon, C.M. Turner, A. Tait, A.R. Tivey, S. Van Aken, D. Walker, D. Wanless, S. Wang, B. White, O. White, S. Whitehead, J. Woodward, J. Wortman, M.D. Adams, T.M. Embley, K. Gull, E. Ullu, J.D. Barry, A.H. Fairlamb, F. Opperdoes, B.G. Barrell, J.E. Donelson, N. Hall, C.M. Fraser, et al. 2005. The genome of the African trypanosome Trypanosoma brucei. *Science*. 309:416-22.

Biebinger, S., S. Rettenmaier, J. Flaspohler, C. Hartmann, J. Pena-Diaz, L.E. Wirtz, H.R. Hotz, J.D. Barry, and C. Clayton. 1996. The PARP promoter of Trypanosoma brucei is developmentally regulated in a chromosomal context. *Nucleic Acids Res*. 24:1202-11.

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

Bingle, L.E., J.L. Eastlake, M. Bailey, and W.C. Gibson. 2001. A novel GFP approach for the analysis of genetic exchange in trypanosomes allowing the in situ detection of mating events. *Microbiology*. 147:3231-40.

Black, S.J., R.S. Hewett, and C.N. Sendashonga. 1982. Trypanosoma brucei variable surface antigen is released by degenerating parasites but not by actively dividing parasites. *Parasite Immunol*. 4:233-44.

Brecht, M., and M. Parsons. 1998. Changes in polysome profiles accompany trypanosome development. *Mol Biochem Parasitol*. 97:189-98.

Breidbach, T., E. Ngazoa, and D. Steverding. 2002. Trypanosoma brucei: in vitro slender-to-stumpy differentiation of culture-adapted, monomorphic bloodstream forms. *Exp Parasitol*. 101:223-30.

Briggs, L.J., P.G. McKean, A. Baines, F. Moreira-Leite, J. Davidge, S. Vaughan, and K. Gull. 2004. The flagella connector of Trypanosoma brucei: an unusual mobile transmembrane junction. *J Cell Sci*. 117:1641-51.

Brun, R., J. Blum, F. Chappuis, and C. Burri. 2010. Human African trypanosomiasis. *Lancet*. 375:148-59.

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

Checchi, F., J.A. Filipe, D.T. Haydon, D. Chandramohan, and F. Chappuis. 2008. Estimates of the duration of the early and late stage of gambiense sleeping sickness. *BMC Infect Dis*. 8:16.

Clamp, M., J. Cuff, S.M. Searle, and G.J. Barton. 2004. The Jalview Java alignment editor. *Bioinformatics*. 20:426-7.

Clayton, C.E. 2002. Life without transcriptional control? From fly to man and back again. *Embo J*. 21:1881-8.

Clement, S.L., and D.J. Koslowsky. 2001. Unusual organization of a developmentally regulated mitochondrial RNA polymerase (TBMTRNAP) gene in Trypanosoma brucei. *Gene*. 272:209-18.

Corpet, F. 1988. Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res*. 16:10881-90.

Czichos, J., C. Nonnengaesser, and P. Overath. 1986. Trypanosoma brucei: cis-aconitate and temperature reduction as triggers of synchronous transformation of bloodstream to procyclic trypomastigotes in vitro. *Exp Parasitol*. 62:283-91.

De Gaudenzi, J., A.C. Frasch, and C. Clayton. 2005. RNA-binding domain proteins in Kinetoplastids: a comparative analysis. *Eukaryot Cell*. 4:2106-14.

de Gee, A.L., P.H. Carstens, P.P. McCann, and J.M. Mansfield. 1984. Morphological changes in Trypanosoma brucei rhodesiense following inhibition of polyamine biosynthesis in vivo. *Tissue Cell*. 16:731-8.

De Lange, T., P.A. Michels, H.J. Veerman, A.W. Cornelissen, and P. Borst. 1984. Many trypanosome messenger RNAs share a common 5' terminal sequence. *Nucleic Acids Res*. 12:3777-90.

Dean, S., R. Marchetti, K. Kirk, and K.R. Matthews. 2009. A surface transporter family conveys the trypanosome differentiation signal. *Nature*. 459:213-7.

Denninger, V., K. Figarella, C. Schonfeld, S. Brems, C. Busold, F. Lang, J. Hoheisel, and M. Duszenko. 2007. Troglitazone induces differentiation in Trypanosoma brucei. *Exp Cell Res*. 313:1805-19.

Desquesnes, M., and A.M. Davila. 2002. Applications of PCR-based tools for detection and identification of animal trypanosomes: a review and perspectives. *Vet Parasitol*. 109:213-31.

Ding, Y., C.Y. Chan, and C.E. Lawrence. 2005. RNA secondary structure prediction by centroids in a Boltzmann weighted ensemble. *Rna*. 11:1157-66.

Ding, Y., and C.E. Lawrence. 2003. A statistical sampling algorithm for RNA secondary structure prediction. *Nucleic Acids Res*. 31:7280-301.

Doerig, C. 2004. Protein kinases as targets for anti-parasitic chemotherapy. *Biochim Biophys Acta*. 1697:155-68.

Domenicali Pfister, D., G. Burkard, S. Morand, C.K. Renggli, I. Roditi, and E. Vassella. 2006. A Mitogen-activated protein kinase controls differentiation of bloodstream forms of Trypanosoma brucei. *Eukaryot Cell*. 5:1126-35.

Engstler, M., and M. Boshart. 2004. Cold shock and regulation of surface protein trafficking convey sensitization to inducers of stage differentiation in Trypanosoma brucei. *Genes Dev*. 18:2798-811.

Engstler, M., T. Pfohl, S. Herminghaus, M. Boshart, G. Wiegertjes, N. Heddergott, and P. Overath. 2007. Hydrodynamic flow-mediated protein sorting on the cell surface of trypanosomes. *Cell*. 131:505-15.

Erondu, N.E., and J.E. Donelson. 1992. Differential expression of two mRNAs from a single gene encoding an HMG1-like DNA binding protein of African trypanosomes. *Mol Biochem Parasitol*. 51:111-8.

Estevez, A.M. 2008. The RNA-binding protein TbDRBD3 regulates the stability of a specific subset of mRNAs in trypanosomes. *Nucleic Acids Res*. 36:4573-86.

Estevez, A.M., T. Kempf, and C. Clayton. 2001. The exosome of Trypanosoma brucei. *Embo J*. 20:3831-9.

Fang, J., and T.F. McCutchan. 2002. Thermoregulation in a parasite's life cycle. *Nature*. 418:742.

Fevre, E.M., B.V. Wissmann, S.C. Welburn, and P. Lutumba. 2008. The burden of human African trypanosomiasis. *PLoS Negl Trop Dis*. 2:e333.

Field, M.C., and M. Carrington. 2009. The trypanosome flagellar pocket. *Nat Rev Microbiol*. 7:775-86.

Frearson, J.A., S. Brand, S.P. McElroy, L.A. Cleghorn, O. Smid, L. Stojanovski, H.P. Price, M.L. Guther, L.S. Torrie, D.A. Robinson, I. Hallyburton, C.P. Mpamhanga, J.A. Brannigan, A.J. Wilkinson, M. Hodgkinson, R. Hui, W. Qiu, O.G. Raimi, D.M. van Aalten, R. Brenk, I.H. Gilbert, K.D. Read, A.H. Fairlamb, M.A. Ferguson, D.F. Smith, and P.G. Wyatt. 2010. N-myristoyltransferase inhibitors as new leads to treat sleeping sickness. *Nature*. 464:728-32.

Furger, A., N. Schurch, U. Kurath, and I. Roditi. 1997. Elements in the 3' untranslated region of procyclin mRNA regulate expression in insect forms of Trypanosoma brucei by modulating RNA stability and translation. *Mol Cell Biol*. 17:4372-80.

Haile, S., A.M. Estevez, and C. Clayton. 2003. A role for the exosome in the in vivo degradation of unstable mRNAs. *Rna*. 9:1491-501.

Hammarton, T.C. 2007. Cell cycle regulation in Trypanosoma brucei. *Mol Biochem Parasitol*. 153:1-8.

Hehl, A., E. Vassella, R. Braun, and I. Roditi. 1994. A conserved stem-loop structure in the 3' untranslated region of procyclin mRNAs regulates expression in Trypanosoma brucei. *Proc Natl Acad Sci U S A*. 91:370-4.

Herbert, W.J., and W.H. Lumsden. 1976. Trypanosoma brucei: a rapid "matching" method for estimating the host's parasitemia. *Exp Parasitol*. 40:427-31.

Hesse, F., P.M. Selzer, K. Muhlstadt, and M. Duszenko. 1995. A novel cultivation technique for long-term maintenance of bloodstream form trypanosomes in vitro. *Mol Biochem Parasitol*. 70:157-66.

Hirumi, H., and K. Hirumi. 1989. Continuous cultivation of Trypanosoma brucei blood stream forms in a medium containing a low concentration of serum protein without feeder cell layers. *J Parasitol*. 75:985-9.

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

Hunt, M., R. Brun, and P. Kohler. 1994. Studies on compounds promoting the in vitro transformation of Trypanosoma brucei from bloodstream to procyclic forms. *Parasitol Res*. 80:600-6.

Irmer, H., and C. Clayton. 2001. Degradation of the unstable EP1 mRNA in Trypanosoma brucei involves initial destruction of the 3'-untranslated region. *Nucleic Acids Res*. 29:4707-15.

Jacobs, S.L., and N.D. Lee. 1964. Determination of Citric Acid in Serum and Urine Using Br82. *J Nucl Med*. 5:297-301.

Janelle, J., M. Koffi, V. Jamonneau, D. Patrel, G. Cuny, and S. Ravel. 2009. Monitoring the pleomorphism of Trypanosoma brucei gambiense isolates in mouse: impact on its transmissibility to Glossina palpalis gambiensis. *Infect Genet Evol*. 9:1260-4.

Jensen, B.C., D. Sivam, C.T. Kifer, P.J. Myler, and M. Parsons. 2009. Widespread variation in transcript abundance within and across developmental stages of Trypanosoma brucei. *BMC Genomics*. 10:482.

Kabani, S., K. Fenn, A. Ross, A. Ivens, T.K. Smith, P. Ghazal, and K. Matthews. 2009. Genome-wide expression profiling of in vivo-derived bloodstream parasite stages and dynamic analysis of mRNA alterations during synchronous differentiation in Trypanosoma brucei. *BMC Genomics*. 10:427.

Kennedy, P.G. 2006. Human African trypanosomiasis-neurological aspects. *J Neurol*. 253:411-6.

Kohl, L., D. Robinson, and P. Bastin. 2003. Novel roles for the flagellum in cell morphogenesis and cytokinesis of trypanosomes. *Embo J*. 22:5336-46.

Kohl, L., T. Sherwin, and K. Gull. 1999. Assembly of the paraflagellar rod and the flagellum attachment zone complex during the Trypanosoma brucei cell cycle. *J Eukaryot Microbiol*. 46:105-9.

Kramer, S., N.C. Kimblin, and M. Carrington. 2010. Genome-wide in silico screen for CCCH-type zinc finger proteins of Trypanosoma brucei, Trypanosoma cruzi and Leishmania major. *BMC Genomics*. 11:283.

Kunz, S., J.A. Beavo, M.A. D'Angelo, M.M. Flawia, S.H. Francis, A. Johner, S. Laxman, M. Oberholzer, A. Rascon, Y. Shakur, L. Wentzinger, R. Zoraghi, and T. Seebeck. 2006. Cyclic nucleotide specific phosphodiesterases of the kinetoplastida: a unified nomenclature. *Mol Biochem Parasitol*. 145:133-5.

Laxman, S., and J.A. Beavo. 2007. Cyclic nucleotide signaling mechanisms in trypanosomes: possible targets for therapeutic agents. *Mol Interv*. 7:203-15.

Laxman, S., A. Riechers, M. Sadilek, F. Schwede, and J.A. Beavo. 2006. Hydrolysis products of cAMP analogs cause transformation of Trypanosoma brucei from slender to stumpy-like forms. *Proc Natl Acad Sci U S A*. 103:19194-9.

Li, C.H., H. Irmer, D. Gudjonsdottir-Planck, S. Freese, H. Salm, S. Haile, A.M. Estevez, and C. Clayton. 2006. Roles of a Trypanosoma brucei 5'->3' exoribonuclease homolog in mRNA degradation. *Rna*. 12:2171-86.

Luo, H., G. Gilinger, D. Mukherjee, and V. Bellofatto. 1999. Transcription initiation at the TATA-less spliced leader RNA gene promoter requires at least two

DNA-binding proteins and a tripartite architecture that includes an initiator element. *J Biol Chem*. 274:31947-54.

Lythgoe, K.A., L.J. Morrison, A.F. Read, and J.D. Barry. 2007. Parasite-intrinsic factors can explain ordered progression of trypanosome antigenic variation. *Proc Natl Acad Sci U S A*. 104:8095-100.

MacGregor, P., and K.R. Matthews. 2010. New discoveries in the transmission biology of sleeping sickness parasites: applying the basics. *J Mol Med*.

Mair, G., E. Ullu, and C. Tschudi. 2000. Cotranscriptional cap 4 formation on the Trypanosoma brucei spliced leader RNA. *J Biol Chem*. 275:28994-9.

Mancini, P.E., and C.L. Patton. 1981. Cyclic 3',5'-adenosine monophosphate levels during the developmental cycle of Trypanosoma brucei brucei in the rat. *Mol Biochem Parasitol*. 3:19-31.

Mathews, D.H., J. Sabina, M. Zuker, and D.H. Turner. 1999. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J Mol Biol*. 288:911-40.

Matthews, K.R. 2005. The developmental cell biology of Trypanosoma brucei. *J Cell Sci*. 118:283-90.

Matthews, K.R., J.R. Ellis, and A. Paterou. 2004. Molecular regulation of the life cycle of African trypanosomes. *Trends Parasitol*. 20:40-7.

Matthews, K.R., and K. Gull. 1994a. Cycles within cycles: the interplay between differentiation and cell division in Trypanosoma brucei. *Parasitol Today*. 10:473-6.

Matthews, K.R., and K. Gull. 1994b. Evidence for an interplay between cell cycle progression and the initiation of differentiation between life cycle forms of African trypanosomes. *J Cell Biol*. 125:1147-56.

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

Maudlin, I., and S.C. Welburn. 1989. A single trypanosome is sufficient to infect a tsetse fly. *Ann Trop Med Parasitol*. 83:431-3.

Mayho, M., K. Fenn, P. Craddy, S. Crosthwaite, and K. Matthews. 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*. 34:5312-24.

McCulloch, R., E. Vassella, P. Burton, M. Boshart, and J.D. Barry. 2004. Transformation of monomorphic and pleomorphic Trypanosoma brucei. *Methods Mol Biol*. 262:53-86.

McLintock, L.M., C.M. Turner, and K. Vickerman. 1993. Comparison of the effects of immune killing mechanisms on Trypanosoma brucei parasites of slender and stumpy morphology. *Parasite Immunol*. 15:475-80.

Mideo, N., V.C. Barclay, B.H. Chan, N.J. Savill, A.F. Read, and T. Day. 2008. Understanding and predicting strain-specific patterns of pathogenesis in the rodent malaria Plasmodium chabaudi. *Am Nat*. 172:214-38.

Milhausen, M., R.G. Nelson, S. Sather, M. Selkirk, and N. Agabian. 1984. Identification of a small RNA containing the trypanosome spliced leader: a donor of shared 5' sequences of trypanosomatid mRNAs? *Cell*. 38:721-9.

Miller, M.R., L. Raberg, A.F. Read, and N.J. Savill. 2010. Quantitative analysis of immune response and erythropoiesis during rodent malarial infection. *PloS Comp. Biol.* Submitted.

Moreira-Leite, F.F., T. Sherwin, L. Kohl, and K. Gull. 2001. A trypanosome structure involved in transmitting cytoplasmic information during cell division. *Science*. 294:610-2.

Morrison, L.J., P. Majiwa, A.F. Read, and J.D. Barry. 2005. Probabilistic order in antigenic variation of Trypanosoma brucei. *Int J Parasitol*. 35:961-72.

Morrison, L.J., S. McLellan, L. Sweeney, C.N. Chan, A. MacLeod, A. Tait, and C.M. Turner. 2010. Role for parasite genetic diversity in differential host responses to Trypanosoma brucei infection. *Infect Immun*. 78:1096-108.

Naula, C., M. Parsons, and J.C. Mottram. 2005. Protein kinases as drug targets in trypanosomes and Leishmania. *Biochim Biophys Acta*. 1754:151-9.

Nolan, D.P., S. Rolin, J.R. Rodriguez, J. Van Den Abbeele, and E. Pays. 2000. Slender and stumpy bloodstream forms of Trypanosoma brucei display a differential response to extracellular acidic and proteolytic stress. *Eur J Biochem*. 267:18-27.

Oberholzer, M., G. Marti, M. Baresic, S. Kunz, A. Hemphill, and T. Seebeck. 2007. The Trypanosoma brucei cAMP phosphodiesterases TbrPDEB1 and TbrPDEB2: flagellar enzymes that are essential for parasite virulence. *Faseb J*. 21:720-31.

Ogbadoyi, E.O., D.R. Robinson, and K. Gull. 2003. A high-order trans-membrane structural linkage is responsible for mitochondrial genome positioning and segregation by flagellar basal bodies in trypanosomes. *Mol Biol Cell*. 14:1769-79.

Parsons, M., R.G. Nelson, K.P. Watkins, and N. Agabian. 1984. Trypanosome mRNAs share a common 5' spliced leader sequence. *Cell*. 38:309-16.

Parsons, M., E.A. Worthey, P.N. Ward, and J.C. Mottram. 2005. Comparative analysis of the kinomes of three pathogenic trypanosomatids: Leishmania major, Trypanosoma brucei and Trypanosoma cruzi. *BMC Genomics*. 6:127.

Paterou, A., P. Walrad, P. Craddy, K. Fenn, and K. Matthews. 2006. Identification and stage-specific association with the translational apparatus of TbZFP3, a CCCH protein that promotes trypanosome life-cycle development. *J Biol Chem*. 281:39002-13.

Perry, K.L., K.P. Watkins, and N. Agabian. 1987. Trypanosome mRNAs have unusual "cap 4" structures acquired by addition of a spliced leader. *Proc Natl Acad Sci U S A*. 84:8190-4.

Pholig, G., S. Bernhard, J. Blum, C. Burri, A. Mpanya Kabeya, J.-P. Fina Lubaki, A. Mpoo Mpoto, B. Fungula Munungu, G. Kambau Manesa Deo, P. Nsele Mutantu, F. Mbo Kuikumbi, A. Fukinsia Mintwo, A. Kayeye Munungi, A. Dala, S. Macharia, C. Miaka Mia Bilenge, V. Kande Betu Ku Mesu, J. Ramon Franco, N. Dieyi Dituvanga, and C. Olson. 2008. Phase 3 trial of pafuramidine maleate (DB289), a novel, oral drug, for treatment of first stage sleeping sickness: safety and efficacy. *abstract. 542 from 57th Meet. Am. Soc. Trop. Med. Hyg*:160.

Picozzi, K., E.M. Fevre, M. Odiit, M. Carrington, M.C. Eisler, I. Maudlin, and S.C. Welburn. 2005. Sleeping sickness in Uganda: a thin line between two fatal diseases. *Bmj*. 331:1238-41.

Ploubidou, A., D.R. Robinson, R.C. Docherty, E.O. Ogbadoyi, and K. Gull. 1999. Evidence for novel cell cycle checkpoints in trypanosomes: kinetoplast segregation and cytokinesis in the absence of mitosis. *J Cell Sci*. 112 (Pt 24):4641-50.

Pollitt, L.C., P. Macgregor, K. Matthews, and S.E. Reece. in review. Malaria and trypanosome transmission strategies: do the same rules apply? .

Price, H.P., M.R. Menon, C. Panethymitaki, D. Goulding, P.G. McKean, and D.F. Smith. 2003. Myristoyl-CoA:protein N-myristoyltransferase, an essential enzyme and potential drug target in kinetoplastid parasites. *J Biol Chem*. 278:7206-14.

Priest, J.W., and S.L. Hajduk. 1994. Developmental regulation of mitochondrial biogenesis in Trypanosoma brucei. *J Bioenerg Biomembr*. 26:179-91.

Queiroz, R., C. Benz, K. Fellenberg, J.D. Hoheisel, and C. Clayton. 2009. Transcriptome analysis of differentiating trypanosomes reveals the existence of multiple post-transcriptional regulons. *BMC Genomics*. 10:495.

Raz, B., M. Iten, Y. Grether-Buhler, R. Kaminsky, and R. Brun. 1997. The Alamar Blue assay to determine drug sensitivity of African trypanosomes (T.b. rhodesiense and T.b. gambiense) in vitro. *Acta Trop*. 68:139-47.

Reuner, B., E. Vassella, B. Yutzy, and M. Boshart. 1997. Cell density triggers slender to stumpy differentiation of Trypanosoma brucei bloodstream forms in culture. *Mol Biochem Parasitol*. 90:269-80.

Robinson, D.R., and K. Gull. 1991. Basal body movements as a mechanism for mitochondrial genome segregation in the trypanosome cell cycle. *Nature*. 352:731-3.

Robinson, D.R., T. Sherwin, A. Ploubidou, E.H. Byard, and K. Gull. 1995. Microtubule polarity and dynamics in the control of organelle positioning, segregation, and cytokinesis in the trypanosome cell cycle. *J Cell Biol*. 128:1163-72.

Rodgers, J. 2009. Trypanosomiasis and the brain. *Parasitology*:1-12.

Rolin, S., J. Hancocq-Quertier, F. Paturiaux-Hanocq, D.P. Nolan, and E. Pays. 1998. Mild acid stress as a differentiation trigger in Trypanosoma brucei. *Mol Biochem Parasitol*. 93:251-62.

Rolin, S., P. Paindavoine, J. Hanocq-Quertier, F. Hanocq, Y. Claes, D. Le Ray, P. Overath, and E. Pays. 1993. Transient adenylate cyclase activation accompanies differentiation of Trypanosoma brucei from bloodstream to procyclic forms. *Mol Biochem Parasitol*. 61:115-25.

Ross, R., and D. Thomson. 1910. A case of sleeping sickness showing regular periodical increase of the parasite disclosed. *Proceedings of the Royal Society, Series B*:411-415.

Rudenko, G. 2000. The polymorphic telomeres of the African Trypanosome trypanosoma brucei. *Biochem Soc Trans*. 28:536-40.

Rurangirwa, F.R., A.J. Musoke, V.M. Nantulya, C. Nkonge, L. Njuguna, E.Z. Mushi, L. Karstad, and J. Grootenhuis. 1986. Immune effector mechanisms involved in the control of parasitaemia in Trypanosoma brucei-infected wildebeest (Connochaetes taurinus). *Immunology*. 58:231-7.

Savill, N.J., W. Chadwick, and S.E. Reece. 2009. Quantitative analysis of mechanisms that govern red blood cell age structure and dynamics during anaemia. *PLoS Comput Biol*. 5:e1000416.

Savill, N.J., and J.R. Seed. 2004. Mathematical and statistical analysis of the Trypanosoma brucei slender to stumpy transition. *Parasitology*. 128:53-67.

Sbicego, S., E. Vassella, U. Kurath, B. Blum, and R. I. 1999a. The use of transgenic Trypanosoma brucei to identify compounds inducing the differentiation of bloodstream forms to procyclic forms. *Mol Biochem Parasitol*. 104:311-322.

Schnaufer, A., G.J. Domingo, and K. Stuart. 2002. Natural and induced dyskinetoplastic trypanosomatids: how to live without mitochondrial DNA. *Int J Parasitol*. 32:1071-84.

Schurch, N., A. Furger, U. Kurath, and I. Roditi. 1997. Contributions of the procyclin 3' untranslated region and coding region to the regulation of expression in bloodstream forms of Trypanosoma brucei. *Mol Biochem Parasitol*. 89:109-21.

Schwede, A., T. Manful, B.A. Jha, C. Helbig, N. Bercovich, M. Stewart, and C. Clayton. 2009. The role of deadenylation in the degradation of unstable mRNAs in trypanosomes. *Nucleic Acids Res*. 37:5511-28.

Scory, S., Y.D. Stierhof, C.R. Caffrey, and D. Steverding. 2007. The cysteine proteinase inhibitor Z-Phe-Ala-CHN2 alters cell morphology and cell division activity of Trypanosoma brucei bloodstream forms in vivo. *Kinetoplastid Biol Dis*. 6:2.

Seebeck, T., K. Gong, S. Kunz, R. Schaub, T. Shalaby, and R. Zoraghi. 2001. cAMP signalling in Trypanosoma brucei. *Int J Parasitol*. 31:491-8.

Seed, J.R., and S.J. Black. 1997. A proposed density-dependent model of long slender to short stumpy transformation in the African trypanosomes. *J Parasitol*. 83:656-62.

Seed, J.R., and S.J. Black. 1999. A revised arithmetic model of long slender to short stumpy transformation in the African trypanosomes. *J Parasitol*. 85:850-4.

Seed, J.R., and J. Sechelski. 1988. Growth of pleomorphic Trypanosoma brucei rhodesiense in irradiated inbred mice. *J Parasitol*. 74:781-9.

Sharma, R., E. Gluenz, L. Peacock, W. Gibson, K. Gull, and M. Carrington. 2009. The heart of darkness: growth and form of Trypanosoma brucei in the tsetse fly. *Trends Parasitol*. 25:517-24.

Sharma, R., L. Peacock, E. Gluenz, K. Gull, W. Gibson, and M. Carrington. 2008. Asymmetric cell division as a route to reduction in cell length and change in cell morphology in trypanosomes. *Protist*. 159:137-51.

Sherwin, T., and K. Gull. 1989. The cell division cycle of Trypanosoma brucei brucei: timing of event markers and cytoskeletal modulations. *Philos Trans R Soc Lond B Biol Sci*. 323:573-88.

Siegel, T.N., D.R. Hekstra, and G.A. Cross. 2008. Analysis of the Trypanosoma brucei cell cycle by quantitative DAPI imaging. *Mol Biochem Parasitol*. 160:171-4.

Siegel, T.N., D.R. Hekstra, L.E. Kemp, L.M. Figueiredo, J.E. Lowell, D. Fenyo, X. Wang, S. Dewell, and G.A. Cross. 2009. Four histone variants mark the boundaries of polycistronic transcription units in Trypanosoma brucei. *Genes Dev*. 23:1063-76.

Siegel, T.N., D.R. Hekstra, X. Wang, S. Dewell, and G.A. Cross. 2010. Genome-wide analysis of mRNA abundance in two life-cycle stages of Trypanosoma brucei and identification of splicing and polyadenylation sites. *Nucleic Acids Res*. 38(15):4946-57.

Simo, G., T. Asonganyi, S.W. Nkinin, F. Njiokou, and S. Herder. 2006. High prevalence of Trypanosoma brucei gambiense group 1 in pigs from the Fontem sleeping sickness focus in Cameroon. *Vet Parasitol*. 139:57-66.

Sturm, N.R., and L. Simpson. 1990. Kinetoplast DNA minicircles encode guide RNAs for editing of cytochrome oxidase subunit III mRNA. *Cell*. 61:879-84.

Sykes, M.L., and V.M. Avery. 2009. Development of an Alamar Blue viability assay in 384-well format for high throughput whole cell screening of Trypanosoma brucei brucei bloodstream form strain 427. *Am J Trop Med Hyg*. 81:665-74.

Szoor, B., I. Ruberto, R. Burchmore, and K.R. Matthews. 2010. A novel phosphatase cascade regulates differentiation in Trypanosoma brucei via a glycosomal signaling pathway. *Genes Dev*. 24:1306-16.

Szoor, B., J. Wilson, H. McElhinney, L. Tabernero, and K.R. Matthews. 2006. Protein tyrosine phosphatase TbPTP1: a molecular switch controlling life cycle differentiation in trypanosomes. *J Cell Biol*. 175:293-303.

Trouiller, P., P. Olliaro, E. Torreele, J. Orbinski, R. Laing, and N. Ford. 2002. Drug development for neglected diseases: a deficient market and a public-health policy failure. *Lancet*. 359:2188-94.

Turner, C.M., N. Aslam, and S.D. Angus. 1996. Inhibition of growth of Trypanosoma brucei parasites in chronic infections. *Parasitol Res*. 82:61-6.

Turner, C.M., N. Aslam, and C. Dye. 1995. Replication, differentiation, growth and the virulence of Trypanosoma brucei infections. *Parasitology*. 111 (Pt 3):289-300.

Tyler, K.M., P.G. Higgs, K.R. Matthews, and K. Gull. 2001. Limitation of Trypanosoma brucei parasitaemia results from density-dependent parasite differentiation and parasite killing by the host immune response. *Proc Biol Sci*. 268:2235-43.

Tyler, K.M., K.R. Matthews, and K. Gull. 1997. The bloodstream differentiation-division of Trypanosoma brucei studied using mitochondrial markers. *Proc Biol Sci*. 264:1481-90.

Ullu, E., K.R. Matthews, and C. Tschudi. 1993. Temporal order of RNA-processing reactions in trypanosomes: rapid trans splicing precedes polyadenylation of newly synthesized tubulin transcripts. *Mol Cell Biol*. 13:720-5.

Van Den Abbeele, J., Y. Claes, D. van Bockstaele, D. Le Ray, and M. Coosemans. 1999. Trypanosoma brucei spp. development in the tsetse fly: characterization of the post-mesocyclic stages in the foregut and proboscis. *Parasitology*. 118 (Pt 5):469-78.

Van den Bossche, P., A. Ky-Zerbo, J. Brandt, T. Marcotty, S. Geerts, and R. De Deken. 2005. Transmissibility of Trypanosoma brucei during its development in cattle. *Trop Med Int Health*. 10:833-9.

Vanhamme, L., and E. Pays. 1995. Control of gene expression in trypanosomes. *Microbiol Rev*. 59:223-40.

Vassella, E., A. Acosta-Serrano, E. Studer, S.H. Lee, P.T. Englund, and I. Roditi. 2001a. Multiple procyclin isoforms are expressed differentially during the development of insect forms of Trypanosoma brucei. *J Mol Biol*. 312:597-607.

Vassella, E., and M. Boshart. 1996. High molecular mass agarose matrix supports growth of bloodstream forms of pleomorphic Trypanosoma brucei strains in axenic culture. *Mol Biochem Parasitol*. 82:91-105.

Vassella, E., R. Kramer, C.M. Turner, M. Wankell, C. Modes, M. van den Bogaard, and M. Boshart. 2001b. Deletion of a novel protein kinase with PX and FYVE-related domains increases the rate of differentiation of Trypanosoma brucei. *Mol Microbiol*. 41:33-46.

Vassella, E., B. Reuner, B. Yutzy, and M. Boshart. 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):2661-71.

Vaughan, S., and K. Gull. 2003. The trypanosome flagellum. *J Cell Sci*. 116:757-9.

Vaughan, S., L. Kohl, I. Ngai, R.J. Wheeler, and K. Gull. 2008. A repetitive protein essential for the flagellum attachment zone filament structure and function in Trypanosoma brucei. *Protist*. 159:127-36.

Veitch, N.J., P.C. Johnson, U. Trivedi, S. Terry, D. Wildridge, and A. MacLeod. 2010. Digital gene expression analysis of two life cycle stages of the human-infective parasite, Trypanosoma brucei gambiense reveals differentially expressed clusters of co-regulated genes. *BMC Genomics*. 11:124.

Vickerman, K. 1965a. Polymorphism and mitochondrial activity in sleeping sickness trypanosomes. *Nature*. 208:762-6.

Vickerman, K. 1965b. Polymorphism and mitochondrial activity in sleeping sickness trypanosomes. *Nature*. 208:762-6.

Vickerman, K. 1985. Developmental cycles and biology of pathogenic trypanosomes. *Br Med Bull*. 41:105-14.

Walrad, P., A. Paterou, A. Acosta-Serrano, and K.R. Matthews. 2009. Differential trypanosome surface coat regulation by a CCCH protein that co-associates with procyclin mRNA cis-elements. *PLoS Pathog*. 5:e1000317.

Wargo, A.R., J.C. de Roode, S. Huijben, D.R. Drew, and A.F. Read. 2007. Transmission stage investment of malaria parasites in response to in-host competition. *Proc Biol Sci*. 274:2629-38.

Wargo, A.R., N. Randle, B.H. Chan, J. Thompson, A.F. Read, and H.A. Babiker. 2006. Plasmodium chabaudi: reverse transcription PCR for the detection and quantification of transmission stage malaria parasites. *Exp Parasitol*. 112:13-20.

Welburn, S.C., K. Picozzi, E.M. Fevre, P.G. Coleman, M. Odiit, M. Carrington, and I. Maudlin. 2001. Identification of human-infective trypanosomes in animal reservoir of sleeping sickness in Uganda by means of serum-resistance-associated (SRA) gene. *Lancet*. 358:2017-9.

Wenzler, T., D.W. Boykin, M.A. Ismail, J.E. Hall, R.R. Tidwell, and R. Brun. 2009. New treatment option for second-stage African sleeping sickness: in vitro and in vivo efficacy of aza analogs of DB289. *Antimicrob Agents Chemother*. 53:4185-92.

WHO. 2000. WHO Report on Global Surveillance of Epidemic-prone Infectious Diseases -African trypanosomiasis *http://www.who.int/csr/resources/publications/CSR_ISR_2000_1tryps/en/index.html*.

WHO. 2003. Global Burden of Disease 2000: version 3. *http://www.who.int/healthinfo/global_burden_disease/estimates_regional_2000_v3/en/index.html*. Accessed 1 June 2010.

WHO. 2006. Human Africa Trypanosomiasis (sleeping sickness): epidemiological update. *Weekly Epidemiological Record*. 81:69-80.

Woodward, R., and K. Gull. 1990. Timing of nuclear and kinetoplast DNA replication and early morphological events in the cell cycle of Trypanosoma brucei. *J Cell Sci*. 95 (Pt 1):49-57.

Ziegelbauer, K., M. Quinten, H. Schwarz, T.W. Pearson, and P. Overath. 1990. Synchronous differentiation of Trypanosoma brucei from bloodstream to procyclic forms in vitro. *Eur J Biochem*. 192:373-8.

Zilberstein, D., and M. Shapira. 1994. The Role Of Ph and Temperature In the Development Of Leishmania Parasites. *Annual Review of Microbiology*. 48:449-470.

Zuker, M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res*. 31:3406-15.

Appendix A: Chemicals and Solutions

Unless otherwise stated, chemicals were purchased from Sigma Aldrich.

Cloning and DNA manipulations

Miniprep Solution I:	50mM Glucose 25mM Tris-HCl pH 8 10mM EDTA pH 8 Autoclaved
Miniprep Solution II:	0.2M NaOH 1% SDS
Miniprep Solution III:	5M KOAc
6x DNA Gel loading buffer:	0.25% bromophenol blue 0.25% xylene cyanol 30% glycerol
1x TAE	40mM Tris-Acetate 1mM EDTA pH8

Northern Solutions

10x MOPS:	23.13g MOPS 50mM Na-Acetate pH 7 10mM EDTA Made up to a final volume of 500ml with dH_2O Autoclaved and stored in the dark at 4°C

RNA Gel Loading Buffer:	150µl formamide
	83µl formaldehyde (37%)
	50µl 10x MOPS
	0.01% Bromophenol blue
	50µl glycerol
	167µl dH_2O
20x SSC:	3M NaCl
	0.3M Tri Sodium Citrate
Hybridisation buffer:	12.5ml 20x SSC
	25ml formamide
	100µl 10% SDS
	10ml 10% DIG Block
Maleic Acid buffer:	100mM Maleic Acid
	150mM NaCl
	pH adjusted to 7.5 with NaOH and autoclaved
10% DIG Block:	10g Blocking Reagent (Roche 14497700) dissolved in
	100ml Maleic Acid buffer
	Autoclaved and stored at -20°C
RNA Wash Buffer:	498.5ml Maleic Acid
	0.3ml Tween-20 (Fisher BP337-100)
RNA Detection Buffer:	100mM Tris-HCl pH 9.5
	100mM NaCl

Trypanosome work

HMI-9 Media:	HMI-9 media was purchased as a powder (Gibco 074-90915N). One bottle of powder was dissolved in 4 litres of dH_2O, to which 15.12g $NaHCO_3$ and 71.5µl β-mercaptoethanol was added and the pH was adjusted to 7.5.
Mowiol:	10% w/v Mowiol (Calbiochem 4759204) 25% w/v Glycerol 0.1M Tris pH8.5

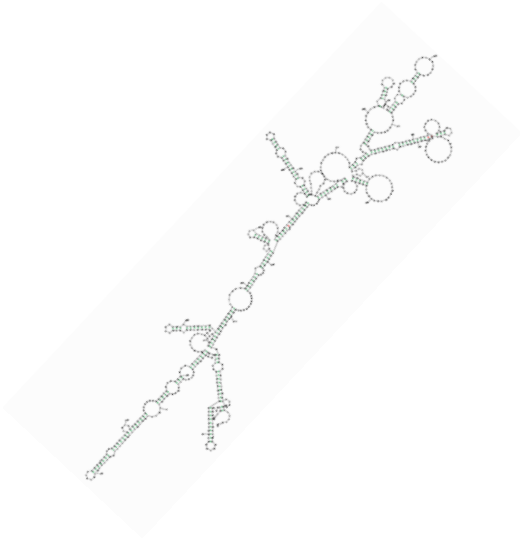
Miscellaneous

10x PBS:	1.37M NaCl 27mM KCl 100mM Na_2HPO_4 18mM KH2PO4 pH to 7.4 with HCl or NaOH
10% SDS	10% w/v sodium dodecyl sulfate in dH_2O

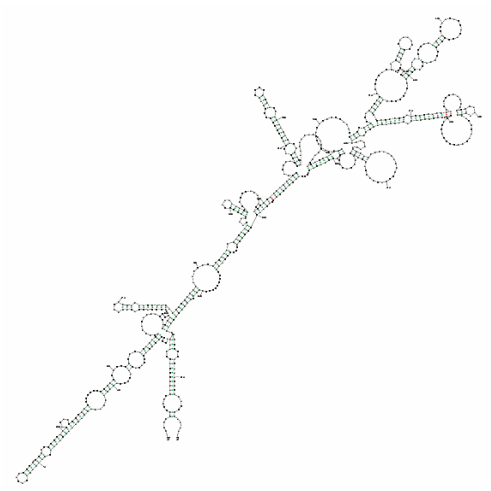
Appendix B: Predicted structures of the PAD1 3'UTR deletion series

Using Sfold analysis (http://sfold.wadsworth.org) (Ding et al., 2005; Ding and Lawrence, 2003), the secondary structure of each of the deletion constructs in the *PAD1* 3'UTR deletion series created in Chapter 3 was predicted using position 710nt as the polyadenylation site. The resulting secondary structures for each deletion are shown below. The results of this analysis are discussed in Chapter 3.

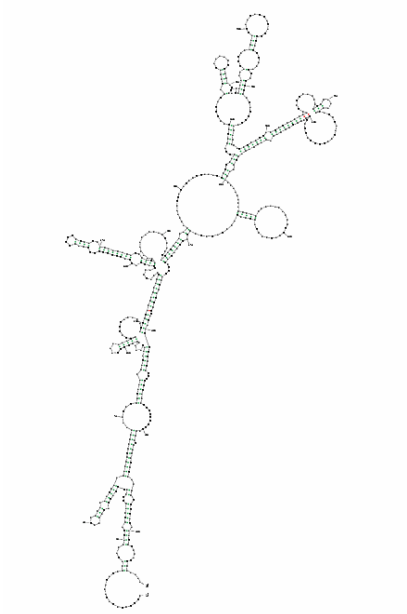
Full length *PAD1* 3'UTR



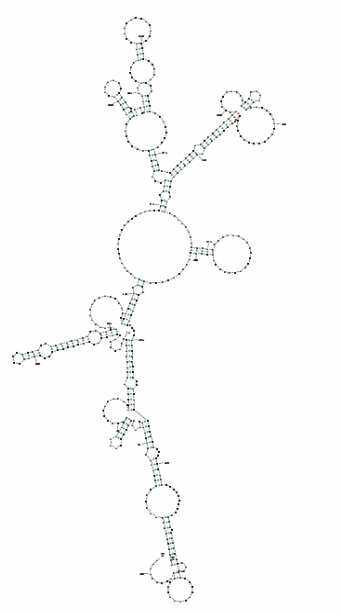
***PAD1* 3'UTR Δ1-21**



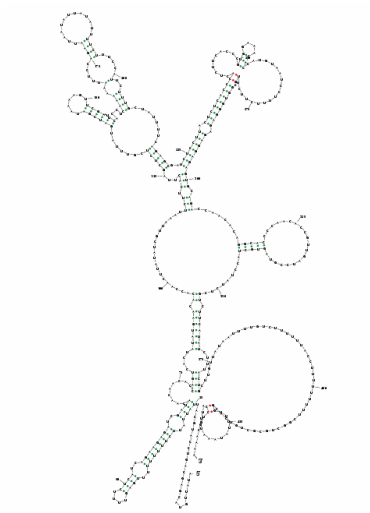
***PAD1* 3'UTR Δ1-145**



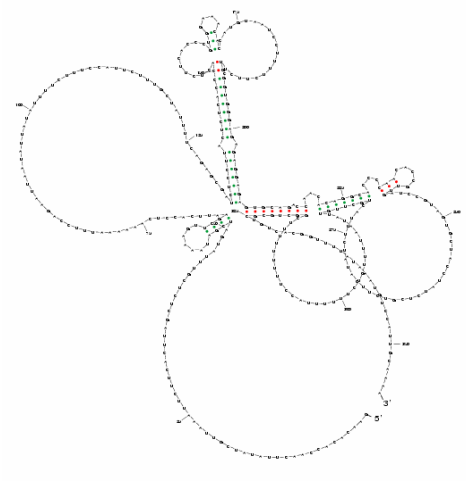
***PAD1* 3'UTR Δ1-207**



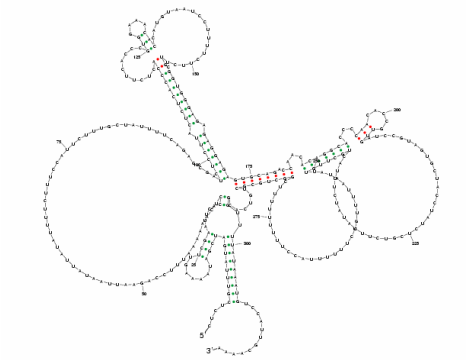
***PAD1* 3'UTR Δ1-265**



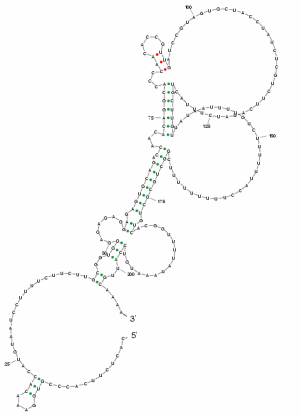
***PAD1* 3'UTR Δ1-354**



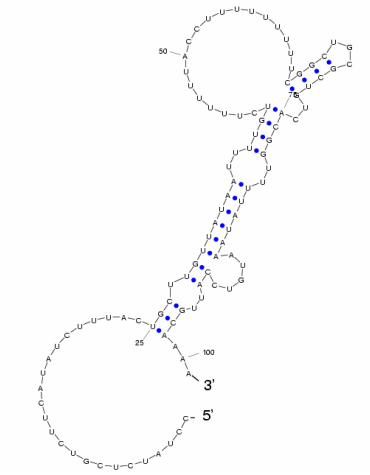
***PAD1* 3'UTR Δ1-391**



***PAD1* 3'UTR Δ1-502**



***PAD1* 3'UTR Δ1-609**



Appendix C: The mathematical model for trypanosome within-host dynamics

The quantitative data generated in Chapter 5 was incorporated into a mathematical model by Dr Nick Savill, University of Edinburgh. Details of how the model is constructed; how the model is fitted; and how the parameters are estimated, **as described by Dr Savill**, are detailed below.

C.1 Construction of the model

The model is constructed as follows. Let the concentration of non-committed slender cells of variant type v at time t be $l_v(t)$. The initial infection by variant type 1 is at time $t=0$. Switching between variants is sequential, i.e., variant v switches to variant $v+1$ and no other. This is different than other models (Lythgoe et al., 2007) where one variant may switch to a number of different variants. We made this assumption in the first instance to reduce the number of parameters in the model. It turns out, however, that our simpler model can give rise to multiple variants being expressed simultaneously (see Chapter 5). It can, therefore, be justified, even though it may not be biologically correct.

Non-committed slender cells replicate at rate α (i.e., a cell-cycle time of $\ln 2/\alpha$). They are cleared by a time-dependent, variant-specific immune response at rate $r_v(t)$. They become committed to differentiate at rate $\beta f(t)$, where $f(t)$ is SIF concentration, and β the differentiation rate. Switching from variant v to variant $v+1$ is at variant specific rate ω_v. Thus the differential equations that describe the dynamics of each non-committed slender variant are:

$$\frac{d}{dt}l_v(t) = (\alpha - \omega_v - \beta f(t) - r_v(t))l_v(t) + \begin{cases} 0 & \text{if } v=1 \\ \omega_{v-1}l_{v-1}(t) & \text{if } v=2,\ldots,\nu \end{cases}$$

where ν is the total number of variant types that appear in a mouse during the experiment. Let $L(t)$ be the total concentration of non-committed slender cells; it is given by:

$$L(t) = \sum_{V=1}^{v} l_v(t)$$

Let the age of differentiated cells since becoming committed to differentiation be a and let $d_v(a,t)$ be the age distribution of differentiated cells of variant type v at time t.

Differentiated cells fall into three classes: replicating, committed slender cells, non-replicating intermediate cells and non-replicating stumpy cells. All committed cells of variant type v are cleared by the immune system at a rate proportional to the clearance rate of non-committed slender cells of the same type, i.e., $r_v(t)$. Thus committed slender cells are cleared by the immune response at rate $\delta_c r_v(t)$. At age τ_c they become intermediate cells. Intermediate cells are cleared at rate $\delta_i r_v(t)$. They become stumpy cells at age τ_i. Stumpy cells are cleared at rate $\delta_s r_v(t)$. They die at age τ_s. Committed slender cells replicate with rate α. Thus the partial differential equations that describe the dynamics of the age distribution of each differentiated variant type v are:

$$\frac{\partial}{\partial t} d_v(a,t) + \frac{\partial}{\partial a} d_v(a,t) = -d_v(a,t) \times \begin{cases} \delta_c r_v(t) - \alpha & \text{if } 0 \le a < \tau_c \\ \delta_i r_v(t) & \text{if } \tau_c \le a < \tau_i \\ \delta_s r_v(t) & \text{if } \tau_i \le a < \tau_s \end{cases}$$

The boundary conditions on these equations are determined by differentiation of non-committed slender cells into age $a = 0$, i.e., $d_v(0,t) = \beta f(t) l_v(t)$, and stumpy death at age τ_s, i.e., $d_v(\tau_s,t) = 0$.

Let $D(a,t)$ be the age distribution of differentiated cells of all variant types at time t. It is given by the sum over all variants:

$$D(a,t) = \sum_{v=1}^{\nu} d_v(a,t)$$

Let $C(t)$ be the total concentration of committed slender cells, let $I(t)$ be the total concentration of intermediate cells, let $S(t)$ be the total concentration of stumpy cells, and let $T(t)$ be the total concentration of cells. These are given by:

$$C(t) = \int_0^{\tau_c} D(a,t)da$$
$$I(t) = \int_{\tau_c}^{\tau_i} D(a,t)da$$
$$S(t) = \int_{\tau_i}^{\tau_s} D(a,t)da$$
$$T(t) = L(t) + C(t) + I(t) + S(t)$$

SIF is produced by non-committed slender cells and committed cells up to age τ_f. SIF is removed at rate γ. Thus the differential equation describing the dynamics of SIF concentration is:

$$\frac{d}{dt} f(t) = L(t) \int_0^{\tau_f} D(a,t) - \gamma f(t)$$

Note that, because SIF is not measured, its concentration is on a dimensionless scale. We can therefore assume that a slender cell produces 1 unit of SIF per hour. SIFs effect on the differentiation of slender cells is quantified through β.

PAD1 is only expressed in differentiated cells. It is scaled relative to the PAD1 expression of mouse 5 on day 4. It rises linearly from 0 up to a level ρ_1 in cells from age 0 to age τ_p respectively. It then rises or falls linearly to a level ρ_2 in cells just before they die at age τ_s. Let $p(a)$ be the relative PAD1 expression of cells of age a, it is given by:

$$p(a)=\begin{cases}\rho_1\dfrac{a}{\tau_p} & \text{if } 0\le a<\tau_p\\ \rho_1+(\rho_2-\rho_1)\dfrac{a-\tau_p}{\tau_s-\tau_p} & \text{if } \tau_p\le a<\tau_s\end{cases}$$

Let $P(t)$ be the total relative PAD1 expression per ml of blood at time t. It is found by multiplying the concentration of cells of age a by their relative PAD1 expression and then integrating over all ages, i.e.:

$$P(t)=\int_0^{\tau_s} p(a)D(a,t)da$$

The immune response against trypanosomes is multifactorial and highly complex, and only qualitatively understood at best. A detailed mathematical model of the immune response is, therefore, of little use when no data is available to fit to. Instead, we use a single variable $r_v(t)$, that determines the clearance rate of each variant type v. We assume that the immune response against variant type v is activated at a rate ψ by replicating slender cells of variant type v, and self-enhances at rate ϕ. We also assume that the initial immune response to the first variant is only activated after a certain time τ_r. The differential equations that describe the immune-mediated clearance rates of each variant type v are:

$$\frac{d}{dt}r_v(t)=\begin{cases}0 & \text{if } t<\tau_r\\ \phi r_v(t)+\psi l_v(t) & \text{if } t\ge\tau_r\end{cases}$$

Naive mice are infected with non-committed slender cells of variant type 1 at a concentration λ. Therefore the initial conditions are $l_1(0)=\lambda$, $l_v(0)=0$ for $v=2,\ldots,\nu$, $d_v(a,0)=0$ for all v and a, $r_v(0)=0$ for all v, and $f(0)=0$. These imply $L(0)=\lambda$, $D(a,0)=0$ for all a and $C(0)=I(0)=S(0)=T(0)=P(0)=0$. All variables, functions and parameters are listed in Table 1.

Independent variables		
t	Time since infection	h
a	Age of differentiated cells	h
Dependent variables		
$l_v(t)$	Concentration of non-committed slender cells of variant type *v*	cells ml^{-1}
$d_v(a,t)$	Age distribution of differentiated cells of variant type *v*	cells $ml^{-1}h^{-1}$
$r_v(t)$	Immune-mediated clearance rate of replicating slender cells of variant type *v*	h^{-1}
$f(t)$	SIF concentration	dimensionless
$L(t)$	Total concentration of non-committed slender cells	cells ml^{-1}
$C(t)$	Total concentration of committed slender cells	cells ml^{-1}
$I(t)$	Total concentration of intermediate cells	cells ml^{-1}
$S(t)$	Total concentration of stumpy cells	cells ml^{-1}
$T(t)$	Total concentration of all cells	cells ml^{-1}
$D(a,t)$	Age distribution of differentiated cells	cells $ml^{-1}h^{-1}$
$P(t)$	Total relative PAD1 expression per ml of blood	ml^{-1}
Functions		
$p(a)$	Relative PAD1 expression of cells of age *a*	$cells^{-1}$
Parameters		
ν	Number of variant types	dimensionless
α	Replication rate of slender cells	h^{-1}
β	Differentiation rate	h^{-1}
γ	SIF removal rate	h^{-1}
δ_c	Proportionality constant for immune-mediated clearance of committed slender cells	dimensionless
δ_i	Proportionality constant for immune-mediated clearance of intermediate cells	dimensionless
δ_s	Proportionality constant for immune-mediated clearance of stumpy cells	dimensionless
τ_c	Lifespan of committed slender cells	h
τ_i	Lifespan of intermediate cells	h
τ_s	Lifespan of stumpy cells	h
τ_p	Duration of initial rise in PAD1 expression	h
τ_f	Duration of SIF production of committed cells	h
τ_r	Time until initial activation of immune response	h
ρ_1	Relative PAD1 expression of cells age τ_p	$cells^{-1}$
ρ_2	Relative PAD1 expression of cells age τ_s	$cells^{-1}$
ϕ	Self-induced growth rate of immune response	h^{-1}
ψ	Committed slender cell-induced growth rate of immune response	h^{-1}
λ	Initial concentration of variant type 1	cells ml^{-1}
ω_v	Switch rate from variant type *v* to variant type *v* + 1	h^{-1}

Table C.1 Variables, functions and parameters used in the model.

C.1.1 Cell-Concentration induced differentiation

Other studies have suggested that differentiation is cell-concentration induced rather than SIF induced (Lythgoe et al., 2007; Turner et al., 1995; Tyler et al., 2001). We tested this hypothesis by removing $f(t)$ from the model and replacing $\beta f(t)$ by $\beta' T(t)$ i.e.:

$$\frac{d}{dt} l_v(t) = (\alpha - \omega_v - \beta' T(t) - r_v(t)) l_v(t) + \begin{cases} 0 & \text{if } v = 1 \\ \omega_{v-1} l_{v-1}(t) & \text{if } v = 2, \ldots, \nu \end{cases}$$

and by changing the boundary condition $d_v(0,t) = \beta' T(t) l_v(t)$.

The predicted dynamics of the model with concentration-induced differentiation and the standardised residuals are shown in Chapter 5. This model gives a poor fit to the data with several outliers and systematic biases particularly around the first peak. We can quantify how much our belief in the two models changes given the data. The ratio of our posterior beliefs in the two models is given by the (integrated) likelihood ratio (also called an evidence ratio or a Bayes factor) of the models given the data multiplied by the ratio of our prior beliefs in the two models before the data were collected, i.e.,

$$\frac{\text{Prob}(\text{model}_1 \mid \text{data})}{\text{Prob}(\text{model}_2 \mid \text{data})} = \frac{\text{Prob}(\text{data} \mid \text{model}_1)}{\text{Prob}(\text{data} \mid \text{model}_2)} \cdot \frac{\text{Prob}(\text{model}_1)}{\text{Prob}(\text{model}_2)}$$

where model_1 is the SIF-induced differentiation model and model_2 is the concentration-induced differentiation model. (The symbol '|' means 'given' or 'conditional upon'). The likelihoods are products over all mice because we assume that the data from each mouse is independent of the data from all the other mice, i.e.,

$$\text{Prob}(\text{data} \mid \text{model}) = \prod_{i=1}^{6} \text{Prob}(\text{data from mouse } i \mid \text{model})$$

If our prior beliefs in the two models are equal, i.e., our prior knowledge base gives us no reason to choose between them, then

$$\frac{\text{Prob}(\text{model}_1)}{\text{Prob}(\text{model}_2)} = 1$$

Of course, our prior knowledge base may lead us to have a stronger belief in one of the models, and therefore this ratio will not equal 1. The model fitting algorithm calculates the likelihoods: these are $\text{Prob}(\text{data} \mid \text{model}_1) = e^{-467}$ and $\text{Prob}(\text{data} \mid \text{model}_2) = e^{-624}$. Therefore the likelihood ratio is

$$\frac{\text{Prob}(\text{model}_1 \mid \text{data})}{\text{Prob}(\text{model}_2 \mid \text{data})} = e^{157} = 10^{68}$$

Thus the data overwhelmingly support the SIF-induced differentiation model irrespective of our initial belief in either model.

C.2 Model fitting, parameter estimation and adequacy of fit

C.2.1 Likelihood

The measurement errors in the ZFP3 and PAD1 Ct values are assumed to be normally distributed. In other studies, the standard deviation of the errors in $\log_{10}$-transformed concentrations derived from qPCR is approximately 0.20 (Mideo et al., 2008; Miller et al., 2010). This means that the standard deviation of the errors in Ct values (which are on a $\log_2$ scale) is approximately $0.2\log_2 10$ which equals 0.66.

The model is fitted to each mouse's $\log_{10}$-transformed parasite concentrations and relative Ct-values for PAD1 expression. Let $\sigma_L = 0.2$ be the standard deviation in the normally distributed errors of $\log_{10}$-transformed parasite concentrations, and let $\sigma_P = 0.66$ be the standard deviation in the normally distributed errors of $\log_2$-transformed PAD1 expression.

For particular numerical values of the model parameters, the model is solved numerically for each mouse. In order to quantify the fit of the model with these parameters to the data we proceed by calculating the log-likelihood of the model solution at each data point. The value of the likelihood is given by the normal probability density function because the errors are normally distributed. So, for an observed parasite concentration $L_{\text{data}}(t)$, at time t above the detection limit (about 10^6 parasites/ml), the log-likelihood is proportional to:

$$\ln\left(\phi\left(\frac{\log_{10}L_{\text{data}}(t)-\log_{10}L_{\text{model}}(t)}{\sigma_L}\right)\right)$$

where $\phi(x)$ is the normal probability density function. For an observed parasite concentration below the detection limit at time t the log-likelihood is proportional to:

$$\ln\left(\Phi\left(\frac{6-\log_{10}L_{\text{model}}(t)}{\sigma_L}\right)\right)$$

where $\Phi(x)$ is the normal cumulative distribution function. For an observed relative PAD1 expression $P_{\text{data}}(t)$, at time t above the detection limit, the log-likelihood is proportional to:

$$\ln\left(\phi\left(\frac{\log_{10}P_{\text{data}}(t)-\log_{10}P_{\text{model}}(t)}{\sigma_P}\right)\right)$$

The lowest $\log_2$ relative PAD1 expression is -7.52, we take this value as the detection limit, although it may be smaller. For a relative PAD1 expression below the detection limit at time t the log-likelihood is proportional to:

$$\ln\left(\Phi\left(\frac{-7.52-\log_{10}P_{\text{model}}(t)}{\sigma_P}\right)\right)$$

The total log-likelihood for a mouse is the sum of the log-likelihoods over all the data for that mouse.

The parameter posterior distribution is found by multiplying the likelihood by the prior distributions, which are given in the next section. Samples from the posterior are drawn using an adaptive population based Markov chain Monte Carlo algorithm (Miller et al., 2010; Savill et al., 2009).

C.2.2 Parameter Priors

The number of variant types ν that arose during the experiment is unknown. On the one hand there must be more than one because of the multiple peaks in parasite concentration, but on the other we, do not want the model to over-fit the data with hundreds of variant types. We therefore chose a prior of $1+N_T(0,5)$ which allows for tens of variant types but penalises the model for having too many. (Note, $N_T(\mu,\sigma)$ is a normal probability density function with mean μ, standard deviation σ but truncated at 0.) Slender cell-cycle time has been estimated to be around 4-5 hours, (Seed and Black, 1997; Vassella et al., 1997) although Savill and Seed (2004) estimated it around 2 hours. We therefore chose a prior on α as $N_T(0.1,0.1)$ in order to incorporate this uncertainty. Savill and Seed (2004) estimated the differentiation rate β to be between 0.5 and $3\times10^{-9}h^{-1}$, we therefore chose a prior of $N_T(10^{-9},10^{-9})$. SIF removal rate γ was estimated to be between 0.2 and $1.4h^{-1}$ (Savill and Seed, 2004), we therefore chose a prior $N_T(0.5,0.5)$. There is no quantitative prior information about immune-mediated clearance of committed cells. It is thought, however, that the immune system is less effective at clearing stumpy cells than slender cells (Engstler et al., 2007). We therefore chose priors of Uniform(0,1) for δ_c, δ_i, δ_s to reflect this uncertainty. Savill and Seed (2004) estimated the committed slender cell lifespan τ_c, to be between 8 and 12h, we therefore chose a prior of $N_T(10,4)$. Vassella *et al* (1997) estimated the time from cell-cycle exit to the mitochondrion metabolic

activity (when cells become morphologically stumpy) to be between 8 and 10h. Savil and Seed (2004) estimated this to be between 3 and 8h. We therefore chose a prior of $N_T(6,6)$ on the lifespan of intermediate cells τ_i. Black *et al.,* (1982) estimated the lifespan of stumpy cells τ_s, to be between 24 and 36h. Turner *et al.,* (1995) estimated it to be between 48 and 72h, and Savill and Seed (2004) around 58h but with very wide confidence intervals. We therefore chose a prior of $N_T(48,24)$ to cover this uncertainty. The initial rise in PAD1 expression is thought to last until the end of the intermediate stage. Given the priors on τ_c and τ_i we set a prior on τ_p of $N_T(10+6,\sqrt{(4^2+6^2)})= N_T(16,7)$. Peak total relative PAD1 expression is about 2^5. This occurs at a peak parasitaemia of around $10^{8.5}$ parasites/ml which is composed mainly of stumpy cells. Therefore relative PAD1 expression per stumpy cell is about $2^5/10^{8.5}=10^{-7}$. Therefore we chose a prior of $N_T(10^{-7},10^{-7})$ on ρ_1 and ρ_2. Savill and Seed (2004) suggested that committed slender cells produced SIF, we therefore set the prior for the duration of SIF production of committed cells τ_f as $N_T(10,4)$. Mice were infected intra-peritoneally with about 10^3 slender cells. We therefore have a prior of $N(3,0.5)$ on $\log_{10}\lambda$. In Lythgoe *et al.,* (2007) switch rates between variants were allowed to vary over six orders of magnitude with an average switch rate of 0.01 per population doubling, which is about $0.0014h^{-1}$ for a cell-cycle time of 7h. We therefore had a very broad prior on $\log_{10}\omega_v$ of $N(-6,2)$. The prior on the initial activation of the immune response τ_r, was taken as $N_T(168,24)$, i.e., around 7 days as is usual in acute infectious diseases of naïve animals. The growth of the immune responses against variants should occur on the order of several days. We therefore take broad priors on ϕ and ψ of $N_T(0.1,0.1)$ and $N_T(10^{-8},10^{-8})$ respectively.