THE MECHANISM OF DUPLICATIVE *VSG* GENE ACTIVATION DURING ANTIGENIC VARIATION IN *TRYPANOSOMA BRUCEI*

PETER BURTON
WELLCOME CENTRE FOR MOLECULAR PARASITOLOGY
THE ANDERSON COLLEGE
UNIVERSITY OF GLASGOW

Submitted for the Degree of Doctor of Philosophy

September 2003

DECLARATION

I declare that this thesis and the results presented in it are entirely my own work except were otherwise stated.

PETER BURTON

ACKNOWLEDGEMENTS

Firstly, I thank Jude and my family for their continual support and encouragement. This has made it possible for me to reach this stage.

I would also like to thank Dave for his supervision, discussion and support throughout this project. Additionally, further gratitude goes to all the members of the lab, past and present, for their help and discussion. I am particularly grateful to Richard for his constant guidance during the project and the reading of this thesis, and to Liam 'just the one' Morrison for help with the rodents.

Last, but not least, I must thank the Elves of Elbrook Court for the music.

CONTENTS

FIGURES LIST

CHAPTER 1
(INTRODUCTION)

CHAPTER 3
(THE MECHANISM OF DUPLICATIVE ACTIVATION OF *VSG* GENES IN THE PLEOMORPHIC ILTAT 1.2 TRYPANOSOME LINE)

CHAPTER 5

(THE ROLE OF *RAD51* IN *T. BRUCEI* ANTIGENIC VARIATION IN THE PLEOMORPHIC ILTAT 1.2 LINE)

TABLE LIST

CHAPTER 2

(MATERIALS AND METHODS)

CHAPTER 3

(THE MECHANISM OF DUPLICATIVE ACTIVATION OF *VSG* GENES IN THE PLEOMORPHIC ILTAT 1.2 TRYPANOSOME LINE)

CHAPTER 4

(*IN VITRO* STUDIES ON THE MECHANISM OF DUPLICATIVE ACTIVATION OF *VSG* GENES)

CHAPTER 5

(THE ROLE OF *RAD51* IN *T. BRUCEI* ANTIGENIC VARIATION IN THE PLEOMORPHIC ILTAT 1.2 LINE)

ABBREVIATIONS

BC	basic copy
BES	bloodstream expression site
BIR	break-induced replication
BLE	bleomycin resistance protein gene
BSF	bloodstream form
bp	base pair(s)
cDNA	complementary DNA
DSB	double-strand break(s)
DMSO	dimethylsulfoxide
EATRO	East African Trypanosomiasis Research Organisation
ELC	expression linked copy
ESAGs	expression site associated genes
ESB	expression site body
ESI	electrospray ionisation
EST	expressed sequence tag
GPI	glycophosphatidylinositol
ILTat	International Laboratory for research on animal diseases, Trypanozoon antigen type
kb	kilobase(s)
MALDI-TOF	Matrix assisted laser deabsorption ionisation - time of flight
Mb	megabase(s)
MES	metacyclic expression site
MMR	mismatch repair
MS	Mass spectrometry
NCBI	National Centre for Biotechnology
NHEJ	non-homologous end joining
ORF	open reading frame
PCR	polymerase chain reaction
RT-PCR	reverse transcriptase polymerase chain reaction
SDSA	synthesis-dependent strand annealing
SRA	serum resistance-associated gene

SSA	single strand annealing
STR	subtelomeric repeats
TREU	Trypanosomiasis Research Edinburgh University
UTR	untranslated region
VSG	variant surface glycoprotein

ABSTRACT

Trypanosoma brucei is a parasitic protozoan of the order Kinetoplastida. In sub-Saharan Africa it causes nagana in livestock and wild mammals, and sleeping sickness in humans. It is transmitted between mammalian hosts by the tsetse fly vector and during its life in both hosts it proceeds through a unidirectional life cycle, involving several morphological and molecular changes as adaptations to life in different environments.

During growth in the mammalian host the parasite undergoes the phenomenon of antigenic variation in order to escape the immune system of the host. A highly immunogenic surface protein known as the variant surface glycoprotein (VSG) is expressed, preventing the immune response from reacting with any other invariant surface proteins. Periodically, the VSG on the surface is altered to one with immunologically distinct epitopes, making the immune response specific to the prior VSG ineffective. *VSG*s, in the bloodstream form, are expressed from specialised polycistronic, telomeric transcription units named bloodstream expression sites (BESs). There are approximately 20 BESs in the genome, only one of which is active at any time. Different *VSG* genes can be activated by turning on the expression of a different BES and silencing the previously active one, a process termed *in situ* switching. There are, however, an estimated 1000 silent *VSG* genes; some of these reside in the subtelomeres of the minichromosomes whereas the rest are located in chromosome-internal arrays. Chromosome-internal and minichromosomal *VSG* genes can only be activated by moving the gene sequence into the active BES, replacing the previously expressed *VSG*. This transposition most frequently occurs through a duplication of the *VSG* gene. It has been proposed to occur by homologous recombination as sequence homologies appear to mark the limits of such events.

The *VSG* switching rate varies depending on the strain examined, ranging from as rapid as ~2×10^{-2} switches/cell/generation to as infrequent as $1 \times 10^{-6} - 1 \times 10^{-7}$ switches/cell/generation. Those strains that switch at the lowest rates have been laboratory adapted by serial syringe passaging in mice over many years and are monomorphic, having lost the ability to naturally differentiate from the rapidly dividing bloodstream form. Those with a rapid *VSG* switch rate are pleomorphic and are naturally competent to complete a full life cycle. Not only is there a difference in the switching rate, however, there is also a difference in the method of switching employed. Monomorphic lines predominately switch *VSG* expression *via* the *in situ* mechanism, whereas at least one pleomorphic line switches

predominately by duplicative transposition. Also, the duplicative transpositions analysed in the pleomorphic ILTat 1.2 line all utilise 70-bp repeat sequence, a repetitive DNA element upstream of almost all *VSG* genes, as the upstream conversion limit, whereas monomorphic lines tend to use chance homologies. Combining these data leads to the hypothesis that the monomorphic lines have lost the ability to drive *VSG* switching at a high rate due to the loss of some factor(s), or the deregulation of a certain recombination pathway.

The main aim of this thesis was to test the hypothesis that pleomorphic trypanosomes switch *VSG* genes by a different mechanism to the monomorphic lines. The ILTat 1.2 line was used as it is amenable to study due to a switching rate (1×10^{-5} switches/cell/ generation) that is low enough to enable analysis of individual switch events, but it remains pleomorphic. It was also demonstrated that it could be cultured *in vitro* and genetically manipulated by DNA transformation.

To determine the influence of DNA sequence homology in individual switch events, the conversion limits of 5 duplicative transpositions of the ILTat 1.22 gene (a telomeric *VSG* gene residing in a metacyclic expression site) were analysed in detail. This gene was chosen as previous restriction mapping of 3 of these switches had determined that the upstream conversion limit lies within a 400-bp fragment containing a very short (115 bp) 70-bp repeat region. After confirming that the upstream conversion limit of the other 2 switch events occurred within this region, a linker-based PCR method was used to sequence the conversion limit. Despite having only a very short repeat region it was used in several different ways. Thus, the *VSG* switching reactions were not site-specific events, but rather recombination events utilising only short lengths of homology. In 2 cases the maximum length of homology used was at most 40-bp. After isolating the region between the ILTat 1.22 basic copy and the telomere by PCR, the downstream conversion limits of these 5 switch events were mapped. In every case the conversion appeared to continue all the way to the telomere. The upstream conversion limit of the duplicative activation of a second telomeric gene, ILTat 1.64, was shown to have most likely occurred within the 70-bp repeat region, which for this gene is approximately half a single repeat. If this is the case then this is the smallest 70-bp repeat to have been demonstrated to be used to catalyse a *VSG* duplicative activation.

Attempts were made to determine upstream sequences of an internal *VSG* gene, ILTat 1.71, and map the conversion limits of its duplicative transposition. However, due to its absence from a phage library these attempts were unsuccessful. An analysis, *via* the *T. brucei*

genome project, of internal *VSG* gene arrays showed that the presence of short 70-bp repeat regions is commonplace, with the average being only 91-bp. It also showed that at least two-thirds of the genes were pseudogenes, making a case for the importance of mosaic gene conversion, a process involving the conversion of only parts of individual *VSG* genes.

As single 70-bp repeat units, or less, can be utilised for catalysing *VSG* duplicative transposition, an assay was developed to study the importance of different regions of this repeat unit. The principle behind the assay was to analyse the rate of events utilising an integrated 70-bp repeat region, and subsequent manipulations of it, to move a selectable marker into the active expression site. This assay was demonstrated to work in the monomorphic line Lister 427, but the recombination occurred at a very low frequency (~ 1 in 5 x 10^7 cells) even with a wild-type repeat region. No gene conversion events could ever be detected in the ILTat 1.2 line, making this assay unusable without further improvements. Preliminary attempts were also made to isolate proteins binding to the 70-bp repeats. Despite obtaining DNA binding proteins, no 70-bp repeat specific proteins were isolated. Nor, in further experiments, could a firm DNA metabolic activity acting specifically on the 70-bp repeats be demonstrated *in vitro*.

Proteins of the RAD51 family are central to homologous recombination in all organisms, and it has been shown that mutation of the best candidate for *T. brucei RAD51* causes a reduction in *VSG* switching frequency in monomorphic trypanosomes. However, as the switch events activating the ILTat 1.22 gene appeared to look more like the *RAD51*-independent mechanism of homologous recombination break-induced replication (*i.e.* small homologies initiate a conversion that continues until the end of the chromosome), this thesis asked whether or not RAD51 is crucial to *VSG* switching in pleomorphic trypanosomes. With difficulty, two independent homozygous mutants of *rad51* were made in the ILTat 1.2 line. They both displayed the expected *in vivo* growth rate, but did produce a relapse peak in a rat host, implying that *VSG* switching was occurring. Further analysis showed that, unlike the monomorphic trypanosomes, mutation of *rad51* did not reduce *VSG* switching frequency. This demonstrates that *VSG* switching, in at least this pleomorphic line, is genetically and mechanistically different to *VSG* switching in the monomorphic Lister 427 line, indicating the importance of working on pleomorphic strains.

CHAPTER 1

INTRODUCTION

1.1 General introduction

Protozoan parasites cause devastation to populations worldwide, delivering disease and death to the human and animal inhabitants. One such parasite is *Trypanosoma brucei*, the causative agent of human sleeping sickness and Nagana in livestock. According to the World Health Organisation, as of March 2001, over 60 million people are at risk from *T. brucei* infection, with 300 000 to 500 000 infected (http://www.who.int/inf-fs/en/fact259.html). The parasite is endemic across sub-Saharan Africa, causing a disease that manifests in two clinically different forms depending on the morphologically indistinguishable sub-species of the infecting trypanosome. *T. brucei gambiense* causes a chronic form of the disease which can take many years to kill the host, whereas *T. brucei rhodesiense* often causes death within the first six months of the infection. The former of these subspecies is common to west and central Africa whereas the latter is common to central and east Africa, a separation roughly following the divide created by the Rift Valley (Welburn, 2001). A third subspecies, *T. brucei brucei*, is not human infective and causes Nagana in the wildlife and livestock found throughout sub-Saharan Africa.

African trypanosomes are unicellular, flagellated protozoan of the order Kinetoplastida. Characteristic of parasites in this order is the presence of a single large mitochondrion. In *T. brucei*, the DNA of this large self-perpetuating organelle, called the kinetoplast, is found at the base of the flagellum (Figure 1.1).

The parasite proliferates in the bloodstream, capillary beds and tissue fluids of the mammalian host and is transmitted between these hosts by its vector the tsetse fly (*Glossina* spp.). The adaptation to transmission by the tsetse fly is evident by an intricate life cycle, which can be broadly categorised into the mammalian stages and the tsetse stages, and also, within each of these stages into replicative and non-replicative transmission forms (Figure 1.2; reviewed in Vickerman, 1985 and Barry and McCulloch, 2001). All of the differentiation steps are thought to be irreversible.

Upon biting, a tsetse fly will release, into the connective tissue of the host, thousands of metacyclic trypanosomes that will shortly find their way into the bloodstream. The metacyclic form is a preadaptation to life in the mammalian host that is first detected in the salivary glands of the tsetse fly. It lies in the G0 stage of the cell cycle so is non-replicative, a feature common to all transmission stages in the parasite life cycle.

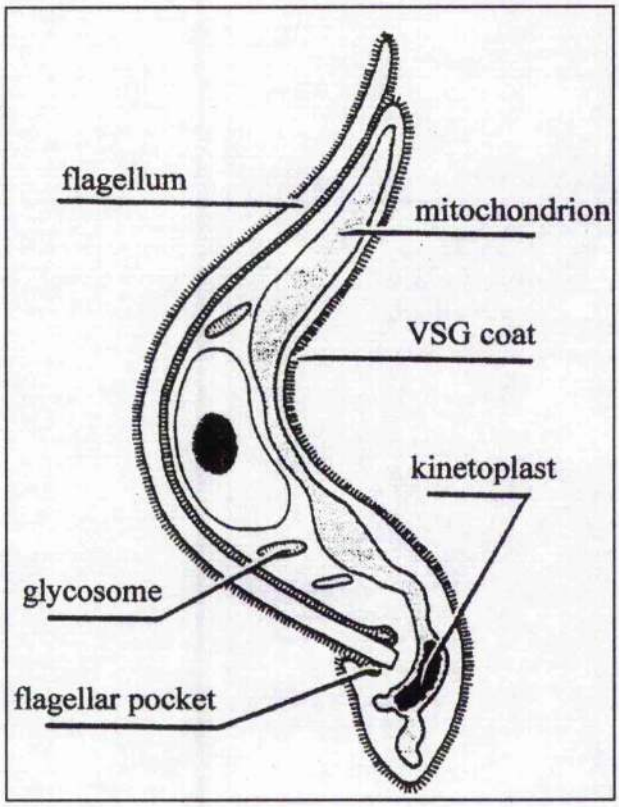


Figure 1.1 A representation of a longitudinal section of a *Trypanosoma brucei* cell. All the major structural features are marked. After Borst (1991).

The variant surface glycoprotein (VSG), a protein covering the surface of the parasite in the mammalian stages of the parasite, is also first detectable on the surface at this stage. Upon entering the mammalian host, the parasite emerges from quiescence and begins dividing rapidly by mitosis, concomitant with a morphological change that provides this stage with the name: the long-slender bloodstream form. Here, the parasite is ubiquitously covered in VSG and has the characteristic population dynamic of a fluctuating parasitemia, where parasite numbers rise and fall as the infection progresses with time. This fluctuating parasitaemia is, in part, due to removal of the parasites by the host immune system, followed by re-growth of a subset of the population that has evaded this response (see Section 1.2). Also responsible for maintaining the parasite at a sublethal level, however, is differentiation into the non-dividing stumpy form (Tyler *et al*, 2001). The cause of this differentiation appears to be density-dependent, and is believed to be initiated by an inducing factor, termed stumpy induction factor (SIF) (Vassella, 1997). Although the putative SIF is yet to be identified, it appears relatively stable, as culture medium that

contained differentiating trypanosomes could induce stumpy differentiation in other cultures (Reuner *et al* 1997). It is also small, being under 500 Da, and appears to act *via* a cyclic AMP (cAMP) pathway to induce growth arrest (Vasella *et al*, 1997). *ESAG4* (expression site-associated gene) is a member of a family of adenylate/guanylate cyclase encoding genes (Alexandre *et al*, 1990; Alexandre *et al*, 1996) that also contains several related genes related to *ESAG4* termed *GRESAG4*. Proteins encoded by these genes are involved in regulating the levels of intracellular cAMP, by converting ATP to cAMP, and are implicated in the differentiation of the bloodstream form to the procyclic forms (Rolin *et al*, 1993). Perhaps the SIF acts as a ligand that interacts with the extracellular domains of the adenylate cyclases, causing an increase of cAMP within the cell that can trigger differentiation via cell cycle regulation (Vasella *et al*, 1997). More recently, laboratory-adapted lines unable to differentiate naturally from the long-slender dividing form (see below) have been differentiated into the stumpy form when treated with the membrane-permeable cAMP derivative 8-(4-chlorophenylthio)-cAMP (Breidbach *et al*, 2003).

Once the cells have become stumpy they can survive only by differentiating in the tsetse fly, suggesting that this life cycle stage is a pre-adaptation to life in the tsetse fly. The long-slender form metabolises glucose efficiently via glycolysis, and has a repressed mitochondrion (Vickerman, 1965). The short-stumpy parasites, however, activate some components of the mitochondrial respiratory chain, including proline and alpha-ketoglutarate oxidases. As the abundant quantities of glucose available to the parasite in the bloodstream of the mammalian host are depleted from the bloodmeal following tsetse fly feeding, an amino acid-based energy metabolism is utilised in the fly.

To date, two other proteins (tbZFP1 and tbZFP2), from a rare CCCH zinc finger family, have been implicated in the control of differentiation from the bloodstream form to the procyclic form. tbZFP1 is transiently enriched during this differentiation step and expressed in the procyclic stage, whereas tbZFP2 appears to be essential for this progression (Hendriks *et al*, 2001).

The differentiation from the stumpy form to the proliferative procyclic stage take place in the endoperitrophic space of the midgut of the tsetse fly. This differentiation step also appears to be synchronous (Ziegelbauer *et al*, 1990; Matthews and Gull, 1994). The morphological changes are several-fold: the cell lengthens, the VSG coat is shed and replaced by a coat of the procyclin proteins, and the mitochondrion enlarges. The procyclin coat, unlike the *VSG* coat, can be heterogenous, as different forms of the protein are present

on the surface at one time. The procyclins vary in that they contain either internal EP dipeptide repeats, or GPEET pentapeptide repeats

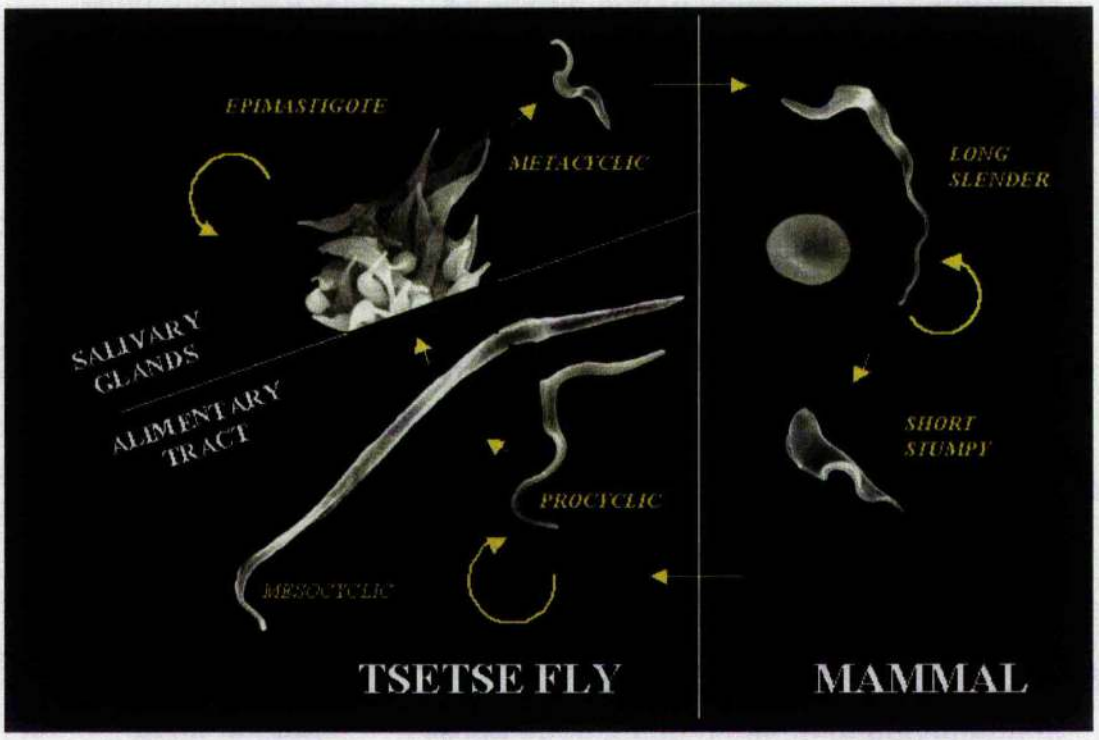


Figure 1.2 The *Trypanosoma brucei* life cycle. Straight arrows represent irreversible differentiations, wheras the curled arrows indicate the replicative stages of the parasite. The white lines demarcate the habitat of the stages of the life cycle. See text for further explanation. From Barry and McCulloch (2001).

(reviewed in Roditi *et al*, 1998). Also, procyclin proteins display a heterogeneity in their carbohydrate components, including *N*-glycosylation of the EP form (Treumann *et al*, 1997), and the GPEET form can be phosphorylated (Bütikofer *et al*, 1997). The function of this coat is still unknown, although potentially it protects against hydrolases in the midgut of the fly, and could have a role in tropism (Roditi *et al*, 1998).

Within the first three days following tsetse feeding, the trypanosome population size decreases by several orders of magnitude (Van Den Abbeele, 1999). A high proportion of undifferentiated long-slender forms in the bloodmeal does not account for the large drop in trypanosome numbers as a population containing ~70 % stumpy forms decreased from an estimated 7.5×10^6 to $1\text{-}2 \times 10^3$ in 3 days (Van Den Abbeele, 1999). Following this drop in number, the parasites colonise rapidly the ectoperitrophic space of the midgut until a stationary number of 5×10^5 per midgut is reached and maintained (Van Den Abbeele, 1999). During expansion, the parasites move towards the anterior of the midgut, coinciding

with an elongation of body length associated with the differentiation to the mesocyclic stage. These mesocyclic trypanosomes eventually traverse through to the foregut and the proboscis. In the proboscis, the parasite then replicates its DNA to create a long 4N epimastigote trypanosome and it is this life cycle form that is predicted to make the journey to the salivary glands. The tetraploid cell then divides asymmetrically to create a long and a short epimastigote trypanosome, each with a 2N DNA content. It is only the short epimastigote trypanosome that proceeds in the life cycle, with the long epimastigote cell presumably failing to survive or differentiate. The proliferating epimastigote parasites then become attached, via their flagella, to the microvilli of the epithelial cells of the salivary gland. This attachment is a prerequisite for the formation of the infective metacyclic cells that coincides with the cessation of cell division, the gain of the surface VSG coat and the repression of the mitochondrion.

1.2 Antigenic variation

Although the ability of the mammalian immune system to deal with infective agents is considerable, many parasites have gained the ability to 'deceive' the immune response and establish chronic infections, this phenomenon is termed antigenic variation. Many infectious agents, through viruses to eukaryotes, have utilised the phenomenon of antigenic variation to extend the lifespan of infections in the host, thus increasing the likelihood and incidence of transmission between these hosts (for a comprehensive review on antigenic variation see the book 'Antigenic Variation', edited by Craig and Scherf, 2003).
One form of antigenic variation involves the expression of a highly immunogenic surface protein that, by forming a dense coat, prevents the immune response from reacting with any other surface entities. This surface protein periodically changes to an immunologically distinct type, thus rendering the immune response to the previously expressed protein ineffective against the parasite expressing the novel type. This form of antigenic variation allows the survival of the parasite population within the host at the expense of the majority of individuals, which are removed by immune responses due to the immunogenic nature of the variation. It is this mechanism that is employed by *T. brucei* (recently reviewed in Barry and McCulloch, 2001; Borst, 2001; Donelson, 2003).
The genus *Trypanosoma* contains a variety of species that are capable of infecting every vertebrate class, but they can be broadly divided, via rDNA and protein coding gene

sequence phylogenetic analysis, into two distinct evolutionary branches, namely Salivaria (*T. brucei, T. equiperdum, T. evansi, T. congolense, T. simiae* and *T. vivax*) and Stercoraria (Haag *et al* 1998; Overath *et* al 2001). One difference between Salivaria and Stercoraria, along with their predominant means of transmission, is that the surface coats of the Stercoraria trypanosomes, such as *T. cruzi* and *T. carassi*, are coated with poorly immunogenic carbohydrate-rich mucin-like glycoproteins that, unlike the highly immunogenic VSGs that cover the Salivarian trypanosomes, do not vary throughout the infection of the mammalian host (Haag *et al*, 1998; Overath *et al*, 2001). Evolutionary divergence then, appears to correspond to the development of a more immunogenic surface protein. The expansion and diversification of a repertoire of variants of this protein allowed, perhaps along with a dedicated mechanism, frequent switching of the surface protein being expressed and evasion of an immune response specified against a single variant. All the Salivarian trypanosomes appear to have adopted a similar mechanism for antigenic variation, based upon varying the VSG being expressed on the surface, a process that was probably derived from a common ancestor.

In different mammalian hosts, the population profiles of *T. brucei* infections are similar. Each 'peak' within the fluctuating parasitaemia contains a number of VSG variants, thus providing a number of opportunities for evasion of the host's immune response (Miller and Turner, 1981; Robinson *et al*, 1999). The *VSG* gene switching that leads to these variants appears to be spontaneous (possibly stochastic) and independent of the immune system, as it occurs also *in vitro* (Doyle *et al*, 1980). Some cow infections lead to self-cure, and this has been proposed to be due to the host becoming immune to all variants presented by the parasite, as reinfection by the same, but not a different, strain can be unsuccessful (Nantulya *et al*, 1984; Barry, 1986). Also, the order of appearance of variants appears to be semi-hierarchical, with certain VSGs appearing earlier and some later in infection (Gray, 1965; Van Meirvenne *et al*, 1975; Capbern *et al*, 1977; Liu *et al*, 1985; Barry, 1986; Robinson *et al*, 1999). A mechanism whereby some genes are activated preferentially over others is a possible explanation. Such a preference could be due to a number of *cis*- and *trans*-acting factors operating imprecisely within a stochastic system, leading to a variable, only partially, predictable order. The activation of some genes preferentially over others has had experimental (Laurent *et al*, 1984; Liu *et al*, 1985; Robinson *et al*, 1999) and mathematical (Frank, 1999) support.

It is important that mechanisms of antigenic variation be considered in the conditions under which they evolved (Barry, 1997). Laboratory-adapted trypanosome lines, generated by short-term syringe passaging in laboratory animals, are monomorphic, displaying only the dividing long-slender bloodstream form which does not transform routinely into the short-stumpy, non-dividing form. The classic example of such a line is the strain used for the majority of experimental work, namely Lister 427 (Cross, 1975). Strains capable of proceeding through a complete life cycle have been termed pleomorphic. One critical phenomenon related to the monomorphic adaptation is that the VSG switching rate is orders of magnitude lower in monomorphic lines than in pleomorphic lines. Pleomorphic lines switch between *VSG* genes being expressed at an average rate as rapid as ~2 x 10^{-2} switches/cell/generation (Turner and Barry, 1989; Turner, 1997), whereas the monomorphic, syringe passaged lines regularly switch at rates as low as 1 x 10^{-6} – 1 x 10^{-7} switches/cell/generation (Lamont *et al*, 1986). Contingency genes operate, in many organisms, by enhancing variation of a specific phenotype; they moderate this variation at a similar rate in different organisms and the phenotypic changes are pre-emptive of an environmental change such as an immune response (Moxon *et al*, 1994). In the case of the trypanosome, the *VSG* system qualifies as a contingency gene system (Barry and McCulloch, 2001; Barry *et al*, 2003). Loss of a regulation of the contingency gene system in the monomorphic lines could explain their background level of *VSG* switching. For example, during the loss of the ability to differentiate into the short stumpy form, a deregulation possibly linked to the cell cycle (for a review of the differences in differentiation between monomorphic and pleomorphic cell lines, see Hendricks *et al*, 2000), a normal pathway used to catalyse *VSG* switching may have been uncoupled from the process, thus leading to a reduction in frequency and alteration mechanism used (Robinson *et al*, 1999). A more thorough discussion of the differences in *VSG* switching between monomorphic and pleomorphic trypanosomes can be found in Chapter 6.

1.3 The bloodstream-stage trypanosome cell surface

In the bloodstream of the mammalian host the trypanosome comes across two, perhaps conflicting, necessities with respect to its surface. One, it must be able to uptake nutrients and factors required for healthy growth *via* either transporters, facilitated diffusion, or

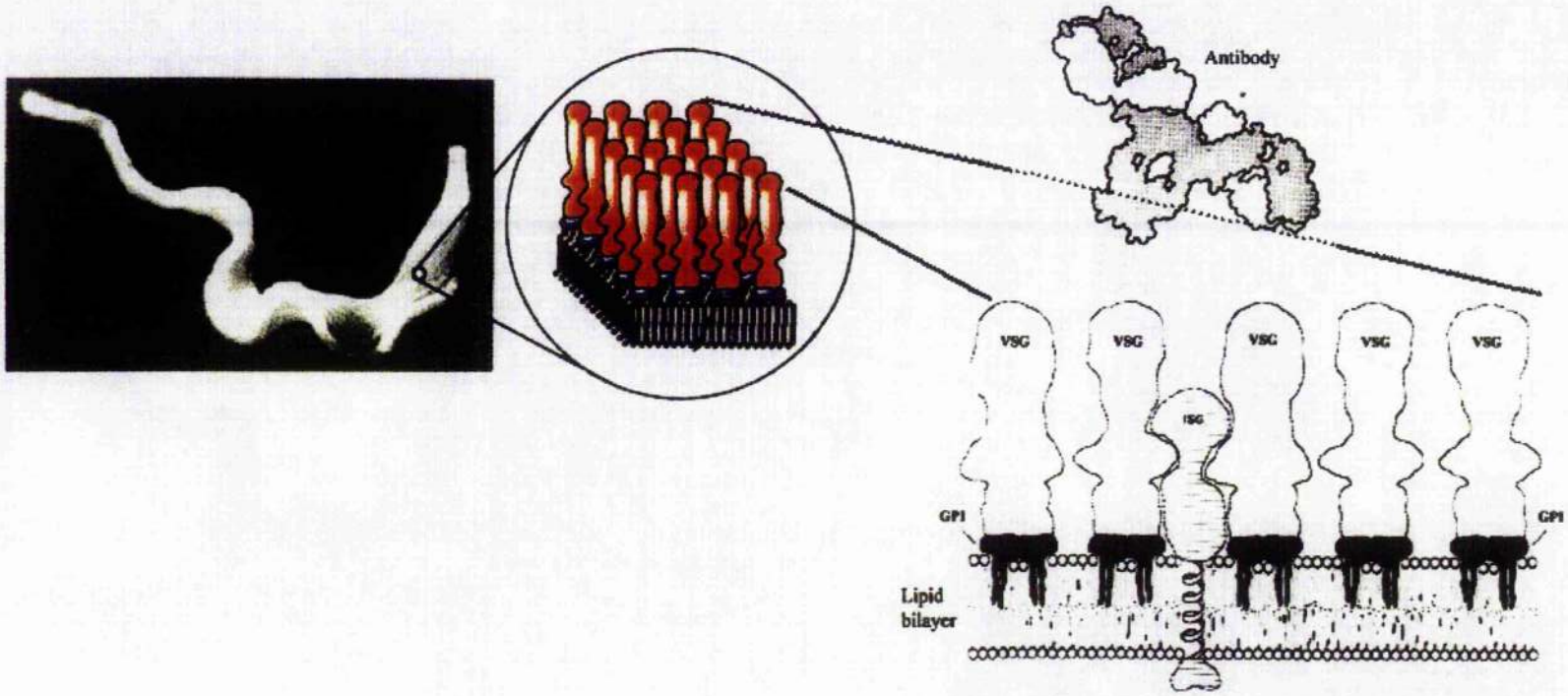
receptor mediated endocytosis. Secondly, these invariant transporters and receptors must be shielded from the immune system.

Protective shielding is achieved by VSG homodimers creating a 12-15 nm thick monolayer barricade that prevents an immune response accessing any invariant surface proteins (Vickerman, 1969) (Figure 1.3). The VSG molecules are anchored to the cell surface by a glycosylphosphatidylinositol (GPI) anchor (Ferguson *et al*, 1988). In fact, the glycosylation of the GPI-anchor varies according to the VSG subclass of the C-terminal domain (see below) (Reviewed in Mehlert *et al*, 1998). This suggests that the carbohydrate acts as a space-filling material to shroud other surface proteins.

A VSG molecule comprises two domains, namely the N-terminal domain of 350-400 residues and the C-terminal domain of 50-100 residues (Carrington *et al*, 1991). The sequences of VSGs are highly divergent, particularly in the N-terminal domain where the variability probably accounts for the antigenic variation of the molecule as opposed to any great structural differences (Blum *et al*, 1993). Exceptions to this variability are the cysteine residues required to form the disulphide bridges in both domains. In the VSGs studied so far, these residues display three patterns for the N-terminus and four for the C-terminus, and the two domains can be found in any combination (Carrington *et al*, 1991). In fact, the pattern of cysteine residues is a common feature amongst other surface proteins in *T. brucei* (Carrington and Boothroyd, 1996). Two independent VSG N-termini, with only 20% sequence identity, shared very similar tertiary structures, including a characteristic structural motif named the 'VSG fold' (Blum *et* al, 1993). This fold is likely to be a common feature amongst other surface proteins such as those encoded by *ESAG6*, *ESAG7* and invariant surface glycoprotein 75 (ISG75) (Carrington and Boothroyd, 1996; Salmon *et al*, 1997). It therefore appears that, despite primary sequence differences, the tertiary structure of different VSGs is similar, allowing each variant to form a complete monolayer. Also, the conserved structural fold amongst other surface proteins suggests a common origin. By mutational analysis, Wang *et al* (2003) recently demonstrated the necessity of these conserved structural motifs for efficient display on the cell surface.

Despite being a dense monolayer, the VSG coat still allows the diffusion of small molecules to transporters on the cell surface and many of these transporters, such as the glucose and nucleoside transporters, are anticipated to be partially hidden by the VSG coat (reviewed in Borst and Fairlamb, 1998). The VSG coat, however, appears to be too dense to allow the penetration of larger, essential molecules such as transferrin and low-density

Figure 1.3 A pictorial representation, at several magnifications, of the surface of the trypanosome. The variant surface glycoprotein (VSG) forms a dense layer on the surface of the trypanosome (depicted in red in the middle picture) and is anchored to the membrane by a glycosylphosphatidylinositol (GPI) anchor (depicted in blue in the middle picture). The cross section (the right hand picture) shows how other proteins, *e.g.* the invariant surface glycoprotein (ISG), may sit in this coat. The size of an antibody is also depicted and portrays how the *VSG* coat shields the invariant protein from the immune response. Adapted from Overath *et al* (1994) and Ferguson (1999).



lipoprotein. To counter this, the receptors for such molecules are localised in the flagellar pocket, an invagination of the plasma membrane adjacent to the base of the flagellum, which serves as the sole site for endo- and exocytosis (reviewed in Landfear and Ignatushchenko, 2001). The transferrin receptor, a heterodimer encoded by *ESAG6* and *ESAG7*, is an example of such a receptor and is predicted, due to its VSG-like structure, to fit perfectly into the VSG monolayer within the pocket. It has been stated that the flagellar pocket is sufficiently open to allow the entry of macromolecules including, conversely, immunoglobulins, however it is sufficiently restrictive to prohibit the cellular arm of the host immune response, thus protecting the trypanosome from the recognition of the invariant surface proteins (Borst and Fairlamb (1998).

ESAG4 is the only other *ESAG* encoded protein that has been firmly identified to be on the membrane, and this location is exclusive to the flagellum (Paindavoine *et al*, 1992). In fact, *ESAG4* is a multigene family and homologues outside of expression sites, known as *GRESAGs* (genes related to *ESAGs*), are present (Pays *et al*, 1989; Alexandre *et al*, 1996) and are also located on the flagellum. Although there is much debate as to whether products of other *ESAGs* are located on the surface, it appears likely the most are (see Pays *et al*, 2001).

1.4 The genome of *T. brucei*

T. brucei has an estimated haploid nuclear genome size of 35 Mb, which can vary between trypanosome isolates by up to 25% (El-Sayed *et al*, 2000). The chromosomes that make up the nuclear genome are categorised into three classes according to size. The largest of them range from 1 Mb to greater than 6 Mb and are termed the megabase chromosomes. There are 11 of these chromosomes, which are the diploid chromosomes of mendelian inheritance. They are named in ascending order of size from I to XI in the TREU927/4 strain (Mellville *et al*, 1998). In TREU927/4, the eleven diploid pairs represent approximately 53.4 Mb of DNA (Melville *et al*, 1998), although they vary greatly in size, by up to 15-20%, between strains and even between homologues from the same strain (Gottesdiener *et al*, 1990; Melville *et al*, 1998). This variation in size does not appear to affect the gene content of the chromosomes, and is likely to result from alterations in repetitive sequence, arrays of repeated housekeeping genes, and arrays of the retrotransposons known as *INGI* (Melville *et al*, 1999; Bringaud *et al*, 2002; El Sayed *et al*,

2003; Hall *et al*, 2003). The second size category of chromosomes are between 200 and 900 kb, are of uncertain ploidy, and are classified as intermediate. Not one of the 401 cDNA and gene clones tested to date hybridise to the intermediate chromosomes, indicating that they may not contain any housekeeping genes (Melville *et al*, 1998). However, not enough genes may have been screened to locate a housekeeping gene due the small size of the chromosomes and thus a reduced number of genes. The minichromosomes make up the third size class of chromosomes; they are approximately 50 – 150 kb in length and are comprised almost entirely of tandem arrays of a 177-bp repeat (Sloof *et al*, 1983; Weiden *et al*, 1992). All of the aforementioned chromosomes are linear and have the $(TTAGGG)_n$ telomeric repeats at their ends (Van der Ploeg *et al*, 1984 Weiden *et al*, 1992). Outside the nucleus, another body of DNA, known as the kinetoplast (kDNA), is located within the matrix of the cell's single mitochondrion. The kDNA network is condensed into a single disc-shaped structure that is found at the base of the flagellum and has an extraordinary structure, apparently unique throughout nature, which involves thousands of topologically catenated circles of DNA (Lukeš *et al*, 2002).

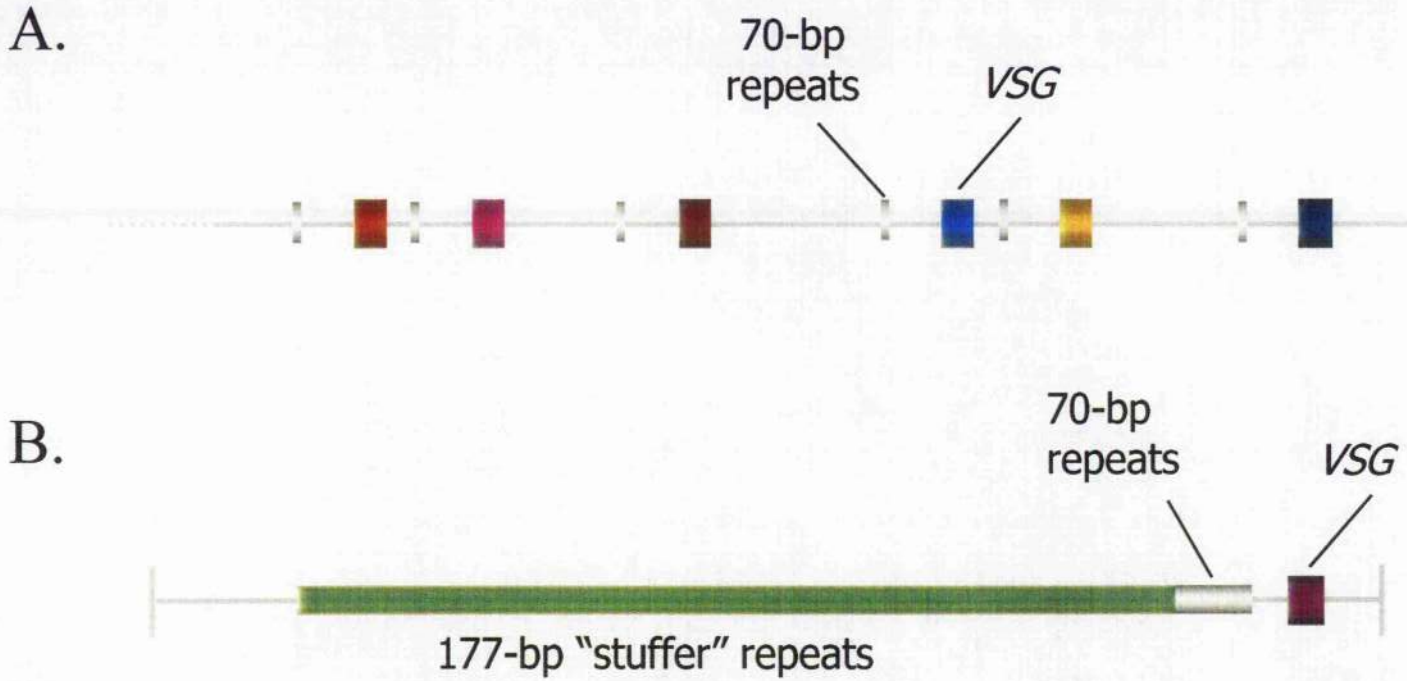
From the nuclear genome it seems likely that all of the housekeeping gene are located entirely on the megabase chromosomes. If this is the case, what function do the mini- and intermediate chromosomes play? Here it is suggested that their sole role within the cell is to act in the process of antigenic variation.

1.5 The *VSG* genes and their expression

An estimate of the number of *VSG* genes within the trypanosome genome falls around the 1000 mark (Van der Ploeg *et al*, 1982). This is an indirect estimate based upon the screening of a library, so confirmation will have to await the full genome sequence. The *VSG* genes have been detected on all classes of chromosomes in the nuclear genome. Most *VSG* genes lie within chromosome-internal arrays (Figure 1.4), whereas the rest reside in subtelomeric loci, the majority of which are at the ends of the 100 or so minichromosomes (Figure 1.4). All of these sequences are flanked upstream by a 70-bp repeat region (the minichromosome *VSG* genes appear to have longer regions (Shah *et al*, 1987).

Transcription of *VSG* genes occurs from only a number of specialised loci, known as expression sites. Two life cycle stages, the metacyclic stage and the bloodstream stage, express *VSG* genes and they differ in the type of expression site used. They have been

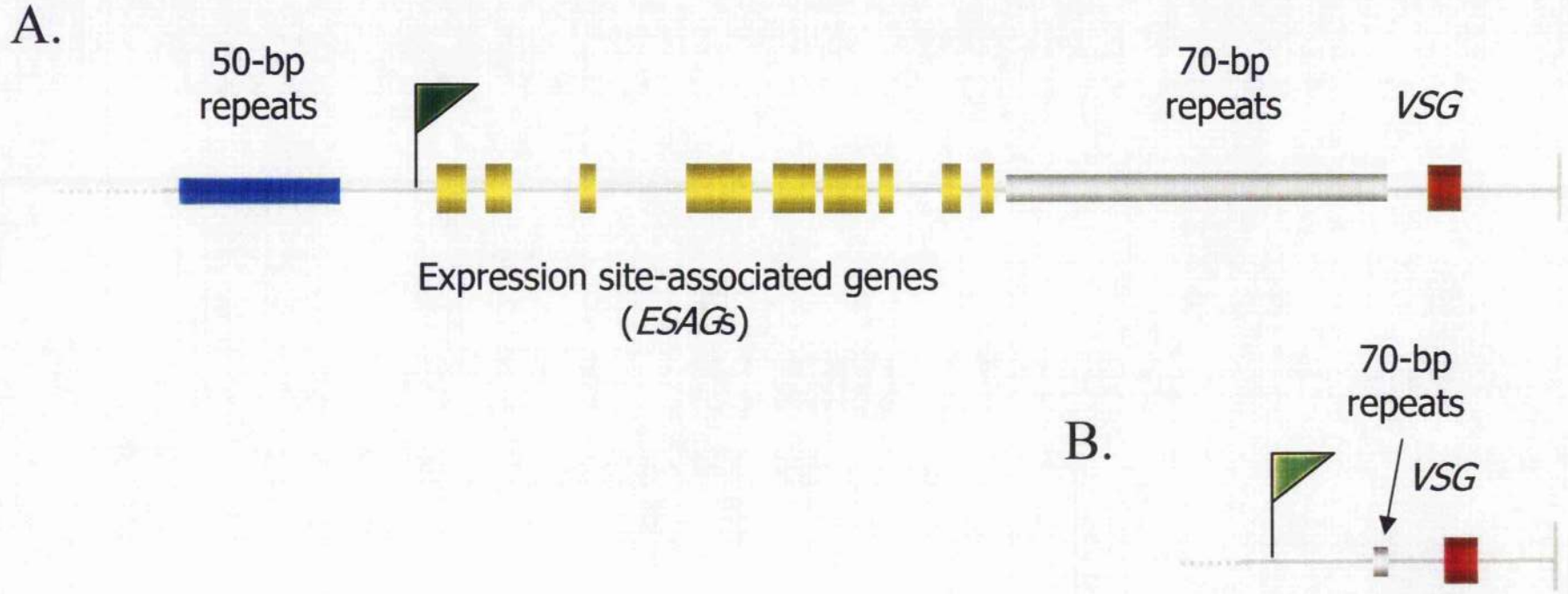
Figure 1.4 The genomic locations of the silent *VSG* genes. **A**. A chromosome internal array of *VSG* genes. Different *VSG* genes are portrayed in different colours, whereas the 70-bp repeats are coloured silver. The dotted line portrays the continuation of the chromosome. **B.** A *VSG* gene at the subtelomeric region of a minichromosome. The purple box is the *VSG* gene, the silver box is the the 70-bp repeat region and the green box represents the 177-bp repeats. The vertical bars portray the end of the chromosome.



named the metacyclic expression sites (MES) and the bloodstream expression sites (BES) (Figure 1.5) in concordance with the life cycle stage in which they are transcribed. BESs appear to be exclusively telomeric and an estimated 20-30 are found within the genome (Cully *et al*, 1985; Johnson *et al*, 1987; Kooter *et al*, 1987; Pays *et al*, 1989). In *T. brucei*, transcription occurs almost exclusively from polycistronic units where one promoter drives the expression of several genes. The BES is no exception, with a promoter approximately 50 kb upstream of the *VSG* gene (Zomerdijk *et al*, 1990). The *VSG* is the most proximal gene to the telomere, with approximately 1 kb of sequence between the gene and the telomere tract (Aline and Stuart, 1989). Within this region there are two common subtelomeric sequences, of which one is similar to the telomere tract sequence itself whereas the other is a short, non-repetitive region called STR for subtelomeric repeat (Aline and Stuart, 1989). Directly upstream of the *VSG* gene is a large repetitive region comprised of many copies of a unit termed the 70-bp repeat (Liu *et al*, 1983; Aline *et al*, 1985). This unit is degenerate and is named after its average size of 76 bp. The size of the repeat array also varies between BESs. The individual repeat unit is comprised of three sections, designated only by their sequence composition. From most distal to the *VSG*, there is first a (T-purine-purine)$_n$ triplet repeat region that is variable in size, followed by a GT-rich region and then by an AT-rich region (Aline *et al*, 1985; for an alignment of many 70-bp repeat sequences see Figures 3.18 and 3.19). No function has been ascribed to any section of these repeats.

Upstream of the 70-bp repeats are a series of genes known as the *ESAG*s. These genes differ in function and number between BESs, with a total of 12 different *ESAG* families identified to date (reviewed in Pays *et al*, 2001). The products of some are membrane-bound receptors, such as the aforementioned transferrin receptor heterodimer (ESAG 6 and ESAG 7) and probably the adenylate cyclase (ESAG 4), whereas others are not, such as ESAG 8 that accumulates in the nucleolus (Hock *et al*, 2000), and interacts with a member of the Puf family of RNA regulatory proteins (Hoek *et al*, 2002). Despite this variation in function and number, *ESAG 6* and *ESAG 7* are the only *ESAGs* with functional copies in every BES sequenced to date (Berriman *et al*, 2002). Several of the *ESAG*s are members of multigene families found throughout the genome, which includes genes that are transcribed outside of the BESs as well as pseudogenes. *ESAG 6*, *ESAG 7* and *ESAG 8*, however, appear to be exclusively found in BESs. One more *ESAG* of note is a gene highly homologous to *VSG* genes called the serum resistance-associated gene (*SRA*) as it confers

Figure 1.5 The *VSG* expression sites. **A.** A bloodstream expression site. **B.** A metacyclic expression site. The *VSG* genes are depicted in red, the 70-bp repeats in silver, all the different expression site-associated genes in yellow, the 50-bp repeats in blue and the promoter regions as green flags. The vertical bar depicts the end of the telomere and the dotted line depicts the continuation of the chromosome.



resistance to human serum (De Greef *et al*, 1994). SRA has been shown to be a lysosomal protein and confers this resistance by interacting with the human-specific serum protein apolipoprotein L-I (Vanhamme *et al*, 2003). The promoter is the next notable sequence upstream of the ESAGs (Zomerdijk *et al*, 1990). In one instance two promoter sequences, both of which can function efficiently, are found in this region (Gottesdiener *et al*, 1992). Although the above mentioned sequences appear to describe the functional elements of the BES, other common elements appear further upstream, most notably a large region of repetitive DNA known as the 50-bp repeats. Although a role is yet to be attributed to this region, it may be the boundary of the regulated BES domain (Sheader *et al*, 2003). Further upstream of the 50-bp repeat region lies another domain conserved between the subtelomeric regions encoding BESs. This is in the form of a long array of the retrotransposon named *ingi* (Kimmel *et al*, 1987). A preferred site of integration for the *ingi* elements appears to be in genes and pseudogenes of the *RHS* (retrotransposon hotspot) family (Bringaud *et al*, 2002). A third of this family appears to contain retroelements inserted in-frame and at the same nucleotide.

The functional length of a BES appears to be around 45 - 60 kb. The extent of homology between the BES containing chromosome ends, however, can extend up to 126 kb (Bringaud *et al*, 2002; Berriman *et al*, 2002). This indicates a significant percentage of the genome dedicated to the expression of *VSG* genes in the bloodstream stage of the life cycle.

The MESs, found also at chromosome ends, differ somewhat from the BESs. They are, unlike their bloodstream counterparts, monocistronic transcription units (Alarcon *et al*, 1994). Up to 27 metacyclic *VSGs* (*MVSG*) are expressed by a metacyclic trypanosome population, so there could be as many MESs (Esser *et al*, 1982; Turner *et al*, 1988).

As with the BES, the coding sequence most proximal to the telomere is the *VSG* gene itself. In between the *VSG* gene and the telomere repeat tract are the same telomere tract-like and STR sequences as found in BESs. Upstream of the *VSG* is often a very short (typically one to two repeat units) region of 70-bp repeat sequence (Matthews *et al*, 1990) followed ~1-3 kb further upstream by a promoter distinct from the BES promoter (Alarcon *et al*, 1994; Nagoshi *et al*, 1995; Ginger *et al*, 2002).

Pseudo-*ESAG*s and *ingi* elements are often upstream of the promoter, perhaps indicating an evolutionary relationship between the BESs and the MESs (Barry *et al*, 1998). In fact, the BESs and MESs have many similarities. Firstly, they are both transcribed, unusually for protein coding genes, by an α-amanatin resistant polymerase (Rudenko *et al* 1989;

Zomerdijk *et al*, 1991). This has now been demonstrated to be RNA polymerase I (RNA PolI; Gunzl *et al*, 2003). Secondly, only one of each form of expression site is transcribed at any one time in a single cell, so each is subject to a similar form of allelic exclusion (Borst, 2002). The form of control exhibited here is unusual for trypanosomes as it appears to be linked to the control of transcription itself, as opposed to the majority of other trypanosome genes where post-transcriptional processes modify expression levels. One possibility that may allow a unique form of transcriptional control is that both expression sites are functionally monocistronic; *i.e.* when an expression site is inactivated a full complement of the inactivated genes are reactivated from another site. This allows a degree of control on the promoter itself that cannot be performed on other polycistronic transcription, units where the genes transcribed by a single promoter are of unrelated function. Thus, specific protein factors acting upon sequence elements within the expression sites may silence or enhance their transcription, or predispose the chromatin to allow or deny access to the factors involved in expression.

In the bloodstream form, the ability to transcribe only one BES at any one time is likely to be linked to a nuclear body known as the expression site body (ESB) (Navarro and Gull, 2002). The ESB appears to contain the active BES and RNA pol I, but does not contain at least the one tested inactive BES. The ESB is also distinct from the nucleolus (Chaves *et* al, 1998; Navarro and Gull, 2002). Sequestration of the active BES into the ESB may allow its sole expression. Initiation of transcription in the bloodstream stages is thought to occur in the majority of BESs, with the RNA elongation being inhibited (Vanhamme *et al*, 2000). A stable transcriptional complex, absent from the inactive BESs, may be present and might be located only in the ESB. The ESB is yet to be sought in the metacyclic life cycle stage, so it is not known whether a similar phenomenon takes place there.

An interesting alternative to explain exclusive expression site expression control is a certain DNA modification. *T. brucei* contains the base β-D-Glucosyl-hydroxymethyluracil (Base J) which is a modification, conserved in kinetoplastid protozoa, of a small fraction of the base thymine in the parasite genome (Van Leeuwen *et al*, 1998). Its presence varies between active and silenced BESs, in that it is present in the 50-bp repeats and the telomeric repeats of both, but not between these flanks in the active BES, and it is not present in internal VSG genes (Van Leeuwen *et al*, 1996; Van Leeuwen *et al*, 1997). Due to these differences, it has been suggested that base J is involved in the silencing of BESs (Van Leeuwen *et al*, 1997). The role of base J may include altering transcription,

suppressing recombination between repetitive sequences or mediating chromatin remodelling, and it has therefore been speculated that these potential mechanisms of repression may have been recruited for control of VSG expression sites (Van Leeuwen *et al*, 1998). More recently, a J-binding protein was isolated from a number of kinetoplastids (Cross *et al*, 1999). It appears that the most likely role for this protein is to facilitate the formation of base J, in areas that already contain basal amounts of the modified base, by activating the thymidine modification enzymes (Cross *et al*, 2002).

Both forms of expression site are also subtelomeric, which perhaps is unsurprising, as contingency genes in pathogenic microbes are often associated with telomeric loci (Barry *et al*, 2003). Subtelomeric location may have helped spread the expression sites across a large number of telomeres, as sequence homogenisation by ectopic recombination between non-homologous chromosomes occurs to a greater extent at subtelomeric locations that non-subtelomeric locations in some organisms (Louis and Haber, 1990; Barry *et al*, 2003). It is also possible that subtelomeric location aids the silencing of expression sites. A telomeric silencing effect is seen in other organisms, in which the status of chromatin can repress genes in telomeric regions. A number of proteins and protein complexes are involved in this silencing, one of which is the Ku 70 and Ku 80 heterodimer in *S. cerevisiae* (Boulton and Jackson, 1998). Ku null mutants in *S. cerevisiae* alleviate the silencing of genes in telomeric locations (Boulton and Jackson, 1998). *T. brucei* Ku null mutants, however, do not alter the transcriptional status of *MVSG*s in the procyclic stage, perhaps indicating the silencing here is not equivalent to the telomere position effect seen in some other organisms (Conway *et al*, 2002). However, a Ku-independent form of telomeric silencing may take place. It is apparent that the role of the telomeric location of expression sites is not fully understood.

The number of expression sites is limited by the number of telomeres and would, of course, have to accommodate both the BESs and the MESs. This limits the number of endogenously expressible *VSG* genes to the relatively few that occur in the expression sites specific for each life cycle stage. In the bloodstream stage this number is in the low tens, thus the chronicity seen for *T. brucei* infections cannot be explicable by the use of just these variants as, at the frequency that VSGs are switched, the host would rapidly become immune to them. It is not surprising, then, that *VSG* genes are on every class of chromosome. The hundred or so minichromosomes appear to house nothing other than the 177-bp repeats, *VSG* genes, their associated 70-bp repeats and subtelomeric sequences

(Figure 1.5 B). Some intermediate chromosomes have been shown to contain BESs, and long chromosome internal arrays of *VSG* genes, as well as BESs, are present on the megabase chromosomes. In fact, intermediate chromosomes have been suggested to have arisen from breakages of megabase chromosomes (Berriman *et al*, 2002). Perhaps, then, the intermediate chromosomes and the minichromosomes have evolved solely to function in the process of antigenic variation, with the former being a repository for expression sites and the latter being a large reservoir of *VSG* genes. If expression of *VSG* genes occurs only from BESs during the bloodstream stage, how then is the majority of the *VSG* gene repertoire accessed? It has been demonstated that this occurs by a form of duplicative transposition of a silent *VSG* gene into the active BES, a process that may be carried out by homologous recombination (McCulloch and Barry, 1999).

1.6 How does *VSG* switching occur?

Broadly, the main means of switching the expression of one *VSG* gene to another can be classified into either transcriptional or recombinational mechanisms. These are discussed in turn, below.

1.6.1 Transcriptional (*in situ*) switching

This phenomenon involves the inactivation of the active BES and concurrent activation of another BES, thus changing the *VSG* gene being expressed. It is still unknown how this (in)activation occurs. DNA rearrangements, including deletions of an upstream duplicated promoter, in activated (Navarro and Cross, 1996) and inactivated BESs (Gottesdiener *et al*, 1991; Gottesdiener *et al*, 1992), have been detected, perhaps suggesting an interaction of telomeres prior to switching. Due to differences in the reported rearrangements, it seems unlikely that any specific rearrangement is necessary for *in situ* switching, and it is more likely that the BESs are brought together and spontaneous recombination occurs between homologous sequences. Also, a study *in vitro* found no DNA rearrangements in nine *in situ* switches, similarly bringing into question whether DNA rearrangements are involved in (in)activation of BESs (Horn and Cross, 1997).

It has been suggested that a form of cross-talk is involved in switching between BESs (Chaves *et al*, 1999; Ulbert *et al* 2002). Cells with drug resistance markers introduced into two BESs displayed, when put on a double drug selection, an unusual double resistant phenotype, where the cells switch between the marked expression sites at the rapid rate of 10^{-1} per cell per generation (Chaves *et al*, 1999). Placing the cells on triple drug selection did not yield cells switching between three BESs, rather it produced cells switching between two BESs and partial promoter activation of the third marked BES, implying cross-talk might only be between two expression sites (Ulbert *et al*, 2002). It was concluded that the phenotype of rapidly switching between two BESs seen in the double drug resistant lines is a natural intermediate in *in* situ switching. It appears that the inactivated expression site remains in a pre-active state, allowing its reactivation at a rapid rate, that is eventually lost so that then its reactivation occurs at the usual low rate (Chaves *et al*, 1999; Ulbert *et al*, 2002).

How this interaction and cross-talk occur is unknown, but one strong candidate is localisation in the ESB (Navarro and Gull, 2002). It is possible that BESs could be brought together due to interactions between homologous DNA sequence regions. It is possible that interaction between BESs may also depend on homology, and it is interesting to note that mutation of the *T.brucei* RAD51 gene, a key player in homologous recombination, reduces the level of *in situ* switching (McCulloch and Barry, 1999). Such interactions could also explain the DNA rearrangements seen in some switching events.

As alluded to earlier, the formation of base J within silent expression sites could be involved in *in situ* switching (Van Leeuwen *et al*, 1997). Whether this modification is a cause or consequence of *in situ* switching is unknown, but it is likely that all inactive BESs contain this modification. Maybe removal of base J, perhaps during replication, allows an inactive expression site to become active. It would be interesting to know if the pre-active BES described by Chaves *et al* (1999) contains base J.

The significance of *in situ* switching as a means of driving antigenic variation is still uncertain. Varying the *VSG* being expressed seems not to be a sufficient explanation. As mentioned, the repertoire of variants available for activation by *in situ* switching is very small, altering which BES is expressed cannot access the majority of the VSG repertoire. Also, *in situ* switching appears not always to be the dominant means of *VSG* activation (Robinson *et al*, 1999). It seems possible that *in situ* switching serves another purpose.

As trypanosomes can infect many different mammalian species, one potential role of *in situ* switching is that each BES, through expression of variants of the ESAGs, confers advantages to survival in different host species. Evidence for this hypothesis comes mainly from studies on the specificities of the transferrin receptor encoded by *ESAG 6* and *ESAG 7* (Bitter *et al*, 1998; Gerritis *et al*, 2002). The transferrin receptors encoded by different BESs differ in an area known as the hypervariable region that may have undergone selection (Salmon *et al*, 1997). Bitter *et al* (1998) and Gerritis *et al* (2002) demonstrated that different BESs express transferrin receptors with different affinities for different host transferrins. When cultured in different sera, trypanosomes expressing a BES encoding a transferrin receptor conferring higher affinity for the transferrin in the serum are selected. The diverse transferrin affinities, however, are not a complete explanation for *in situ* switches, as culturing of trypanosomes in some sera (horse, goat and pig) resulted in selection of an *in situ* switch that could not be blocked by the addition of bovine transferrin, the preferred substrate for the BES active at the onset of growth (Gerritis *et al*, 2002). It is also apparent that the amount of transferrin available in the host is far in excess of any limitations imposed by transferrin receptors with lower affinities (Steverding *et al*, 2003). The limitations imposed by chronic infections are still unknown. The transferrin levels in the host may become limiting as infection proceeds and the animal becomes starved of nutrients or pathological changes occur. Also, antibodies generated against an isoform of the transferrin receptor during a chronic infection (Hobbs and Bothroyd, 1990) may compete with transferrin for binding, thus necessitating a switch to an immunologically distinct receptor (Steverding *et* al, 1995; Bitter *et al*, 1998; Steverding *et al*, 2003).

Of course, *ESAG*s other than *ESAG6* and *ESAG7* may vary between BESs in both sequence and availability and could play an important role in host specificities. The transferrin receptor could be one component of a more complete battery of receptors adapted to different hosts. One problem with the host-specific BES hypothesis is that *T. equiperdum*, a species of trypanosome with only one host species, the horse, retains a number of BESs (Longacre *et al*, 1983a; Baltz *et al*, 1986). Therefore, the ability to infect different host species cannot explain the presence of multiple BESs in *T. equiperdum*. However, there are recent claims that there is less diversity in different *ESAG6* genes from *T. equiperdum*, which would strengthen the claim of species-specific transferrin receptors (Isobe *et al*, 2003). However, an alternative to a host-specific explanation for the limited diversity is

that *T. equiperdum* does not undergo sexual exchange, perhaps reducing the variability of the *ESAG6* genes. One important point to note is that it is still unknown how many of the BESs can actually be used to drive expression of a *VSG*. If the number is small, the relevance of host range may be brought into question.

One other possible reason for *in situ* switching is that the inactive BESs are required to act as regions for the formation of mosaic *VSG* genes, which are formed from multiple recombination events with many *VSG* genes and pseudogenes. As one recombinational exchange may not suffice to change all exposed epitopes, a full switch may require progressive accumulation of products of successive recombination events in an inactive BES followed by activation *via in situ* switching.

1.6.2 Recombinational switching

The dominant mechanism of *VSG* switching is duplicative transposition (Robinson *et al*, 1999). This is where the duplicated copy of another *VSG* gene replaces the expressed *VSG* in the active BES. It is the only means of activating the majority of the *VSG* repertoire, *i.e.* the chromosome-internal and subtelomeric minichromosomal *VSG* genes, as they have no means of being expressed from their endogenous locus.

Activation events by duplicative transposition have been observed to vary in detail, generally in the amount of the expression site sequence replaced by the donating locus (see below). It is not just the *VSG* gene sequence that is copied into the BES during a duplicative activation, but also associated companion sequence (Pays *et al*, 1983). Reciprocal reactions have also been detected, where sequence is exchanged between loci, thus resulting in a conservative process involving no loss of sequence (Rudenko *et al*, 1996).

The majority of work on the duplicative transposition of *VSG* genes has been observational, involved in the mapping of the limits of sequence transposition. Downstream of the *VSG* gene in the active BES, areas for the duplication of sequence to begin or terminate are constrained by the amount of available homologous sequence, such that conversion limits have been detected only within the 3' end of the *VSG* gene (Michels *et al*, 1983; Pays *et al*, 1985; Donelson *et al*, 1983; Liu *et al*, 1983; Liu *et al*, 1985), within sequences immediately downstream of the *VSG* gene (Timmers *et al*, 1987) and, in the case of some duplicative activations of telomeric *VSG* genes, all the sequence up to, and

including, the telomere tract has been copied (Scholler *et al*, 1989; De Lange *et al*, 1983). The upstream limits of sequence conversion display more variability. Here the conversion limit has been seen to occur within the sequence immediately adjacent to the *VSG* gene, known as the co-transposed region (CTR) (Michiels *et al*, 1983; Pays *et al*, 1985; Donelson *et al*, 1983), within the 70-bp repeat region (Liu *et al*, 1983; De Lange *et al*, 1985; Florent *et al*, 1987; Lee *et al*, 1987; Shah *et al* 1987; Kooter *et al*, 1988; Matthews *et al*, 1990; Le Count *et al*, 2001) and within regions upstream of this repetitive region (Lee *et al*, 1987; Kooter *et al*, 1988).

One feature evident of all these above conversion limits is that they appear to be in regions of homology between the BES and the donating silent *VSG* locus. Activations involving conversion limits within the CTR have been shown to be in regions that share 70% homology between the recipient and donor of the conversion (Michiels *et al*, 1983). The events in which limits occur within the 70-bp repeat region are manifestly in regions of homology, and conversion limits occurring in regions upstream often occur in the *ESAG*s that are shared between expression sites (Kooter *et al*, 1988). One duplicative activation had an upstream conversion limit within *ESAG2*, and this sequence was 90% homologous with the same gene in the donating locus (Kooter *et al*, 1988). Within the downstream conversion limits, the 3' ends of *VSG* genes have the highest degree of homology with other *VSG* genes, and the short length of sequence between the *VSG* genes in subtelomeres and the telomere tract is similar to that in other chromosome ends. It is apparent that homology between the recipient (*i.e.* active BES) and the donor (a silent *VSG* locus) is important in catalysing the conversion process. This leads to the hypothesis that homologous recombination is involved in these duplicative activations. Another form of recombination involved in *VSG* duplicative activation is that of mosaic gene formation. This involves the formation of a single functional *VSG* gene from the recombination of several other *VSG* genes or pseudogenes. For this to yield a successful switch, all relevant VSG epitopes must be changed, and the tertiary structure of the final protein must be compatible with producing a *VSG* monolayer. Mosaic events have been detected late in infections and therefore have been considered a relatively rare means of activation (Longacre and Eisen, 1986; Roth *et al*, 1986; Pays, 1989; Thon *et al*, 1989; Thon *et al*, 1990; Kamper and Barbet, 1992). The importance of mosaic events cannot be determined until the number of *VSG*s that are pseudogenes is known, as only then will it be known how much of the *VSG* repertoire can only be expressed *via* mosaic gene formation

1.7 Homologous recombination in Eukaryotes

Before describing how homologous recombination might drive the duplicative transposition of *VSG* genes, it is necessary to provide a background based upon the model organisms in which homologous recombination has been studied.

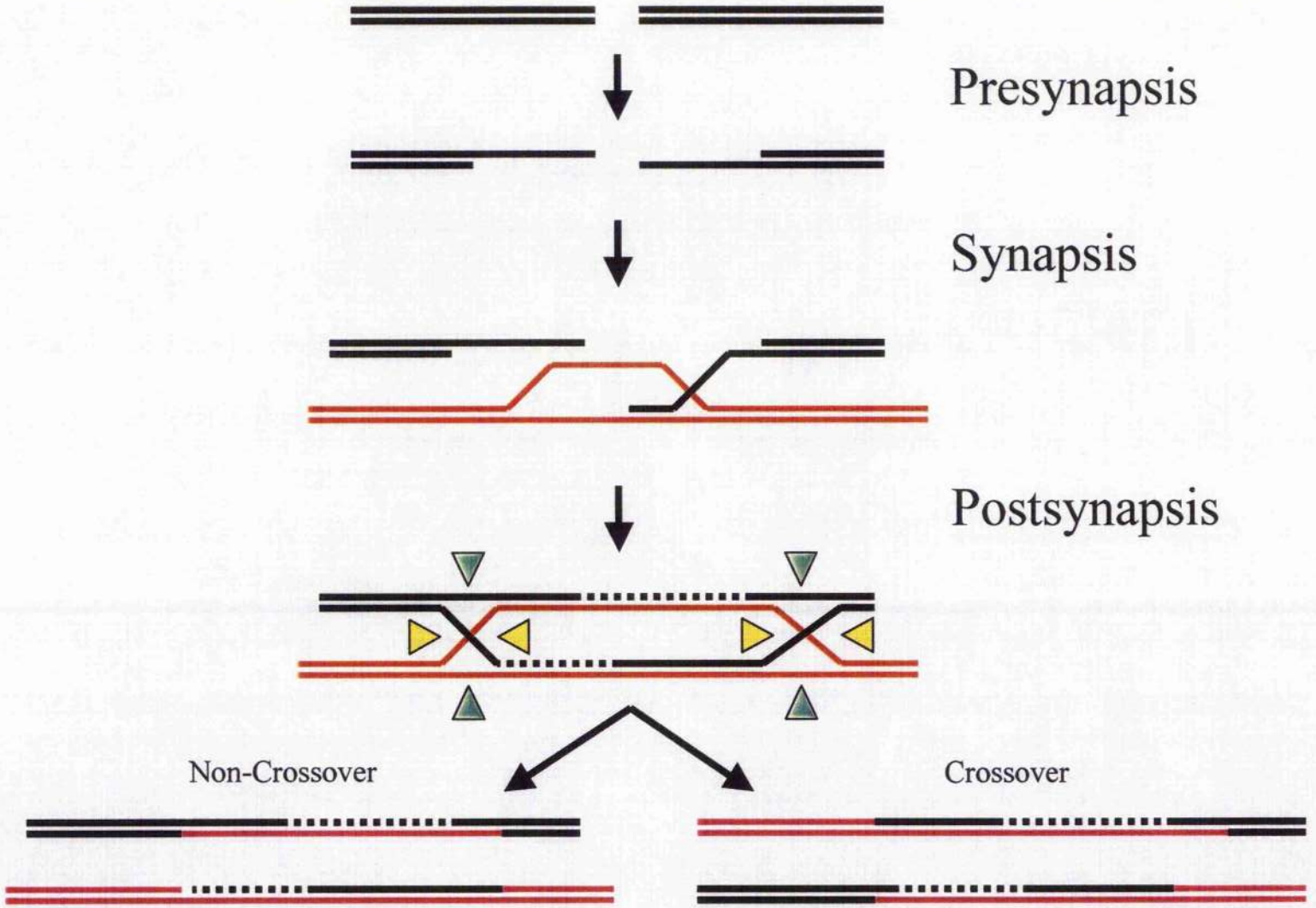
1.7.1 Models of mitotic homologous recombination

The recombination of DNA is a fundamental biological process that occurs in every kingdom of life, and the principles by which it occurs have been conserved (for a recent review see Cromie *et al*, 2001). The main roles of recombination are to repair DNA double-strand breaks (DSBs), maintain genomic stability, generate diversity, and restart collapsed replication forks. Failure to repair DNA damage can lead to cell death. From the necessity for an organism to cope with DSBs, many models of homologous recombination have been based upon their repair. One form of such repair is non-homologous end joining (NHEJ), which involves the ligation of the two ends formed after formation of a DSB, often using microhomologies of 2-4 bp (reviewed in Lewis and Resnick, 2000). NHEJ uses an almost exclusive set of factors and is very unlikely to catalyse *VSG* switching. In fact, when *KU70* and *KU80*, genes crucial to NHEJ in other species, were mutated in *T. brucei*, no effect on *VSG* switching was detected (Conway *et al*, 2002).

Homologous recombination is a largely distinct form of DSB repair, and can be considered as a set of discrete catalytic steps, which are summarised as three different models in Figures 1.6 – 1.8. The first stage, known as presynapsis, involves at least one side of the DSB being processed into a recombinogenic substrate. The 5' strand of the duplex is resected to reveal a 3' single strand overhang (White and Haber, 1990). In the next stage, synapsis, this 3' overhang invades a homologous duplex and promotes replication, and therefore duplication, of the invaded template. Postsynapsis is the third stage and may involve the termination of the reaction by either the formation of four-stranded branched structures, known as Holliday junctions (Holliday, 1964), between both sides of the DSB and the template duplex, or expulsion of the invaded template followed by rejoining with the other side of the DSB. These alternatives, known respectively as the DSB repair model (Szostak *et al*, 1983) (Figure 1.6) and synthesis dependent strand annealing (SDSA)

(termed by Nassif *et al*, 1994) (Figure 1.7), are the two main models for gene conversion, a process described as the nonreciprocal transfer of information from an intact region of DNA to another DNA molecule. One other means of repairing DSBs does not involve both ends of the DSB. A single end invasion could establish a replication fork that could replicate the remainder of the template DNA molecule, potentially extending to the end of the chromosome. This process is known as break-induced replication (BIR) (Figure 1.8).
The DSB repair model (Szotak *et al*, 1983) proposed the double Holliday junction structure, initially during meiotic recombination, as a means of explaining the observation of crossovers being associated with some gene conversion events (Orr-Weaver and Szostak, 1983). This model involves the invasion of both ends of the DSB into a homologous template, followed by processing of the Holliday junctions by a specialised resolvase that can generate crossover or non-crossover products (Figure 1.6). Holliday junction resolvases (such as the RuvABC resolvasome from *E. coli*, reviewed in West, 1997) may cut the crossed strands or the non-crossed strands, and if the two Holliday junctions are cut differently then a crossing over of flanking sequence will occur, but if not then a parental strand configuration remains.
Models that can explain a lack of crossover events in mitotic recombination (Klar and Strathern, 1984; Sugawara *et al*, 1995; Paques and Haber, 1995) come under the broad heading of SDSA. Several alternative hypotheses can explain this recombination pathway (Figure 1.7) (reviewed in Pâques and Haber, 1999). The principle is that conservative replication occurs, resulting in all the newly synthesised DNA being in the recipient molecule. Firstly, the reaction could be essentially the same as the DSB repair model, except without Holliday junctions. The invading strands, after synthesis, simply dissociate from the template and anneal with each other (Figure 1.7 A). A variation of the SDSA model has only one 3' overhang invading a homologous template, synthesising DNA until sequence that is homologous to the other end of the DSB is generated, and allowing simple annealing to occur (Figure 1.7 B) (Ferguson and Holloman, 1996; Paques *et al*, 1998; Paques and Haber, 1999)). Neither of these hypotheses allows crossing over to take place, but one variation does (Figure 1.7C). If the D-loop created by the strand invasion, and displaced by DNA synthesis and branch migration, anneals with homologous sequence at

Figure 1.6 The double strand break repair model (Szotak *et al*, 1983). Here a DSB is inflicted upon a duplex piece of DNA (black lines), and presynapsis follows which creates a recombinogenic 3' overhang which then invades a homologous duplex (red lines) during synapsis. In the process of post-synapsis DNA synthesis repairs any lost sequence and the formation of Holliday junctions allows their resolution, leading to non-crossover (cutting with green or yellow arrowheads) or crossover (cutting with the green and yellow arrowheads independently on each Holliday junction) products. The dotted lines indicate newly synthesised sequence.

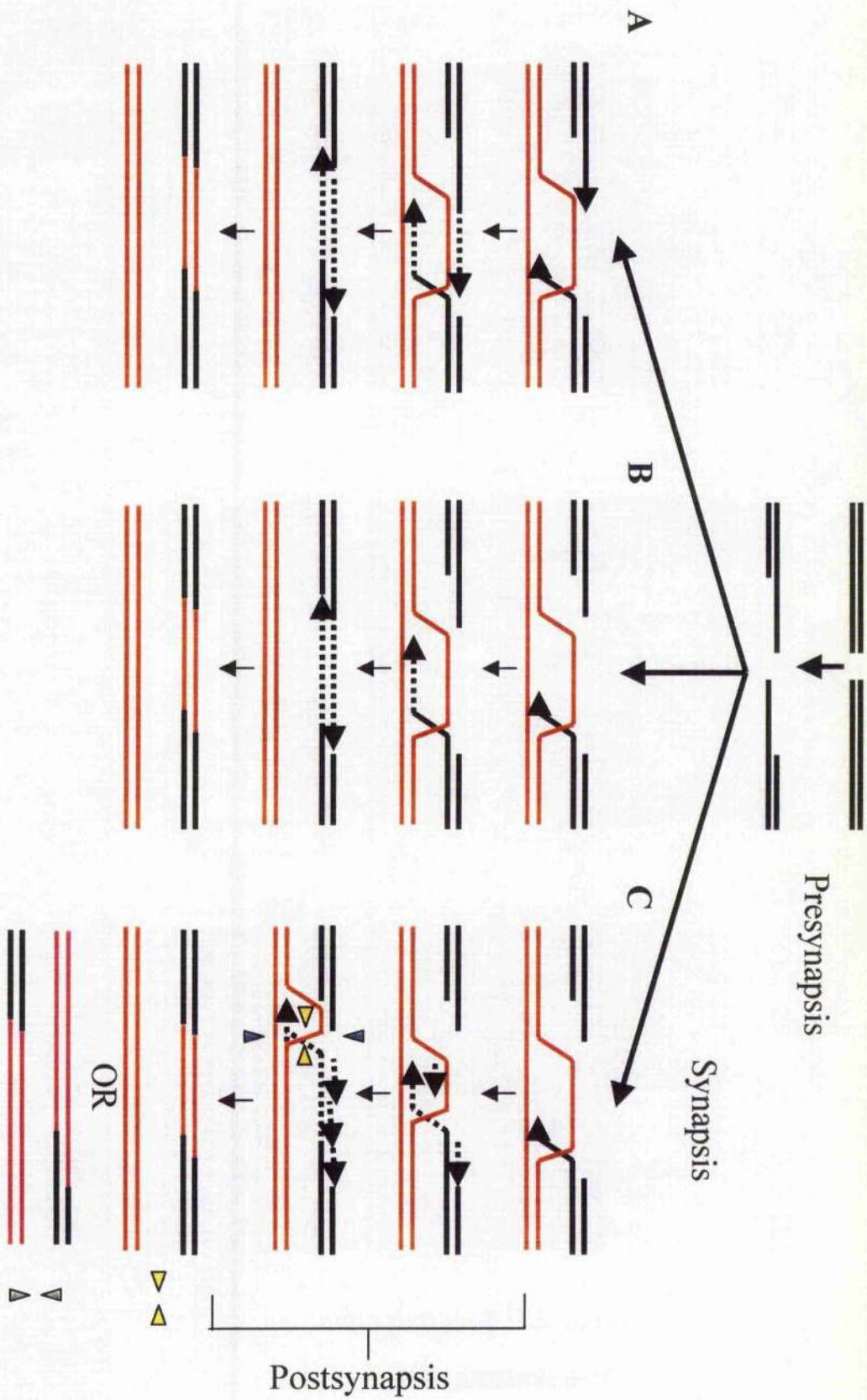

Presynapsis
Synapsis
Postsynapsis
Non-Crossover
Crossover

the other end of the DSB, one or two Holliday junctions may be formed, allowing crossing over (Figure 1.7 C – only one Holliday junction shown) (Ferguson and Holloman, 1996; Paques and Haber, 1999). Evidence supporting SDSA comes from a study involving the segregation of small palindromic inserts flanking a DSB site during meiosis, which escape the mismatch repair system. The segregations predicted by the DSB repair model were rare and were best explained by a SDSA model (Gilbertson and Stahl, 1996). Also, the error-prone replication of DNA during the repair of HO endonuclease induced DSBs, potentially caused by the error-prone polymerase Pol ζ (Holbeck and Strathern, 1997), leads to the recipient receiving all of the mismatches, in a manner that is independent of the mismatch repair system (McGill *et al*, 1998).

One fact that is critical in understanding all these recombination models is that mutants deficient in components of lagging strand synthesis are severely impaired in DSB repair and gene conversion (Holmes and Haber, 1999). It is therefore likely that the DNA synthesis involved in DSB repair establishes a replication fork, rather than proceeding by leading strand synthesis alone, as some of the earlier models of DSB repair imply (Holmes and Haber, 1999).

Figure 1.8 illustrates the final model of homologous recombination-dependent DSB repair: BIR. In *E. coli*, recombination induced by DNA damage can initiate replication at sites separate from normal replication origins, in a process termed inducible stable DNA replication (reviewed in Kogoma, 1997; Cox *et al*, 2002). Similar instances of large regions of conversion initiated at a DSB have been seen in yeast (Malkova *et al*, 1996). If the second end of a DSB fails to engage, replication initiated by the invasion of one end could continue to the end of the chromosome. In a model system where a DSB was induced in haploid *S. cerevisiae* cells 12 kb from the left end of chromosome *III*, repair of the broken chromosome must restore a telomere. 60% of surviving cells had catalysed repair *via* NHEJ, which re-ligates the broken telomere-containing fragment onto the chromosome, whereas 35% had copied the entire right end of chromosome *III* using a 72 bp region of homology to initiate repair (Bosco and Haber, 1998). This occurred at a ~100 times higher frequency than *de novo* telomere addition. BIR is, however, a relatively rare phenomenon in wild-type cells, resulting in only about 2% of the recombinational repair events initiated by a DSB (Malkova *et al*, 1996). If, as suggested by the use of lagging strand synthesis in gene conversion reactions (Holmes and Haber, 1999), BIR and gene conversion events replicate DNA during repair in similar fashions, why does the second end of the DSB ever

Figure 1.7 Synthesis dependent strand annealing model of double strand break (DSB) repair (Reviewed in Paques and Haber, 1999). In all cases presynapsis generates a 3' recombinogenic overhang. **A.** Both 3' overhangs invade the homologous duplex and initiate DNA synthesis, simple annealing of the two newly synthesised strands creates a non-crossover product with all newly synthesised DNA in the recipient of the DSB. **B.** As (A) except only one 3' overhang invades and initiates DNA synthesis. Once homology with the other side of the break is reached, strand annealing, followed by DNA synthesis, initiated by the 3' strand of the other side of the DSB, allows completion of repair resulting in a non-crossover product with all newly synthesised DNA in the recipient of the DSB. **C.** Strand invasion by one 3' overhang from the DSB allows the formation of a replication fork with both leading- and lagging-strand synthesis. Branch migration of the D-loop created during invasion followed by annealing of the second strand of the invaded homologous duplex allow the formation of a Holliday junction that can be cleaved to give non-crossover (yellow arrowheads) or crossover (green arrowheads) products. The dotted arrows indicate DNA synthesis (based on Paques and Haber, 1999).


A
B
C
Presynapsis
Synapsis
OR
Postsynapsis

get involved? It is likely that the replication fork set up for the repair of DSBs is different from an origin-induced fork in terms of stability, perhaps by lacking certain factors, resulting in frequent dissociations and leading to engagement of the second end of the DSB (Kraus *et al*, 2001). In evidence, the repair of a DSB in a plasmid, in one assay, occurred by extensive DNA replication from a chromosomal template. The plasmid readily dissociated from the invaded duplex, and religated to a circular form *via* NHEJ (Kraus *et al*, 2001). This indicates the instability of recombination-dependent replication in yeast. Rejoining of a plasmid after repair of an induced DSB by BIR can also terminate by a reaction known as single-strand annealing (SSA) (Ira and Haber, 2002). (SSA is normally an intrachromosomal reaction where extensive resection from a DSB reveals homologous sequences flanking it, which then anneal to each other. This is a less conservative reaction, in that the DNA between the two homologous sequences has been deleted in the rejoined molecule (reviewed in Paques and Haber, 1999))

1.7.2 The proteins of mitotic gene conversion

The main proteins involved in catalysing gene conversion, in yeast, are in the *RAD52* epistasis group (Game and Mortimer, 1974). The genes of this group, in *S. cerevisiae* are *RAD50*, *RAD51*, *RAD52*, *RAD53*, *RAD54*, *RAD55*, *RAD57*, *RAD59*, *MRE11* and *XRS2* (NBS1 in humans). This group, however, represents a diverse set of functions and their role, along with other factors, will be discussed in terms of the stages of the recombination process during gene conversion. A model of the interaction of these proteins with each other and the DNA is presented in Figure 1.9. The proteins involved in eukaryotic recombination have been studied mostly in the budding yeast *S. cerevisiae*, which is the basis of the ensuing discussion unless otherwise stated. Although there are many differences in the phenotypes of homologous recombination mutants between budding yeast and vertebrate cells, these commonly reflect variations in the form of repair utilised, rather than differences in the role of the proteins in repair.

Figure 1.8 The break-induced replication model of double strand break (DSB) repair (reviewed in Kraus *et al*, 2001). A 3' recombinogenic overhang from one side of the DSB, generated in presynapsis, invades a homologous duplex. A replication fork is set up, with leading- and lagging-strand synthesis that replicates the homologous duplex until the end of the chromosome, thus resulting on a non-reciprocal conversion event. The dotted arrows indicate DNA synthesis.

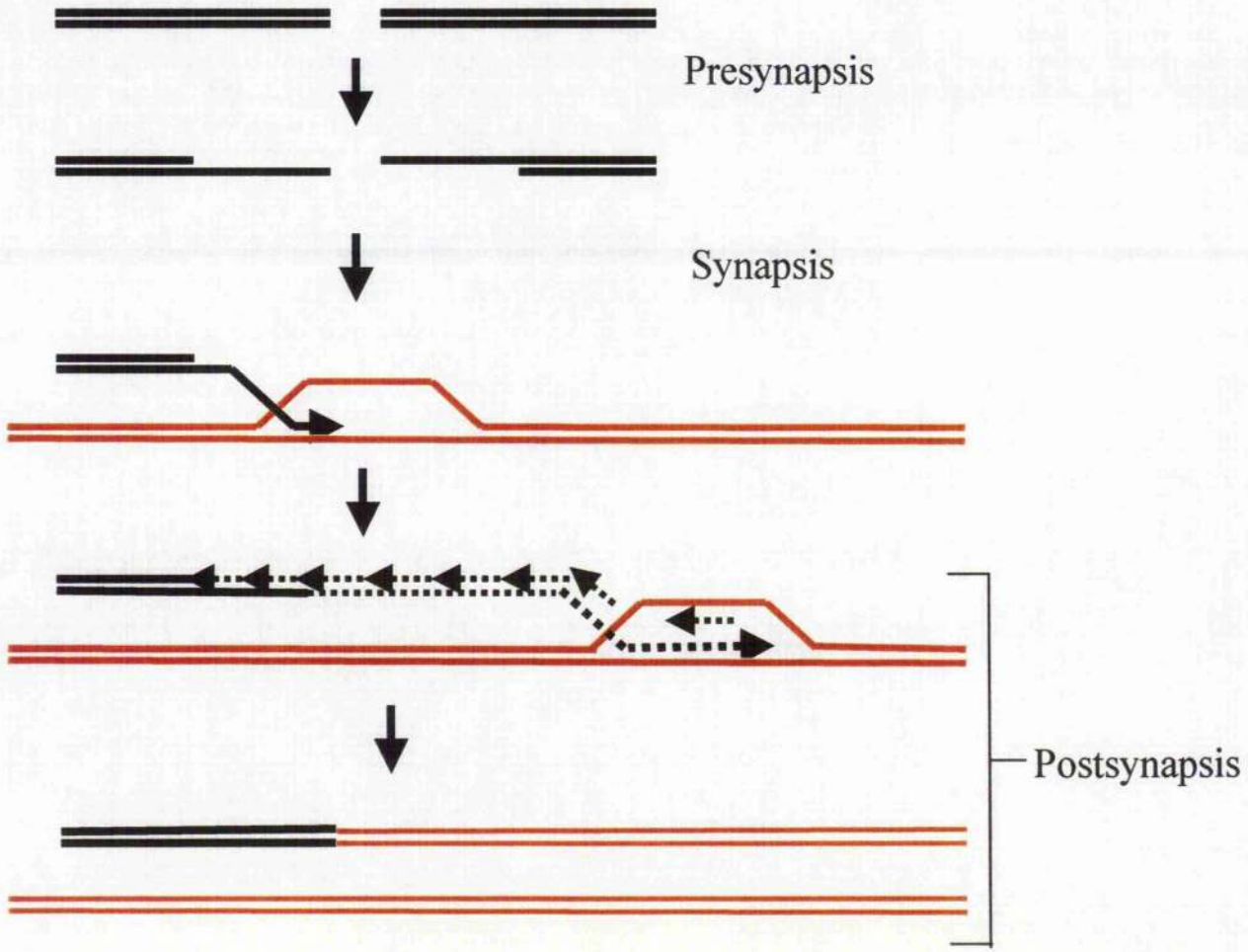
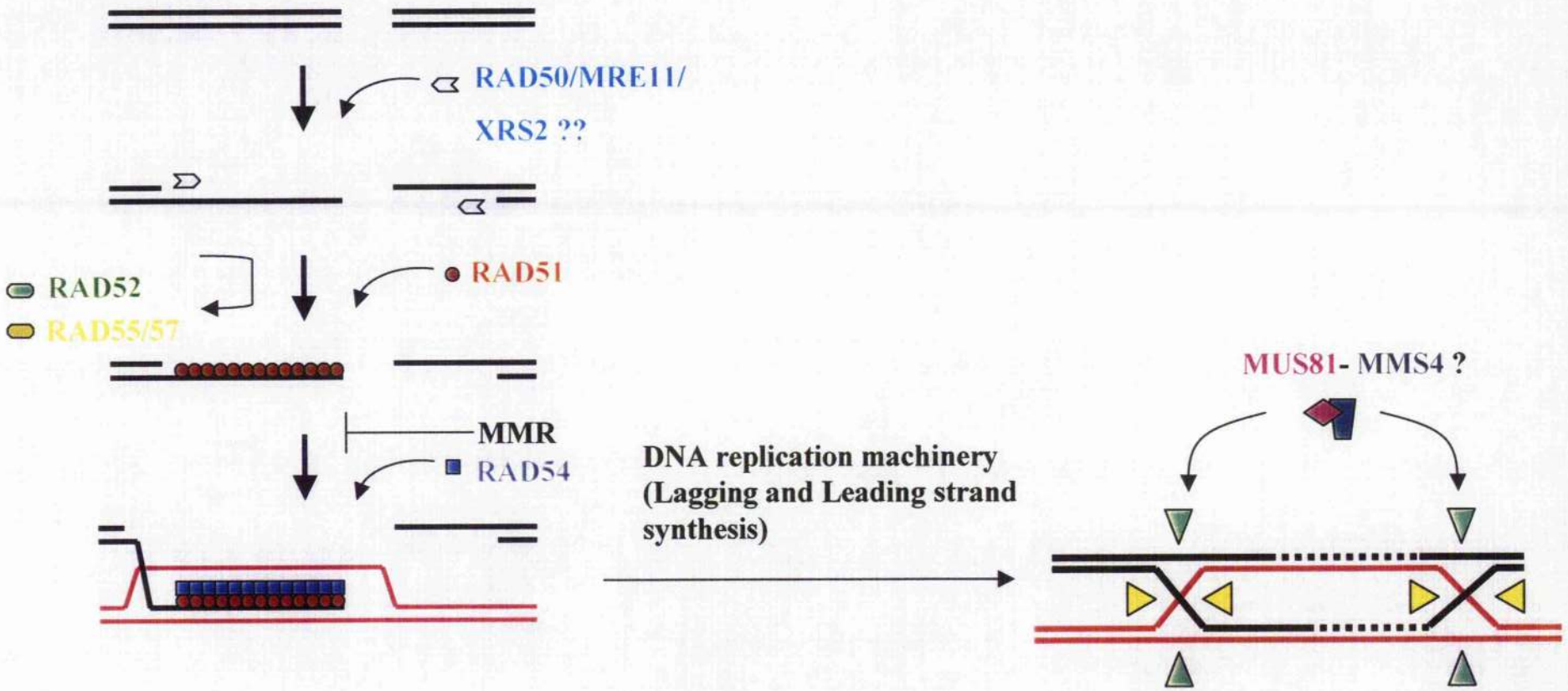


Figure 1.9 The proteins involved in homologous recombination. The black lines indicate the duplex that has suffered the double strand break, the red lines indicate the homologous duplex of the repair template, the dotted lines indicate newly synthesised DNA and the green and yellow arrowheads indicate potential cutting sites for the resolution of Holliday junctions. See text for further explanation of the proteins involved.



1.7.2.1 Presynapsis

The protein(s) that drive the resection of the DNA duplex end, to reveal a recombinogenic 3' overhang, are still poorly understood in eukaryotes. One set of proteins proposed to play a key role in presynapsis is the Mre11, Rad50 and Xrs2 (MRX) complex. The 5' to 3' resection of DNA duplex ends is attenuated in mutants of the MRX complex during repair of HO-endonuclease induced DSBs in mitotic cells (Ivanov *et al*, 1994; Tsubouchi and Ogawa, 1998; Lee *et al*, 1998). The MRX complex, particularly Mre11, has many nuclease activities (Paull and Gellert, 1998; Trujillo *et al*, 1998; Moreau *et al*, 1999; Usui *et al*, 1998), none of which, however, correspond to the expected 5' to 3' exonuclease activity for revealing 3' DNA overhangs. If this is analogous to the bacterial proteins that create 3' overhangs from DSBs (named RecBCD reviewed in Eggleston and West, 1997), perhaps a helicase activity could unwind the duplex and the observed MRX single strand endonucleases activity could cleave the 5' strand to reveal a 3' overhang (Haber, 1998). Alternatively, the MRX complex might not itself process the break directly, but coordinate the process (Bressan *et al*, 1999).

The MRX complex has a multitude of other roles, including maintaining telomere length and genomic stability, the creation of DSBs during meiosis, DNA replication, NHEJ and DNA damage signalling (reviewed in Haber, 1998; D'Amours and Jackson, 2002; Hopfner *et al*, 2002; Connelly and Leach, 2002), thus making the assessment of its contribution to particular phenotypes difficult.

The strand invasion step (synapsis) whereby the single strand 3' overhang invades a homologous duplex requires the prior assembly of a nucleoprotein filament upon the single-stranded DNA. In eukaryotes this protein is Rad51 (Shinohara *et al*, 1992), and is homologous to bacterial RecA, archaeal RadA and T4 bacteriophage UvsX (reviewed in Baumann and West, 1998; Thacker, 1999; Gasior *et al*, 2001).

Mutations of yeast *rad51* reduce recombinational repair and the frequency of spontaneous gene conversion by approximately 20-fold (Aboussekhra *et al*, 1992; Shinohara *et al*, 1992). Mutation of *rad51* in vertebrate cells also leads to an accumulation of chromosomal breaks as a precursor to cell death (Sonoda *et al*, 1998). *in vitro*, Rad51 exhibits homologous pairing and strand exchange activities in an ATP-dependent manner (Sung, 1994) and forms a nucleoprotein filament on single- and double-stranded DNA (Ogawa *et al*, 1993; Sung and Robberson, 1995). Despite the Rad51 and RecA proteins sharing only

~30% sequence homology (Shinohara *et al*, 1992) the nucleoprotein filaments they form are very similar, suggesting a structural conservation (Ogawa *et al*, 1993). Also, *rad51* null point mutations are almost entirely (75%) within residues that are conserved between a variety of RecA homologues (Chanett *et al*, 1996).

The nucleoprotein filament formed by Rad51 *in vitro* causes the DNA to be underwound, extending the DNA from 12 bp per helical turn to 18.6 bp, consistent with 3 basepairs per monomer, the same as seen with RecA nucleoprotein filaments (Ogawa *et al*, 1993; Sung and Robberson, 1995). Three monomers of Rad51 per basepair of DNA are also required to give an optimal strand exchange rate (Sung and Robberson, 1995). Although Rad51 binds DNA by itself, the reaction is stimulated, or facilitated, by certain protein factors, namely RPA (replication protein A), Rad52, and the Rad55-Rad57 heterodimer. RPA is a heterotrimeric single-stranded binding complex that is believed to increase the efficiency of Rad51 nucleoprotein filament formation by removal of secondary structures in the ssDNA substrate (Sung and Robberson, 1995; Sugiyama *et al*, 1997). The most critical co-factor is Rad52p, mutation of which obliterates almost all recombination. Rad52 forms a complex with Rad51 (Sung, 1997) and most likely facilitates the formation of the Rad51 nucleoprotein filament by interacting with RPA, as opposed to replacing RPA, to allow binding of Rad51 (New *et al*, 1998; Shinohara and Ogawa, 1998). Human Rad52 also stimulates Rad51-dependent strand exchange (Benson *et al*, 1998), but the enzyme is not so critical for homologous recombination in vertebrate cells (Yamaguchi-Iwai *et al*, 1998). This difference may lie in a redundancy found in vertebrate cells but not in yeast.

The Rad55-Rad57 heterodimer also functions as a mediator complex in strand exchange, mutation of either reduces the levels of spontaneous mitotic recombination (Rattray and Symington, 1995) and increases sensitivity to ionising radiation (Johnson and Symington, 1995). The heterodimer is proposed to help load Rad51 onto the single-stranded DNA as its absence delays this loading (Sugawara *et al*, 2003).

In vertebrate cells five Rad51 paralogues have been identified and shown to form two distinct complexes (Masson *et al*, 2001). Mutants of these proteins in DT40 cells display reduced growth rates and chromosomal instability (Takata *et al*, 2001). These complexes may play an early role in homologous recombination, perhaps being involved in formation of the Rad51-ssDNA presynaptic nucleoprotein filament, and Rad51 foci formation.

1.7.2.2 Synapsis

The Rad51p-ssDNA nucleoprotein filament is capable of interacting with other DNA molecules. Homologous alignment of sequences is likely to occur *via* random collision, as opposed to binding followed by sliding along the duplex DNA (Adzuma, 1998). The main problem faced by a strand exchange protein when searching for homology in a duplex is the inward facing Watson-Crick bonds. Either there is transient unpairing of the target duplex followed by checking for complementarity with the ssDNA bound in the nucleoprotein filament, or a triple-helix is formed between the duplex and the ssDNA, allowing a novel form of homology searching (reviewed in Rao and Radding, 1995). When a region of sufficient homology is found, the structure formed during the search is converted to a stable strand exchange intermediate by pairing the bases from the ssDNA with one strand of the homologous duplex target DNA. The joint molecule can be extended by branch migration of the invading ssDNA. In humans and yeast, Rad51-directed branch migration can occur in a 3' to 5', or 5' to 3', orientation with respect to the single strand (Sung and Robberson, 1995; Baumann and West, 1997; Namsaraev and Berg, 1998; Namsaraev and Berg, 2000). Branch migration is reduced or inhibited by the insertion of heterologous sequence in the single- or double-stranded substrate (Sung and Robberson, 1995; Namsaraev and Berg, 2000; Holmes *et al*, 2001).

Rad51p alone is unable to form a D-loop structure *in vitro*, and is capable only of exchanging strands between a duplex molecule and ssDNA suggesting the role of other proteins in initiating or promoting synapsis. One such protein is Rad54, a member of the Swi2/Snf2 family of chromatin remodelling proteins (reviewed in Pazin and Kadonaga, 1997). Inclusion of Rad54 allows the Rad51-ssDNA nucleoprotein filament to form D-loops *in vitro* (Petukhova *et al*, 1998), and Rad54 also interacts, *in vitro* and *in vivo*, with Rad51 (Jiang *et al*, 1996). Rad54 appears to alter the conformation of DNA in a manner suggestive of a transient separation of strands or unwinding (Petukhova *et al*, 1999; Tan *et al*, 1999), and is dependent on the interaction of Rad54 with the Rad51-ssDNA nucleoprotein filament (Van Komen *et al*, 2000; Mazin *et al*, 2000). Recently Rad54 has been shown to catalyse bi-directional nucleosome distribution, a process stimulated by the Rad51-nucleoprotein filament (Alexeev *et al*, 2003).

1.7.2.3 Postsynapsis

Less is known about the processes of postsynapsis. Firstly, the invaded strand must promote DNA synthesis. Holmes and Haber (1999) have shown that factors involved in lagging strand synthesis are required for successful gene conversion events at the mating-type switching locus, suggesting that a replication fork similar to an origin-initiated fork may synthesise DNA during repair. After DNA synthesis, rejoining of DNA molecules must take place and this may occur by the formation and resolution of Holliday junctions, as in bacteria. One candidate for a Holliday junction resolvase in eukaryotes is Mus81, a protein conserved from yeast to humans (reviewed in Haber and Heyer, 2001) that is in a complex that can cleave Holliday junctions *in vitro* (Boddy *et al*, 2001; Chen *et al*, 2001). In the fission yeast *Schizosaccharomyces pombe*, the protein is found in a complex with Emel and mutation of either results in an identical phenotype (Boddy *et al*, 2001). A similar relationship is seen in *S. cerevisiae* with Mms4, a protein weakly homologous to Emel (Kaliraman *et al*, 2001). It has been suggested, however, that the complex is not a resolvase that would function in the postsynaptic phase of gene conversions, and rather it is primarily involved in the restart of collapsed replication forks (Haber and Hayer, 2001; Doe *et al*, 2002).

Mammalian cell-free extracts have a branch migration activity, which cofractionates with a Holliday junction resolvase activity (Constantinou *et al*, 2001). This suggests that, in mammalian cells at least, there is a resolvasome present that couples branch migration with Holliday junction resolution. Most recently Constantinou *et al* (2002) have demonstrated that human cell free extracts contain two distinct endonucleases that can cleave Holliday junctions. The activities that promote the dissociation of DNA strands, and reannealing of dissociated strands, in the absence of Holliday junction formation are unknown.

1.7.3 Proteins involved in BIR

Overall the means by which BIR occurs is, in theory, similar to that of gene conversion. The end of the DNA duplex at the break still needs to be processed to a recombinogenic 3' overhang, which then needs to invade a homologous duplex and initiate replication and, potentially, a Holliday junction has to be resolved at the site of invasion. The only major

difference is that the second end of the DSB does need not be involved. Surprisingly, then, BIR can occur in the absence of some factors that are required for gene conversion (reviewed in Kraus *et* al, 2001). Most unusually, the main strand exchange protein rad51p, critical for gene conversion, is not required for BIR (Malkova *et al*, 1996; Signon *et al*, 2001). Genetic inactivation of *RAD51*, *RAD54*, *RAD55* and *RAD57* abolishes gene conversion but not BIR, whereas *RAD*52 is essential for both (Malkova, 1996; Signon *et al*, 2001). Mutations in *rad50*, *rad59*, or *tid1* , however, did not significantly affect gene conversion in this study, except when coupled with *rad51* or *rad54* mutations, where both BIR and gene conversion are eliminated. It appears, then, that a *RAD51*- and *RAD54*-independent pathway of BIR is present that relies on *RAD59*, *TID1*, and *RAD50* (and presumably *MRE11* and *XRS2*) as the ability to perform BIR is removed when the mutation of any of these genes is coupled with a *rad51* mutant. How conversion by BIR occurs in the absence of *RAD51* and its cofactors is still unidentified, but in one case a ~200 bp cis-acting element is required 34 kb upstream of the break site (Malkova *et al*, 2001). The strand invasion step in *RAD51*-independent BIR may require the Mre11-Rad50-Xrs2 (NBS1) complex, as human Mre11 has ssDNA annealing activities (de Jager *et al*, 2001), or other candidates with ssDNA annealing activities such Rad52 and Rad59 (Mortensen *et al*, 1996; Shinohara *et al*, 1998; Petukhova *et al*, 1999; Davis and Symington, 2001). It is possible that one, or more, of these proteins manages to catalyse the pairing, allowing the initiation of replication.

One fundamental difference, recently uncovered by Ira and Haber (2002), between gene conversion and *RAD51*-independent BIR, relates to the homology requirements for conversion. *RAD51*-dependent repair requires approximately 100 bp or more to be efficient, whereas the *RAD51*-independent BIR requires as little as 30 bp. Thus, for BIR to initiate in *rad51* mutant cells much shorter stretches of homology are required. Little is known of any *RAD51* dependent pathway of BIR.

1.7.4 The role of mismatch repair

Mismatch repair (MMR) maintains genome stability by correcting base mismatches arising through replication errors or chemical attack (for reviews see Kolodner and Marsischky, 1999; Hsieh, 2001). It also plays a crucial role in regulating homologous recombination (Evans and Alani, 2000).

In *S. cerevisiae*, MMR requires MSH2 as a heterodimer with either MSH6, to recognise mismatches and small insertion/deletion loops, or MSH3 to bind larger insertion/deletion loops; repair of single base mismatches appears to involve only the former complex. *S. cerevisiae* has at least three other MMR heterodimers as well: MLH1 with either PMS1, MLH2 or MLH3. It is thought that mismatch recognition is achieved by the MSH heterodimers, followed by an interaction with the MLH heterodimers and other factors to remove and repair mismatches. All these proteins appear to be common to most eukaryotes, although the number in each species can vary.

MMR proteins can inhibit recombination between heterologous sequence (Datta *et al*, 1996; Datta *et al*, 1997; Elliott and Jasin, 2001 and reviewed in Evans and Alani, 2001). It influences the chances of completing a recombination reaction and the extent of crossing over, as well as possibly in the search for homology during recombination itself. These anti-recombinogenic processes have been implicated in blocking gene transfer between bacterial species and inhibiting recombination between divergent repeats, thus reducing genomic rearrangements (Abuin *et al*, 2000). The general consensus is that, as divergence increases, the probability of forming successful heteroduplexes decreases exponentially in a MMR dependent manner. The mechanisms by which mismatch repair exerts this influence on recombination are not yet understood, nor its influence on the different pathways of homologous recombination.

In *T. brucei* there are orthologues of at least 5 MMR genes, and those that have been functionally analysed show a role in both MMR and the inhibition of recombination between heterologous sequences (Bell *et al*, in press; Bell and McCulloch, in press).

1.8 The potential role of homologous recombination in the duplicative transposition of *VSG* genes

By basing the duplicative activation of *VSG* genes on the models of homologous recombination, one can make certain predictions of its molecular nature. *VSG* genes located in chromosome internal arrays can, in theory, be activated only by gene conversion. It is predicted that chromosome-internal *VSG* gene conversion events would never involve crossover events, as this could lead to translocations. Thus the model of SDSA, which allows for the conservative replication seen in *VSG* gene conversions, is more probable

(Barry and McCulloch, 2001; Borst, 2001). *VSG* genes located in telomeric loci can be activated either by gene conversion or BIR, as neither reaction would lead to loss of critical sequence between the *VSG* gene end and the telomere in the BES.

As alluded to earlier, the conversion limits of duplicative transpositions of *VSG* genes seem always to be in regions of homology. Several studies have implicated the 70-bp repeat region in the BES, located ~1-2kb upstream of nearly all *VSG* genes, as the upstream limit of *VSG* duplicative transposition (Liu *et al*, 1983; De Lange *et al*, 1985; Florent *et al*, 1987; Lee *et al*, 1987; Shah *et al* 1987; Kooter *et al*, 1988; Matthews *et al*, 1990; Le Count, 2001). It has been proposed that this region could act consistently as a region of homology required for conversions, as it is found upstream of all *VSG* genes and in very large arrays upstream of the *VSG* genes in the BESs (Barry, 1997).

It has been found that, in monomorphic trypanosome lines (laboratory adapted strains that switch at a rate up to 4-5 orders of magnitude lower than the pleomorphic strains), *VSG* duplicative transpositions tend to involve fortuitous sequence homologies, as opposed to the pleomorphic lines' consistent use of the 70-bp repeats (Shah *et al* 1987; Matthews *et al*, 1990; Le Count, 2001). In fact, removal of the 70-bp repeats from the active BES in a monomorphic line made no difference to the switching frequency (McCulloch *et al*, 1996); this experiment remains to be performed in a pleomorphic line. So data on *VSG* switching infer a dedicated mechanism utilising the 70-bp repeats to drive *VSG* switching at a rapid rate which has been lost during laboratory adaptation in the monomorphic lines. The monomorphic lines would, as the switching rate indicates, presumably switch *VSG*s *via* spontaneous recombination.

How might the 70-bp repeats be involved in VSG switching? A rapid switching rate, one that is far higher than the expected background level of sequence shuffling, is likely to involve an increased level of recombination. If, as postulated by models of gene conversion, a DSB is the initiating event, it could occur in the 70-bp repeat region. The sequence may act as a site for a specific endonuclease that would create a DSB, or be bound by proteins that block replication, leading to fork arrest and subsequent collapse. These proposals are merely speculation, however, as no such endonuclease has been observed.

An analogy of a break initiating a gene conversion event is the mating-type switch, in *S. cerevisiae*, which begins with a DSB created by the HO-endonuclease. This leads to the duplication of sequence into the locus that received the break, thus creating a switch in the

mating-type (reviewed in Haber (2002)). The standard homologous recombination factors necessary for gene conversion are required for the event, thus necessitating only an initiating event to create a switch.

If a DSB break does indeed initiate the *VSG* switch and it is repaired by the use of silent *VSG* donors *via* homologous recombination machinery, then mutation of proteins involved in the process should affect the frequency of switching. After a break is created, the DNA end would require processing into a recombinogenic 3' overhang in both gene conversion and BIR events. The Mre11-Rad50-Xrs2 complex is suggested to perform this role; however, inactivation of *T. brucei MRE11* did not result in a reduction of *VSG* switching or an alteration in the frequency of use of different mechanisms (Robinson *et al*, 2002). DNA transformation efficiency, however, was reduced, but only by 4 –10 fold (Robinson *et al*, 2002; Tan *et al*, 2002). The mutants were also sensitive to phleomycin, a DNA damaging agent that causes DSBs and, to a lesser extent, the alkylating agent methyl methanesulphonate (MMS) (Robinson *et al*, 2002; Tan *et al*, 2002). *mre11* mutants also display increased frequency of gross chromosomal rearrangements (GCRs) (Robinson *et al*, 2002), a phenomenon seen in *mre11* mutants in other organisms (Chen and Kolodner, 1999; Myung *et al*, 2001). As homologous integration can still occur, it seems that if the Mre11 complex is catalysing the exonucleolytic degradation of the 5' strand at the DSB, it is not the only factor catalysing this process, agreeing with some speculations that Mre11 is not directly involved with this process.

It is still difficult to reconcile that *mre11* mutants are reduced in their ability to repair DSBs and integrate linear DNA with the fact that they are not reduced in their ability to duplicate *VSG* genes, a mechanism expected to occur *via* a DSB. One interesting possibility stems from the mutants' lack of a strong sensitivity to MMS (Robinson *et al*, 2002). MMS is an alkylating agent that creates adducts in the DNA that potentially lead to a DSB if a replication fork is arrested by the lesions and subsequently collapses. Mre11 might not be directly involved in the repair of DSBs induced this way. If this kind of damage is a predominant cause of spontaneous recombination, a reduction of *VSG* switching rate in *mre11* mutants might not be detected in monomorphic lines. Also, the assay used to measure *VSG* switching frequencies is not very sensitive, being able to detect changes of only an order of magnitude; a reduction may be slight and therefore undetectable.

As Mre11 has been implicated in *RAD51*-independent BIR, the potential of this pathway to catalyse the conversion of telomeric *VSG* genes is perhaps reduced. However, as only 2% of conversions in wild-type budding yeast cells use any BIR pathway (Malkova *et al*, 1996), reduction of this event in a pathway of *VSG* gene conversion relying on spontaneous recombination may be undetectable. Thus, the use of Mre11 in a pathway of telomeric *VSG* gene conversion may still occur in the high switching pleomorphic lines, where telomeric gene conversion predominates over any other reaction (Robinson *et al*, 1999).

Another set of factors predicted to be involved in the duplication of *VSG* genes is the main recombinase RAD51 and its co-factors RAD52, RAD54, RAD55 and RAD57. They would catalyse the invasion of a free DSB-induced DNA end from the active BES into a silent *VSG* gene locus, prompting synthesis of DNA and subsequent conversion of the BES *VSG* gene. As predicted, *T.brucei rad51* null mutants have a reduced *VSG* switching rate (McCulloch and Barry, 1999), which is consistent with other homologous recombination and repair defects in this mutant, including an increased population doubling time, a reduced DNA transformation efficiency and increased DNA damage sensitivity. One unexpected property of the reduced switching frequency is that recombinational and *in situ* switching events were affected equally. An effect upon transcriptional regulation of *VSG* genes would not be expected for the inactivation of a gene central to homologous recombination. It is unlikely that the reduction in both pathways is due to the observed reduction in growth rate and it is more likely that a residual pathway, which is *RAD51*-independent, was catalysing the remaining *VSG* switching events. How lack of *RAD51* affects both means of switching is still unknown, but the authors suggested that either a recombinational event is required for *in situ* switching to occur, or a transient, RAD51-mediated interaction between BESs is required. The former suggestion is unlikely, as DNA rearrangements are not consistently found during *in situ* switching (Horn and Cross, 1997), however smaller rearrangements cannot be discounted. Whether or not RAD51 is involved in such reactions is still undetermined. An alternative possibility is that the absence of RAD51 affects other proteins that are involved in *in situ* switching, thus having an indirect effect (McCulloch and Barry, 1999). Again, an important point to note is that the *rad51* null mutations were performed in the low switching monomorphic lines, so it is possible that the effect seen on switching is not an effect upon the genuine mechanism of switching, but rather upon the background

spontaneous recombination operative in monomorphic trypanosomes. By assuming, as the evidence suggests, that *VSG* switching in rapid switching pleomorphic lines is driven by a mechanism acting on the 70-bp repeats as the point of homology that initiates *VSG* gene conversion, then one can make predictions based upon the models of RAD51 acting on such a DNA substrate in yeast. In favour of a role for RAD51, studies upon RecA and Rad51 have shown that these proteins prefer to bind *in vitro* to GT-rich oligonucleotides (Tracy and Kowalczykowski, 1996; Tracy *et al*, 1997; Biet *et al*, 1999). These oligonucleotides are very similar to the GT-rich region within the 70-bp repeat so could be a favourable target for the protein in the trypanosome nucleus. 70-bp repeats also contain a (T-purine-purine)$_n$ repeat tract that is commonly TAA (Aline *et al*, 1985). A (TAA)$_{90}$ tract has been shown to be predisposed to strand separation *in vitro* (Ohshima *et al*, 1996), perhaps making strand invasion by Rad51 more amenable. However, it would also make the 70-bp repeats more amenable to any recombination events requiring the invasion of a homologous duplex, regardless of RAD51 (*e.g.* BIR).

Against a role for RAD51 is the degeneracy of the 70-bp repeats. Rad51 requires a large stretch of homology, greater than 100 bp, to be efficient in catalysing gene conversion *in vivo* (Ira and Haber, 2002). This is also true in *T.brucei* (Bell and McCulloch, in press). Heterology within strand exchange intermediates, catalysed by Rad51 aborts the reaction *in vitro* (Sung and Robberson, 1995; Namsaraev and Berg, 2000; Holmes *et al*, 2001), and heterology *in vivo* is known to reduce recombination efficiency (Evans and Alani, 2000). Thus, it appears that, despite their having features favourable to binding by Rad51, the degeneracy of 70-bp repeats may not favour subsequent strand exchange. Recently, in both yeast and trypanosomes, a *RAD51*- independent pathway of homologous recombination has been identified that appears to require only short stretches of homology (Ira and Haber, 2002; Conway *et al*, 2002). In the case of yeast, 30 bp is sufficient to catalyse recombination (Ira and Haber, 2002), whereas in trypanosomes 24 bp is sufficient to catalyse the integration of exogenous linear DNA into the parasite genome (Conway *et al*, 2002). In both instances, the presence of RAD51 seemed somewhat to inhibit the use of short sequences to act as a substrate in recombination. In trypanosomes, the use of the short sequences was analysed in greater detail, and it was found that, in some insertions, as few as 7-13 bp were used to target integration, and these stretches could include some heterology (Conway *et al*, 2002). Thus, this *RAD51*-independent pathway could catalyse *VSG* switching in the absence of RAD51 in the monomorphic mutants. Moreover, due to

its requirement for only short regions of homology, it may be the dominant mechanism of *VSG* switching in pleomorphic lines.

In trypanosomes, the reduction in efficiency of the integration of heterologous, exogenous, linear DNA is alleviated to some extent by the genetic inactivation of mismatch repair genes (Bell *et al*, in press). Null mutation of mismatch repair genes in monomorphic lines, however, does not affect *VSG* switching rates (Bell and McCulloch, in press). These data fit with the hypothesis that the mismatch repair system is restricting the use of heterologous sequence during recombination and thus, if spontaneous recombination is catalysing *VSG* switching in monomorphic lines, mismatch repair may restrict the use of the 70-bp repeats, as is seen by their lack of involvement in switching events (Rudenko *et al*, 1996)).

No other genes involved in homologous recombination have been investigated in trypanosomes. One would predict that removal of the co-factors of RAD51 (RAD54, RAD55, RAD57 and potentially other RAD51-like) would have an effect similar to that of the mutation of *rad51*, as they are in the same pathway. The key recombination factor in yeast, *RAD52*, is yet to be identified in trypanosomes by genome searching or degenerate PCR (Richard McCulloch, personal communication). Its absence, however, is not too surprising as it is not apparent in *Caenorhabditis elegans* or *Drosophila melanogaster* and does not play as critical a role in DSB repair in mammalian cells (Rijkers *et al*, 1998; Yamaguchi-Iwai *et al*, 1998).

1.9 Objectives

i) to elucidate the use of sequence homology, with regard to the 70-bp repeats *in vivo*, in driving the conversion of *VSG* genes during antigenic variation.

ii) to create an assay whereby the 70-bp repeat sequence homology requirements could be tested, and to determine if the 70-bp repeats are bound by specific proteins or acted upon, specifically, by enzymes involved in DNA metabolism.

iii) to determine the role of *RAD51* in pleomorphic lines and whether the *RAD51*-independent pathway is capable of catalysing *VSG* switching in these cell lines.

CHAPTER 2

MATERIALS AND METHODS

REAGENT ABBREVIATIONS

APS	ammonium persulfate
B & W	bind and wash
BSA	bovine serum albumin
CBSS	Carter's Balanced Salt Solution (1 x): 0.023 M HEPES, 0.12 M NaCl, 5.41mM KCl, 0.55 mM $CaCl_2$, 0.4 mM $MgSO_4$, 5.6 mM Na_2HPO_4, 0.035 M glucose, 0.04 mM phenol red, pH to 7.4
CIP	calf intestinal phosphatase
DAPI	4, 6-diamidino-2-phenylindole
DEPC	diethyl pyrocarbonate: Used at 0.1% to remove RNAase
dH_20	distilled water
DMSO	dimethyl sulphoxide
dNTP	deoxynucleoside triphosphate
FITC	fluorescein isothiocyanate
PBS	phosphate buffered saline (Sigma, Ltd.)
PSG	phosphate/ sodium chloride/ glucose buffer (1 x): 0.06 M Na_2HPO_4, 3.6 mM NaH_2PO_4, 46 mM NaCl, 55mM glucose, pH 8
RT	reverse transcriptase
SDS	sodium dodecyl sulphate
SOB	Bacterial media (per Litre): 20 g bacto-tryptone, 5 g bacto-yeast extract, 0.5 g NaCl
SOC	SOB + 20 mM glucose
SSC	sodium chloride/ sodium citrate solution (1 x): 0.15 M NaCl, 0.015 M $Na_3C_6H_5O_7$
TAE	TRIS/ acetate/ EDTA buffer (1 x): 0.04 M TRIS base, 0.04 M glacial acetic acid, 1 mM EDTA

TE	10 mM Tris.Cl, 1mM EDTA
TEMED	N, N, N', N'-tetramethyl-ethylenediamine

2.1 Trypanosome strains and their growth

The two trypanosome strains, ILTat 1.2 and Lister 427, and derivatives thereof, were routinely grown in this study. *T. brucei* bloodstream form cells of strain Lister 427, expressing MITat1.2a (VSG 221), were derived from an unknown number of syringe passages through rodent hosts over many decades. Lister 427 switches the variant surface glycoprotein (VSG) being expressed at about $1 \times 10^{-6} - 1 \times 10^{-7}$ switches/cell/generation, and is monomorphic, displaying only the bloodstream form stage routinely.
The line colloquially named ILTat 1.2 (I.L.R.A.D. – International Laboratory for Research in Animal Diseases – *Trypanozoon* antigen type) is a line of the EATRO 795 stock (a field isolate from bovine blood; Uhembo, Central Nyanza Province, Kenya; 1964) that has been subjected to some serial syringe passaging and is more virulent than its parent line. ILTat 1.2 is capable of tsetse transmission, is pleomorphic and has a switching rate of approximately 1×10^{-5} switches/trypanosome/generation.
In vitro growth was achieved in the Lister 427 line by using HMI-9 medium (Hirumi and Hirumi, 1989), whereas the ILTat 1.2 line was grown in HMI-9, supplemented with 2.75 % methylcellulose. Trypanosomes grown in female Wistar rats (250 g), or female ICR mice (25 g), were infected via interperitoneal injection.

2.2 Trypanosome isolation and stabilate manufacture

To obtain parasites from the blood of the mammalian host, the animal was exsanguinated by cardiac puncture into 5 % sodium citrate anticoagulant in CBSS. Either a Percoll (Sigma) density gradient (Grab and Bwayo, 1982) or a DEAE cellulose column was used to purify the trypanosomes. The former method involved mixing the trypanosome-containing blood with a Percoll stock solution (100 ml of 100 % Percoll with 8.55 g of sucrose and 2 g of glucose, made up to pH 7.4 with HEPES) at a 3 : 5 ratio, followed by centrifugation at 17500 x g for 15 min at 4 °C in a JA20 (Beckman) rotor with a fixed angle of 34°. Purified trypanosomes could be collected from a discrete band at the top of the gradient with a disposable plastic pipette. Trypanosomes collected by this method were washed several times in large volumes of CBSS or PSG to remove any trace of percoll. The latter method involved passing the blood from a rat through DEAE cellulose, equilibrated in PSG, at a

pH of 8.0. Fractions were collected containing the pure trypanosomes, and harvested by centrifugation at 1500 x *g*.

For analysis of VSG switching rates, a buffy coat was made to allow isolation of trypanosomes. Here, 0.4 ml of exsanguinated mouse blood (from 1 ml, containing 150 µl of CBSS with 5% citrate) was centrifuged at 5000 rpm in a microcentirfuge for 5 min. The centrifugation separates the blood into red blood cells as the bottom layer and plasma as the top layer, with the trypanosomes and the white blood cells being found in the interphase. This middle layer was removed with a needle (19 gauge) and syringe.

Trypanosome stabilates were prepared from culture by mixing 900 µl of log-phase culture, with 100 µl of 100 % glycerol in cryotubes (Nunc). The sample was frozen gradually to -70°C over 24 hours, and then transferred to liquid nitrogen for long-term storage.

2.3 Transformation of trypanosomes

2.3.1 Transformation of the monomorphic Lister 427 bloodstream form line

5 x 10^7 trypanosomes were electroporated in 0.5 ml of Zimmerman Post-Fusion medium (132 mM NaCl, 8 mM Na_2HPO_4, 1.5 mM KH_2PO_4, 0.5 mM Mg acetate, 0.09 mM Ca acetate, pH 7.0) supplemented with 1% glucose, at 1.4 kV and 25 µF capacitance using a Biorad Gene Pulser II. Approximately 5 µg of DNA, that had been restriction digested, phenol/chloroform extracted and ethanol precipitated, was routinely used for all transformations. Cells were recovered in 10 ml HMI-9 for 18 h before selection on drug. For selection, 1 x 10^7 cells were diluted in 18.5 ml of HMI-9 medium (~8.33 x 10^6 cells.ml^{-1}) containing the appropriate antibiotic, and plated out over 12 wells, at 1.5 ml per well, in multi-well dishes. Transformants were allowed to grow for 5-14 days, from which transformation efficiencies could be calculated. The antibiotics, and the concentration, used for selection were hygromycin (5.0 µg.ml^{-1}, Roche) and G418 (2.5 µg.ml^{-1}, Sigma).

2.3.2 Transformation of the pleomorphic ILTat 1.2 bloodstream form line

Due to the great difficulty of obtaining transformants of this line two methods were used:
Method (*i*): Trypanosomes were initially grown in culture or in rats. 3 x 10^7 were then electroporated in 0.5 ml of Zimmerman Post-Fusion medium supplemented with 1% glucose, at 1.4 kV and 25 μF capacitance using a Biorad Gene Pulser II. Approximately 5 μg of DNA, which had been restriction digested, phenol/chloroform extracted and ethanol precipitated, was used routinely for all transformations. Cells were recovered in 10 ml HMI-9 for 18 h before selection on drug. For selection, 3 x 10^6 cells were diluted in 18.5 ml of HMI-9 medium (~1.67 x 10^5.ml^{-1}) containing the appropriate antibiotic, and plated out over 12 wells, at 1.5 ml per well, in multi-well dishes. Transformants were allowed to grow for 10-21 days. The inefficiency of this method prevented calculation of transformation efficiencies. Puromycin (0.5 - 1.0 μg.ml^{-1}, Calbiochem), hygromycin (5.0 μg.ml^{-1}, Roche) and phleomycin (1 - 2 μg.ml^{-1}, Cayla) were the antibiotics used for selection.
Method (*ii*) (based upon a personal communication from Eric Vassella): Trypanosomes were initially grown in culture. 2 x 10^7 were then electroporated as in method (i). After the electroporation, the cells were recovered for 30 min in 10 ml of medium, and then centrifuged at 583 x *g* for 10 min at room temperature. These cells were then resuspended and recovered in 130 ml of medium for 18 h prior to selection. Drug selection was achieved by adding the drug at the appropriate concentration to the 130 ml of medium. Transformants were allowed to grow for 10-21 days. Hygromycin (5.0 μg.ml^{-1}, Roche), Puromycin (0.5 - 1.0 μg.ml^{-1}, Calbiochem) and phleomycin (1 - 2 μg.ml^{-1}, Cayla) were the antibiotics used for selection.

2.4 Immunofluorescence

2.4.1 Preparation of trypanosome slides for immunofluorescence

Blood smears were prepared by using the inclined end of a glass slide to draw along a drop of blood (~7 μl) across along a separate slide. After air-drying for approximately 5 min the

slide was fixed by submersion of the slide in methanol for 5 min, and was then air-dried and stored at 4 °C in sealed polythene bags containing silica gel.
Trypanosomes from culture were harvested by centrifugation at 583 x *g* for 10 min at room temperature and washed three times in PBS before resuspension in PBS. 10 µl spots of the resuspension were placed on slides and allowed to air dry. Slides were fixed by submersion in methanol for 5 min, and then air-dried and stored at 4 °C in sealed polythene bags containing silica gel.

2.4.2 Immunofluorescence of the trypanosome VSG coat

Prepared slides were marked into wells for antibody containment with an acrylic paint pen (Mark-Tex corp., BDH). Well contents were then rehydrated with 50 µl 1 x PBS, 1% BSA solution for 15 min. The 1 x PBS, 1% BSA solution was carefully removed with a pipette and 50 µl of the primary antibody dilution was added (see Table 2.1) and incubated for 30 min in a humid chamber. The slides were then rinsed in 1 x PBS, 1 %BSA and washed by submersion in 1 x PBS, 1% BSA for 5 min. 50 µl of a 1 : 50 dilution of a FITC conjugated secondary antibody, specific to host from which the primary antiserum was developed, was added to each well and incubated for 30 min in a dark humid chamber. The slides were then washed twice in the dark, each for 5 min in 1 x PBS. Vectorshield (Vector Laboratories) containing DAPI (1 µg of DAPI/ml) was then added, and sealed with a glass coverslip. Fluorescence was then observed by UV light microscopy.
The dilutions seen in Table 2.1 were used after titration previously with trypanosomes of known variable antigen type. The rat and rabbit antisera were obtained from Liam Morrison.

Variable antigen type	Dilution of antiserum	Species from which antiserum was derived
ILTat 1.2 (Day 6 antiserum from ILTat 1.2 infected rabbit)	1 : 100	Rabbit
ILTat 1.21	1 : 40	Rat
ILTat 1.22	1 : 40	Rat
ILTat 1.25	1 : 1250	Rabbit
ILTat 1.64	1 : 40	Rat
ILTat 1.67	1 : 40	Rat
ILTat 1.71	1 : 40	Rat
ILTat 1.73	1 : 1750	Rabbit

Table 2.1 Description of the primary antisera used for identification of VSGs on the surface of trypanosomes.

2.5 Analysis of VSG switching frequency

The method used to analyse the frequency of VSG switching here is based upon that used by McCulloch and Barry (1999). In the present study, the switching frequencies of the ILTat 1.2 wild-type, and *RAD51* heterozygous and homozygous mutants, derived from the wild-type, were analysed. Clones for the wild-type and each independent heterozygous and homozygous mutant line, were generated by plating out a dilution of 25 trypanosomes in 20 ml of medium over 96 wells. Two clones were taken for each line. Each clone was cultured, and 2 x 10^5 trypanosomes were injected into mice, with 2-3 mice being used per clone. These mice were then cured by injection of cymelarsan (Rhone Merieux; 5 $mg.kg^{-1}$) 3 days later. These immunized mice were then infected (between 7 and 35 days post-curing) with 4 x 10^7 trypanosomes grown in culture and resupended, after centrifugation at 583 x *g* for 10 min at room temperature, in 250µl HMI-9. Twenty-four hours after infection, the mice were exsanguinated and the trypanosomes were isolated by the buffy coat method (Section 2.2). Trypanosomes from the buufy coat were then mixed with 20 mls of medium per 0.4 ml of mouse blood and plated out over 96 wells. Up to 31 days were given for cells to grow and, by assuming that each well represents an individual switch, the frequency of *VSG* switching can be calculated indirectly (see Chapter 5).

2.6 Isolation of genomic DNA

5 - 40 x 10^7 trypanosomes were harvested from culture by centrifugation at 1620 x *g* for 10 min at room temperature, or from blood by separation on a Percoll gradient followed by centrifugation with the same conditions. The trypanosomes were then resuspended in digestion buffer (50 mM Tris.Cl (pH 8.0), 1 mM EDTA, 100 mM NaCl). SDS was added to 1% in the presence of 100 $\mu g.ml^{-1}$ proteinase K, and the preparations were incubated at 37 °C overnight to lyse the trypanosomes and digest proteins. The DNA was phenol/chloroform extracted, ethanol precipitated and resuspended in 50 mM Tris.Cl (pH 8.0), 1 mM EDTA.

2.7 Polymerase chain reaction (PCR)

PCRs were routinely set up in a 50 µl volume. *Taq* (ABgene - 2.5 units per reaction), Herculase (Stratagene – 2.5 units reaction) and *Pfu* (Stratagene – 1.25 units per reaction) polymerases were commonly used to amplify the DNA. For each polymerase the 5 µl of 10 x buffer and manufacturer's recommended magnesium concentration was added, along with 1 µl of 10 mM dNTPs and 2 µl of each primer at 5 µM. The reaction conditions varied depending upon the particular PCR. PCR products were routinely purified using the Stratagene PCR purification kit. A list of the oligonucleotides used for PCRs or otherwise are listed in Table 2.2 at the end of this chapter.

2.8 DNA digestion, electrophoresis and Southern blotting

2.8.1 Restriction enzyme digestion of DNA

Restriction enzyme digestion of plasmid DNA was routinely performed for 2 h at the specified temperature using commercial restriction enzymes (NEB). Genomic DNA, however, was digested for a period ranging from 4 hours to overnight.

2.8.2 Gel electrophoresis

DNA was separated on 0.6 – 2 % agarose (Seakem LE, BMA) gels, at 100 v in 1 x TAE running buffer, using a commercial 1 kb size standard marker (GibcoBRL, Life Technologies). For Southern blotting , separations of digested genomic DNA were run at 30 v overnight. The agarose gels routinely contained 0.2 $\mu g.ml^{-1}$ ethidium bromide in order to visualise the DNA under UV light.
DNA extraction from gel fragments was performed using the Qiagen gel extraction kit.

2.8.3 Southern blotting

Agarose gels to be Southern blotted were initially photographed on a UV transilluminator with a ruler parallel to the gel so that the sizes of bands detected by the hybridisation of radioactively labelled DNA (see section 2.9) could be measured. To nick the DNA, the gel was placed in 0.25 M HCl for 15 min and then rinsed with distilled water. Following this the DNA was denatured by placing the gel in denaturation solution (0.5 M NaOH, 1.5 M NaCl) for 30 min. After rinsing with distilled water, the gel was placed in neutralising solution (1 M Tris-HCl pH 8.0, 1.5 M NaCl) for a further 30 min. By wet blotting using 20 x SSC as the transfer buffer (Sambrook *et al.*, 1989), the DNA was transferred to a nylon membrane (Hybond-XL). Blots were routinely left for 24 h and then the DNA was crosslinked to the membrane using a UV spectrolinker (Stratagene).

2.9 Radiolabelling and hybridisation of DNA

2.9.1 Probe manufacture by random nonamer radiolabelling of DNA

Fragments of DNA used for probes were gel-extracted. Radiolabelling of these DNA fragments was performed using a commercial kit (Prime-It II kit, Stratagene). Twenty-five nanograms of the purified DNA fragment were mixed with 10 µl of random-sequence nonamers i.e. oligonucleotides and sterile, distilled water, in a total reaction volume of 36 µl. The mixture was then heated to 95 °C for 5 min to denature the DNA and cooled to allow the random oligomers to anneal. 10 µl 5 x primer buffer, 3 µl α-^{32}P labelled dCTP

(20 μCi) and 1 μl Klenow ($5U.\mu l^{-1}$) were then added, mixed, and incubated at 37 °C for 8 min. The resulting probes were then purified from nucleoside triphosphates by passage through Microspin columns (Amersham). After purification, the probes were denatured at 95 °C for 5 min before use.

2.9.2 Hybridisation of radiolabelled DNA probes

Nylon membranes (Hybond XL), after the transfer of DNA via Southern blotting, were placed into glass hybridisation tubes. Approximately 25 ml of Church-Gilbert solution (0.342 M Na_2HPO_4, 0.158 M $NaH_2PO_4.2H_2O$, 0.257 M SDS and 1mM EDTA) was added to the tubes and the membranes were prehybridised for at least 1 h at 65°C in a rotating hybridisation oven. The denatured probe was then added to the Church-Gilbert solution in the tube, and the hybridisation was left for approximately 18 h at 65°C. After hybridisation, the membranes were washed, for 30 min each, in a rotating oven at 65°C, with the following series of solutions (50 ml solution used per wash): 5 x SSC, 0.1% SDS (twice – once using solutions at room temperature); 2 x SSC, 0.1% SDS; 0.2 x SSC, 0.1% SDS. The filters were finally rinsed in 0.1 x SSC and heat-sealed in plastic. To visualise the hybridisation the membranes were placed with medical photographic film (Konica Medical Corporation) in an autoradiography cassette at –80 °C for 4-168 h, depending on the strength of signal. An alternative method of visualisation was to place the membrane with a phosphorimage screen (Fuji) at room temperature for 1 to 48 h, again depending on the strength of the signal.

2.9.3 Stripping of hybridised nylon-membranes

Nylon membranes hybridised with radiolabelled DNA were stripped with boiling 0.1% SDS. After adding this solution to the membrane, the solution was allowed to cool to room temperature. This was repeated, then the membrane was rinsed in 2 x SSC before reuse.

2.10 Cloning of DNA fragments

2.10.1 Cloning of DNA fragments using T4 DNA ligase

DNA fragments for cloning were either purified by gel extraction or by phenol:chloroform extraction. Some restriction enzyme digest products required an overhang to be filled in prior to ligation. This was achieved for 5' overhangs by using Klenow polymerase (NEB) in the presence of 10 mM dNTPs, after heat-inactivation of the restriction enzyme(s) used to cleave the DNA. The reaction was performed at 37 °C for 30 min. Prior to ligation the vector rountinely was treated with Calf-intestinal phosphatase (CIP) (Roche Diagnostics), for 30 min at 37 °C, to remove 5' phosphate groups, thus preventing self religation. After CIP treatment, the vector was purified either by phenol:chloroform extraction or from an agarose gel, after electrophoresis, with the Qiagen gel extraction kit.
Ligation of DNA fragments into a plasmid vector (pBluescript (Stratagene) – or derivatives thereof) were performed at room temperature for 4 h or at 16 °C overnight. Ligations were commonly performed in a 10 µl volume with 0.4 units of T4 DNA ligase (NEB).
With each cloning attempted, a variety of insert:vector ratios were prepared (*e.g.* 1,3,5 and 7 µl of purified product to 1 µl vector DNA), to maximize the likelihood of ligation.

2.10.2 Cloning of PCR products into TOPO Vector (Invitrogen)

PCR products generated by use of *Taq* DNA polymerase (AB gene) were suitable for cloning into this vector, as this enzyme produces a 3' single adenosine overhang. Herculase or *Pfu* polymerase-amplified PCR products require this overhang to be generated subsequently. To achieve this, 2.5 units of *Taq* polymerase were added at the end of the PCR and the reaction was heated at 72 °C for 15 min.
0.5 – 4 µl of PCR product was incubated with 1 µl of salt solution and 1 µl of TOPO vector, in a total volume of 6 µl made up with water, for 5 min at room temperature. 2 µl of this reaction was then used to transform prealiquoted TOP 10F' (Invitrogen) cells (see below).

2.10.3 Transformations and plasmid retrieval

The heat shock method was used to transform competent *E. coli* Xl-1 blue MRF' (Stratagene) or TOP10F' cells. 4 µl of ligation product was added to 60 µl of Xl-1 blue MRF' cells, or alternatively 2 µl of a TOPO cloning was added into 25 µl aliquots of TOP10F' cells. These were then mixed gently, and left on ice for 30 min. A heat shock was then performed at 42 °C for approximately 45 s, and the reaction tube was transferred immediately to ice for 2 min. Afterwards, 1 ml of L-Broth medium was added to the cells, which then were incubated at 37 °C for 1 h. The cells were then centrifuged for 1 min at 10 000 rpm in a microcentrifuge and resuspended in 100 µl L-broth. This resuspension was spread over L-agar plates containing 0.27 M ampicillin (Sigma, Ltd). Transformed colonies were re-streaked on fresh L-agar plates (also containing ampicillin at the same concentration). A single colony was selected from this plate and used to inoculate 5 ml L-broth (supplemented with ampicillin to a final concentration of 10 $mg.ml^{-1}$). Plasmids were prepared from 1.5 - 3 ml of the overnight culture using the Qiagen Miniprep kit, or from 200 - 400 ml of culture using the Qiagen Maxiprep kit.

2.11 Reverse transcription polymerase chain reaction (RT-PCR)

2.11.1 RNA isolation

$3 \times 10^5 - 2 \times 10^7$ trypanosomes were harvested from culture medium by centrifugation at 1620 x *g* for 10 min, then were washed twice in PBS and harvested by the same centrifugation procedure. RNA was then isolated using the Qiagen RNeasy mini kit (following the manufacturer's protocol).

2.11.2 Reverse Transcription

Before cDNA preparation, the RNA was treated with DNAase to remove any risk of false positives during further PCR reactions. 1 µg of RNA was incubated, for 15 min at room temperature, with 1 unit of DNAase (Invitrogen) and1 µl of 10 x DNAase buffer (200 mM Tris-HCl (pH 8.4), 20 mM $MgCl_2$, 500 mM KCl), in a total volume of 10 µl made up with DEPC-treated water. The reaction was terminated by the addition of 1 µl of 0.25 mM

EDTA and heating to 65°C for 20 min. The RNA was then ready for cDNA preparation by the Superscript™ First-Strand Synthesis System for RT-PCR kit (Invitrogen). 50 ng of random hexamers were added to 5 µl of the DNAase treated RNA, with 1 µl of 10 mM dNTPs in a volume of 10 µl made up with DEPC-treated water, and heated to 65 °C for 5 min and then incubated on ice for 1 min. 4 µl 25 mM $MgCl_2$, 2 µl 0.1 M DTT, 2 µl of 10 x RT buffer and 1 µl of RNaseOUT recombinant ribonuclease inhibitor were added to the RNA solution and it was incubated for 2 min at 25 °C. 50 units of Superscript™ II reverse transcriptase were added and the reaction contents was mixed and warmed to 42 °C for 50 min; the enzyme was then heat inactivated at 70 °C for 15 min and the tube was then chilled on ice. To remove any remaining single-stranded RNA, 1 µl of RNAase H (3.8 $U.\mu l^{-1}$) was added and the reaction was incubated at 37 °C for 20 min. This cDNA was then suitable as a template for PCR, with 2 µl commonly used in 50 µl PCR reactions. For each cDNA preparation an identical reaction was also set up with the same DNAase-treated RNA, but without reverse transcriptase. This acted as a negative control for the presence of DNA contamination in subsequent PCRs.

2.12 Phenol/chloroform extraction

For extraction of a DNA sample an equal volume of phenol/chloroform (at a 1:1 mixture) (Sigma) was added and the solution was mixed thoroughly by gentle inversion (for genomic DNA) or vortexing (for other forms).The two phases were separated by centrifugation at maximum speed in a microcentrifuge for 10 min at room temperature, after which the aqueous layer was removed and transferred to a fresh eppendorf tube. To remove traces of phenol an equal volume of chloroform was added and the tube contents was mixed by inversion or vortexing.The phases were then separated by centrifugation at maximum speed in a microcentrifuge for 5 min at room temperature, the aqueous layer was removed and added to 1/10 the volume of 3 M sodium acetate (pH 5.2). Two volumes of 100% ethanol were then added and the tube contents was mixed thoroughly. The tube was then incubated at –20 °C for a minimum of 20 min, after which the DNA was harvested by spooling or pelleted by centrifugation at maximum speed in a microcentrifuge for 30 min at 4 °C. The pellet was then washed in 1.5 ml 70% ethanol, air-dried, and resuspended in an appropriate volume of buffer (usually TE) or water.

2.13 Trypanosome nuclear extract preparation

Nuclear extracts were prepared exactly by the procedure of Bell and Barry (1995). Initially, female Wistar rats were infected with ILTat 1.2 trypanosomes by interperitoneal injection. These were then sacrificed and exsanguinated by cardiac puncture when at a parasitaemia of 10^9 trypanosomes.ml^{-1}. Trypanosomes were purified from the blood cells by anion exchange chromatography using DEAE cellulose. The purified trypanosomes were harvested by centrifugation at 583 x *g* for 10 min at 4 °C, and washed twice in cold 1 x PBS. After washing, the pelleted trypanosomes were resuspended in two packed cell volumes of Buffer A (10 mM Hepes (pH 7.9 at 4 °C), 10 mM KCl, 0.5 mM DTT). The cells were then lysed with 40 strokes of a Dounce homogeniser (A type pestle). To pellet the nuclei, the homogenate was centrifuged at 3700 x g for 5 min at 4 °C and the supernatant was discarded. The nuclei were then resuspended in Buffer C (20 mM Hepes (pH 7.9), 25 % glycerol, 0.42 M NaCl, 0.2 mM EDTA, 0.5 mM DTT) at 10^{10} cells.ml^{-1} and gently mixed for 30 min at 4 °C. This nuclear lysate was then centrifuged at 25 000 g for 30 min at 4°C. The supernatant was then dialyzed against 50 volumes of Buffer D (20 mM Hepes (pH 7.9), 20 % glycerol, 100 mM KCl, 0.2 mM EDTA, 0.5 mM DTT). The resulting dialysate was centrifuged at 25 000 *g*, at 4 °C for 20 min. The supernatant was snap-frozen in liquid nitrogen as 100 µl aliquots. The nuclear extracts were stored at -80 °C.
Note that all buffers used in this preparation contained protinase inhibitors : 0.5 mM PMSF, 10 µg.ml^{-1} leupeptin, 10 µg.ml^{-1} aprotinin and 10 µg.ml^{-1} pepstatin A.

2.14 Protein gel electrophoresis and visualisation

All protein samples were fractionated on SDS-polyacrylamide gels made up to the desired percentage using 30% acrylamide (Anachem) (Sambrook *et al*, 1989). APS and TEMED facilitated the polymerisation of the acrylamide and the gel was cast between two vertical glass plates and run in 1 X SDS-running buffer (0.18 M Glycine, 0.023 M Tris, 0.003 M SDS) at 200 V.
Protein gels were then washed twice in 50 ml of 10% methanol / 7% acetic acid and stained overnight, whilst gently rocking, in sypro-ruby (Biorad). The gel was then washed twice in 10% methanol / 7% acetic acid and visualised on a Typhoon phosphorimager.

2.15 Isolation of proteins bound to 70-bp repeat DNA from trypanosome nuclear extracts

The method used for this experiment was derived from Gabrielsen *et al* (1989), who used magnetic streptavidin coated magnetic beads, linked to biotinylated DNA, to isolate DNA binding transcription factors from yeast cell extracts.

Here, complementary oligonucleotides (1.22 70bp Repeat S and 1.22 70bp Repeat AS), corresponding to both strands of the 70-bp repeat region upstream of the ILTat 1.22 *VSG* gene were obtained. These were annealed to each other by mixing in TE, at equal molar ratios, and placing in a boiling water bath for 5 mins and allowing to cool gradually to room temperature. The sense strand oligo was biotinylated at the 5' end.The same was performed with oligonucleotides corresponding to the ILTat 1.64 promoter (gift from Michael Ginger).

Streptavidin-coated magnetic beads (Dynal) were mixed gently by inversion, and 1 mg was seperated from its suspension with a magnet (Dynal). These beads were washed twice in 100 µl of 2 x B&W buffer (10 mM Tris-HCl (pH 7.5), 1 mM EDTA and 2 M NaCl) and finally resuspended in 100 µl of B&W Buffer (10 $mg.ml^{-1}$), using the magnet to separate the beads from the wash solution each time.

The biotinylated DNA was immobilised to the magnetic beads by adding an equal volume of DNA to the bead resuspension (*i.e.* 100 µl (5 µg)). These were incubated at room temperature for 15 min, whilst being mixed by gentle rotation. The beads were separated, from the solution, by magnetism for 2 min (the supernatant after incubation was kept for analysis by agarose gel electrophoresis to confirm that the DNA had been immobilised onto the beads), and washed 3 times with 1 x B&W buffer (5 mM Tris-HCl (pH 7.5), 0.5 mM EDTA and 1 M NaCl). The DNA-immobilised beads were resuspended in 200 µl of Buffer A (20 mM tris-HCl (pH 8), 1 mM EDTA, 15% glycerol and 0.05% Nonidet P-40). 1 – 5 mg of total protein from a trypanosome nuclear extract were then mixed with 200 µg of pBluescript DNA (100 µg of which was linear), as competitor DNA to remove non-specific DNA binding proteins, in a volume of 1480 µl made up with Buffer A supplemented to 80 mM $(NH_4)S0_4$, for 5 min at 37°C. 20 µl of the DNA immobilised bead solution was then added and incubated for 10 min at 37 °C. The beads were then washed twice in 0.5 ml of Buffer A suplemented to 75 mM NaCl. Proteins bound to the DNA were then eluted in 100

μl Buffer A supplemented to 0.33 M NaCl. The elutant was then size-fractionated on a 12% polyacrylamide gel and visulaised with sypro-ruby. Bands of interest were excised.

2.16 Matrix-assisted laser desorption/ionisation – time of flight (MALDI-TOF) and Electrospray ionisation – mass spectrometry – mass spectrometry (ESI-MS-MS) analysis of protiens

Excised gel bands containing the protein of interest were sent to either Dr. Kathyrn Lilley from the Cambridge Centre for Proteomics (University of Cambridge, Department of Biochemistry Building O, Downing Site, Cambridge, CB2 1QW, England) for ESI-MS-MS sequencing or Dr. Chistopher Ward (Joseph Black Building, University of Glasgow, G12 8QQ) for MALDI-TOF analysis. The MALDI-TOF data was analysed at Protein Prospector (http://jpsl.ludwig.edu.au/ucsfhtml3.4/msfit.htm) using the NCBI (National Centre for Biotechnology Information) database. The ESI-MS-MS analysis yielded short stretches of protein sequence which were used to perfom BLAST (Basic Local Alignment Search Tool) searches on the NCBI (http://www.ncbi.nlm.nih.gov/BLAST) and Sanger Centre websites (http://www.sanger.ac.uk/Projects/T_brucei/Toolkit/ blast_server.shtml).

2.17 Analysing the DNA metabolic activity of trypanosome nuclear extracts upon 70-bp repeat DNA

DNA was prepared from pBluescript and pNP101 (the plasmid pCRScript (Stratagene) with 2 kb of 70-bp repeats, from upstream of the ILTat 1.21 *VSG* gene, cloned into the *Srv*I site) by initially preparing the plasmid from 400 ml of L-Broth. The plasmids were then digested with *Sca*I for 4 h at 37 °C. The linearised plasmid DNA was then phenol/chloroform-extracted, ethanol-precipitated and dissolved in water.
210 ng of each prepared DNA was incubated separately with either bloodstream or procyclic nuclear extract (0.2 $mg.ml^{-1}$) in a total volume of 700 μl (5 mM Tris-HCl (pH 7.5), 5 mM $MgCl_2$, 1 mM DTT, 1mM ATP, 30 mM KOAc and 50 $\mu g.ml^{-1}$ BSA). 100 μl samples were removed, and phenol/chloroform extracted, at the following time intervals: 0.5, 1, 5, 10, 20, and 40 min. The DNA was then ethanol-precipitated and dissolved in 30 μl of water. These samples were then size-fractionated on a 0.8% agarose gel. The gel was

Southern blotted and the blot was hybridised overnight with a radiolabelled, pBluescript derived, DNA fragment overnight. The hybridisation was visualised with a phosphoimager and densiometric analysis was performed using the Imagequant program (Molecular Dynamics).

OLIGO NAME	OLIGO SEQUENCE (5' – 3')
1.22 Probe 5'	GCGTTAGCTGCATACGCATC
1.22 Probe 3'	GGAACCGTCTTGCTAATGGC
*Ssp*I EXP PR 5'	ATGGACAAGCCACAGCAGAC
*Ssp*I EXP PR 3'	ACCATTGTTTCAAGGGTGGA
1.22 *Ssp*I Walk Pr1	GACCATTGTTTCAAGGGTGG
1.22 *Ssp*I Walk Pr2	GGCGCAGATTTTATGGCTAC
TELO specific primer	TAACCCTAACCCTAACCCTAACCC
1.22 TELO 1	TATGGCTTCATTCAGCTCCG
VAT I N Probe 5'	TTAGCCGTAACTCAAGTGCC
VAT I N Probe 3'	CTGCTCAGCATTAGCAGCAT
VAT I 1	GGCACTTGAGTTACGGCTAA
1.71 CTR Probe 3'	ATTTACTGAATGATAATCCG
1.71 CTR Probe 5'	GTACTAAAGTGGCCATATAG
1.71 Probe 1 Pr 1	CATATTTTACATGTTAGTGG
1.71 Probe 1 Pr 2	TGTACGCCTTTTTTCACAGC
1.71 Probe (Map) 5'	AATGGATTACCGCGAGTGGC
1.71 Probe (Map) 3'	GGTCACACAATTCCTCCAGC
1.64 P1 5'	CGTTTCCGAAACGAGCAGGG
1.64 P1 3'	CATATCGTAAGGAAACCCCC
1.64 P2 5'	CTGTAATAGAAGTAAGTGAG

Table 2.2 A list of Oligonucleotides used in this study.

1.64 P2 3'	TTCTTATTCACATGTTGGCG
ILTat 1.2 N-term 3'	GCCTAGCAAGCAGCCTTTG
ALDO 1 *Sac*I	CCCCCGAGCTCGCTTTGGAGCATAGTTTCGC
ALDO 1 *Sac*II	CCCCCCCGCGGCGTGCGCTACACAGCTTGAG
ALDO 2 PsPI	CCCCCCTGCAGCTCAAACAGGCACGGAAGCC
ALDO 2 *Kpn*I	CCCCCGGTACCAAAGCGCTTCGAGCAAGAAC
1.22 70bp repeats P1/2	CCCCCGCGGCCGCTAATAATAATAATAATAGGA
1.22 70bp repeats P2/2	CCCCCGCGGCCGCTTACTACATCATCATCATC
221 CTR 5'	CCCCCCTCGAGAATGCATTGGCACACTTTCCGC
221 CTR 3'	AGGGACACTCGAGTATACAG
1.2 CTR 5' 1/2	CTCTCCTCGAGCTTAAACCGGTAAGTCAGCG
1.2 CTR 3' 1/2	CTCTCCTCGAGGAGGTCCTTTTCGTCTGTCT
221 5'CTR *Not*I 5'	CCCGCGCGGCCGCAGTTAAGACAATGGTATGTG
221 5'CTR *Not*I 3'	GAACCGCGGCCGCACTCTATCTCTCTTCTTTC
ALDO Test S1	TTTTTGGTAGTCGGACTGTG
HYGRO U1	CATCAGCTCATCGAGAGC
HYGRO 5' Switch	CAGCGTCTCCGACCTGATGC
VSG 221 N-Term	TTGGGCTAGGACCAAGACGG
1.22 70bp Repeat S	TAATAATAATAATAATAGGAGAGTGTTGTGAGTGTGT ATATACGAATCTTATAATAAGAGAAGCAGTAATAAT AATAATAATGATAATAATGATGATGATGATGATAG
1.22 70bp Repeat AS	CTATCATCATCATCATCATTATTATCATTATTATTATT ATTACTGCTTCTCTTATTATAAGATTCGTATATACACA CTCACAACACTCACCTATTATTATTATTATTA
Ko 5'	GATGATGCCGCAGTGGATGCGG
Ko 3'	GCACTTCTCCGCTTTCGGAGTG
Rad51d6	TGGTGACGCTGCCGGTGGGC
Rad51u3	TTTCCAAGATGCATCTGCCG

Table 2.2 continued.

CHAPTER 3

THE MECHANISM OF DUPLICATIVE ACTIVATION OF *VSG* GENES IN THE PLEOMORPHIC ILTAT 1.2 TRYPANOSOME LINE

3.1 Introduction

A trypanosome has at its disposal a variety of mechanisms to vary the current *VSG* (variant surface glycoprotein) gene being expressed. Transcriptional (*in situ*) switches involve activation of a different expression site concurrent with silencing of the previously active one, whereas recombination events can involve gene conversions and telomere conversions (grouped as duplicative transpositions), reciprocal recombination or mosaic gene formation (possibly a gene conversion event). All of these routes of activation have been detected in monomorphic lines which, due to continual passaging in the laboratory, no longer have the ability to differentiate, naturally, through different life cycle stages. These strains switch their VSG surface coat at a level equivalent to a background mutation rate (10^{-6} to 10^{-7} switches/ trypanosome/generation), unlike the unadapted pleomorphic lines which can switch antigen type as frequently as 10^{-2} to 10^{-3} switches/trypanosome/generation. This suggests that prolonged adaptation to the laboratory has caused the loss, or reduction, of this enhanced phenotypic variation. That is, perhaps even a specific mechanism dedicated to the variation of *VSG* expression may be absent from, or diminished in, monomorphic lines.

Evidence for the loss of a specific mechanism comes from the difference in the preference of VSG switching mechanisms shown by the monomorphic and pleomorphic lines. *in situ* switching tends be the most frequent mechanism of activation in monomorphic lines (Liu *et al*, 1985), whereas duplicative transposition predominates in the high switching unadapted lines (Robinson *et al*, 1999).

There is also increasing evidence that the actual mechanism of duplicative transposition may differ between these two lines. When monomorphic lines duplicate a *VSG* gene, replacing the one present in the active expression site, there is a tendency to use chance, variable regions of homology between the sequences involved (Michiels *et al*, 1983; Pays *et al*, 1985; Lee *et al*, 1987; Kooter *et al*, 1988). However, all the analysed duplicative activations of *VSG* genes in pleomorphic lines utilise the 70-bp repeat region as the upstream limit of conversion (Shah *et al*, 1987; Matthews *et al*, 1990, Le Count *et al*, 2001). This implies that there is a mechanism dedicated to *VSG* switching, involving this common repetitive region found upstream of almost all *VSG* genes, which has been lost during the laboratory adaptation of the monomorphic lines.

The aim of this work was to elucidate further the mechanism adopted by the pleomorphic line ILTat 1.2 to activate *VSG* genes by duplicative transposition, thereby testing the hypothesis that the 70-bp repeats are directly involved in this recombinational event. This was achieved by determining, in sequence detail, the conversion limits of several independent activations of the ILTat 1.22 *VSG* gene. Attempts to study activations of other *VSG* genes were also made, as well as utilising the genome project for a study of 70-bp repeats upstream of internal *VSG* genes.

3.2 The characterisation of the duplicative transpositions of the ILTat 1.22 *VSG* gene

The aim of this initial experiment was to characterise, in detail, the conversion limits of five independent activations of the ILTat 1.22 *VSG* gene. ILTat 1.22 is a telomeric *VSG* gene which is situated in its metacyclic expression site (MES; Figure 3.1). It is also frequently activated, however, in the long-slender bloodstream stage during infections of mice, rats and rabbits (Miller and Turner, 1981; Matthews *et al*, 1990; Robinson *et al* 1999). The activation of this *VSG*, in those bloodstream stage cases examined, has always been by duplication into the active expression site, as opposed to an *in situ* activation of the metacyclic expression site in which it resides (Matthews *et al*, 1990; Robinson *et al* 1999), as has been seen for some metacyclic *VSG* genes (Alacorn *et al*, 1994).

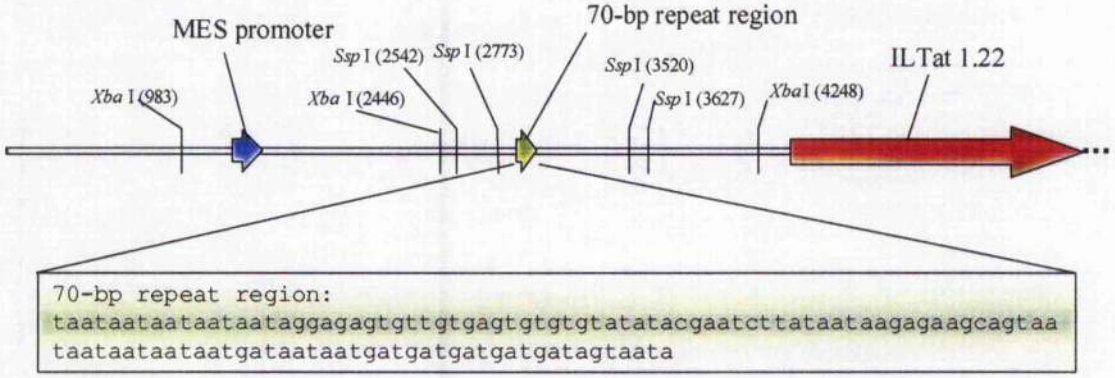


Fig 3.1 A map of the ILTat 1.22 locus. The sequence depicted is the 115bp 70bp-repeat region of this locus. The dotted region indicates the telomere. MES denotes metacyclic expression site.

What makes the study of duplicative activations of the ILTat 1.22 gene interesting is that it has only a very short array of 70-bp repeats upstream (115 bp long) that appears to be used as the upstream limit of conversion (Matthews *et al*, 1990). In the event of the 70-bp array

being used in duplication, it is also possible, due to the absence from the ILTat 1.22 array of the *Ssp*I restriction enzyme recognition site present in nearly all 70-bp repeat units, to clone and sequence the exact point of conversion for each activation under study (see below). Thus, it should be possible to answer the following questions: how short can be the homologous tracts involved in *VSG* switching? When given a limited amount of homology (*i.e.* a repeat array of only 115 bp), do pleomorphic lines have a preference to utilise a particular region or site within the repeat, or a particular repeat from the expression site? Also, being a sub-telomeric gene, do the downstream conversion limits lie within the telomere, or are sequence homologies in the 3' end of the gene used for classical gene conversion reactions?

3.2.1 Derivation of the ILTat 1.22 expressing lines

The activations of ILTat 1.22 under study were originally derived from an ILTat 1.2 expressing line which, *via* partial laboratory adaptation, is within the *T. brucei* EATRO 795 stock (a field isolate from bovine blood; Uhembo, Central Nyanza Province, Kenya; 1964), or the EATRO 795 stock itself. The ILTat 1.2 line was derived by animal passaging an unknown number of times until clones expressing ILTat 1.2 were stabilated at ILRAD (International Laboratory for Research on Animal Diseases, Nairobi, Kenya). This is a pleomorphic line that has a switching rate of approximately 1 x 10^{-5} switches/trypanosome/ generation. However, after fly transmission this rate varies, and in some clones can be as high as 1 x 10^{-2} switches/ trypanosome/generation (Turner and Barry, 1989; Turner 1997). The five activations of ILTat 1.22 have been named, for convenience, activations 1 to 5. Activations 1 to 3 were obtained from the first patent parasitaemia of mouse infections initiated by metacyclic trypanosomes or by bloodstream trypanosome injection (Matthews *et al*, 1990), and correspond to ILTat 1.22 expressors 1.22a, 1.22d and 1.22g respectively (Matthews *et al*, 1990). Activations 4 and 5 were isolated from the second relapse peak of a chronic rabbit infection initiated by ILTat 1.2 bloodstream form trypanosomes (Robinson *et al*,1999)

3.2.2 Analysis of the mechanism of switching of five independent activations of ILTat 1.22 *VSG* gene

To determine whether each of the five activations had occurred by duplication, 2 μg of genomic DNA from *T.brucei* having undergone each of these activations was digested with *Hin*dIII. The digests were size-fractionated on a 0.6% agarose gel and Southern blotted. The blot was then probed with an ILTat 1.22 N-terminus coding region-specific probe, which had been generated by PCR (Primers: 1.22 Probe 5' and 1.22 Probe 3' - conditions: 1 cycle of 96 °C for 5 min, 30 cycles of 96 °C for 1min, 55 °C for 1 min and 72 °C for 1min, and 1 cycle of 72 °C for 10 min) from ILTat 1.2 genomic DNA and purified by gel extraction. In every case, the ILTat 1.22 specific probe hybridised to two bands, indicating that activation is associated with duplication (Figure 3.2), with the ILTat 1.22 gene being single copy in the ILTat 1.2 line (Matthews *et* al, 1990; Robinson *et al*, 1999). One band, constant in size through all lines, corresponds to the restriction fragment associated with the 1.22 basic copy (BC) locus, whereas the second, larger, band varied in size between the activations and corresponded to the duplicated expression linked copy (ELC).

3.2.3 Restriction mapping of the upstream conversion limits of ILTat 1.22 activations 4 and 5

Restriction mapping has previously shown that the upstream conversion limit of activations 1 to 3 occurs within a 420 bp restriction fragment of the 1.22 BC locus, between the *Pvu*II and *Pst*I sites surrounding the 70-bp repeat region (Figure 3.3) (Matthews *et al*, 1990).

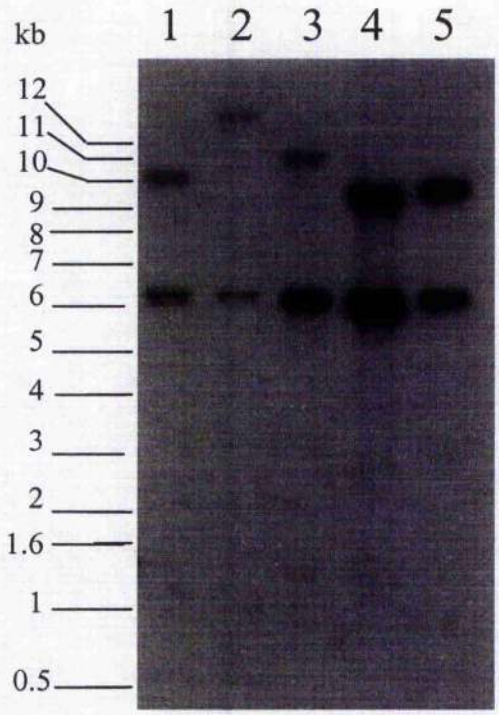
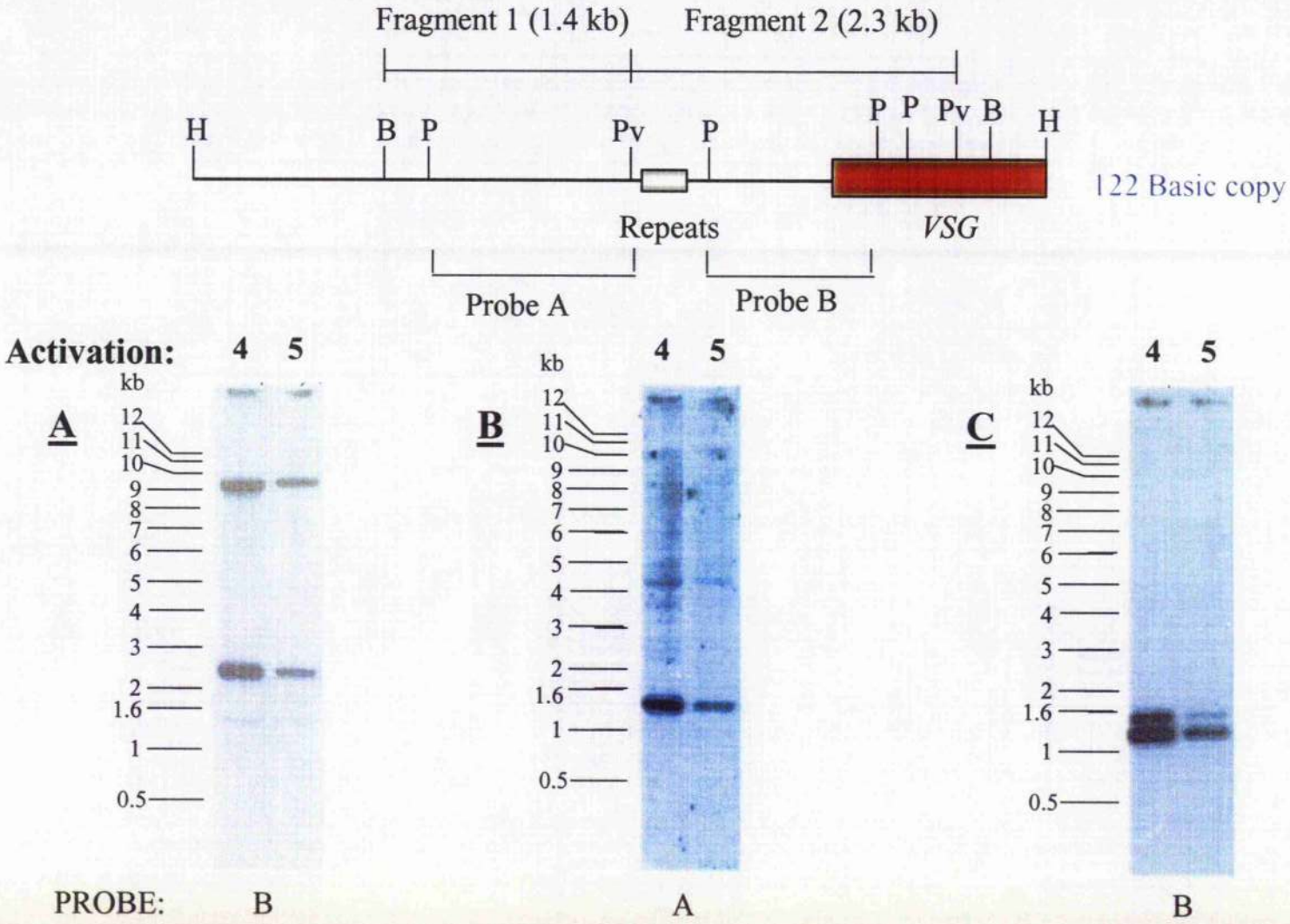


Fig 3.2 –ILTat 1.22 is activated by duplication. Genomic DNA from all five ILTat 1.22 activations was digested with *Hin*dIII. These digests were fractionated on a 0.6 % agarose gel, Southern blotted and hybridised with a 1.22 N-Terminus coding region specific probe. ELCs of the 1.22 gene are present in all the lines under study. Lines 1-5 are respectively activations 1, 2, 3, 4 and 5. The blot was washed to 0.1 X SSC, 65 °C.

To map the ILTat 1.22 activations 4 and 5, the same restriction mapping approach was employed. Genomic DNA from activations 4 and 5 was doubly digested with *Bam*HI and *Pvu*II, size-fractionated on a 0.6% gel and probed with 1.22-probe B (Figure 3.3, panel A). Probe B hybridised to two bands in both activations, the BC (2 kb – fragment 2) and the ELC (9.5 kb), indicating that this area has been duplicated during activation. This blot was then stripped and re-probed with probe A (Figure 3.3, panel B). Probe A hybridises to only one band (1.4 kb - fragment 1) in both samples, indicating that the duplication limit does not lie directly upstream of the 70-bp repeats. Together, these Southern blots imply that the upstream conversion limit of these two activations must lie somewhere within fragment 2 indicated in Figure 3.3. In a *Pst*I digest of these activations, which was size-fractionated on a 0.6% agarose gel, Southern blotted and hybridised with probe B, one dominant band is present with a slightly larger, fainter band (Figure 3.3, panel C).

Fig 3.3 – The 70-bp repeat region is always associated with the upstream conversion limit of 1.22 activations. Panel A is a *Bam*HI and *Pvu*II double digest hybridised with Probe B. Panel B is the same blot as panel A but stripped and then hybridised with Probe A. Panel 3 is a *Pst*I digest hybridised with Probe B. Probes A and B were generated from the plasmid pMG-7.1-1 by digesting with *Pst*I and *Pvu*II and then gel-extracting the appropriate bands. The restriction map depicted is also derived from pMG-7.1-1. All blots were washed to 0.1 X SSC, 65 °C.



We believe that the fainter band represents either a cross reacting band elsewhere in the genome, or more likely, a partial digest created by the presence of the base-J modification within silent MESs. Base-J modifications reduce the number of successful cuts at *Pst*I sites (Borst and van Leeuwen, 1997) allowing hybridisation, in this case, to occur to a fragment generated by a cut at the known *Pst*I site ~100bp downstream of the most frequently cut site. If this is the case, then the dominant single band seen indicates that the map of the ELC and the BC is identical in this region *i.e.* the single band must be a doublet of the ELC and BC region, as it is known that this region is duplicated (Figure 3.2). The conclusion of these hybridisations is that the conversion limits of these two activations, like activations 1 to 3, most likely lies within the 420 bp *Pst*I/*Pvu*II fragment that contains the 70-bp repeat region of the 1.22 BC locus.

3.2.4 Cloning of the upstream conversion limits

To obtain accurate sequence data, it was fortuitous that the ILTat 1.22 BC has an extremely short 70-bp repeat region (115 bp). Particularly useful, however, was the lack, in this region, of the *Ssp*I restriction site that is common to most 70-bp repeat units. If the upstream conversion limit of these activations occurs within this 70-bp repeat region, as appears to be the case (Matthews *et al* 1990; Figure 3.3), it follows that, after the duplications of the ILTat 1.22 gene, the limit of conversion in each case should lie between the most downstream *Ssp*I site donated by the BES and the most upstream *Ssp*I site donated by the duplicated ILTat 1.22 locus. This should enable cloning of these DNA fragments that contain the upstream conversion limits of the ILTat 1.22 activations (Fig 3.4).
To determine whether this would be possible, *Ssp*I digestion was performed on genomic DNA from each ILTat 1.22 activation, and products were then size-fractionated on a 2% agarose gel, Southern blotted and probed with a ILTat 1.22 co-transposed region (CTR) specific probe generated by PCR (Primers: *Ssp*I EXP PR 5 and *Ssp*I EXP PR 3 - Conditions: 1 cycle of 96 °C for 5 min, followed by 30 cycles of 96 °C for 1min, 55 °C for 1 min and 72 °C for 1min, and followed by 72 °C for 10 min) and gel-extracted. In each case, two bands were detected, one of a consistent size that corresponds to the BC and one variable in size that corresponds to the ELC (Figure 3.5). This indicates that a clonable fragment containing the upstream conversion limit is available.

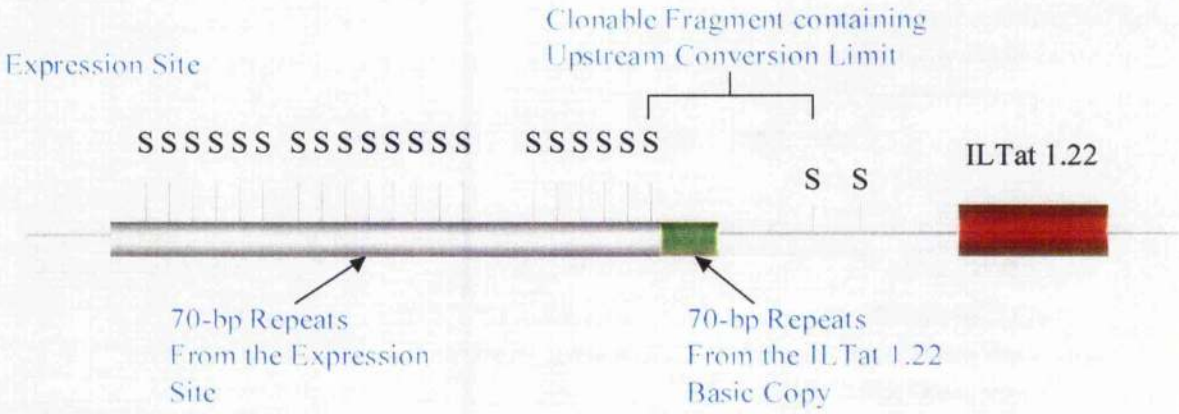


Figure 3.4 A representation of the ILTat 1.22 *VSG* gene after undergoing activation by duplicative transposition into the active expression site using the 70-bp repeat region as the upstream conversion limit. The clonable fragment is indicated. S, *Ssp*I.

Activation 3, however, appeared to have 2 ELCs, which could have occurred if the trypanosome population was not truly clonal and there had been two independent switches to ILTat 1.22 within that population, or alternatively, if a second duplication event of ILTat 1.22 had occurred subsequently within this line.

Initial attempts to clone the upstream conversion limits involved generating a plasmid library. Ten µg of genomic DNA from each of the ILTat 1.22 activations was digested with *Ssp*I, and the digests were ligated to *Eco*RV digested pBluescript and transformed into supercompetent MRF' XL1-Blue *Escherichia coli*. A range of 50 to 600 colonies grew per transformation. Colony lifts onto nitrocellulose membranes were performed and these lifts were probed with the ILTat 1.22 CTR probe. Despite screening 3000 colonies, no positive colonies were detected, so an alternative approach was attempted.

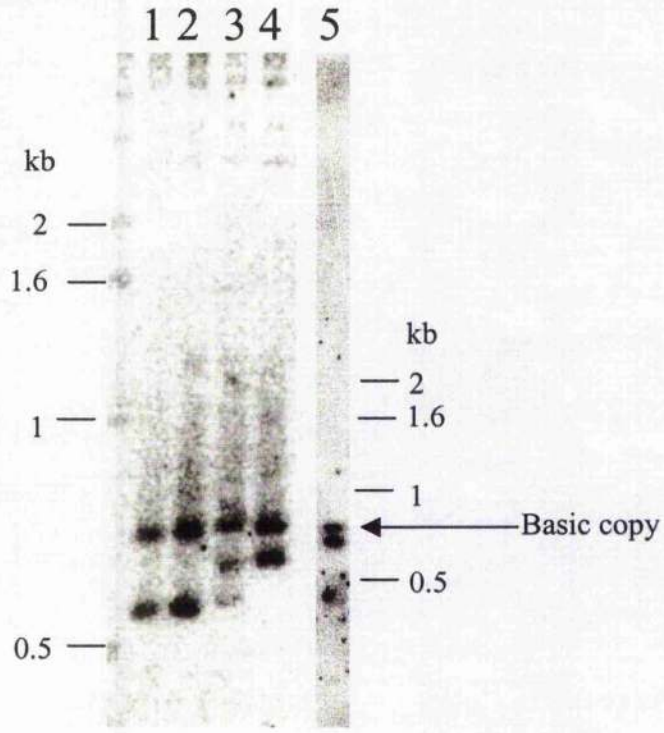


Fig 3.5 – Each ILTat 1.22 activation contains an expression-linked *Ssp*I fragment that is predicted to contain the upstream conversion limit. Genomic DNA from each activation was digested with *Ssp*I and size fractionated on a 2% agarose gel, Southern blotted and probed with a region specific to the ILTat 1.22 co-transposed region. The blot was then washed to a stringency of 0.1 x SSC, 65°C.

The method of choice was a linker-based PCR method. Ten μg of genomic DNA from each activation was digested with *Ssp*I, then, depending on whether the Vectorette (Sigma) (activations 1 and 2) or TOPO-Walking (Invitrogen) (activations 3, 4 and 5) methodology was used, the digested DNA was, respectively, ligated to a blunt-ended linker or primer extended with one of two ILTat 1.22 CTR specific primers (Primers: 1.22 *Ssp*I Walk Pr1 and 1.22 *Ssp*I Walk Pr2) followed by attachment of the TOPO-linker. PCR, with Taq DNA polymerase, was then performed to amplify the region between the linker and the ILTat 1.22 CTR specific primer (Average conditions: 1 cycle of 96 °C for 5 min, followed by 30 cycles of 96 °C for 1min, 50 °C for 1 min and 72 °C for 1min, and followed by 72 °C for 15 min), thus amplifying a region of the ILTat 1.22 BC and the upstream conversion limit associated with the ELC (Fig 3.6). These PCR products match the sizes predicted from the Southern blot of the region (Fig 3.5), also confirming the presence of two ELCs in activation 3. The PCR reactions for each activation were replicated at least once and cloned into TOPO-Vector. Multiple clones from each duplicate PCR of the BC and ELC of each switch were sequenced. The PCR and sequencing of multiple clones was repeated to

remove the possibility of sequencing errors. Also, for those activations that appeared identical (see below), another PCR was conducted with fresh materials to avoid the risk of contamination. It should be noted that the size of the cloned product was identical to the PCR product, so there was no detectable loss of sequence due to rearrangements within the *E. coli* host.

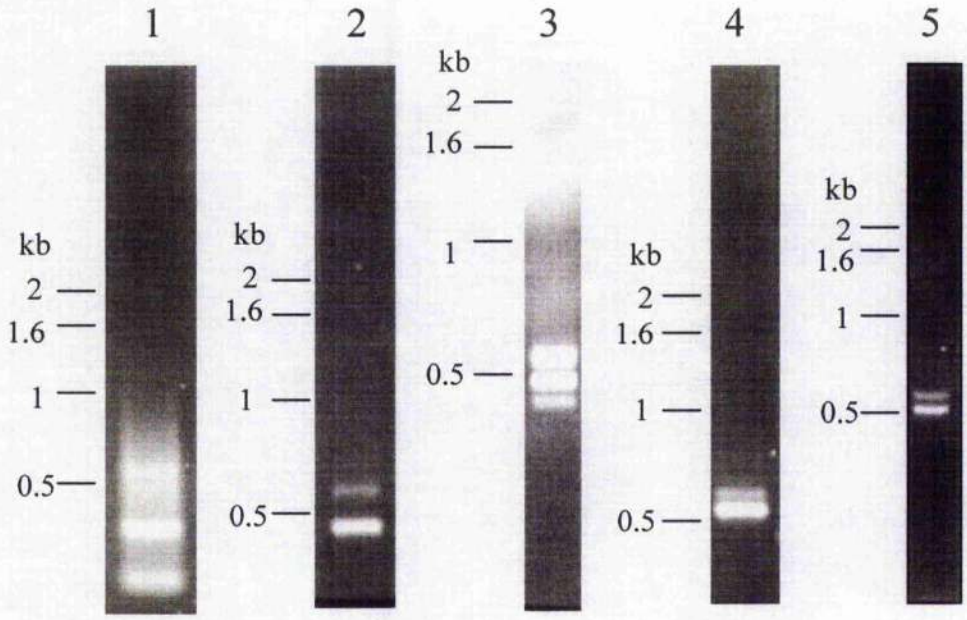
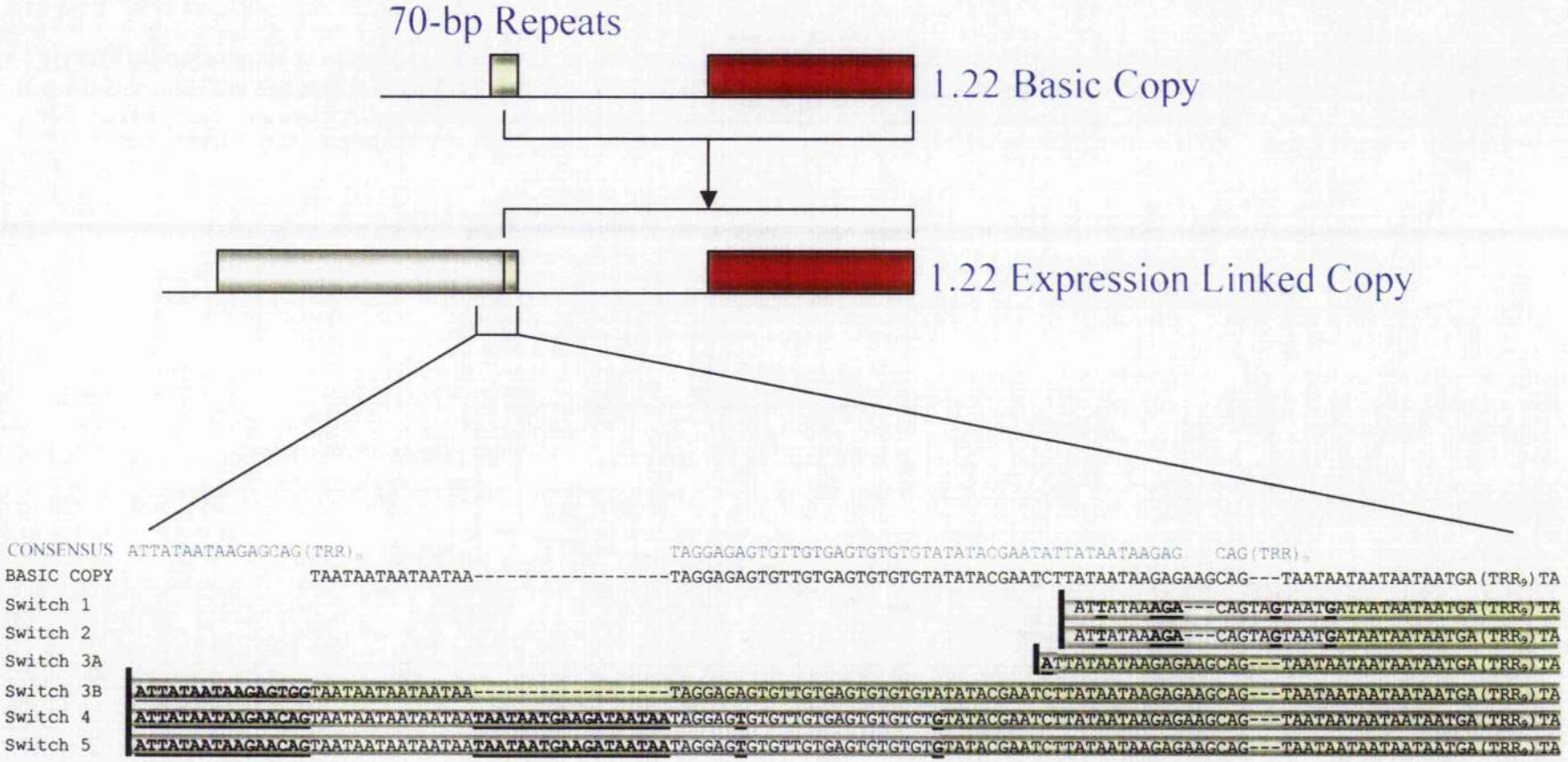


Fig 3.6 – The results of the linker based PCR of the basic copy (BC) region containing the 70-bp repeats of ILTat 1.22, and the conversion limit within the expression-linked copy. PCRs were conducted according to the protocol of Vectorette for activations 3, 4, and 5, whereas for activations 1 and 2 TOPO-Walker methodology was used. The largest bands are the BCs and the smaller bands are the expression-linked copies (see text for explanation). Activations 2 – 5 used the same 1.22 cotransposed region primer (1.22 *Ssp*I Walk Pr1) whereas activation 1 used a primer further upstream (1.22 *Ssp*I Walk Pr2), hence the smaller PCR products.

3.2.5 Sequence analysis of the upstream conversion limits

An alignment of the upstream conversion limits of the ILTat 1.22 gene activations is shown in Figure 3.7. It appears that, despite having a very short region of potential homology with which to recombine (115 bp), the reactions have utilised different regions and different lengths of the ILTat 1.22 *VSG* gene-associated 70-bp repeat region. The different lengths of duplicated ILTat 1.22 70-bp repeat DNA are 42, 64, 75 and 115 bp.

Fig 3.7 – Sequence data of the upstream conversion limit of six independent switching events to the 1.22 *VSG* gene. Bold and underlined letters indicate differences between the basic copy (BC) repeats and the expression-linked copy (ELC) repeats. Those repeats depicted in green are believed to be derived from the ILTat 1.22 basic copy locus, whereas those in grey are novel and are believed to come from the expression site. The bold vertical line indicates the *Ssp*I site at which the PCR product ends. Note that the (TRR)n tract indicated and all sequence further downstream is identical in all BCs and ELCs. For further explanation see text. The consensus sequence is derived from Aline *et al,* 1985.



Different 70-bp repeat units from the ES have also been used indicating, as suspected, that the process does not involve a site-specific recombination event. Instead, it probably relies upon homologous recombination, as previously suggested by the reduction of *VSG* switching frequencies in *rad51* -/- null mutants (McCulloch and Barry, 1999). That homologous recombination is the likely mechanism is also indicated by the fact that the integrity of the repeat unit is preserved i.e. in all switching events the previously defined arbitrary regions (Aline *et al*, 1985) of the repeat unit are conserved over the crossover points.

Activations 1 and 2 reveal that the maximum region of 100% identity (if only 70-bp repeat sequence was used) between the BES and the BC was 42 bp, suggesting that small regions of homology are acceptable for *VSG* switching. As these two activations were identical, yet had occurred independently, it also appears that the use of such short homology is not an unusual occurrence.

Activations 4 and 5 are identical and are derived from the same relapse peak in a single rabbit infection, but were isolated on different days. They may, however, represent the presence in the rabbit, on different days, of trypanosomes that were the progeny of the same activation event.

3.2.6 Characterisation of the ILTat 1.22 subtelomeric region

In order to map the downstream conversion limits of these ILTat 1.22 activations, sequence needed to be obtained of the ILTat 1.22 BC subtelomeric region.

A PCR-based method was used involving a telomeric-specific primer based upon the $(TTAGGG)_n$ repeat motif of telomeres (TELO specific primer) and a ILTat1.22 N-terminus coding region specific primer (1.22 TELO 1) to amplify the subtelomeric region of the ILTat 1.22 BC locus (conditions: 1 cycle of 96 °C for 5 min, followed by 30 cycles of 96 °C for 1min, 45 °C for 1 min and 72 °C for 10min, and followed by 72 °C for 20 min) (Figure 3.8). Using genomic DNA from ILTat 1.2 *T. brucei*, a 2.3 kb DNA fragment was amplified. This fragment was cloned into the TOPO vector, sequenced and confirmed as being derived from the ILTat 1.22 *VSG* gene and extending to the telomere repeat sequence. The intervening sequence demonstrates that the previously described common subtelomeric regions are present in this locus (Aline and Stuart, 1989) (Fig 3.9).

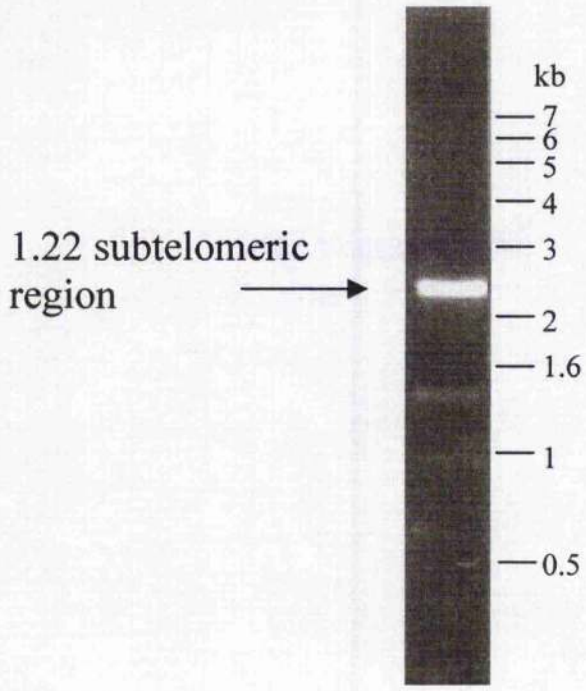


Fig 3.8 PCR amplification of the subtelomeric region of the ILTat 1.22 basic copy. The primers used were specific to the ILTat 1.22 N-terminus coding region and to the *T. brucei* telomere repeat tract. The dominant 2.3 kb band is the ILTat 1.22 subtelomeric region, with the faint, smaller products probably being generated by the telomeric repeat specific primer annealing to telomere-like sequence upstream of the actual telomere tract.

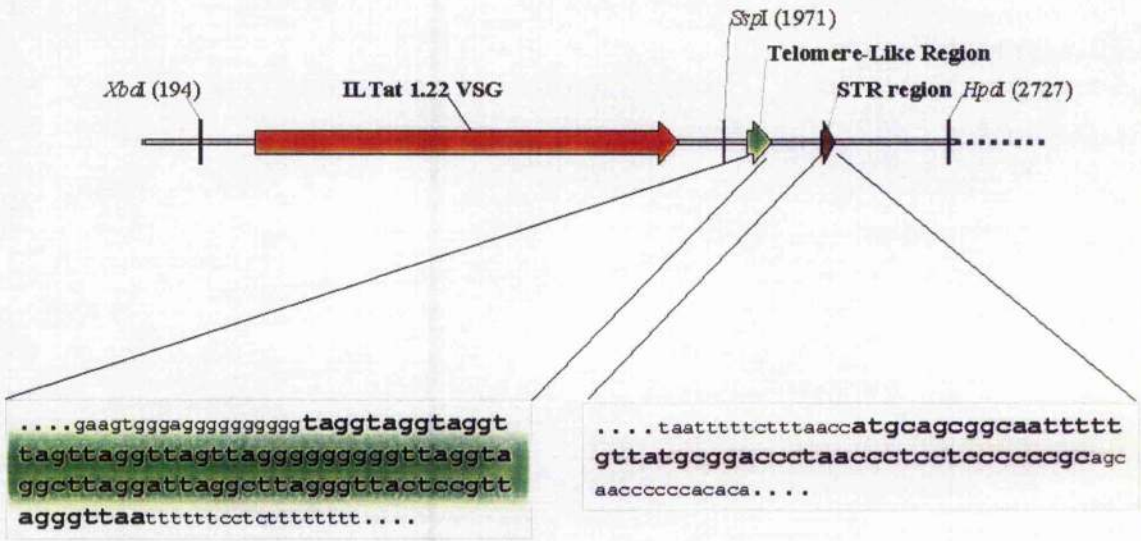


Fig 3.9 –A map of the ILTat 1.22 subtelomeric region. The sequences depicted in bold match those previously defined as subtelomeric regions by Aline and Stuart (1989). The dotted line indicates the telomere tract. STR - subtelomeric repeats.

3.2.7 Restriction mapping of the downstream conversion limits of the ILTat 1.22 activations

DNA from each activation was doubly-digested with *Xba*I and *Hpa*I. These enzymes cut, respectively, downstream of the 70-bp repeat region and ~10 bp upstream of the start of the telomere tract of the ILTat1.22 locus (Figure 3.9). The DNA digests were size-fractionated on a 0.6% agarose gel, Southern blotted and hybridised with an ILTat 1.22 N-terminus coding region-specific probe. Only one band was seen on autoradiogram exposure indicating that the restriction maps of the ILTat1.22 BC and ELC are identical all the way up to the telomere tract (Figure 3.10). This indicates that the *VSG* gene conversion events must either use the telomere tract itself as homology or alternatively, and more likely, initiate with a double-strand break (DSB) in the 70-bp repeat region of the bloodstream expression site and copy the ILTat 1.22 BC all the way to the end of the telomere. This may be analogous to the break-induced replication model of conversion (Paques and Haber, 1999). Of course, it cannot be ruled out that the activations utilise the very few base pairs between the *Hpa*I site and the telomere tract.

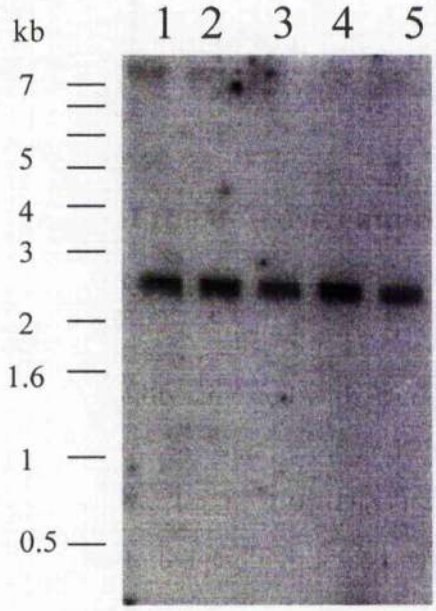


Fig – 3.10 The downstream conversion limits of all five 1.22 activations lie beyond the subtelomeric region. DNA from all the activations was digested with *Xba*I and *Hpa*I, size fractionated on a 0.7% agarose gel and hybridised with the ILTat 1.22 N-terminus coding region specific probe. The blot was washed to the stringency 0.1 X SSC, 65°C.

3.3 Characterisation of the locus of the chromosome internal *VSG* ILTat 1.71

Within the two relapse peaks of several rabbit infections analysed by Robinson *et al* (1999), only one internal *VSG* appeared to have been activated, and that occurred during the second relapse peak. In order to characterise the upstream conversion limit of this activation, sequence surrounding the *VSG* gene had to be first obtained.

3.3.1 Attempts at obtaining sequence surrounding ILTat 1.71 via a genomic lambda library

A 14 kb lambda clone (λ 3/3) was isolated from a lambda library (*Sau*3A partial digests in λ-GEM-12) that hybridised through 4 rounds of screening with a probe specific to the entire length of ILTat 1.71 cDNA (The library construction and initial screening was performed by Nils Burman). DNA was prepared from the λ 3/3 clone and 1 μg was digested singly and doubly with three restriction enzymes recognising and cutting 6-bp sites. These DNAs were size-fractionated on a 0.6% agarose gel, Southern blotted and probed with a sequence specific for the N-terminus coding region of the ILTat 1.71 gene, generated by PCR (primers: VAT I N Probe 5' and VAT I N Probe 3' - Conditions: 1 cycle of 96 °C for 5 min, followed by 30 cycles of 96 °C for 1min, 58 °C for 1 min and 72 °C for 1min, and followed by 72°C for 10 min)) followed by gel extraction. Bands that appeared specific to the N-terminus coding region of ILTat 1.71 appeared to be present (Figure 3.11 A). To authenticate λ 3/3 as containing ILTat 1.71, 2 μg of ILTat 1.2 genomic DNA was digested, singly and doubly with three restriction enzymes recognising and cutting 6-bp sites, size-fractionated on a 0.6% agarose gel, Southern blotted and probed with the same ILTat 1.71 N-terminus specific probe as above (Figure 3.11 B). If the ILTat 1.71 gene is present in the λ 3/3 clone, then there should be correspondence, in terms of the size of the bands hybridising to the ILTat 1.71 N-terminus coding region specific probe, between the Southern blots of the digests of λ 3/3 and the ILTat 1.2 genomic DNA. Only one ILTat 1.71 N-terminus coding region hybridising band, the 5.5 kb band seen in the *Eco*RI digests,

is of the same size when the six λ 3/3 clone digests are compared with the same digests of the ILTat 1.2 genomic DNA.

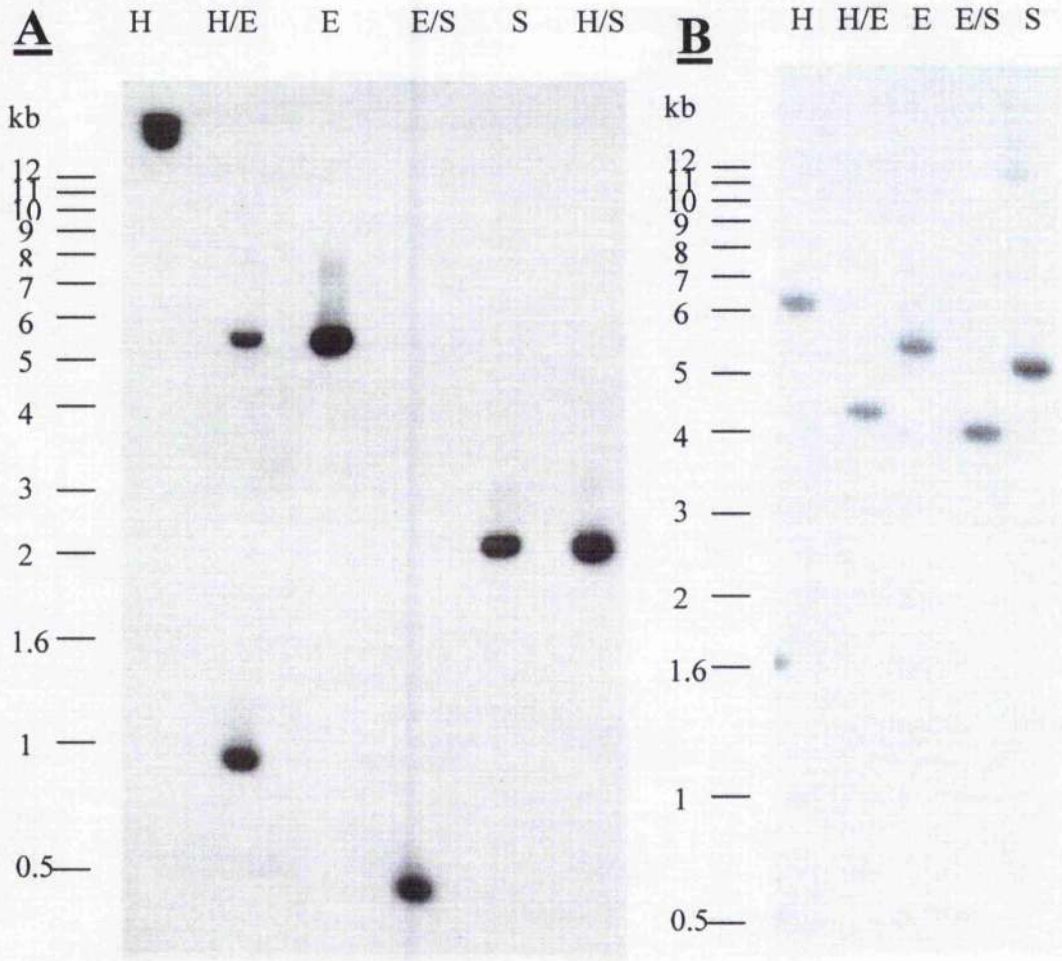


Fig 3.11 The restriction maps of λ3/3 and ILTat 1.2 genomic DNA do not match exactly with ILTat 1.71 sequence. **A.** Lambda clone 3/3 DNA was digested singly and doubly with a series of restriction enzymes, size fractionated on a 0.6% agarose gel and Southern blotted. This was then probed with an ILTat 1.71 N-terminus coding region specific probe. **B.** ILTat 1.2 genomic DNA was treated in exactly the same manner as the Lambda DNA. H, *Hind* III; E, *Eco*RI,; S, *Sac*I. These blots were washed to the stringency 0.1 X SSC, 65°C

This perhaps indicated that the λ 3/3 clone does not contain the ILTat 1.71 *VSG* gene but instead a similar *VSG* gene. However, if the gene is close to the lambda arms then the bands hybridising to the ILTat 1.71 N-terminus probe in the restriction digests would be altered by the presence of restriction sites within these arms. To check this possibility, sequence was required from the λ 3/3 clone.

Fresh DNA was prepared from the λ 3/3 clone and 3 μg of DNA was digested with *Eco*RI. This, when size-fractionated on a 0.8% agarose gel, yielded five DNA fragments of 5.5, 3.9, 2.0, 1.7, and 0.7 kb, and fragments >12 kb that correspond to the lambda phage arms. The pUC18 vector was also digested with *Eco*RI and then treated with calf-intestinal phosphatase for 30 min at 37 °C. Both the digested lambda and vector DNA were then phenol : chloroform extracted. The lambda DNA digested fragments were then shotgun cloned, as a pool of fragments, into pUC18 by ligating for approximately 20 h at 16°C. Clones of all five of the *Eco*RI digested lambda DNA fragments were obtained. Two independent clones of the 5.5 kb DNA fragment were prepared and sequenced, as this was the band that hybridised to the ILTat 1.71 N-terminus coding region specific probe. Within this 5.5 kb, at one end, was the open reading frame of the N-terminus coding region of a *VSG* gene that did not, however, match exactly the sequence of the ILTat 1.71 cDNA. The DNA was 82% identical over a 285 bp stretch (Figure 3.12) and the predicted protein sequences were approximately 76% identical. This explains the conflicting data of an ILTat 1.71 N-terminus coding region specific probe hybridising to the λ 3/3 clone DNA digests yet having a restriction map different from that of the genomic DNA, that is, λ 3/3 contains a *VSG* gene similar enough to ILTat 1.71 to allow hybridisation of a probe specific to the N-terminus coding region.

```
ILTat 1.71      ATGTCGAAGGCAGCGCTTTTAACGGCAGTGTTTGCCTTAGCCGTAACTCAAGTGCCATTA 60
λ 3/3           ATGTCAAAGGCAGCAATTTTAACGGCAGCGTTTGCCTTATCCGTAACTCAAGTGCCATTA 60
                ***** ********  ************ ********** ********************

ILTat 1.71      CCAAAAGTCAAAGCAAACCCTGTTGGATTAAAGGCAGCTGGGGCAAACAGACTATGCGAG 120
λ 3/3           CCAAAAGTAACAGAAAACCCTGTTGGATTAAAGGCAGCTGGAGCAAACAGATTGTGCGAG 120
                ******** * ** ***************************** ********* * ******

ILTat 1.71      ATTTCGAAAAAAATGAAAGATTACACAGCGTACGTCAAAGACAAGACAAGCAACGCTCTG 180
λ 3/3           ATTTTGAAAAAAATGAAAGATTACTTAGCATACGTCAAAACCAAAACAAGCAGCGCTCTG 180
                **** ******************  *** *********  *** ******* *******

ILTat 1.71      GCGGCGGTAGCAAAAGCAATGCAAACCAGGGACAAAATGCTGCTAATGCTGAGCAGAGAG 240
λ 3/3           ACGACACTAGCAAAAGCAATGCAAACTCGGAACAAGATACAACTATTGCTGAGCACGCAG 240
                 ** *  ******************  ** **** ** *  *** *********   **

ILTat 1.71      AAGAAGCAACCCAATGTAAAGACAAGCGTC-ATATTTGCCCTAAC 284
λ 3/3           CGGAAACAACCGACTCCAAAGTCAAGCGTCCATATTTGCCCTAAC 285
                  *** ***** * *  **** ******** ***************
```

Figure 3.12 An alignment of the N-terminus of the ILTat 1.71 *VSG* gene and the putative *VSG* gene sequence from the 5.5 kb *Eco*RI fragment from the λ 3/3 clone. An asterisk indicates identity between the two sequences.

As the probe originally used to screen for the ILTat 1.71 VSG gene in the lambda genomic library was the entire cDNA sequence of the gene, it is possible that cross-reactions were

likely to be isolated, as the C-terminal coding sequences of *VSG* genes can be similar. To test the prevalence of cross reacting sequences in the genome, ILTat 1.2 genomic DNA was digested, singly and doubly, with a series of restriction enymes with 6-bp recognition sequences, size-fractionated on a 0.7% agarose gel, Southern blotted and probed with a sequence specific to the whole of the ILTat 1.71 cDNA (obtained from Nicholas Robinson). Use of the whole cDNA as a probe indicated there is a large number of cross reacting sequences present in the genome, most likely being other *VSG* gene sequences including the gene present in the λ 3/3 clone (Figure 3.13).

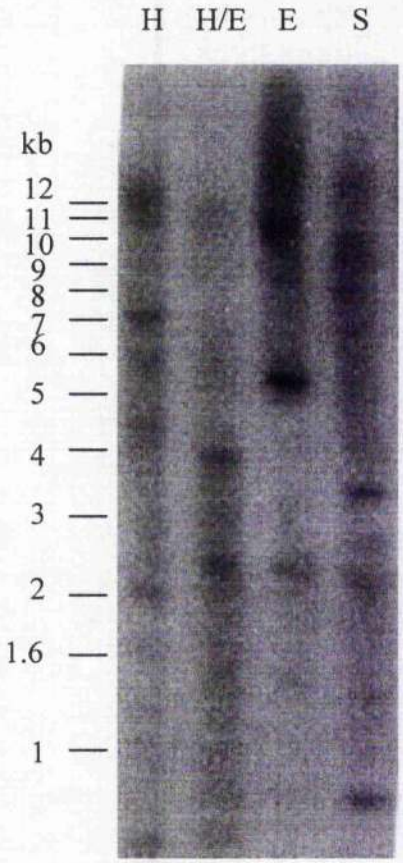


Fig 3.13 ILTat 1.71 cDNA cross-reacts throughout the genome. Genomic DNA from ILTat 1.2 expressing trypansomes was digested singly and doubly with a series of restriction enzymes, and products were then size fractionated on a 0.6% agarose gel, Southern blotted and probed with the whole cDNA sequence of ILTat 1.71. The blot was washed to the stringency of 0.1 X SSC, 65°C.

In a final attempt to obtain the sequence of the ILTat 1.71 locus, the lambda-genomic library was re-screened with a probe specific to the N-terminus coding region of ILTat 1.71, to reduce the level of cross reaction. The library was plated out, grown at 37°C to near confluence, and then blotted to a nylon membrane and hybridised with an ILTat 1.71 N-terminus coding region specific probe. This yielded only what appeared to be very

weakly hybridising plaques, which were probably due to cross-reactions. When two of these lambda clones were taken to a further secondary screen, a low level of hybridisation was again seen, indicating that these clones were merely displaying cross-reactions.
It appears that, despite there being *VSG* genes within this library, including at least one that is very similar in sequence to ILTat 1.71, the ILTat 1.71 gene itself is not present, so this approach is inappropriate for obtaining flanking sequences.

3.3.2 Genomic walking upstream of the ILTat 1.71 *VSG* gene

As the ILTat 1.71 locus was not present within the lambda library, a genomic walking approach similar to that used to obtain the unknown sequence of the ILTat 1.22 activations (see above) was employed. Genomic DNA from an ILTat 1.2 expressing line was digested with *Ssp*I, single primer extension was performed with a primer specific to the ILTat 1.71 N-terminus, and then the DNA was joined to a linker of known sequence according to the TOPO-Walker protocol. *Ssp*I was chosen as it usually cuts in every 70-bp repeat sequence, making it possible that all the sequence up to the 70-bp repeats could be generated. This would allow a better approach for a second round of walking over a potential array of 70-bp repeats (*i.e.* smaller distance for PCR amplification). The linked DNA was then amplified with one primer specific to the ILTat 1.71 N-terminus coding region (VAT I 1) and another specific to the unique linker sequence using the Herculase DNA polymerase (Conditions - 1 cycle of 96 °C for 5 min, followed by 30 cycles of 96 °C for 1min, 40 °C for 1 min and 72 °C for 10 min, and followed by 72 °C for 20 min). A 800 bp product was generated which was cloned into the TOPO-vector and sequenced. When this sequence was aligned with sequence from the cDNA of the ILTat 1.71 gene, a 100 % identical overlap was seen and thus 700 bp of novel ILTat 1.71 CTR (co-transposed region) sequence was generated. To walk further, a different primer was used (1.71 CTR PROBE 3') and the same procedure was carried out, except this time the genomic DNA was digested with *Mbo*I, an enzyme with a four base-pair recognition sequence, to reduce the size of DNA fragment to amplify (conditions - 1 cycle of 96 °C for 5 min, followed by 30 cycles of 96 °C for 1min, 40 °C for 1 min and 72 °C for 10 min, and followed by 72 °C for 20 min). A 500 bp sized PCR product was generated which was cloned into the TOPO-vector and sequenced. When this sequence was aligned with the sequence generated previously, a 100% identical overlap was seen and thus approximately 1 kb, in total, of novel sequence

had been generated of the ILTat 1.71 CTR. A map of this region can be seen in Figure 3.15.

To confirm that this newly identified sequence was genuinely linked to the ILTat 1.71 *VSG* gene, a series of digests were performed on ILTat 1.2 genomic DNA. These were then size-fractionated on a 0.6% agarose gel, Southern blotted and then probed with a region PCR amplified (primers: 1.71 CTR Probe 5' and 1.71 CTR Probe 3' - conditions: 1 cycle of 96 °C for 5 min, followed by 30 cycles of 96 °C for 1min, 60 °C for 1 min and 72 °C for 1min, and followed by 72 °C for 10 min) from the plasmid containing the newly generated ILTat 1.71 CTR sequence. The restriction maps using the 1.71 CTR specific probe (Figure 3.14) are the same as those using the ILTat 1.71 N-terminus specific probe (Figure 3.11 B), indicating that the regions are linked. There were a number of bands cross-reacting with the ILTat 1.71 CTR probe, again indicating there are very similar sequences present in the genome.

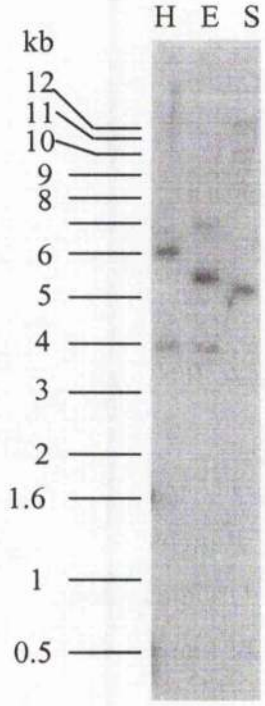


Figure 3.14 Restriction enzyme mapping of the newly generated ILTat 1.71 co-transposed sequence (CTR) demonstrates linkage with the ILTat 1.71 *VSG* gene (see Figure 3.11 B). ILTat 1.2 genomic DNA was digested with a series of restriction enzymes, size-fractionated on a 0.6% Agarose gel and Southern blotted. This was then probed with a DNA fragment amplified from the newly generated ILTat 1.71 CTR sequence (see text). H, *Hin*d III; E, *Eco*RI,; S, *Sac*I. This blot was washed to the stringency 0.1 X SSC, 65°C

3.3.3 Attempted mapping of the ILTat 1.71 conversion limit

Within the newly derived sequence of the ILTat 1.71 CTR there is a potential, very degenerate and very short, 70-bp repeat region (Figure 3.15). The area was identified due to a high frequency of surrounding *Ssp*I restriction sites. This seems to be a signature for very short 70-bp repeat sequences throughout the genome (see Section 3.5). To determine whether activation of this *VSG* gene utilised this region, genomic DNA from ILTat 1.2 and ILTat 1.71 expressing clones was digested with the enzyme *Xmn*I. Products were run on a 0.6 % agarose gel, Southern blotted and probed with a region amplified (primers: 1.71 Probe 1 Pr 1 and 1.71 Probe 1 Pr 2 - Conditions: 1 cycle of 96°C for 5 min, followed by 30 cycles of 96°C for 1min, 45°C for 1 min and 72°C for 30 seconds, and followed by 72°C for 10 min) from immediately upstream of the degenerate 70-bp repeat and then stripped and re-probed with a region amplified (primers: 1.71 Probe (Map) 5' and 1.71 Probe (Map) 3' Conditions: 1 cycle of 96°C for 5 min, followed by 30 cycles of 96°C for 1min, 45°C for 1 min and 72°C for 30 seconds, and followed by 72°C for 10 min) from immediately downstream of the degenerate 70-bp repeat region (Figure 3.16). For each probe fragment, an extra fragment was apparent in the ILTat 1.71 derived genomic DNA compared with the ILTat 1.2 genomic DNA. Most likely, the extra fragment corresponds to the ELC suggesting that the *VSG* gene conversion had extended across both regions probed, and thus activation did not use this highly degenerate 70-bp repeat. However, it is not possible to say whether further 70-bp repeats lie upstream which are used in conversion. A number of attempts were made to isolate further upstream sequence by PCR-based genomic walking (see section 3.3.2), but these were all unsuccessful; perhaps indicating presence of an array of repeats that are difficult to amplify.

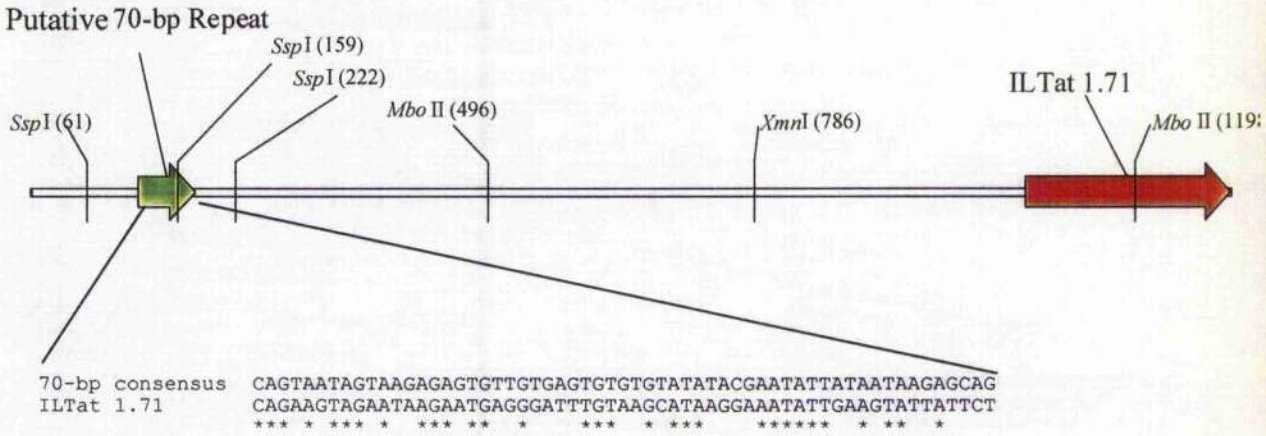


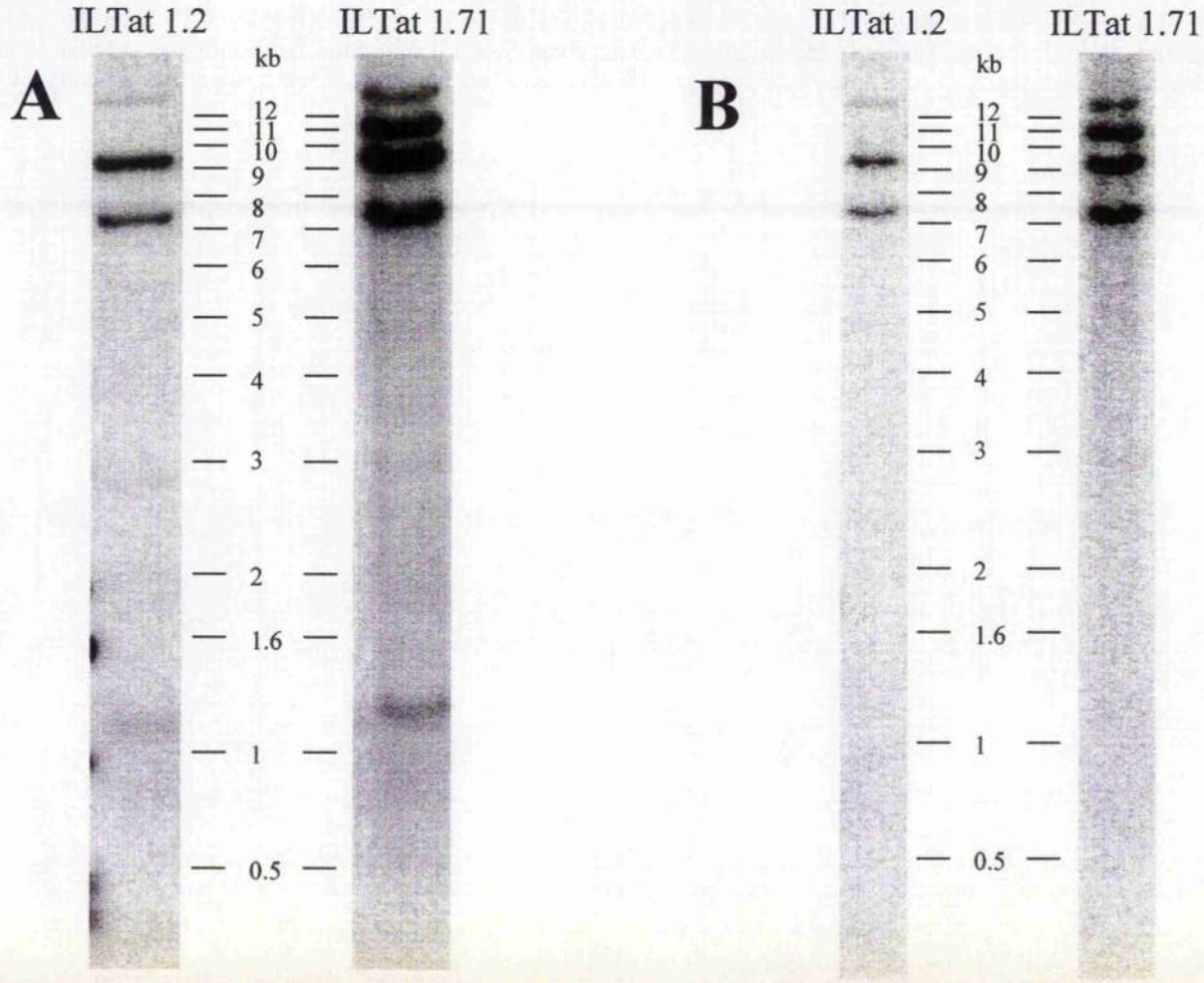
Figure 3.15 Part of the ILTat 1.71 co-transposed region. Below the restriction map is an alignment of the potentially degenerate 70-bp repeat region with part of the 70-bp repeat consensus sequence derived from Aline *et al* (1985). Only the N-terminus coding region of the ILTat 1.71 gene is depicted here.

3.4 Restriction mapping of the upstream conversion limit of a duplicative activation of the metacyclic ILTat 1.64 gene

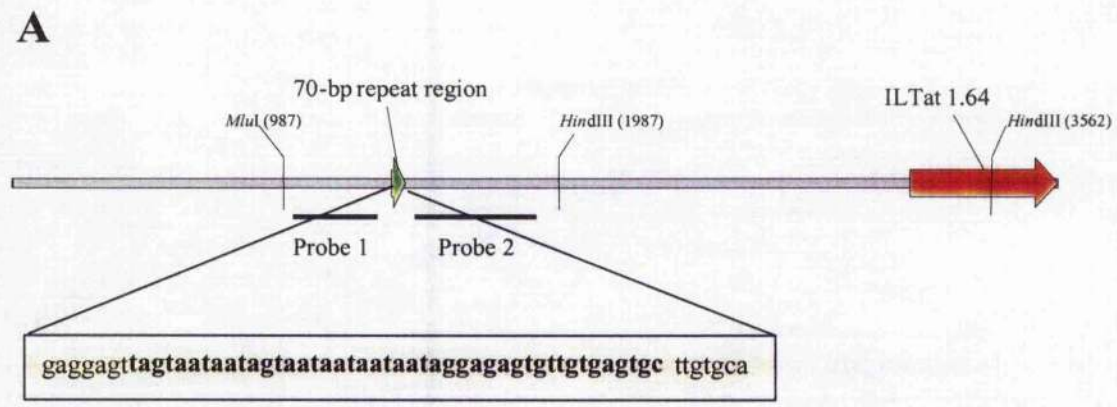
Previously, approximately 4 kb of sequence was obtained from upstream of the ILTat 1.64 *VSG* gene by Allyson Lewis and Michael Ginger (personal communication). Within this region, a very short, incomplete single 70-bp repeat is present 1.8 kb upstream of the *VSG* gene (Figure 3.17 A). Also, an activation of ILTat 1.64, by duplicative transposition, had been previously isolated from the first relapse peak of a rabbit infection (Robinson *et al*, 1999).

To map the upstream conversion limit of this activation, genomic DNA from ILTat 1.2 and ILTat 1.64 expressing trypanosomes was doubly digested with *Hin*dIII and *Mlu*I. These digests were split into equal amounts, size fractionated on a 0.6% agarose gel and then Southern blotted. The DNAs from both ILTat 1.2 and ILTat 1.64 expressing trypanosomes were hybridised with probe 1, a region PCR amplified from genomic DNA that is directly upstream of the 70-bp repeat (Primers: 1.64 P1 5' and 1.64 P1 3' - conditions: 1 cycle of 96 °C for 5 min, 30 cycles of 96 °C for 1min, 45 °C for 1 min and 72 °C for 30 seconds, and 1 cycle of 72 °C for 10 min), and probe 2, a region PCR amplified from genomic DNA

Fig 3.16 The duplicative activation of ILTat 1.71 does not utilises a sequence homologous to an incomplete 70-bp repeat as the upstream limit of conversion. The genomic DNAs, indicated above each blot, were digested with *Xmn*I and size fractionated on a 0.6% agarose gel and southern blotted. **A.** The Southern blots were probed with Probe 1 (a region upstream of the 70bp repeats) **B.** The Southern blots were probed with Probe 2 (a region downstream of the 70-bp repeats). All blots were washed to a stringency of 0.1 x SSC, 65°C.



that is directly downstream of the 70-bp repeat (Primers: 1.64 P2 5'and 1.64 P2 3' - conditions: 1 cycle of 96°C for 5 min, 30 cycles of 96°C for 1min, 45°C for 1 min and 72°C for 30 seconds, and I cycle of 72°C for 10 min (Figure 3.17 A). In both the ILTat 1.2 and ILTat 1.64 expressing trypanosomes, probe 1 hybridised to only one band (1 kb), indicating that either this region is single-copy or the restriction maps for the ELC and the BC are identical due to all of this region having been copied (Figure 3.17 B). Probe 2, however, hybridised to two bands in the ILTat 1.2 expressing line and three bands in the ILTat 1.64 expressing line (Figure 3.17 C). The 1 kb band is present in both lines and is of the expected size for the region from the BC, the same as that seen in Figure 3.17 B for probe 1. The second fragment seen in both lines is approximately 4.9 kb and is presumably a previous duplication of the probed region, perhaps to another expression site, or a very similar region elsewhere in the genome. This is likely to be the case as the *VSG* gene sequence itself is single copy in the ILTat 1.2 line (Robinson *et al*, 1999; Michael Ginger personal communication). The third fragment, only present in the ILTat 1.64 expressing line and approximately 8.5 kb, appears to be linked to its activation, meaning that there is a difference in the restriction maps between the BC and the ELC for this area. The faint band seen at 3.1 kb in both lines is presumably some form of cross- reaction with a similar sequence elsewhere. As seen before for the ILTat 1.71 gene, sequences similar to both the *VSG* gene and the surrounding regions are present elsewhere in the genome. As the same restriction enzymes were used for both probes and the region which hybridised to probe 1 is single copy, it can be concluded that the upstream conversion limit of the ILTat 1.64 duplicative transposition is downstream of probe 1 and upstream of probe 2. This region between the probes contains the incomplete single 70-bp repeat, which therefore could be a sufficient sequence for relatively frequent activation. However, it cannot be discounted from these data that other sequences present in this region could have been used as the limit of conversion.



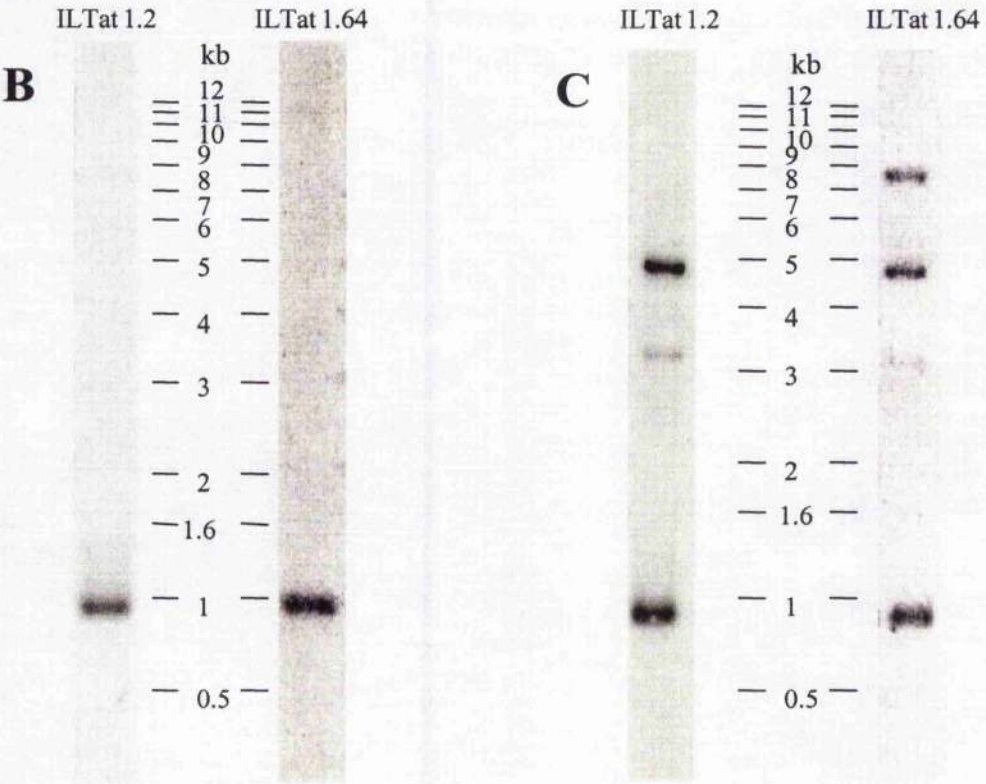


Fig 3.17 The duplicative activation of ILTat 1.64 probably utilises a very short, incomplete 70-bp repeat as the upstream limit of conversion. **A.** This is a representation of the restriction map of the ILTat 1.64 basic copy gene, and the probes used to map the activation of the gene. The bold sequence depicted below is the 70-bp repeat sequence whereas the unbold text is the flanking sequence. **B.** The genomic DNAs indicated above each blot were digested doubly with *Hin*dIII and *Mlu*I and size fractionated on a 0.6% agarose gel, these were then Southern blotted. These Southern blots were probed with Probe 1 (a region upstream of the 70-bp repeats) **C.** The genomic DNAs indicated above each blot were digested doubly with *Hin*dIII and *Mlu*I and size fractionated on a 0.6% agarose gel, these were then Southern blotted. These Southern blots were probed with Probe 2 (a region downstream of the 70-bp repeats) All blots were washed to a stringency of 0.1 x SSC, 65°C.

3.5 A Genomic analysis of the 70-bp repeat regions upstream of internal *VSG* genes from chromosomes *IX* and *X*

Sequence from 14 contigs generated by the ongoing *T. brucei* genome sequencing project and derived from chromosomes *IX* and *X* were obtained by searching the Sanger Institute's GeneDB web site (http://www.genedb.org/genedb/tryp/index.jsp) for contigs containing sequences with *VSG* gene homology. Only thirteen contigs are shown as one appeared to contain a metacyclic expression site (Table 3.1)). These were then annotated by eye and using a BLAST (Basic Local Alignment Search Tool) (http://www.ncbi.nlm.nih.gov/BLAST) for *VSG* gene and 70-bp repeat sequence. In total, 54 regions homologous to *VSG* genes were identified from 234,784 bp of sequence. Of these, only 18 represented potentially full-length ORFs, whereas the other 36 were shorter than 1.4 kb, suggesting that they cannot encode a complete *VSG*, or had internal stop codons or frameshifts. Of all of these *VSG* sequences, 38 had 70-bp repeat sequence upstream, 7 did not, and the remaining 9 were too close to the end of the contig to consider. The 70-bp repeat sequences found upstream formed an almost identical consensus sequence to that previously defined when aligning 70-bp repeats from within an array (Figure 3.18 and Figure 3.19; Aline et al, 1985). The unexpected finding, however, comes from their size. The 70-bp repeat regions here are very short, ranging from 24 to 266 bp in length and averaging at 91 bp, which is in sharp contrast to the large, several kilobase, arrays upstream of minichromosome and expression site associated *VSG* genes. They are however very similar to those seen upstream of metacyclic *VSG* genes. Assuming that these 18 complete *VSG* genes can be activated by duplication, this supports the idea that *VSG* switching can act on very short 70-bp repeat regions, similar to the duplicative activations of the ILTat 1.22 gene which involved the use of a 70-bp repeat region of only 115 bp in length see section 3.2.5). This may imply that the use of short regions of homology during VSG switching is a common occurrence, perhaps by a mechanism distinct from standard homologous recombination.

```
                1                                                  50
1       (1) --------------------------------------------------
2       (1) ----------------------------------TAATAATAATAATA
3       (1) --------------------------------------------------
4       (1) --------------------------------------------------
5       (1) --------------------------------------------------
6       (1) --------------------------------------------------
7       (1) --------------------------------------------------
8       (1) --------------------------------------------------
8b      (1) CAG----------------------------------------------
9       (1) --------------------------------------------------
9b      (1) CAG----------------------------------------------
10      (1) --------------------------------------------------
11      (1) --------------------------------------------------
12      (1) --------------------------------------------------
13      (1) --------------------------------------------------
14      (1) --------------------------------------------------
15      (1) --------------------------------------------------
16      (1) --------------------------------------------------
16b     (1) CAG----------------------------------------------
17      (1) --------------------------------------------------
18      (1) --------------------------------------------------
19      (1) --------------------------------------------------
20      (1) --------------------------------------------------
20b     (1) CAG----------------------------------------------
21      (1) CAG----------------------------------------------
22      (1) --------------------------------------------------
23      (1) CAG----------------------------------------------
24      (1) --------------------------------------------------
24b     (1) CAG----------------------------------------------
24c     (1) CAG----------------------------------------------
25      (1) --------------------------------------------------
25b     (1) CAG----------------------------------------------
26      (1) CAG----------------------------------------------
27      (1) CAG----------------------------------------------
27b     (1) CAG----------------------------------------------
28      (1) CAG----------------------------------------------
29      (1) CAG----------------------------------------------
30      (1) --------------------------------------------------
30b     (1) --------------------------------------------------
31      (1) CAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA
32      (1) --------------------------------------------------
33      (1) CAA----------------------------------------------
34      (1) --------------------------------------------------
34b     (1) CAG----------------------------------------------
35      (1) --------------------------------------------------
36      (1) --------------------------------------------------
36b     (1) --------------------------------------------------
37      (1) ----------------------------------------------TA
38      (1) --------------------------------------------------
Consensus (1)
```

Figure 3.18 A alignment of the $(TRR)_n$ region of the 70-bp repeat regions obtained from sequences of chromosomes IX and X. The grey boxes indicate conserved bases, whereas the black box indicates a base conserved through all sequences. The numbers indicate sequences from different 70-bp repeat regions, the letter afterwards indicates that it is a separate repeat sequence from the same numbered 70-bp repeat array. The consensus sequence was generated using the AlignX program from Vector NTI (Informax).

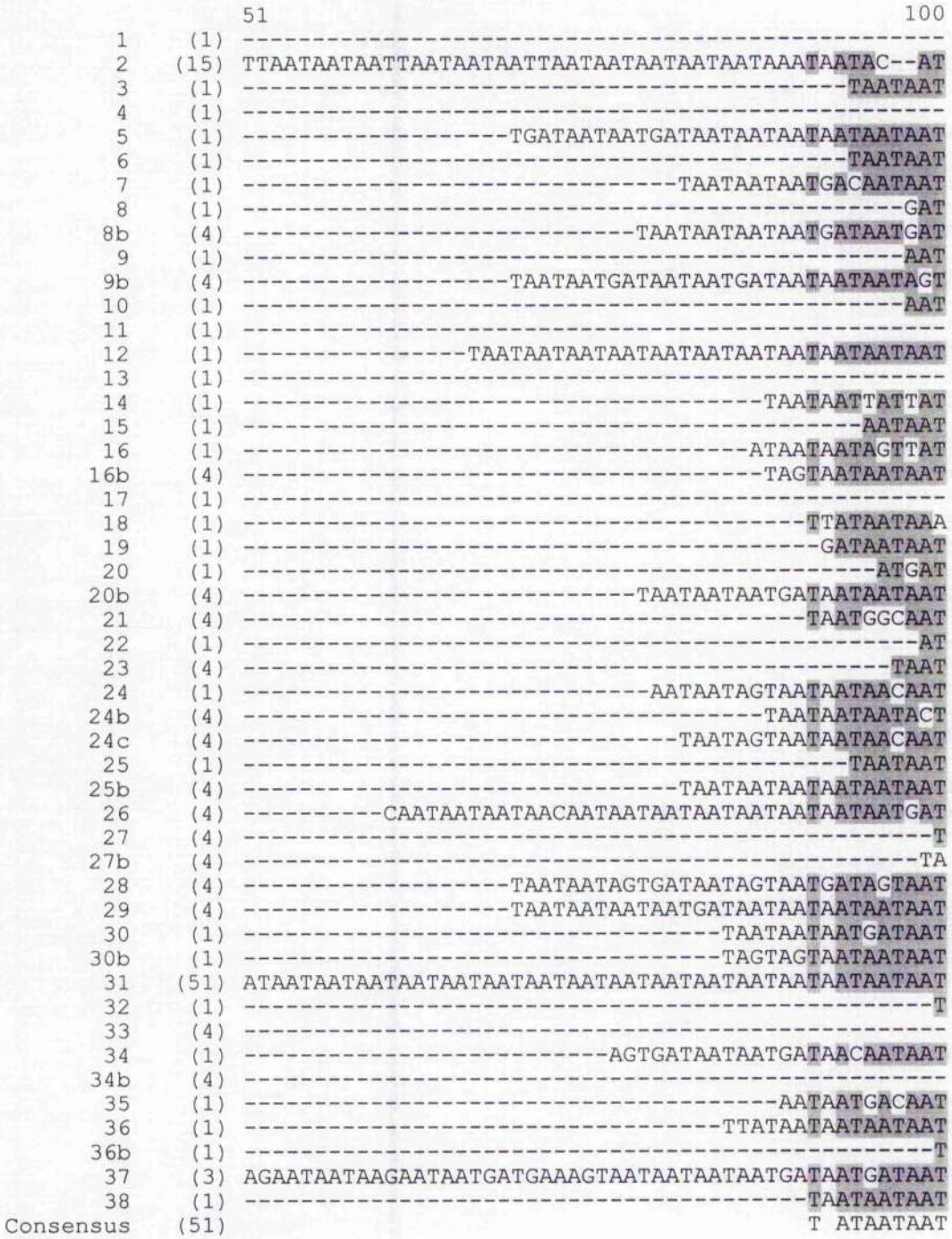


Figure 3.18 Continued.

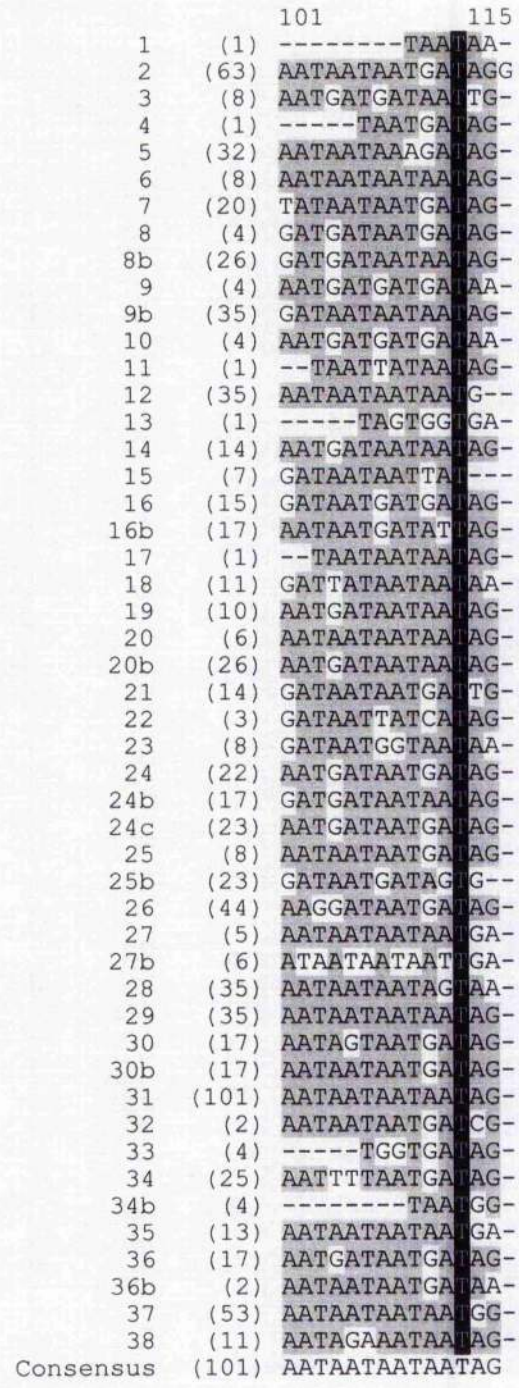


Figure 3.18 Continued.

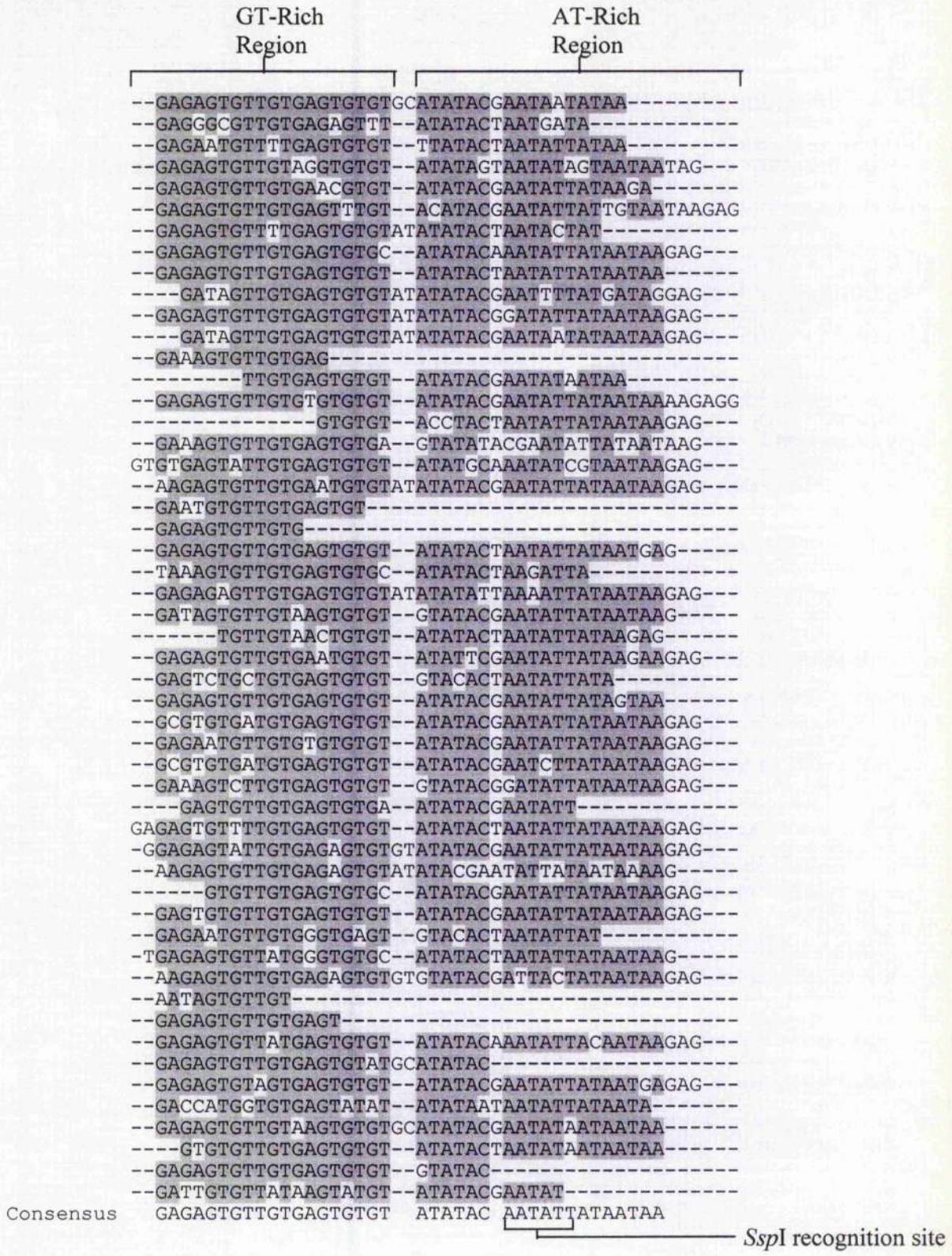


Figure 3.19 A line up of the GT- and AT- rich regions of the 70-bp repeat regions obtained from sequences of chromosomes IX and X. The grey boxes indicate conserved bases. The consensus sequence was generated using the AlignX program from Vector NTI (Informax).

The deficit of complete *VSG* genes may also elevate the significance of mosaic gene formation in *VSG* switching, which also utilises short stretches of homology (Thon *et al*, 1989; Thon *et al*, 1990). The paucity of intact *VSG* genes within arrays has also been seen recently, with 6 out of 7 *VSG* genes in a small array in chromosome II being pseudogenes. One other interesting observation concerning the 70-bp repeats is that those found upstream of internal genes appear to be surrounded by a disproportionate number of *Ssp*I restriction enzyme recognition sites. As this enzyme recognition site is present in the consensus sequence for 70-bp repeats, one explanation could be that the arrays used to be longer and have now become degenerate.

Name of contig	**Number of regions containing homology to *VSG* genes**
Tryp9.0.000860	3
Tryp9.0.000804	3
Tryp9.0.000714	3
Tryp9.0.000676	4
Tryp9.0.000627	1
Tryp9.0.000606	3
Tryp9.0.000604	3
Tryp10.0.003747	4
Tryp10.0.003445	2
Tryp10.0.000381	7
Tryp10.0.002136	3
Tryp10.0.003770	1
Tryp10.0.000839	17

Table 3.1 A list of the contigs obtained from http://www.genedb.org/genedb/tryp/index.jsp and the number of *VSG* analysed from each.

3.6 Conclusions

This work confirms that duplicative transposition-based activation of the ILTat 1.22 *VSG* gene in pleomorphic lines, in contrast to what happens with most *VSG* genes in

monomorphic lines, primarily utilises the 70-bp repeats as the upstream limit of conversion. The upstream conversion limits occurs at several points along the 70-bp repeat tract, indicating a homologous recombination reaction as the underlying mechanism, as opposed to a site-specific recombination event. Also, this reaction appears to require only a short region of homology, with an upper limit in two cases being just 42 bp (the relevance of this is discussed in detail in Chapter 6).

It seems that initiation primarily occurs in the 70-bp repeats of the BES, and copying the ILTat 1.22 BC to the end of the telomere occurs in a manner, similar to the break-induced replication model of conversion previously formulated in yeast (Pâques and Haber, 1999). This type of reaction has been suggested before for activating telomeric *VSG* genes (de Lange, 1983; Scholler *et al*, 1989).

The activation of ILTat 1.64 *VSG* was very similar, in terms of sequence requirements, to the activations of ILTat 1.22 *VSG*. If the activation of ILTat 1.64 did use the 70-bp incomplete 70-bp repeat region then this would be, to my knowledge, the smallest 70-bp repeat region to be utilised in a *VSG* gene conversion reaction (although it cannot be discounted that other sequences were responsible).

A genomic analysis of internal *VSG* genes demonstrates that these small 70-bp repeat units are common and thus poses interesting questions regarding the activation of *VSG* genes. Also, only approximately one third of the *VSG* genes characterised appeared to be intact, thus raising the possibility that mosaic gene formation is more important as a mechanism of antigenic variation than previously believed.

CHAPTER 4

IN VITRO STUDIES ON THE MECHANISM OF DUPLICATIVE ACTIVATION OF *VSG* GENES

4.1 Introduction

Within pleomorphic lines, such as ILTat 1.2, it is apparent that *VSG* duplicative transposition switches are more common than transcriptional, *in situ*, activations (Robinson *et al*, 1999). It is also evident that the 70-bp repeats are an important sequence element, perhaps playing a crucial role in the duplicative transposition of *VSG* genes. In fact, very short arrays, and possibly even less than a single repeat unit, are capable of acting as the upstream conversion limit for duplications, and these short repeat units are widely associated with *VSG* genes (see Chapter 3). Little is known about the nature of the repeat unit, yet much can be speculated. The triplet repeat region, a $(TRR)_n$ run which is often TAA, has been shown to have a low melting temperature which could be advantageous to a recombination reaction as it would be easier for the invading strand to enter the homologous duplex (Ohshima *et al*, 1996). The GT rich region may preferentially be bound by Rad51, as yeast Rad51 and *E. coli* RecA favourably bind similar sequences from a pool of random oligos (Tracy and Kowalczykowski, 1996; Tracy *et al*, 1997; Biet *et al*, 1999). Finally, the AT rich region has sequences that are predicted to form stem and cruciform structures (Liu *et al*, 1983; Aline *et al*, 1985).
Despite these possibilities, there has been little work on these repeats or their influence on *VSG* switching. The 70-bp repeats, however, have been removed from the active bloodstream expression site (BES; McCulloch *et al*, 1997). The removal of the repeats made no measurable difference to the frequency or mechanism of *VSG* switching. This study was made in the monomorphic line Lister 427, however, which does not regularly use the 70-bp repeats for conversion unlike the pleomorpic ILTat 1.2 line, so there are still many interesting questions about the sequence and genetic requirements of *VSG* duplicative transpositions in relation to the 70-bp repeats.
In the work described in this chapter, attempts were made to create a gene conversion assay that would allow various aspects of the 70-bp repeats to be studied. These aspects include their influence on gene conversion and whether they are more recombinogenic than other sequences homologous to the expression site. Efforts were made to study the influence of the repeats in a strain mutated in the recombination protein Rad51. Attempts were also made, *in vitro*, to determine if the 70-bp repeats are cleaved by a specific endonuclease, as has been postulated

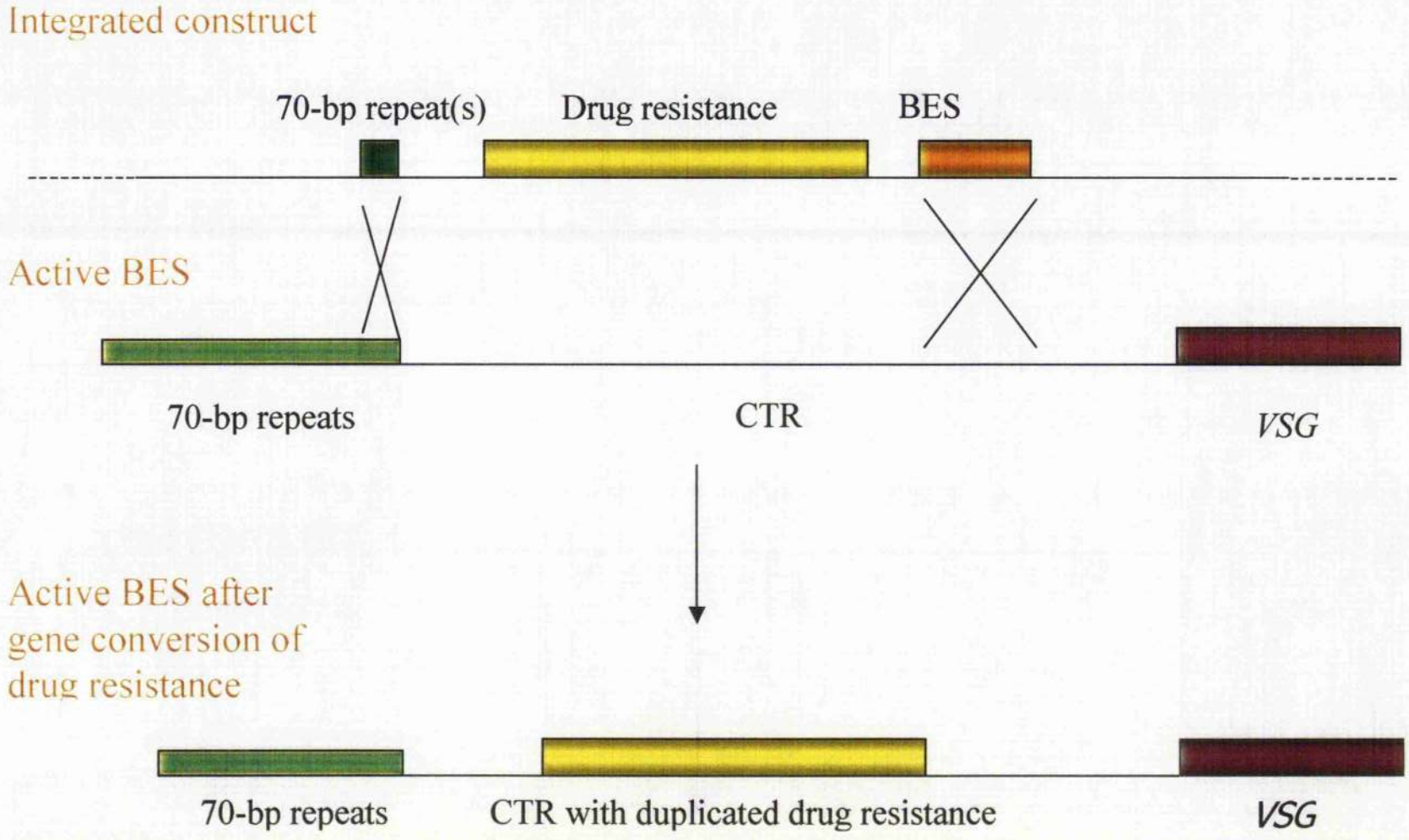
to be an initiating route for *VSG* switching *in vivo* (Barry, 1997), and to identify any protein factors that may bind specifically to the 70-bp repeats.

4.2 An assay to assess the influence of the 70-bp repeat sequence on recombination with the active bloodstream expression site

The aim of this approach was to create an assay in which to test the sequence elements of the 70-bp repeats required to copy regions of DNA into the active BES. The principle behind the assay is to place a drug resistance gene between a 70-bp repeat sequence and another region that is homologous to the active BES. In theory, after integration of this construct into a region of the genome, selection at a high concentration of drug should select any trypanosomes that have gene-converted the drug resistance gene into the BES, due to the high level of expression of the locus relative to RNA Pol II-directed transcription elsewhere in the genome. The movement presumably would occur by a gene conversion reaction using the sequences downstream that are homologous to the BES, and the 70-bp repeats (Figure 4.1). Thus, the frequency of the reaction would be quantifiable and comparisons could be made with reactions involving alterations of the 70-bp repeat sequences, such as length or removal of the AT-, GT- rich or $(TRR)_n$ regions of a single repeat.

The 70-bp repeat region used was from the ILTat 1.22 metacyclic expression site (see Chapter 3), as this sequence is known to be utilized in the duplicative transposition of the ILTat 1.22 *VSG* gene and is also only 115 bp in length and therefore easily manipulated. The downstream region homologous to the active BES was from the co-transposed region (CTR) of this active BES. This was chosen, as opposed to a complete *VSG* gene, so questions could be asked solely about the 70-bp repeat sequences without the influence of any sequences from a *VSG* gene, which have also been shown, in some instances, to be involved in duplicative transpositions (Donelson *et al*, 1983; Liu *et al*, 1983; Michels *et al*, 1983; Liu *et al*, 1985; Pays *et al*, 1985). The locus chosen for the construct to act as a donor was an intergenic region between two copies of the aldolase gene (Gen Bank accession number X52586; Vijayasarathy *et al*, 1990), which is a transcribed chromosome-internal region. Efforts were made to set up this assay in both the monomorphic Lister 427 and the pleomorphic ILTat 1.2 lines.

Figure 4.1 The principle behind the assay designed to test the influence of the 70-bp repeats on moving sequence via gene conversion into a bloodstream expression site. The crossed lines indicate where recombination between the two molecules is predicted to occur. BES, Bloodstream expression site; CTR, co-transposed region.



4.2.1 Isolation of sequence from the ILTat 1.2 co-transposed sequence

As only *VSG* coding sequence has previously been described from the ILTat 1.2 BES, it was necessary to isolate sequence from the CTR so that the assay in this strain would be directly comparable with that in the Lister 427 line. To acquire this sequence, a genomic walking approach was taken, utilising the TOPO-walker procedure (Invitrogen). Firstly, genomic DNA was digested with *Ssp*I, then primer extension was performed using an ILTat 1.2 *VSG* N-terminus coding region-specific primer, followed by addition of a unique linker. PCR was then performed using a primer specific to the ILTat 1.2 N-terminus coding region (ILTat 1.2 N-term 3') and another specific to the linker. An 800 bp PCR product was generated with Herculase polymerase (Stratagene) which was cloned into TOPO-vector (Invitrogen) and sequenced. When aligned with the ILTat 1.2 *VSG* gene sequence a 100% identity was seen, thus 400 bp of novel CTR sequence had been generated.

4.2.2 Generation of constructs

The starting construct from which the sequence modifications were made, pRM322, contains the neomycin phosphotransferase gene, that confers resistance to the drug G418, flanked by actin and calmodulin processing sequences cloned into pBluescript SK.
To allow integration of the construct into the intergenic region between the aldolase genes, the construct must contain targeting flanks homologous to this locus. To design primers to amplify these flanks, DNA sequence of the aldolase locus was obtained from the Gen Bank database (accession number X52586). A 360 bp upstream targeting flank was PCR-amplified, with herculase DNA polymerase, using the ALDO 1 *Sac*I and the ALDO 1 *Sac*II primers (conditions - 1 cycle of 96 °C for 5 min, 30 cycles of 96 °C for 1min, 55 °C for 1 min and 72 °C for 1min, and 1 cycle of 72 °C for 10 min), which respectively have the *Sac*I and *Sac*II restriction enzyme recognition sites incorporated on their 5' ends. The PCR product was cloned into *Sac*I and *Sac*II digested pRM322, and the directional cloning was checked by sequencing with a primer specific to the T3 promoter. This new plasmid construct was named pPB1001.

The downstream targeting flank was then amplified, with herculase DNA polymerase, using the ALDO 2 PsPI and the ALDO 2 *Kpn*I primers (conditions - 1 cycle of 96°C for 5 min, 30 cycles of 96°C for 1min, 55°C for 1 min and 72°C for 1min, and 1 cycle of 72°C for 10 min), which respectively had the *Pspom*I (an *Apa*I isoschizomer) and *Kpn*I restriction enzyme recognition sites incorporated on their 5' ends. The PCR product was cloned into *Pspom*I and *Kpn*I digested pPB1001, and the directional cloning was verified by sequencing with primers specific to the T7 promoter. This new plasmid construct was named pPB1002.

To clone the 70-bp repeat region from upstream of the ILTat 1.22 *VSG* gene, PCR-amplification was performed using the 1.22 70bp repeats P1/2 and 1.22 70bp repeats P2/2 primers (conditions - 1 cycle of 96 °C for 5 min, 30 cycles of 96 °C for 1min, 35 °C for 1 min and 72 °C for 30 seconds, and 1 cycle of 72 °C for 10 min), both of which had *Not*I restriction enzyme recognition sites incorporated into their 5' ends. The template for this reaction was the plasmid pMG7-1.1 which contains a portion of the ILTat 1.22 metacyclic expression site (Figure 3.3). The PCR product was cloned into *Not*I digested pPB1002. *Not*I digestion determined those with successful ligations and sequencing using the T3 promoter-specific primer determined those in the correct orientation. This construct was named as pPB1003.

The next sequences to be cloned into the construct were the regions of the CTR from the 221 (Bernards *et al*, 1985) and ILTat 1.2 BESs, the BESs active in the Lister 427 and ILTat 1.2 lines respectively. A 500 bp fragment of the 221 CTR was generated by PCR, using herculase DNA polymerase, with the 221 CTR 5' and 221 CTR 3' primers (conditions - 1 cycle of 96 °C for 5 min, 30 cycles of 96 °C for 1min, 55 °C for 1 min and 72 °C for 1min, and 1 cycle of 72 °C for 10 min), the former having a *Xho*I recognition site within it and the latter having one incorporated at its 5' end. The PCR product was cloned into *Xho*I digested pPB1003 and checked for the successful ligation by restriction enzyme digestion, and by PCR to check for orientation. This construct was named pPB1004 (Figure 4.2). The amplification of part of the ILTat 1.2 CTR was performed using the 1.2 CTR 5' 1/2 and 1.2 CTR 3' 1/2 primers (conditions - 1 cycle of 96 °C for 5 min, 30 cycles of 96 °C for 1min, 55 °C for 1 min and 72 °C for 1min, and 1 cycle of 72 °C for 10 min), both of which had *Xho*I recognition sites incorporated at their 5' ends. This product was cloned into the *Xho*I site of pPB1003 in a manner identical to the cloning of the 221 CTR fragment. This new plasmid construct was named pPB1005 (Figure 4.2).

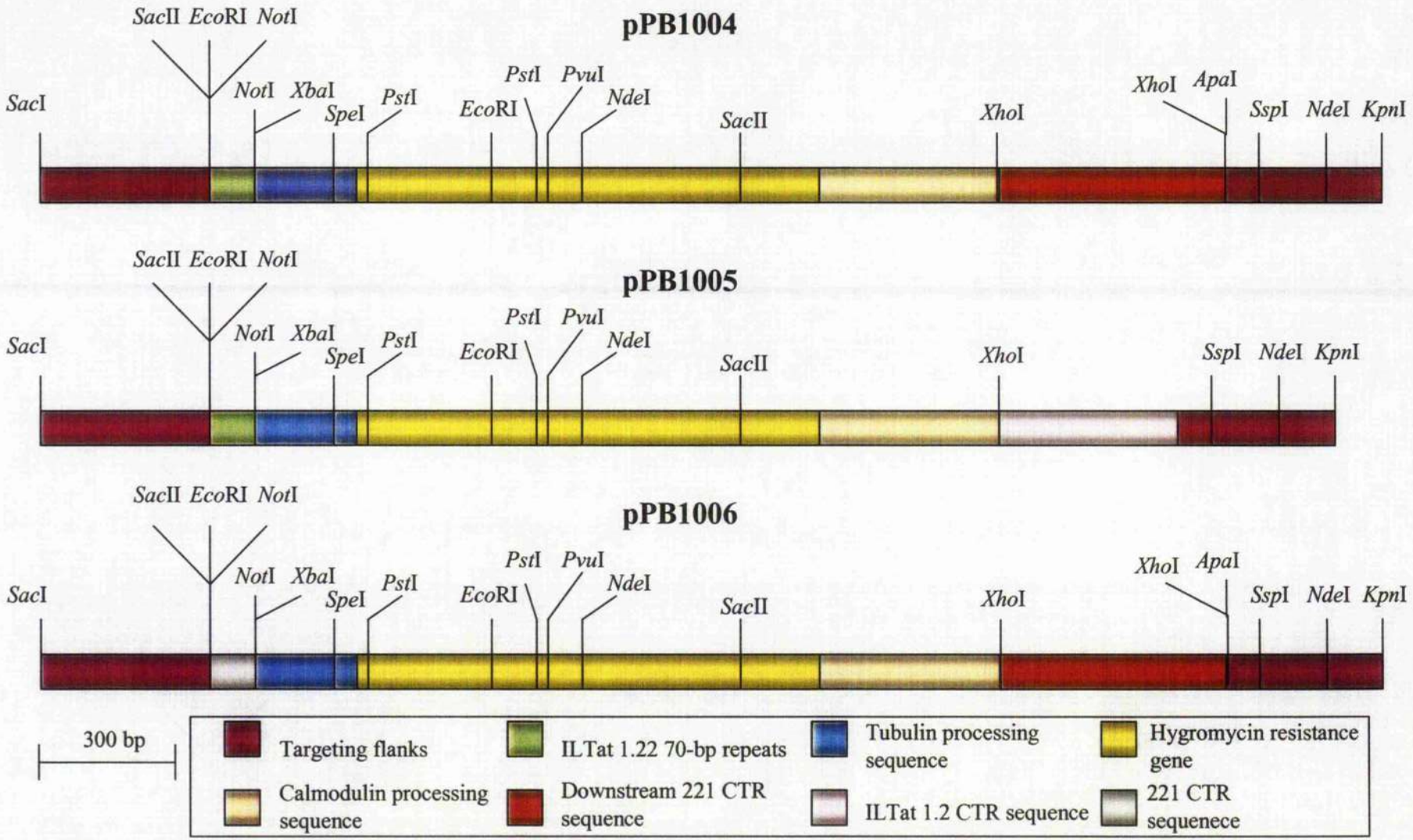
It was deemed important to determine if another sequence, substituted for the 70-bp repeat region, could be used as efficiently as the ILTat 1.22 70-bp repeat sequence at catalysing the gene conversion reactions duplicating sequence from the construct integrated into the aldolase intergenic region to the active BES. A construct was therefore prepared with a region derived from the active BES CTR that was identical in length to the ILTat 1.22 70-bp repeat region but not 70-bp repeat sequence or sequence already present in the construct. This was achieved by PCR amplification from Lister 427 genomic DNA using the 221 5'CTR *Not*I 5' and 221 5'CTR *Not*I 3' primers (conditions - 1 cycle of 96°C for 5 min, 30 cycles of 96°C for 1min, 55°C for 1 min and 72°C for 30 seconds, and 1 cycle of 72°C for 10 min), both of which had a *Not*I site incorporated into their 5' ends. The PCR product was purified and ligated to *Not*I digested pPB1004 using the approach described for the generation of pPB1002. This new plasmid construct was named pPB1006 (Figure 4.2).

4.2.3 The integration of the constructs

The bacterial cells containing the constructs pPB1004, pPB1005 and pPB1006 were grown and the plasmids were purified. Thirty μg DNA of each was digested at 37°C for 24 h with *Sac*I and *Kpn*I, and then phenol/chloroform extracted. Lister 427 trypanosome cells were grown to a density of 2 x 10^6 cells.ml^{-1} and ILTat 1.2 cells to a density of 7.5 x 10^5 cells.ml^{-1}. A total of 5 x 10^7 Lister 427 cells and 3 x 10^7 ILTat 1.2 cells were transformed with 5 μg of the digested DNA. The Lister 427 cells were transformed with pPB1004, pPB1005 and pPB1006, whereas the ILTat 1.2 cells were transformed only with PB1005. After electroporation and recovery overnight in 10 ml of medium, 1 x 10^7 Lister 427 cells or 3 x 10^6 ILTat 1.2 cells were plated over 12 wells and selected at 5 μg.ml^{-1} hygromycin (Roche Diagnostics Ltd.).

Two transformations for each of the constructs pPB1004, pPB1005 and pPB1006 were performed on the Lister 427 line. For pPB1004, 22 transformants were obtained from 24 wells, indicating a high transformation efficiency of approximately1 transformant in 10^6 cells. This was also seen for the transformation of pPB1005 and pPB1006, where 20 and 21 transformants were seen from 24 wells respectively. The transformation of the ILTat 1.2 cells

Figure 4.2 Restriction maps of the constructs created for analysing gene conversion events into the active expression site. CTR, co-transposed region.



was much less efficient, a general phenomenon with transformation of this pleomorphic line, with only 4 wells in total growing out from 9 attempted transformations. Transformants were then cloned in 96 well dishes and stabilated. Genomic DNA was isolated from the clones and digested with *Xho*I. The restriction enzyme-digested DNA was size-fractionated on a 0.6% agarose gel, Southern blotted and probed with a fragment of the hygromycin resistance gene open reading frame obtained from Richard McCulloch (Figure 4.3). In each case, a single band of the expected size (5 kb; Figure 4.3 A) can be seen, indicating the correct integration of the construct. The strains resulting from integration of the plasmids were named 427.pPB1004, 427.pPB1005, 427.pPB1006 and 1.2.pPB1005, after the original strain and the name of the plasmid.

4.2.4 A comparison of the hygromycin resistance levels expressed from the aldolase intergenic region and the active 221 expression site

To elucidate the concentration of hygromycin that could be used to select for gene conversions of the hygromycin resistance gene from the aldolase intergenic regions into the active BES, a comparison of the drug resistance level produced from each locus was necessary. To do this, the 427.pPB1005 strain was chosen as it was less likely to yield gene conversions of the hygromycin resistance gene to the active BES, thus not affecting drug resistance measurements. Two clones of these transformants were compared with the 3174 line of Lister 427, which contains a copy of the hygromycin resistance gene within the active BES (McCulloch *et al*, 1997).

1×10^3 cells were inoculated into 3 ml of HMI-9 medium in 6-well dishes containing varying concentrations of hygromycin. Four days later the wells were examined for trypanosome survival. All transformants of pPB1005 were dead after 4 days at 175 $\mu g.ml^{-1}$ hygromycin, whereas the 3174 line was still growing at only a slightly increased doubling time (Table 4.1). Thus, 175 $\mu g.ml^{-1}$ was judged to be a suitable drug concentration for selection of gene conversion events into the active BES.

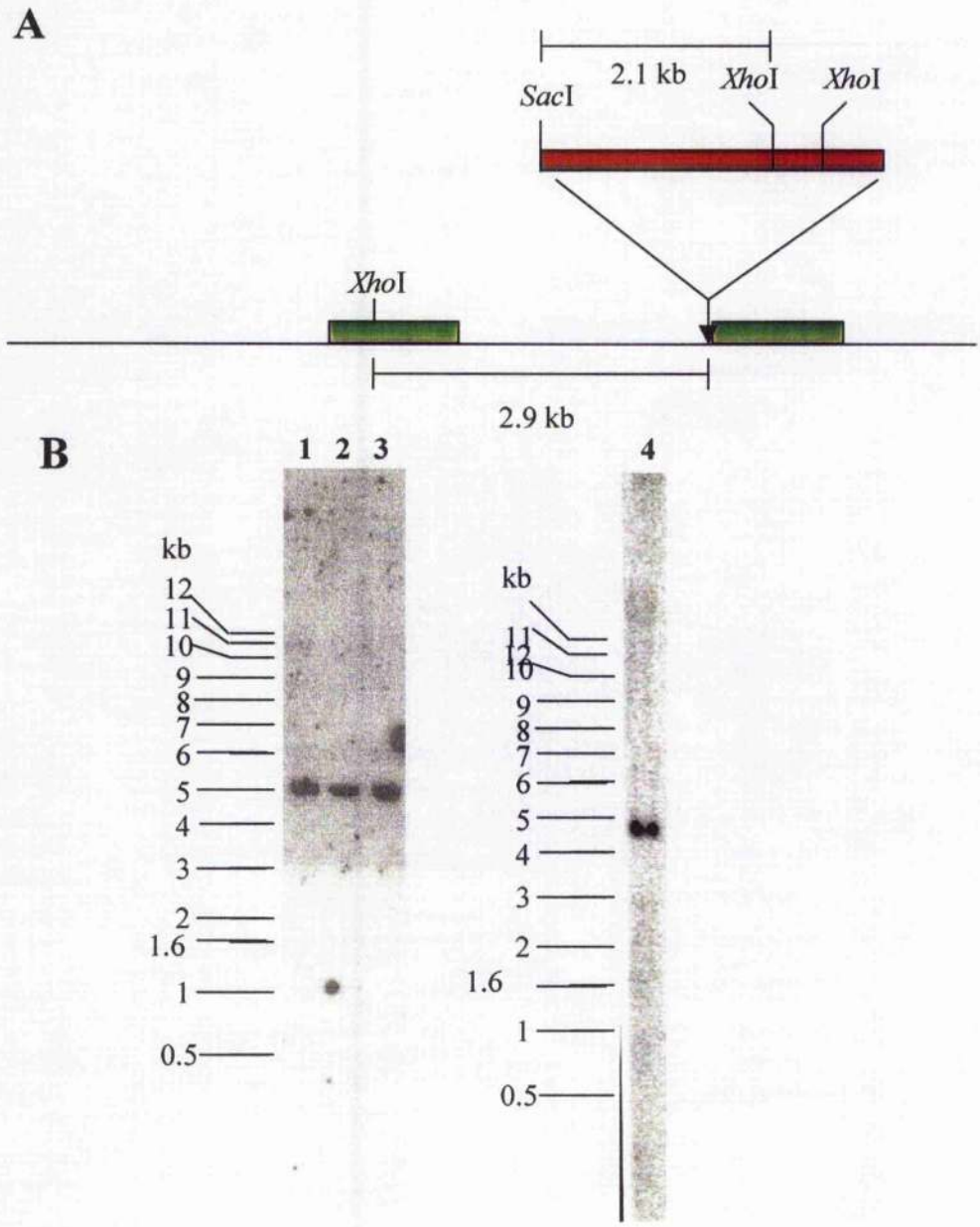


Figure 4.3 Integration of the pPB1004, pPB1005 and pPB1006 constructs. **A**. A representation of the integration of the constructs into the aldolase intergenic region. The red rectangle can represent all three constructs, the green blocks indicate the aldolase genes, and the arrow indicates the place of insertion within the genome. The 2.9 kb distance is the distance between the most proximal *Xho*I site in the genome and where the constructs are inserted, not inclusive of the targeting flank. The 2.1 kb distance is that from the *Sac*I site of the construct to the next downstream *Xho*I site, inclusive of the targeting flank. **B**.Genomic DNA was digested with *Xho*I and then size-fractionated on a 0.6% agarose gel, Southern blotted and probed with a region of the hygromycin resistance gene open-reading frame. Lanes 1 to 4 are strains 427.pPB1004, 427.pPB1005, 427.pPB1006 and 1.2.pPB1005 respectively.

Hygromycin (μg.ml^{-1}) / Cell Line	0	30	60	90	100	125	150	175	200
427.pPB1005.1	++++	++++	++++	+++	++	+	+	X	X
427.pPB1005.2	++++	++++	++++	+++	++	++	+	X	X
3174	++++	++++	++++	++++	++++	++++	+++	+++	+++

Table 4.1 A comparison of the hygromycin resistance levels expressed from the aldolase intergenic region and the active 221BES. 1 x 10^3 cells were placed in 3 ml of medium and grown for 4 days, after which survival was judged. ++++ denotes the highest level of anticipated growth (assuming an 8 hour doubling time) and +++, ++, + denote progressively less growth. X – no living cells.

4.2.5 Measuring the frequency of gene conversions of the hygromycin resistance gene from the aldolase intergenic region to the active bloodstream expression site

Measurement of the frequency of putative gene conversion events duplicating the hygromycin resistance gene from the aldolase intergenic region into the active expression site was undertaken by placing a known number of cells at a high hygromycin concentration, allowing survival of only those that had converted the gene into the BES. Measuring the number of clones that grow provides an estimate of the frequency of gene conversion events.

Two clones of the 427.pPB1004 strain were grown to densities of 1.5 x 10^6 trypanosomes.ml^{-1}. Cells were plated out at 1 x 10^6 trypanosomes/well in HMI-9 medium containing hygromycin at a concentration of 175 μg.ml^{-1} in 6-well dishes (5 ml HMI-9/ well). The majority of cells died during the first 3 – 4 days of incubation with the antibiotic, and cells with putative gene conversions were given the opportunity to grow for up to 2 more weeks.

A total of 1.96 x 10^8 cells (196 wells), over 3 experiments were incubated under selection conditions and a total of only 4 wells produced 175 μg.ml^{-1} resistant cells. Assuming that each well showing positive growth contained a single switch, this gives an approximate frequency of one gene conversion event in 4.9 x 10^7 trypanosomes.

It was necessary to confirm that the lines that had become resistant to a high concentration of drug had duplicated the hygromycin resistance gene and, if so, had undergone a gene conversion reaction copying the gene into the active BES. Genomic DNA, from each of three

of the highly resistant lines (named 427.pPB1004.R1, 427.pPB1004.R2 and 427.pPB1004.R3), was digested with *Hin*dIII, along with genomic DNA from the original Lister 427 strain transformed with PB1004 (427.pPB1004). Following size-fractionation on a 0.6% agarose gel, and Southern blotting, the genomic DNA was probed with a region of the hygromycin resistance gene coding region.

For the DNA from the original integration of the pPB1004 construct, a single hybridizing fragment can be seen (Figure 4.4). In the lanes corresponding to the highly resistant cells, this band was accompanied by a larger band of varying size (Figure 4.4). If these fragments represent gene conversion events then the variation in size could be due to different 70-bp repeat units, from the BES array, being used as the recipient sequence in each clone. The length variation cannot be due to variations in telomere length as there are *Hin*dIII recognition sites within the 221 *VSG* gene. To confirm that these fragments do indeed represent gene conversion events into the active 221 BES, a PCR approach was undertaken using the same genomic DNA preparations, as well as Lister 427 genomic DNA to control for the presence of any PCR products in the absence of a hygromycin resistance gene. Two PCR reactions were set up to determine if the hygromycin resistance gene was present in the aldolase intergenic region and the 221 BES (See Figure 4.5 for the primers and conditions used). In all cases, the hygromycin resistance gene was present in the aldolase intergenic region, but in the highly resistant lines it was also in the 221 BES (Figure 4.5). Thus, it appears that the highly resistant lines indeed have arisen through gene conversion events that had duplicated the hygromycin resistance gene into the active BES as predicted by the assay design. These events appear, however, to be too infrequent to permit study of gene conversions using sequences associated with *VSG* switching. It was necessary to attempt this assay in a strain which has a more rapid *VSG* switching rate, to see if this corresponds to an increased frequency of gene conversion. ILTat 1.2 *T. brucei* cells switch their VSG coat at least an order of magnitude more frequently than the Lister 427 line. If these events are initiated by an induced DSB in the 70-bp repeats upstream of the *VSG* gene in the active BES, it follows that the gene conversion events in this assay should also increase in frequency as they presumably use the 70-bp repeat sequence as the upstream conversion limit (see Chapter 3 for more thorough discussion of the use of 70-bp repeats in *VSG* switching in the ILTat 1.2 line). To test this, and hopefully increase the frequency of gene conversion events, the same experiment was performed in the ILTat 1.2 line

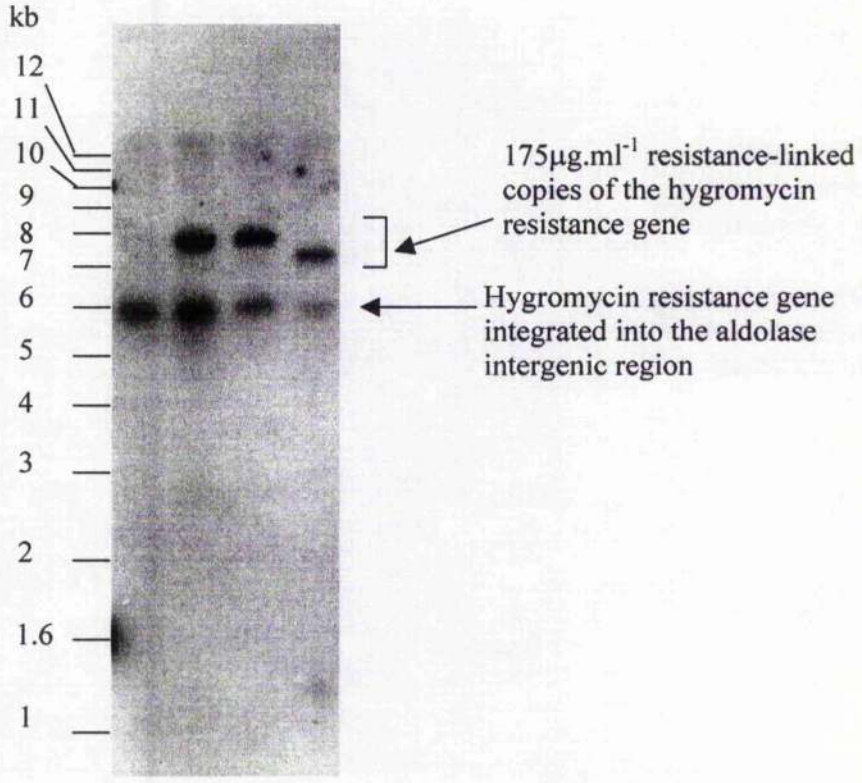
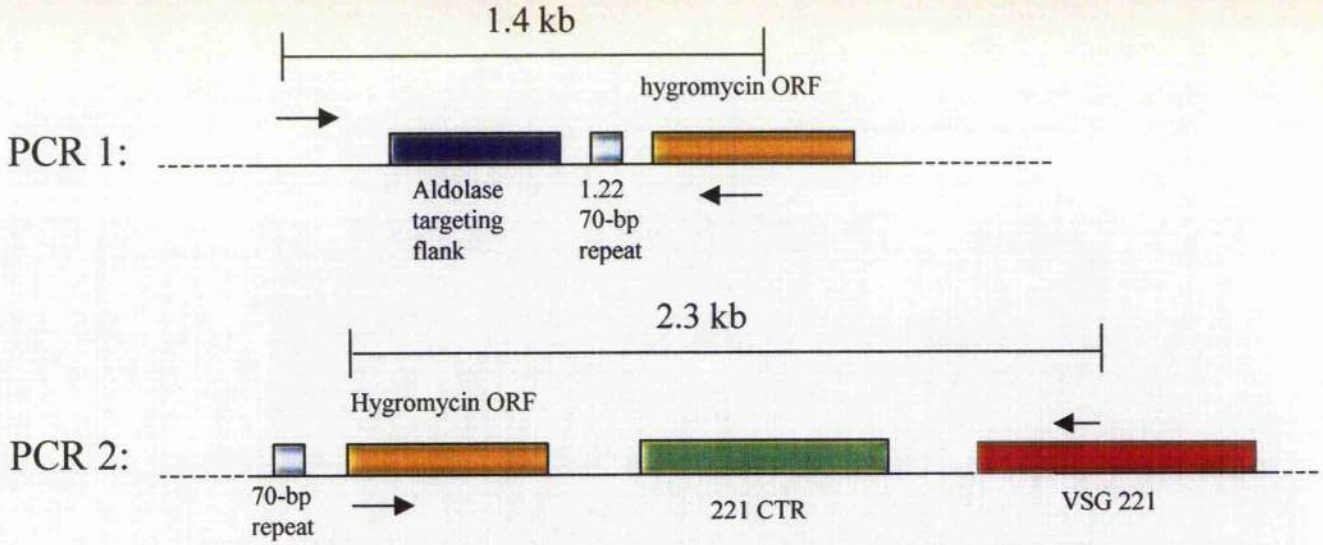


Figure 4.4 High level of hygromycin resistance is linked to the duplication of the hygromycin resistance gene. Genomic DNA was digested with *Hin*dIII, size-fractionated on a 0.6% agarose gel, Southern blotted and probed with a region of the hygromycin resistance gene open reading frame. The blot was washed to a stringency of 0.1 x SSC, 65°C. Lanes 1 to 4 correspond, respectively, to the 427. pPB1004 transgenic line and resistant lines 1 to 3.

(see sections 4.2.2 and 4.2.3 for details of the construct and cell line used). A variety of experimental approaches, based upon the experiment in the Lister 427 line, were performed (summarized in Table 4.2). However, no highly resistant lines (gene conversion events) were ever obtained. At the time of doing these experiments, protocols for transformation of this pleomorphic line were still in the preliminary stages of development, and it was therefore necessary to check that all components of the construct had integrated successfully. This was confirmed with PCR, using a primer specific to a region upstream of the 5' targeting flank of the construct, and other primers specific to several regions of the construct (see Figure 4.6 for the primers and conditions used). The hygromycin resistance gene and the ILTat 1.2 CTR sequence appeared to be linked to the aldolase intergenic locus (Figure 4.6). The presence of the ILTat 1.22 70-bp repeat region is unclear, as a ladder of bands is seen in all lanes

Figure 4.5 PCR demonstration of gene conversion into the active 221 bloodstream expression site (BES) in the *T.brucei* cell lines resistant to a high concentration of hygromycin. **A.** A representation of the two PCRs used to determine whether the hygromycin resistance gene was present in the aldolase intergenic region and the 221 active BES. PCR 1 uses primers specific to sequence upstream of the upstream aldolase targeting flank (ALDO test S1) and within the hygromycin resistance gene (HYGRO U1), resulting in a 1.4 kb product . PCR 2 uses primers specific to sequence from the hygromycin resistance gene (HYGRO 5' Switch) and the 221 *VSG* gene (VSG 221 N-Term), resulting in a 2.3 kb product. Primers are represented by arrows. **B.** The PCR products were detected after electrophoresis on a 0.8 % agarose gel containing ethidium bromide. Lanes 1, 3, 5, 7, 9 and 11 represent PCR 1, whereas lanes 2, 4, 6, 8, 10 and 12 represent PCR 2. Lanes 1 and 2 are controls omitting the DNA template, lanes 3 and 4 are the line before high drug selection (427.pPB1004), lanes 5 and 6 are resistant line 1, lanes 7 and 8 are resistant line 2, lanes 9 and 10 are resistant line 3 and lanes 11 and 12 are 427 to show that the primers do not amplify from trypanosome DNA which has not been transformed with the pPB1004 construct. Conditions for all the PCR reactions: (Conditions - 1 cycle of 96°C for 5 min, followed by 30 cycles of 96°C for 1min, 55°C for 1 min and 72°C for 2.5 min, and followed by 72°C for 10 min)

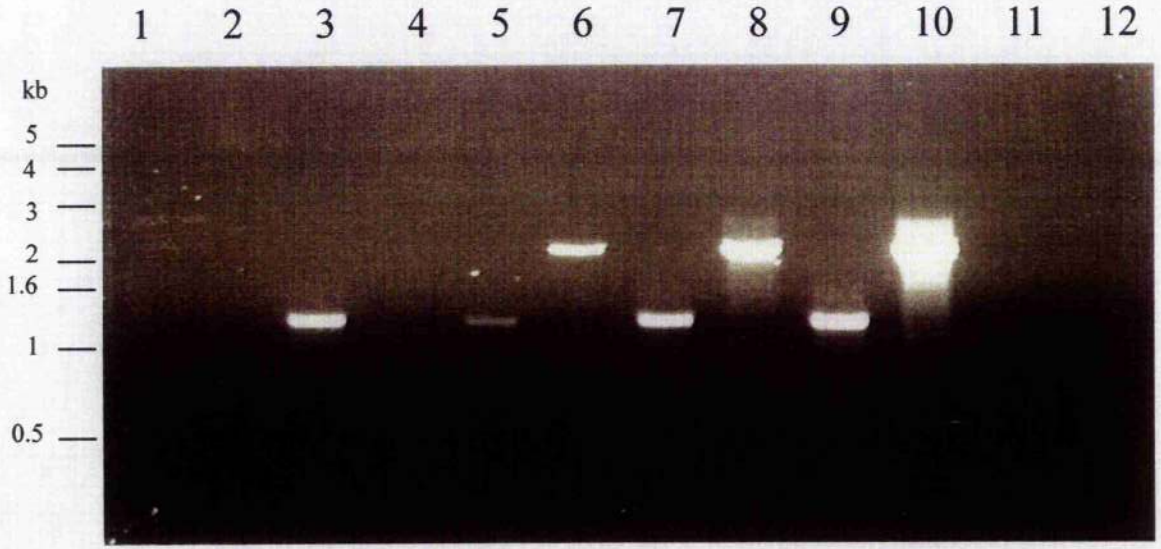
A



Tests integration into aldolase locus

Tests presence of hygromycin duplication into VSG 221 Locus

B



representing this primer (Figure 4.6, lanes 4, 8 and 12). This, however, was presumably due to the presence of many repeats of this sequence throughout the genome. It seems highly unlikely that this region had not integrated when all the other sequences of the construct did.

Another possibility for not generating highly resistant lines is that the ILTat 1.2 *VSG* gene was no longer being expressed in these transformants, thus not providing the downstream homology of the CTR to allow successful gene conversion reactions. To test for this, immunofluorescence was performed using anti-serum from a rabbit infected for six days with ILTat 1.2 trypanosomes, and antiserum from a rat infected with ILTat 1.22 trypanosomes (obtained from Nick Robinson and Liam Morrison). 1.2.pPB1005 grown in culture and ILTat 1.2 and ILTat 1.22 trypanosomes grown in a mouse were monitored with both antisera and visualized by using FITC-conjugated secondary antibodies specific to the primary antibody isotype. In both the cultured transformant and the original ILTat 1.2 trypanosomes, the ILTat 1.2 specific antibody bound to all trypanosomes of each line, whereas the ILTat 1.22 specific antibodies bound to neither. The exact reverse was the case for the ILTat 1.22-expressing trypanosomes. It can be concluded that the transformants were still expressing the ILTat 1.2 *VSG* gene.

It seems more likely that, rather than gene conversion events failing to occur, there was a problem with growth of resistant cells. This could have been due to a longer population doubling time, lower viability of the ILTat 1.2 line *in vitro*, or differentiation of the cells to the non-dividing, stumpy stage. These reasons are exemplified by the greatly reduced transformation efficiency when compared with the Lister 427 line. It cannot be ruled out, however, that the ILTat 1.2 BES is somewhat different from the 221 ES and has different sequence requirements for maintaining an active state, and therefore cannot tolerate the integration and loss of sequence created by gene conversion of the hygromycin resistance gene.

To see if a region of homology to the expression site other than the 70-bp repeat could catalyse gene conversion events at the same frequency, the strain 427.pPB1006 was used. This strain contained a region homologous to the active BES CTR that is identical in length to the ILTat 1.22 70-bp repeat used previously. Two clones of the 427.pPB1006 strain were grown to densities of 1.5×10^6 trypanosomes.ml^{-1}, and 1×10^6 trypanosomes were plated out in 5 ml of medium at a concentration of 175 μg.ml^{-1} hygromycin in 6-well dishes. The cells again died

Experiment No.	**Growth prior to selection**	**No. of wells**	**No. of cells per well**	**Drug concentration ($ug.ml^{-1}$)**	**Volume of medium per well (ml)**
1 (Clone 1)	*in vivo*	9	1×10^6	175	5
		9	1×10^7	175	5
2 (Clone 1)	*in vitro*	9	1×10^6	175	5
3 (Clone 2)	*in vitro*	12	1×10^6	175	5
4 (Clone 1)	*in vitro*	6	1×10^6	175	5
		6	1×10^6	150	5
5 (Clone 1)	*in vitro*	1	1×10^6	200	10
		1	1×10^6	150	10
6 (Clone 2)	*in vitro*	5	1×10^6	50	5
		5	1×10^6	125	5
		5	1×10^6	150	5
		5	1×10^6	175	5
		5	1×10^5	50	5
		5	1×10^5	125	5
		5	1×10^5	150	5
		5	1×10^5	175	5
7 (Clone 2)	*in vivo*	6	1×10^6	25	5
		6	1×10^6	50	5
		6	1×10^6	75	5
		6	1×10^6	125	5
		6	1×10^6	150	5
		6	1×10^6	200	5
		6	1×10^6	250	5

Table 4.2 A summary of the experiments attempted to assay gene conversion events in the ILTat 1.2 line. See text for details.

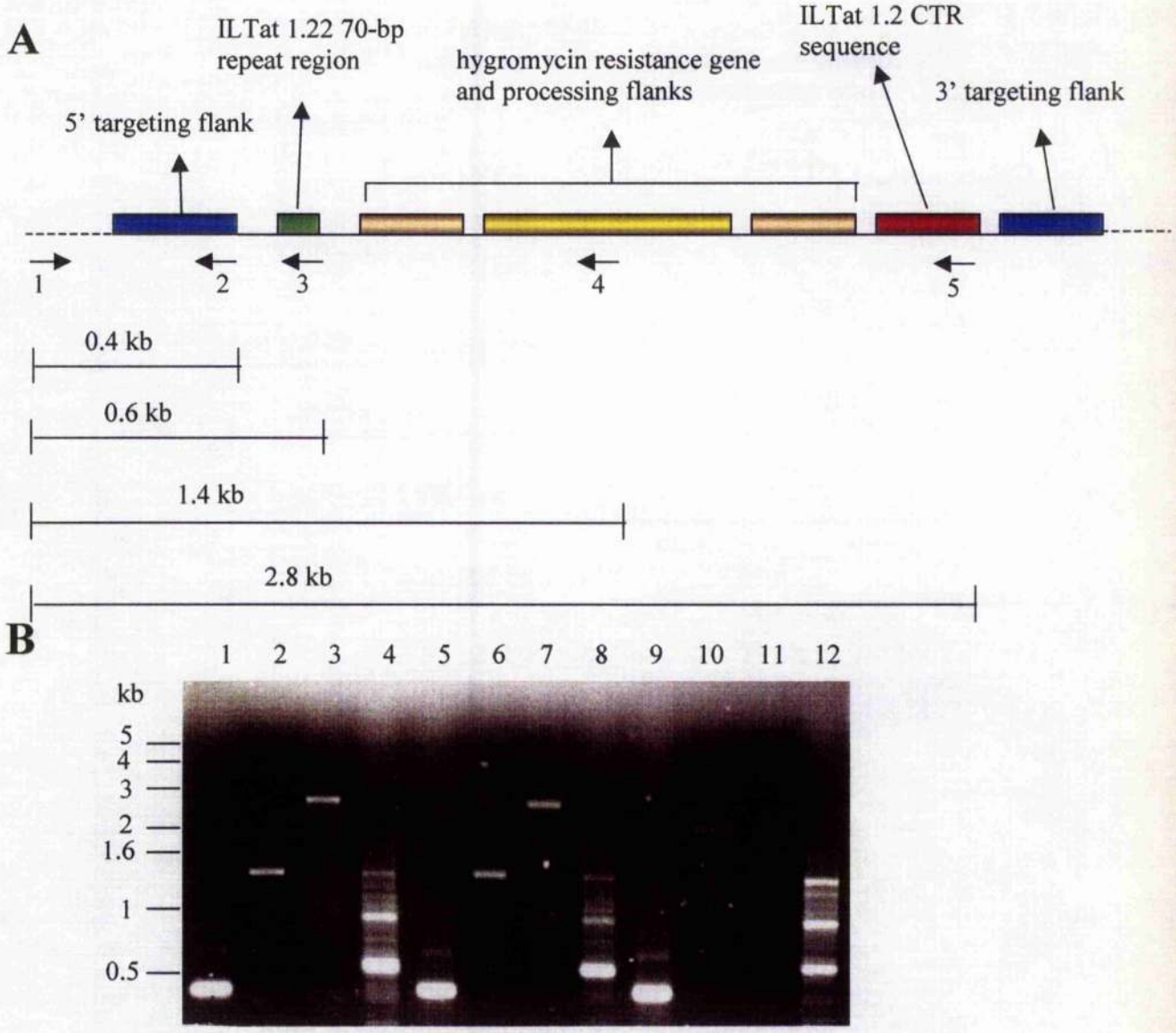


Figure 4.6 All regions of pPB1005 construct integrated successfully into the aldolase intergenic regions in the ILTat 1.2 line. **A.** map of pB1005 integrated into the aldolase intergenic region. The dotted lines indicate the surrounding genomic sequence into which the construct has integrated and the numbered arrows represent the primers used in the PCRs depicted in B. The bars below the diagram represent the expected size of the DNA fragment produced using a PCR involving primer 1 and each other depicted primer **B.** Ethidium bromide stained agarose gel of the PCRs. Lanes 1 to 4, template DNA from 1.2.pPB1005.1; lanes 5 to 8, template DNA from 1.2.pPB1005.2 and lanes 9 to 12, template DNA from ILTat 1.2. Lanes 1, 5 and 9 are reactions using primers 1 and 2, Lanes 2, 6 and 10 are reactions using primers 1 and 4, lanes 3, 7 and 11 are reactions using primers 1 and 5 and lanes 4, 8 and 12 are PCR reactions using primers 1 and 3. Primer 1, ALDO Test S1; primer 2, ALDO 1 *Sac*I; primer 3, 1.22 70bp repeats P2/2; primer 4, HYGRO U1; primer 5, 1.2 CTR 3' 1/2. Conditions - All PCRs except those involving Primer 3: 1 cycle of 96°C for 5 min, 30 cycles of 96°C for 1min, 55°C for 1 min and 72°C for 3 min, and 1 cycle of 72°C for 10 min. Those PCRs involving primer 3: 1 cycle of 96°C for 5 min, 30 cycles of 96°C for 1min, 35°C for 1 min, 72°C for 1min, and 1 cycle of 72°C for 10 min.

between 3 and 4 days later. A total of 1.26 x 10^8 trypanosomes were placed on selection over 2 separate experiments, but no cells resistant to the high concentration of hygromycin grew. This presumably does not reflect the lack conversion events seen with pPB1005, as Lister 427 cells are competent for catalyzing similar conversion events (see above, 427.pPB1004). Perhaps the conversion reaction occurring here in the 427.pPB1004 line, as perhaps also catalyses VSG switching in the Lister 427 strain, is occurring *via* background spontaneous homologous recombination. Thus, in 427.pPB1004 there are more sequences (*i.e.* the 70-bp repeats in the BES) to a act as the recipient for gene conversion than in 427.pPB1006, where only a single region of homology occurs. The difference in the number of homologous targets could account for the difference in frequencies of gene conversion seen in the two strains. Alternatively, the 70-bp repeat sequences may be inherently more prone to recombination, a property predicted by its sequence (Oshima *et al*, 1996). Unfortunately, as the rates are so low it is impossible to make any statistically significant firm conclusions about the nature of the conversion events.

4.2.6 The role of RAD51 in gene conversion

To attempt to investigate the role of RAD51 in the gene conversion reactions involving the 70-bp repeats (see section 4.2.5), pPB1004 was integrated into a Lister 427 bloodstream form *rad51* -/- mutant (McCulloch *et al*, 1999; Conway *et al*, 2002a). Nine transformations were performed, in the same manner as for the Lister 427 transformations described above (section 4.2.4) but, due to the reduced transformation efficiency of this *rad51* -/- line, only nine transformants were recovered after selection on 5 $\mu g.ml^{-1}$ hygromycin. Thus, the transformation efficiency for this line was, as expected, 1 transformant in 1 x 10^7 trypanosomes.

Genomic DNA was prepared from each of these transformants and digested with *Xho*I, size-fractionated on a 0.6% agarose gel, Southern blotted and probed with a fragment of the hygromycin resistance gene (Figure 4.7). None of the transformants had integrated the construct into the aldolase intergenic region. Aberrant integrations into non-targeted sequences using very short regions of homology have been described in *T. brucei rad51* -/- mutants (Conway *et al*, 2002a).

To determine if these integrations had utilised the ILTat 1.22 70-bp repeat region and the 221 CTR sequence for targeting the integrations, PCRs were performed using primers specific to the hygromycin resistance gene (HYGRO 5' Switch) and the 221 *VSG* gene (VSG 221 N-Term) (Figure 4.8 B) (conditions - 1 cycle of 96 °C for 5 min, 30 cycles of 96 °C for 1min, 55 °C for 1 min and 72 °C for 1min, and 1 cycle of 72 °C for 10 min). To control for DNA integrity, primers were used also to amplify a region of the RNA polymerase I large subunit (Figure 4.8 A) (The primers and conditions were identical to that in Rudenko *et al*, 1996). The PCRs revealed that four of the nine transformants had integrated into the 221 BES, whereas the others had integrated elsewhere in the genome.

Due to the absence of integrations of pPB1004 into the expected aldolase intergenic region in this strain, a comparison of the rates of gene conversion events with 427.pPB1004 could not be performed, as the undefined loci into which the integrations occurred may differ in the ability to recombine with the BES.

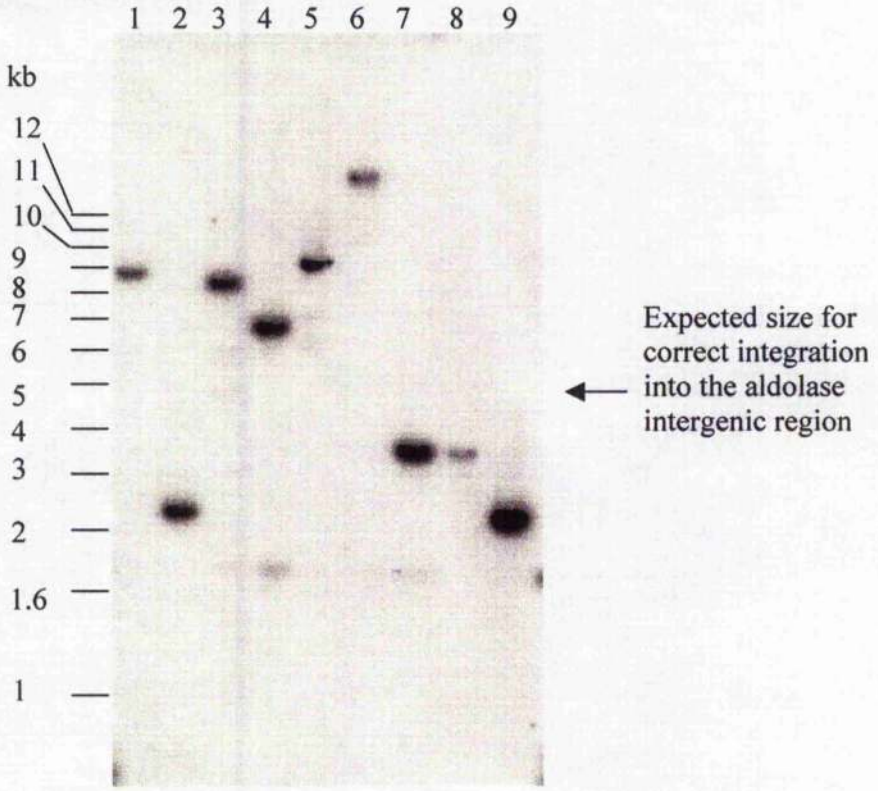


Figure 4.7 Integration of pPB1004 did not occur in the expected locus in *rad51* mutant cell lines. DNA from the nine transformations was digested with *Xho*I and size-fractionated on a 0.6 % Agarose gel, Southern blotted and probed with sequence specific for the hygromycin resistance gene. Lanes 1 to 9 are transformations 1.1, 2.1, 3.1, 5.1, 6.1, 8.1, 9.1 9.2 and 9.3 respectively. The blot was washed to a stringency of 0.1 x SSC, 65°C.

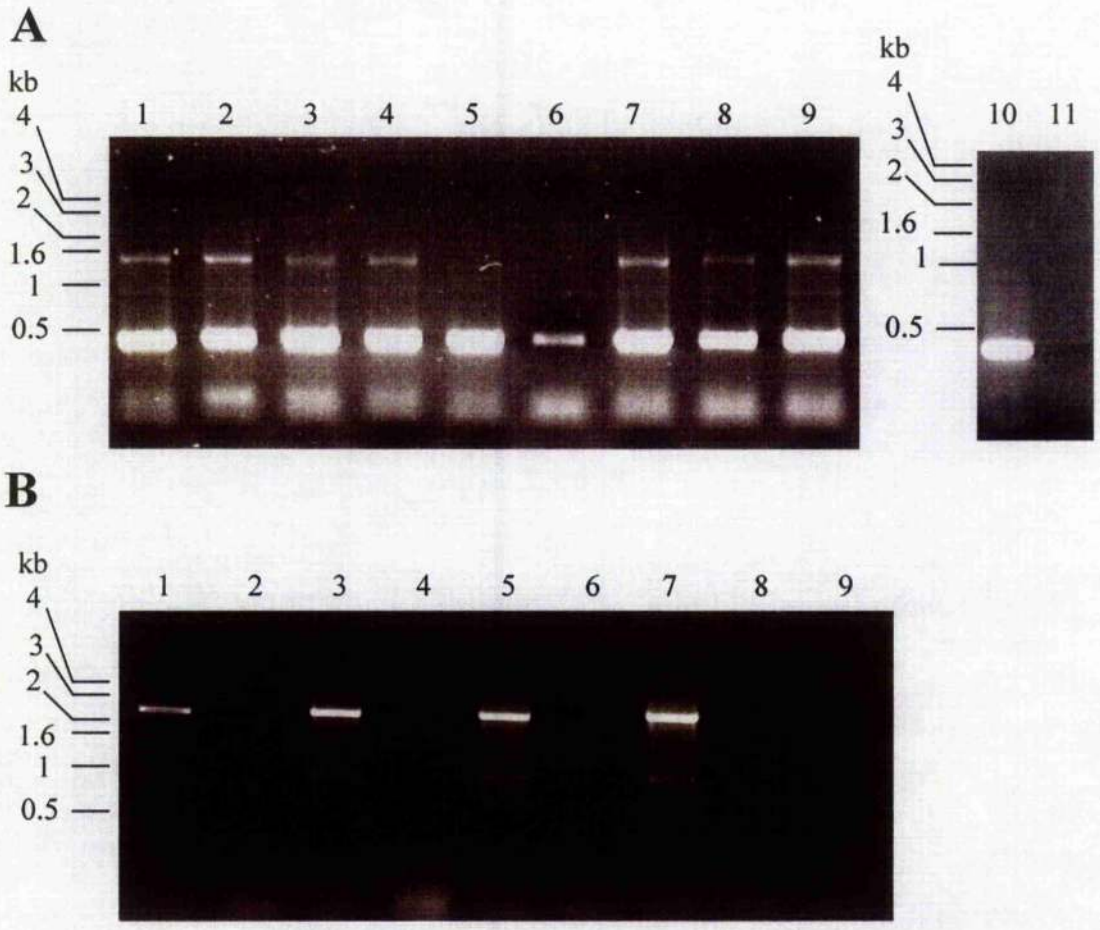


Figure 4.8 Nearly half of all pPB1004 integrations in a *rad51* -/- background were into the 221 bloodstream expression site (BES). **A.** The genomic DNA from each transformant is a valid template for PCR, as PCR amplification of a region of the RNA polymerase I large subunit from the DNA of the transformants gave rise to the expected 450 bp sized product in each case. Lanes 1 to 9 represent template DNA from transformants 1.1, 2.1, 3.1, 5.1, 6.1, 8.1, 9.1, 9.2 and 9.3. Lane 10 represented 427 genomic DNA as a positive control for the primers, and the PCR shown in lane 11 had no DNA added as template for a negative control of the primers. **B.** PCR amplification using primers specific to the hygromycin resistance gene (HYGRO 5' Switch) and the 221 *VSG* gene (VSG 221 N-Term) gave rise to the expected size product of 2.3 kb for integration of the pPB1004 construct into the 221 BES, from the DNA of four out of 9 transformants. The DNA samples are the same as those in A. The pictures depicted are from ethidium bromide stained 0.8 % agarose gels.

4.3 A preliminary study on the *in vitro* activity of nuclear extracts on the 70-bp repeat sequence

To test whether or not a specific nuclease acts upon the *T. brucei* 70-bp repeats or if they have a particular instability, the action of *T. brucei* nuclear extracts upon the 70-bp repeats was

determined *in vitro*. This was performed primarily to assess whether an activity resembling a double-strand endonuclease, a proposed initiator of recombinational *VSG* switching, could be detected (Barry, 1997).

1x 10^{10} ILTat 1.2 trypanosomes were grown in rats to a density of 10^9.ml^{-1}, purified using DEAE cellulose chromatography and a nuclear extract was prepared according to the protocol described by Bell and Barry (1995); a similar extract, prepared from procyclic stage EATRO 795 cells grown *in vitro* was also used (a gift from Michael Ginger). The protein concentration of each nuclear extracts was determined by comparing various dilutions of the extract with a standard curve generated using known concentrations of bovine serum albumin *via* the coomassie protein assay (Pierce). The bloodstream form nuclear extract was at 4 mg.ml^{-1}, whereas the procyclic form nuclear extract was at 3 mg.ml^{-1}.

Substrate DNA for this assay was prepared by digesting pBluescript KSII, and pCRScript containing ~2 kb of 70-bp repeats (pNP101 created by Nicholas Robinson) with *Sca*I. This cuts within the backbone of each plasmid, linearising them and making the ends identical to one another. The linearised pNP101 product contained the 70-bp repeat region towards the centre. Both DNA samples were phenol:chloroform extracted, ethanol precipitated and resuspended in dH_2O, 210 ng of each was incubated with 140 µg of nuclear extract in a 700 µl total volume (200 µg.ml^{-1}) at 37 °C. 100 µl of the reaction was removed at set time points of 0.5, 1, 5, 10, 20 and 40 min and added immediately to phenol:chloroform. These reactions were then phenol:chloroform extracted and ethanol precipitated, resuspended in water, size fractionated on a 0.8% agarose gel and Southern blotted. The membranes were probed with a region of the pBluescript backbone, obtained by doubly digesting the pBluescript with *Sca*I and *Ssp*I and purifying a 300 bp fragment. This was repeated three times for each plasmid, with both bloodstream and procyclic stage nuclear extracts. An example of a hybridized Southern blot of this experiment can be seen in Figure 4.9.

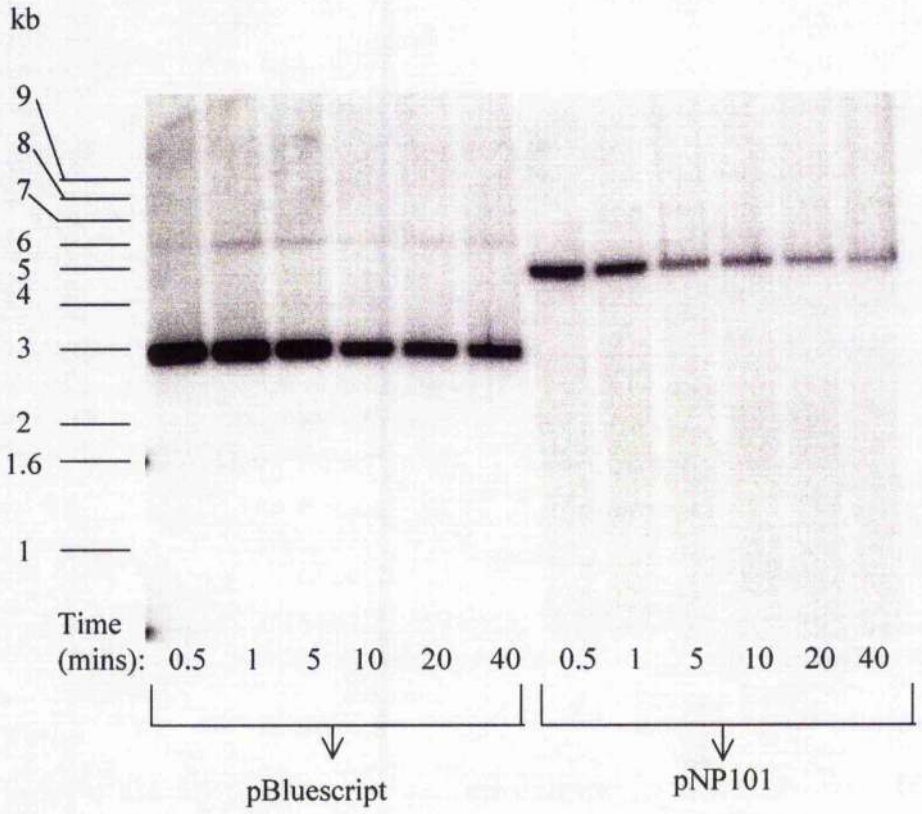


Figure 4.9 An example Southern blot of the assay employed to study the effect of *T. brucei* nuclear extracts upon the 70-bp repeat containing DNA. DNA (either linearised pBluescript or pNP101) was incubated with *T.brucei* nuclear extract for set periods of time. The DNA was then size-fractionated on a 0.6% agarose gel, Southern blotted and probed with a fragment of the pBluescript backbone (see text for details). The example seen here is incubation of the DNA with an ILTat 1.2 bloodstream form nuclear extract. The blot was washed to a stringency of 0.1 x SSC, 65°C.

In all cases, for both plasmids and each extract, the linear DNA was degraded over time and no distinct bands were seen at lower sizes than the linear plasmids as would be predicted by an endonuclease cutting within the 70-bp repeat region (Figure 4.9). However, if a specific endonuclease did act upon the DNA, it is possible that other enzymes, such as exonucleases, may act upon the DNA, making any putative cleaved products undetectable. Also, such an enzyme specific for the 70-bp repeats may cut at different points in the array, leading to a diffuse range of product sizes rather than single discrete bands. Finally, such a nuclease may not be functional in the conditions presented here, and other cellular factors, such as the sub-nuclear localization of the expression site (Navarro and Gull, 2002), may be critical for

regulating the activity. Thus, this experiment by no means rules out the presence of such an activity.

To determine whether the 70-bp repeat-containing DNA sequence is degraded more rapidly than non repetitive DNA, the rate of degradation of the plasmids was quantified using Image-quant analysis (Figure 4.10). The procyclic stage derived extract degraded the plasmid containing the 70-bp repeat sequence more rapidly than the plasmid without the repetitive sequence. With the bloodstream form derived nuclear extract, however, the overall degradation of both plasmids was more rapid, and only a minor, putative, difference in the degradation rates can be seen between the two plasmids. The more rapid degradation of plasmids containing 70-bp repeats could be due to the sequence itself being less stable, perhaps making it more susceptible to enzymes that degrade DNA, or a specific process might act upon the 70-bp repeat sequence. It appears, if the latter is correct, that this is not specific to a particular life cycle stage. This activity might be more evident in the procyclic- derived extract because the general nuclease activity is lower, or because the extract was prepared from a strain that has a very high switching rate (*i.e.* there is a specific mechanism acting upon the repeats that is at a higher concentration in lines with a higher switching rate). This is, however, highly speculative, and there are more parsimonious explanations. The pNR101 plasmid is nearly twice the size of pBluescript, perhaps making it more a more prone to non-specific DNA exonuclease and endonuclease degradation. The preferential degradation of the repeats seen here may also be a general phenomenon seen with many forms of repetitive DNA, and not specific to the 70-bp repeats. One interesting control (not done) could be to determine whether or not a nuclear extract prepared from an organism that does not contain such repeats or undergoes antigenic variation, such as *Leishmania* sp., would have the same activity. If such a closely related organism did not contain this activity, it would lend credence to the idea that the activity seen in the present study was genuine and specific.

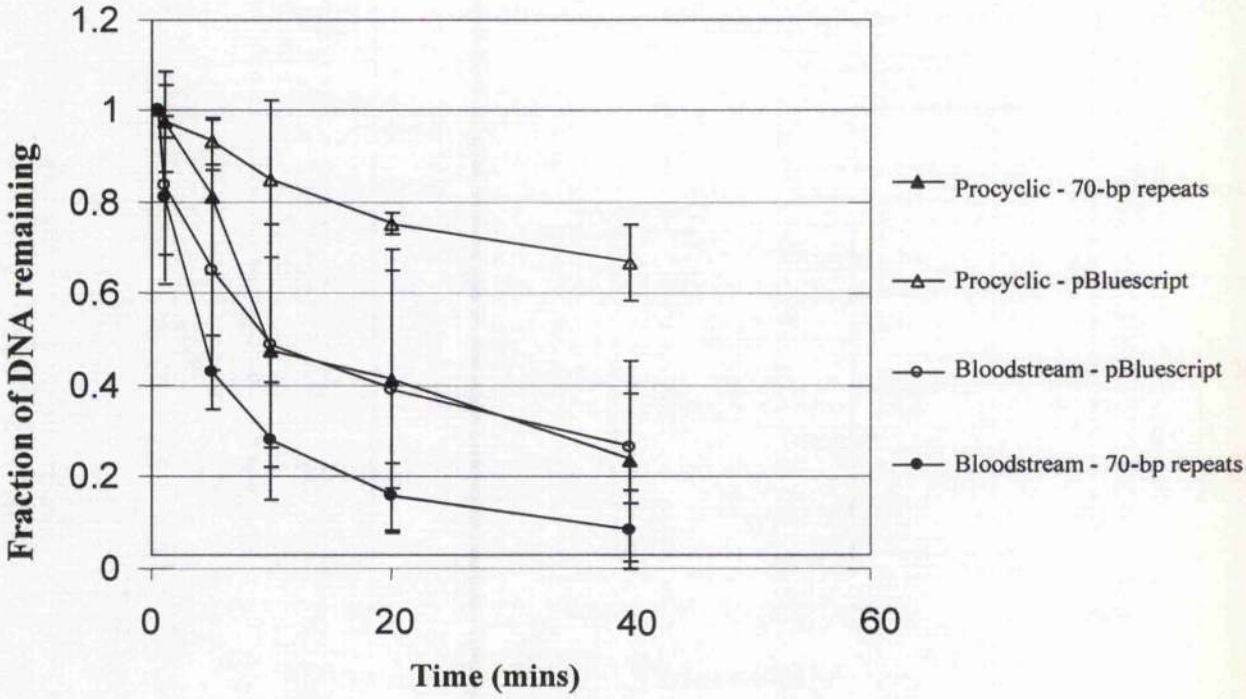


Figure 4.10 A plasmid containing a 70-bp repeat array shows an increased rate of degradation when incubated with a trypanosome nuclear extract. This graph shows the percentage degradation, over time, of plasmids incubated with either a bloodstream form or procyclic form nuclear extract (0.2 $mg.ml^{-1}$). The legend depicts whether or not the plasmid contains the 70-bp repeat array and the life cycle stage from which the extract used was obtained. Time points represent the average % of DNA remaining from 3 experiments, and the error bars represent standard deviation.

4.4 Preliminary attempts to identify proteins that bind to 70-bp repeat sequence

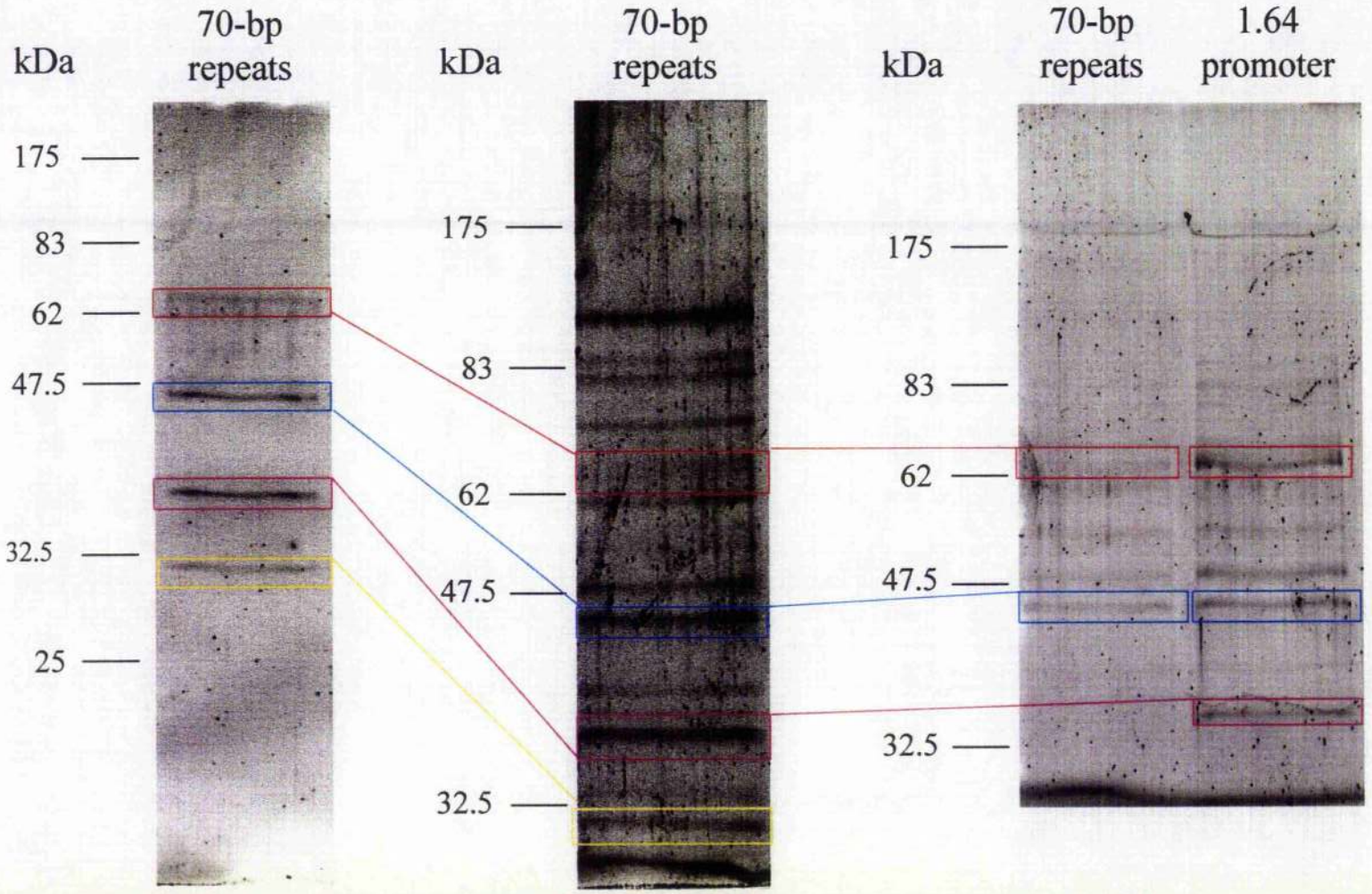
In an attempt to isolate any proteins that specifically bind the 70-bp repeats, DNA precipitation from the nuclear extracts was performed. Biotinylated double-stranded DNA, containing the ILTat 1.22 70-bp repeat region (annealed oligonucleotides: 1.22 70bp Repeat S and 1.22 70bp Repeat AS) was bound to streptavidin-coated magnetic beads (Dynal). The DNA coated beads were then incubated with 5 mg of total protein from an ILTat 1.2 bloodstream form nuclear extract and 250 μg of double stranded competitor DNA (both

digested and undigested pBluescript). Salt washes removed any excess protein, and proteins bound to the DNA were eluted with a high salt concentration (see section 2.15). The eluates were then electrophoresed on a 12% poly-acrylamide gel and stained with Sypro-Ruby (BioRad). This experiment was performed in triplicate and, as a control, DNA containing the ILTat 1.64 metacyclic promoter (annealed oligonucleotides, a gift from Michael Ginger), a sequence of almost identical length to the ILTat 1.22 70-bp repeat region, was treated in exactly the same manner.

Although distinct protein bands were observed in each experiment, those obtained with the 70-bp repeat DNA were neither specific, as they were also seen with the ILTat 1.64 promoter DNA (Figure 4.11), nor reproducible (in the case of band 4). However, matrix-assisted laser desorption/ionisation – time of flight (MALDI-TOF) analysis (Table 4.3) and electrospray ionisation – mass spectrometry – mass spectrometry (ESI-MS-MS) sequencing was performed on the proteins (Table 4.4). In all cases apparent contamination with human keratins was revealed by the sequence but in addition, DNA binding proteins were also detected. The MALDI-TOF analysis was somewhat inconclusive as a lot of peptide fragment masses were shared between the different proteins, indicating either degradation of the samples or contamination from another source. One protein was identifiable from the unique fragments, however. This was derived from band 2, and two fragments showed identity to *T. cruzi* AP endonuclease, confirming the sequencing data. Even with this band, unidentifiable fragments were obtained. However it should hopefully be possible, with further refinement and optimization, to isolate specific binding proteins with this assay; it was, however, outside the time-scale of this study.

This experiment does not exclude the possibility of proteins binding specifically to the 70-bp repeats, as the assay may not be sensitive, or specific, enough. Any specific protein, if associated with the initiation of *VSG* switching, would probably be of low abundance, particularly given that this nuclear extract was prepared from ILTat 1.2, a strain with a somewhat reduced *VSG* switching rate. Unfortunately, it is probably not feasible to obtain nuclear extracts from bloodstream forms of the high switching EATRO 795 strain, as they generate low parasitaemia in rodent hosts. To improve the assay in the future, it would be preferable to use longer arrays of 70-bp repeat DNA, since this may represent a more natural

Figure 4.11 The 70-bp repeat region from the ILTat 1.22 locus binds the same pattern of proteins as the putative ILTat 1.64 metacyclic promoter, indicating the binding of non-specific proteins. The label above each gel indicates the DNA substrate used. The red, blue, purple and yellow boxes respectively represent bands 1, 2, 3 and 4 in the text and Tables 4.3 and 4.4. All proteins were size-fractionated on 12% polyacrylamide gels (each gel represents a separate experiment), stained with Sypro-Ruby for 24 h and visualised on a phosphoimager.



substrate for any specific binding protein, due to abundance of these repeats in the BESs, the proposed site of *VSG* switching initiation.

Band	**Mass/charge (m/z) of peptide fragments after trypsin digestion**
1	1045.5505 1277.6427 1794.7556 2225.1561 2230.2664 2283.2340
2	1045.5504 **1176.6183** 1235.1561 **1268.6186** 1377.3422 1381.6713 1794.7892 2225.1229 2230.2045
3	870.5838 1794.7531 2225.1730
4	1045.5331 1794.7528 2225.1341 2283.1863

Table 4.3 MALDI-TOF analysis of four protein bands isolated from incubation of a trypanosome nuclear extract with the ILTat 1.22 70-bp repeat region. Underlined text indicates peptide shared between all fragments, bold text indicates fragments corresponding to *T. cruzi* AP endonuclease and normal text indicates fragment unique to each band.

Band	Sequence (determined by ESI-MS-MS)	Homologies from genome	Protein name
2	VANAVDTFR	AVNAVDTFR	AP Endonuclease
4	VGDVENSSAPR	VGDVENSPASR	A novel protein with a DNA binding domain similar to the zinc-finger nick sensing domain of human PARP-1

Table 4.4 Tandem ESI-MS-MS sequencing of the trypsin digested proteins isolated from bands 2 and 4 (see text and Figure 4.10 for details). Underlined letters indicate differences between the sequences obtained from the mass spectroscopy and that of the database.

4.5 Conclusion

In this chapter an assay was developed to examine the role of the 70-bp repeats in the duplication of sequence from a chromosomal-internal locus into the active BES. This was performed in the monomorphic Lister 427 strain and the pleomorphic ILTat 1.2 strain. Within the monomorphic line, duplication of sequence from the chromosomal-internal locus to the active BES was observed. This duplication probably involved a short region of 70-bp repeats as the upstream conversion limit, although this was never formally verified. The frequency at which duplication occurred was, unfortunately, too low, and lacked sufficient reproducibility to be utilised for further experimentation. No such gene conversion events were seen when the same assay was employed in the pleomorphic line ILTat 1.2. This could have been for numerous reasons, which are discussed in Section 4.2.6. Although this assay was capable of generating detectable gene conversion events involving the active BES, it needs considerable refinement. A fluorescence-based assay, whereby cells undergoing a conversion event activate the expression of a fluorescent protein detectable by fluorescence-associated cell analysis, may be more appropriate if the sensitivity of detection and switching rate of the trypanosome strain under study were high enough. The gene conversion events detected upon high drug selection when the 70-bp repeat sequence was present were not detected, under the same selection

procedure, when the repeat sequence was replaced with a region of similar length from the active 221 BES CTR. The lack of conversion could be again for several reasons, as discussed in Section 4.2.6. It does, however, emphasize that with such low rates of recombination, the assay is not a feasible method to analyse the role of sequences involved in gene conversion. Attempts were also made to employ this assay in a *rad51* mutant background, but because of the impairment of recombination as a result of this mutation no transformants that had integrated the construct into the aldolase intergenic locus were recovered. Instead, the construct appeared to integrate into non-targeted loci or, at a frequency of approximately 50%, into the active BES. Thus, due to the possible effects of locus upon gene conversion, the assay could not be performed in this background. The nature and relevance of these aberrant integrations are discussed further in Chapter 6.

No detectable 70-bp repeat-specific endonuclease activity within bloodstream or procyclic stage nuclear extracts prepared from pleomorphic trypanosomes was observed in the conditions tested here. This does not discount such an activity, as extensive optimisation and titration of reaction conditions was not performed. A possible susceptibility to nuclease degradation from a nuclear extract of plasmids containing 70-bp repeats was observed. This preference could have been for many reasons other than the presence of the repeats, however, as discussed in Section 4.3.

A method was developed, based upon precipitation of protein from a nuclear extract using streptavidin-conjugated magnetic beads attached to biotinylated DNA, to isolate 70-bp repeat-specific DNA binding proteins. The method was sensitive enough to retrieve DNA binding proteins; however, it lacked specificity and no protein specifically binding to the 70-bp repeats could be identified. Again this does not discount the existence of such a factor, as a thorough examination of binding conditions is required.

CHAPTER 5

THE ROLE OF *RAD51* IN *T.BRUCEI* ANTIGENIC VARIATION IN THE PLEOMORPHIC ILTAT 1.2 LINE

5.1 Introduction

RAD51 is the main recombinase in all eukaryotes. It is the protein responsible for catalysing the invasion of single-stranded DNA into a homologous duplex during homologous recombination in mitotic cells. In trypanosomes, disruption of the gene encoding this protein results in phenotypes consistent with the loss of an important factor involved in homologous recombination and repair of single-strand and double-strand breaks (McCulloch and Barry, 1999). These include an increased population doubling time, reduced DNA transformation efficiency and sensitivity to DNA damaging agents. Also, a reduction in the frequency of *VSG* switching, a mechanism proposed to occur *via* homologous recombination, was detected. This was the first formal identification of a gene involved in antigenic variation.

As those experiments were performed in a monomorphic line (Lister427), *RAD51* provides the first opportunity to search for a difference in the recombination pathways utilised to switch *VSG* genes by the monomorphic *T. brucei* lines, and the pleomorphic lines. A difference in pathways is anticipated, as monomorphic lines switch the expressed *VSG* at a rate similar to that expected for background spontaneous mutation ($1 \times 10^{-6} - 1 \times 10^{-7}$ switches/cell/generation; Lamont *et al*, 1986), whereas the pleomorphic lines can switch much more rapidly (up to 2×10^{-2} switches/cell/generation; Turner and Barry, 1989; Turner, 1997). There is also a mechanistic difference, as monomorphic lines tend to switch *via* transcriptional, *in situ*, means (Liu *et al*, 1985), whereas duplicative transposition predominates in pleomorphic cells (Robinson *et al*, 1999). What role RAD51 might play in these differences is unknown. By mutating *RAD51* in pleomorphic lines, it may be possible to obtain a clearer picture of the role of the protein in the process of antigenic variation.

Homologous recombination pathways other than that driven by RAD51 are present in the trypanosome (McCulloch and Barry, 1999; Conway *et al*, 2002; see Section 4.2.7). It is possible that these pathways play a more crucial role in the catalysis of *VSG* duplicative transposition in pleomorphic lines than in the monomorphic lines. This hypothesis is derived from the observation that short stretches of homology are used for duplicative transposition of *VSG* genes (Campbell *et al*, 1984; De Lange *et al*,1985; Shah *et al*, 1987; Chapter 3) in pleomorphic lines, and a likely candidate for catalyzing these reactions is the homologous recombination pathway exposed by the *rad51* null mutants (Conway *et al*, 2002). It is

alternatively possible that *RAD51* may be even more critical for *VSG* gene switching in the high switching parasites, since duplicative transposition assumes increased importance in at least one pleomorphic line (Robinson *et al*, 1999).

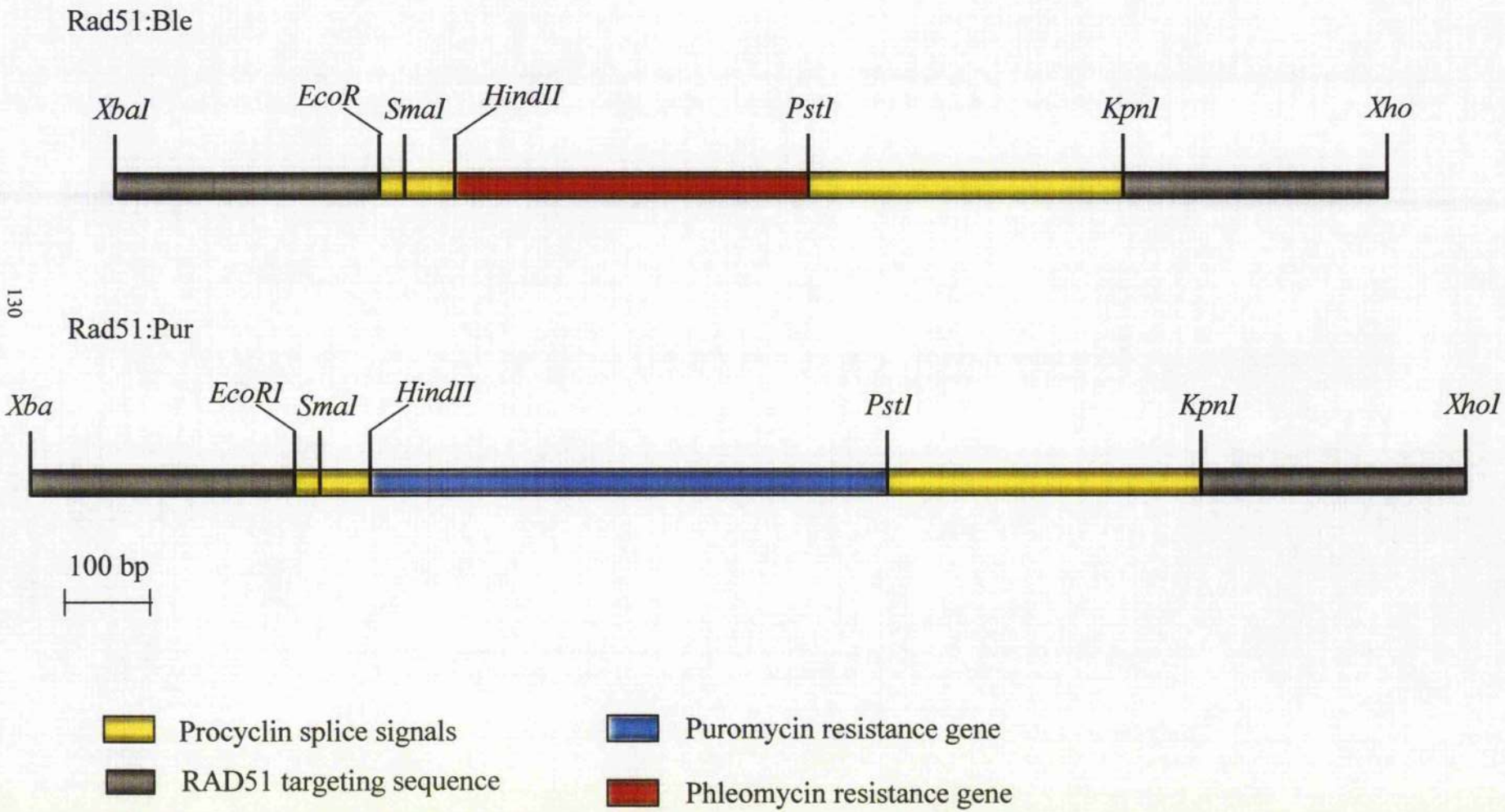
An ideal pleomorphic line for study is the ILTat 1.2 line. This line has been shown to be capable of growing in modified culture medium (Chapter 4; Robinson *et al* 2002), the VSG switching rate is not too frequent (~ 1 in 1 x 10^{-5} switches/cell/generation) but is still in the range associated with pleomorphic lines, and analyses have been performed upon the mechanisms of VSG switching utilised (Robinson *et al*, 1999). In this chapter, heterozygous and homozygous *RAD51* mutants were generated in the ILTat 1.2 line, and analyses of growth and *VSG* switching frequency were performed.

5.2 Generation of *rad51* mutants in the pleomorphic ILTat 1.2 line

The constructs (RAD51::BLE and RAD51::PUR) used originally to generate *rad51* mutants in the monomorphic Lister 427 line were utilised in this study (Figure 5.1) (McCulloch and Barry, 1999). After integration of the constructs, 391 bp of the *RAD51* open reading frame is deleted, resulting in loss of most of the amino acid residues conserved between species, and involved in recombination catalysis. DNA for transformation was prepared by doubly digesting RAD51::BLE and RAD51::PUR with *Xba*I and *Xho*I. The DNA was phenol:choloroform extracted and ethanol precipitated, and approximately 5 μg of DNA was used for each transformation.

Trypanosomes were prepared for the first round of transformations by growing the ILTat 1.2 line bloodstream form (BSF) in HMI-9 medium supplemented with 2.75 % methylcellulose, without allowing the cell density to reach higher than 7.5 x 10^5 trypanosomes.ml^{-1}. This was done in order to reduce the risk of differentiation of long-slender dividing trypanosomes to the non-dividing short stumpy form within the population. These trypanosomes were transformed and selected as in method (*i*) from Section 2.3.2. Variations from this method included alteration of the density of trypanosomes under antibiotic selection from 3 x 10^6 to 1 x 10^7 trypanosomes, in 18.5 ml, over 12 wells. It was only the lower density, however, that gave rise to a transformant. Trypanosomes were selected at 2 μg.ml^{-1} phleomycin and 1 μg.ml^{-1}

Figure 5.1 Restriction enzyme maps of the two constructs utilised to disrupt the two alleles of the *RAD51* open reading frame (ORF). These constructs were obtained from Richard McCulloch and were used previously to disrupt the ORFs of *RAD51* in the monomorphic 427 line (McCulloch and Barry, 1999).



puromycin, respectively, for the constructs RAD51::BLE and RAD51::PUR.

Two independent lines were eventually generated that were resistant, separately, to puromycin and phleomycin. To confirm these as heterozygous for mutation of one allele of the *RAD51* locus, a PCR was performed on genomic DNA extracted from each transformant and from wild-type cells. PCR primers were used that would amplify a region slightly larger than the open reading frame of the *RAD51* locus (primers: Ko 5' and Ko 3'; conditions - 1 cycle of 96°C for 5 min, followed by 30 cycles of 96°C for 1min, 55°C for 1 min and 72°C for 2 min, and followed by 72°C for 10 min; PCRs perfomed in 5% DMSO). As the mutation caused here is not a complete deletion of *RAD51*, the primers still amplify from a mutated allele, but give rise to a PCR product different in size from that of the wild-type locus, due to the insertion of a drug-resistance gene and the associated processing flanks.

Although the PCR produces a single product of the expected size for the wild-type ILTat 1.2 line (Figure 5.2 - lane 1), in the drug-resistant transformants two products can be seen, one corresponding to the remaining wild-type allele and the other, larger product, corresponding to the size expected for the insertion of the respective drug-resistant gene cassette (Figure 5.2 - lanes 2 and 3). Thus, it appears that these two transformant lines hold independent heterozygote disruptions of the *RAD51* locus and these were named *RAD51* +/- P1 (puromycin resistant) and *RAD51* +/- B1 (phleomycin resistant) respectively.

Second round transformations were performed on each of these heterozygous mutants. Construct RAD51::BLE was transformed into the *rad51* +/- P1 line, and RAD51::PUR into the *RAD51* +/- B1 line. The DNA was prepared in a manner identical to that used for the first round transformations and double drug selection was generally performed using 2 $\mu g.ml^{-1}$ phleomycin and 1 $\mu g.ml^{-1}$ puromycin. Because of difficulties in generating double-resistant parasites under these conditions, other selections were attempted using single selection with the same concentration of one drug, and halving the concentrations used for both the single and double drug selections. Trypanosomes used for transformation were routinely grown up in HMI-9 medium supplemented with 2.75% methylcellulose, and again were never allowed to reach a density over 7.5×10^5 trypanosomes.ml^{-1}. Transformation attempts where trypanosomes initially had been grown in a rat, and isolated from the buffy coat after exsanguination of the blood by cardiac puncture, were also performed.

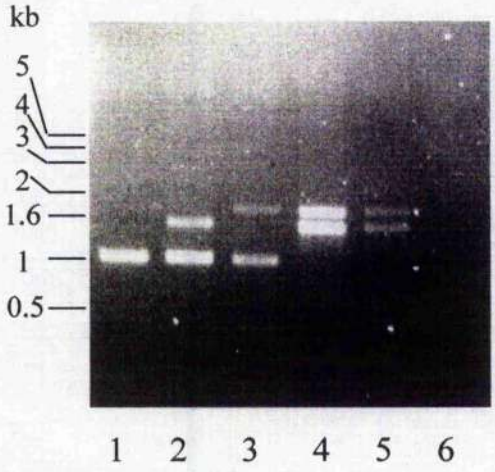


Figure 5.2 PCR analysis of the disruption of the *RAD51* alleles. Genomic DNA was prepared from the single and double drug resistant cells lines generated after transfection with the constructs RAD51:BLE and/or RAD51:PUR. A PCR was then performed upon each of the DNAs using primers, lying outside of the coding sequence, designed to amplify the whole of the *RAD51* ORF, thus detecting the integration of the drug resistant marker into the locus. The PCR products were size fractionated on a 0.8% agarose gel. Lane 1 represents the ILTat 1.2 wild-type, lane 2 the phleomycin resistant line (*RAD51* +/- B1), lane 3 the puromycin resistant line (*RAD51* +/- P1), lanes 4 and 5 represent the lines resistant to both phleomycin and puromycin, *rad51* -/- B1/P2 and *rad51* -/- P1/B2 respectively, and lane 6 is a no-DNA control for contamination.

The recovery of independent double drug resistant transformants was considerably more difficult to achieve than with the first round transformations and, over the course of a year, only two independent, double drug resistant, second round transformants were isolated.

The first such transformant was derived from the *RAD51* +/- P1 line. The transformation protocol was the same as that used for the first round transformation, although many more independent transformations had to be attempted, and several more days on drug selection were required before cells were recovered.

This method did not achieve success with the transformation of the *RAD51* +/- B1 line. The generation of a double drug resistance, second round transformant of this line was achieved by use of method (*ii*) from Section 2.3.2. Here, the *RAD51* +/- B1 line was grown, prior to transformation, in HMI-9 supplemented with 2.75% methylcellulose. 2 x 10^7 trypanosomes were then electroporated and were allowed to recover in 10 ml of HMI-9 for 30 min. These were then harvested by centrifugation and were allowed to recover for 18 h in 130 ml of HMI-9 supplemented with 2.75% methylcellulose. A single transformant was eventually detected after 14 days of selection upon both drugs at the regular concentration.

The difficulty in obtaining transformants is likely to be a general problem with the ILTat 1.2 line, as a low transformation efficiency has been seen with another construct (see section 4.2.3)

For verification of these drug resistant transformants as genuine homozygous disruptions of the *RAD51* locus, the same PCR approach used to verify the heterozygous mutants was performed. Here, the loss of the wild-type band and the presence of two bands, corresponding to the insertion of the two different drug resistance genes into the separate alleles of the *RAD51* locus, was observed (Figure 5.2 – lanes 4 and 5).

To confirm that these lines are genuine *rad51* mutants, an RT-PCR (reverse transcriptase - PCR) approach was taken to confirm the absence of *RAD51* mRNA in the mutants. 2 x 10^7 cells of each heterozygote and homozygote mutant line were grown in culture and RNA was isolated. After treatment with DNAse, cDNA was prepared using random hexamers and Superscript™ II RT. For each cDNA, an RT negative control was also made to control for any genomic DNA contamination. PCR was then performed on each cDNA, and its RT negative control, with primers specific to a region of the RNA polymerase I gene (described previously in Rudenko *et al*, 1996). For the cDNA from each line, a product of the correct size was seen in the RT positive lane but not the RT negative lane, indicating that each cDNA was a functional template and that there was no DNA contamination (Figure 5.3 A). PCR was then performed with primers that amplify the whole of the open reading frame of *RAD51* (Primers: rad51d6 and rad51u3; conditions - 1 cycle of 96 °C for 5 min, 30 cycles of 96 °C for 1min, 55 °C for 1 min and 72 °C for 2 min, and 1 cycle of 72 °C for 10 min; PCRs perfomed in5 % DMSO). With the cDNA from either of the two independent heterozygous lines as a template, a product the size of the *RAD*51 open reading frame can be seen, indicating that full length *RAD51* mRNA is still being produced in these lines (Figure 5.3 B – lanes 1 and 2). With the cDNA from the two independent homozygous lines, however, no product is seen, indicating that these are true *rad51* mutants (Figure 5.3 B – lanes 3 and 4). Unfortunately, reintegration of a functional copy of the *RAD51* gene into the null mutants was not attempted due to time restraints and the envisaged difficulty in achieving successful transformations.

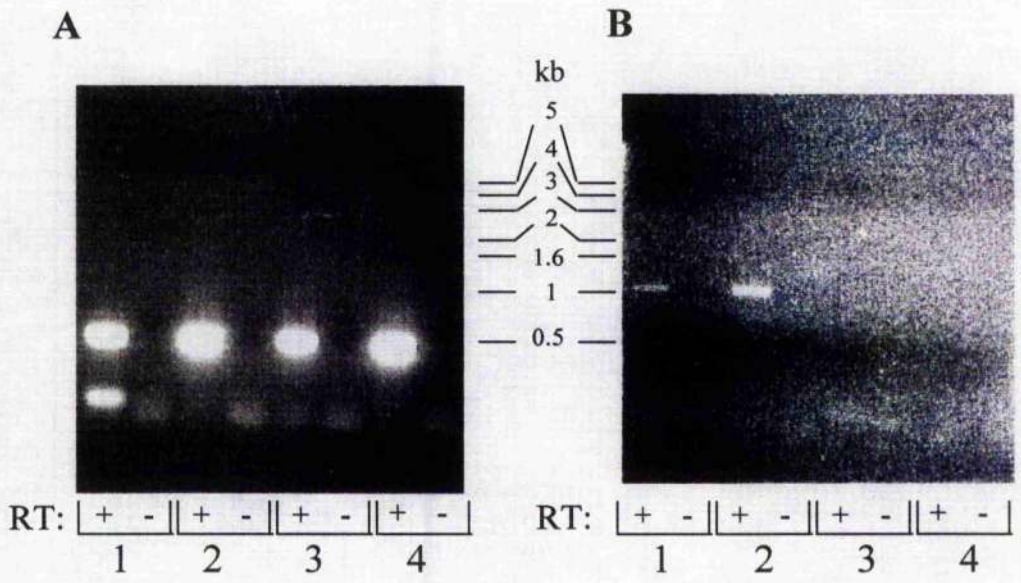


Figure 5.3 Reverse transcriptase-PCR (RT-PCR) confirmation of the disruption of the *RAD51* alleles. **A.** Positive control for the cDNAs. cDNA was made for each independent heterozygous and homozygous *rad51* mutant using random hexamers (Superscript II kit – Invitrogen), along with a RT negative control for DNA contamination. PCR was then performed on each cDNA, and the respective RT –ve controls, with primers designed to amplify a region of the RNA POLI gene. A product of expected size was detected from each of the cDNA containing reactions, but not from the RT–ve controls, thus confirming that each cDNA is suitable for PCR and free from genomic DNA contamination. **B.** Detection of the expression of *RAD51*. A PCR was then performed on each cDNA, and the respective RT –ve controls, using primers designed to amplify the entire *RAD51* coding sequence (Primers: rad51d6 and rad51u3). A product of the expected size was detected in each of the heterozygote cDNA containing reactions but not in the homozygote cDNA containing reactions, nor in any of the RT-ve controls, showing that expression of full length *RAD51* had been eradicated by the integration of the drug resistance markers. The PCR products were size fractionated on a 0.8% agarose gel. In both A and B, the + and – signs indicate whether or not RT was present when making the cDNA. Lanes 1 represents PCR products from the phleomycin resistant line (*RAD51* +/- B1), lanes 2 the puromycin resistant line (*RAD51* +/- P1), lanes 3 and 4 represent respectively the lines resistant to both phleomycin and puromycin, *rad51* -/- B1/P2 and *rad51* -/- P1/B2.

5.3 Recloning and an analysis of plating efficiency

For analyses of growth and switching, it was necessary to reclone the heterozygote and homozygote mutant lines, along with the ILTat 1.2 wild-type line itself. This was required to produce clonal cell lines to reduce the variability of phenotypes possibly associated with a polyclonal population. It was also necessary, for the calculation of *VSG* switching frequencies (see Section 5.5), to assess the plating efficiency of these lines. Each line was grown from a stabilate to a density of less than 2 x 10^5 trypanosomes.ml^{-1}. For each independent line, 25 and

80 trypanosomes were plated out separately over two 96 well plates in a total of 20 ml of HMI-9 medium, supplemented with 2.75 % methylcellulose. The number of wells showing *T. brucei* growth after 21 days revealed the plating efficiency of the cell line. It can be seen that, in each case, the plating efficiency was approximately 10% (Table 5.1), and there appeared to be no difference between the heterozygous and homozygous *RAD51* mutant cell lines. In each case, 3 clones were taken for further analysis.

Trypanosome Line	**No. of positive wells for 25 trypanosomes plated**	**No of positive wells for 80 trypanosomes plated**	**Plating efficiency (%)**
ILTat 1.2 wt	3	8	11
RAD51 +/- B1	2	6	7.8
RAD51 +/- P1	2	9	9.5
rad51 -/- P1/B2	4	4	10.5
rad51 -/- B1/P2	3	8	11

Table 5.1 Plating efficiency of the ILTat 1.2 wild type and the heterozygous and homozygous *RAD51* mutants.

5.4 Analyses of growth

5.4.1 *in vitro* growth

Mutation of *rad51* in monomorphic lines resulted in an increased population doubling time (McCulloch and Barry, 1999), so a similar phenotype was expected from the pleomorphic *rad51* mutant lines. To analyse growth *in vitro*, two clones from each line were grown in medium from stabilate and, after recovery, 5 x 10^5 trypansomes were passaged into 10 ml of fresh medium, resulting in a starting density of 5 x 10^4 trypanosomes.ml^{-1}. Cell density was measured every 24 h for 3 days (after 96 h the number of cells was in decline) with a bright-line haemocytometer (Sigma). The experiment was performed in duplicate for each clone, giving a total of four data sets for each independent line. The results were pooled and plotted

on a semi-logarithmic scale graph (Figure 5.4), and doubling times were calculated (Table 5.2). It appears that the heterozygous mutations had no effect upon growth as they show no visible difference from the wild-type line. The growth of one homozygous mutant line (*rad51* -/- B1/P2) was affected, showing an increased doubling time as expected (Table 5.2). The other line, however, appeared not to have such an apparent growth-related phenotype, at most showing only a very minor increase in population doubling time (Table 5.2).

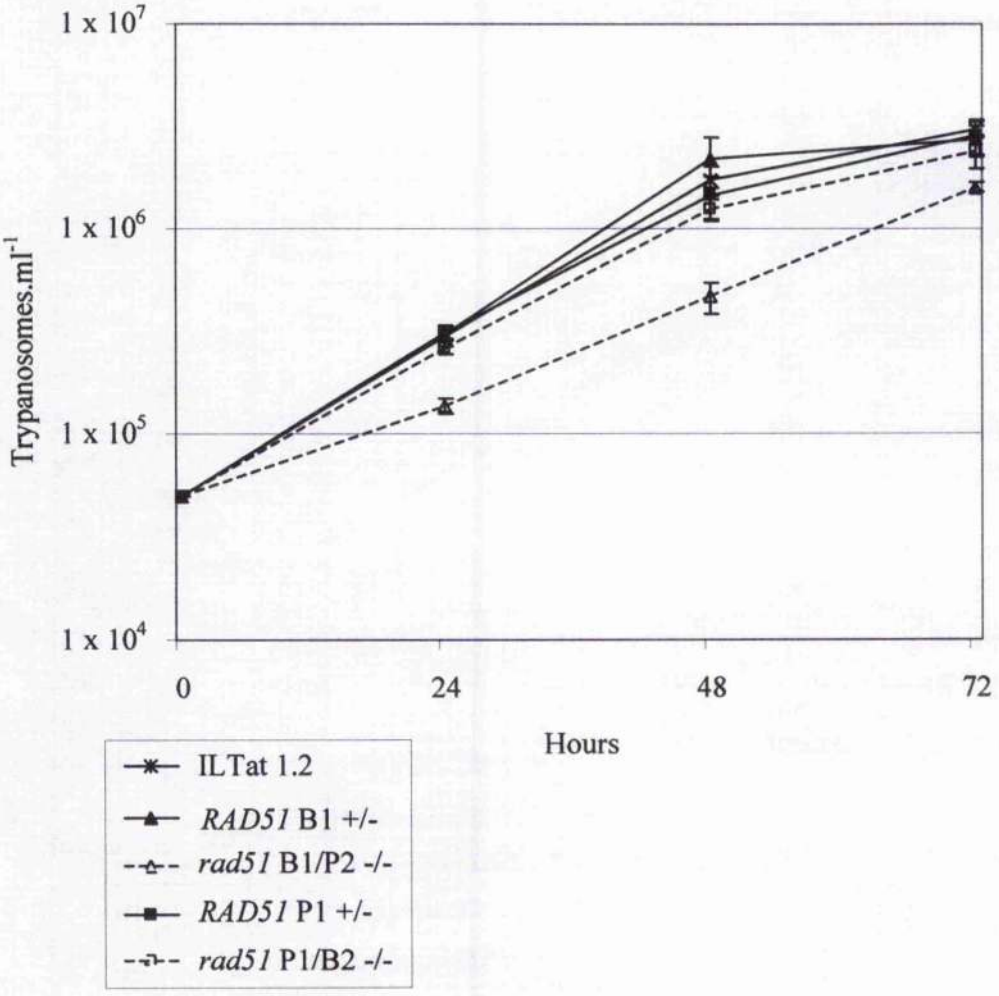


Figure 5.4 *in vitro* analysis of *RAD51* mutant growth. 10 ml of medium was inoculated with 5 x 10^5 trypanosomes and density was measured every 24 h using a bright-line haemocytometer. The data are the means of 4 separate data sets, from 2 clones from each line. The vertical lines represent standard deviations and the strains used are given in the figure key.

The reason for this difference between the null mutant lines could be related to the lack of adaptation to growth in culture medium by the parent ILTat 1.2 line. It is clear that the wild-type ILTat 1.2 line did not grow as well as the monomorphic Lister 427 line and it is possible that the lines under study here contain variations in their ability to grow *in vitro*. To validate this argument, and to confirm further a growth related phenotype, it was necessary to perform an analysis of growth *in vivo*, a more natural habitat for the parasites.

Strain: ***in vitro***	**Average population doubling time (hours)**
ILTat 1.2	9.47
RAD51 +/- B1	8.89
RAD51 +/- P1	9.97
rad51 -/- B1/P2	15.02
rad51 -/- P1/B2	10.4
in vivo	
ILTat 1.2	7.00
RAD51 +/- B1	8.01
RAD51 +/- P1	7.14
rad51 -/- B1/P2	10.49
rad51 -/- P1/B2	12.68

Table 5.2 Population doubling times calculated for ILTat 1.2 *RAD51* mutants and their wild-type progenitor.

5.4.2 *in vivo* growth

in vivo growth was analysed by infecting mice, *via* interperitoneal injection, with 2 x 10^5 trypanosomes previously grown in culture medium, thus providing an initial cell density of approximately 1 x10^5 trypanosomes.ml^{-1} in each mouse. The density of trypanosomes within the blood was then measured every 24 h by extracting a small volume of blood from the tail of each mouse into a heparin-coated capillary tube (Hawksley). 1µl of this blood was diluted in 99 µl of 80 mM ammonium chloride, and the trypanosomes were counted with a

haemocytometer. The experiment was performed in duplicate for each clone of each independent cell line. These results were pooled and plotted on a semi-logarithmic scale graph (Figure 5.5) and doubling times were calculated (Table 5.2). The results from these growth curves demonstrate that the homozygous null mutants have an apparent difference in growth from the heterozygote and wild-type lines. The null mutant grew slower than the heterozygote mutants and wild-type line, with an increase in population doubling time of approximately 2.5 – 5.7 h (Table 5.2). Another interesting growth phenotype observed was that the homozygous mutant trypanosome numbers declined to an undetectable level after 4 days, whereas the heterozygous and wild-type lines continued to grow to very high cell densities. This phenotype is unusual but was also detected in the *rad51 -/-* mutants made in the monomorphic line (R. McCulloch, personal communication). In the monomorphic line this 'early killing' of null mutant trypanosomes was alleviated, however, if the mice were immunosuppressed prior to the cultivation of the trypanosomes. It is perhaps therfore likely that the *rad51* homozygous mutants, showing lower numbers during infection, are more easily cleared by the immune system.

Thus, as expected, the *rad51 -/-* null mutants display a growth phenotype of an increased population doubling time. In fact, both the null mutant lines appeared to grow at similar rates *in vivo*, suggesting that the differences seen during *in vitro* growth are indeed simply a result of their ability to adapt to growth in culture medium.

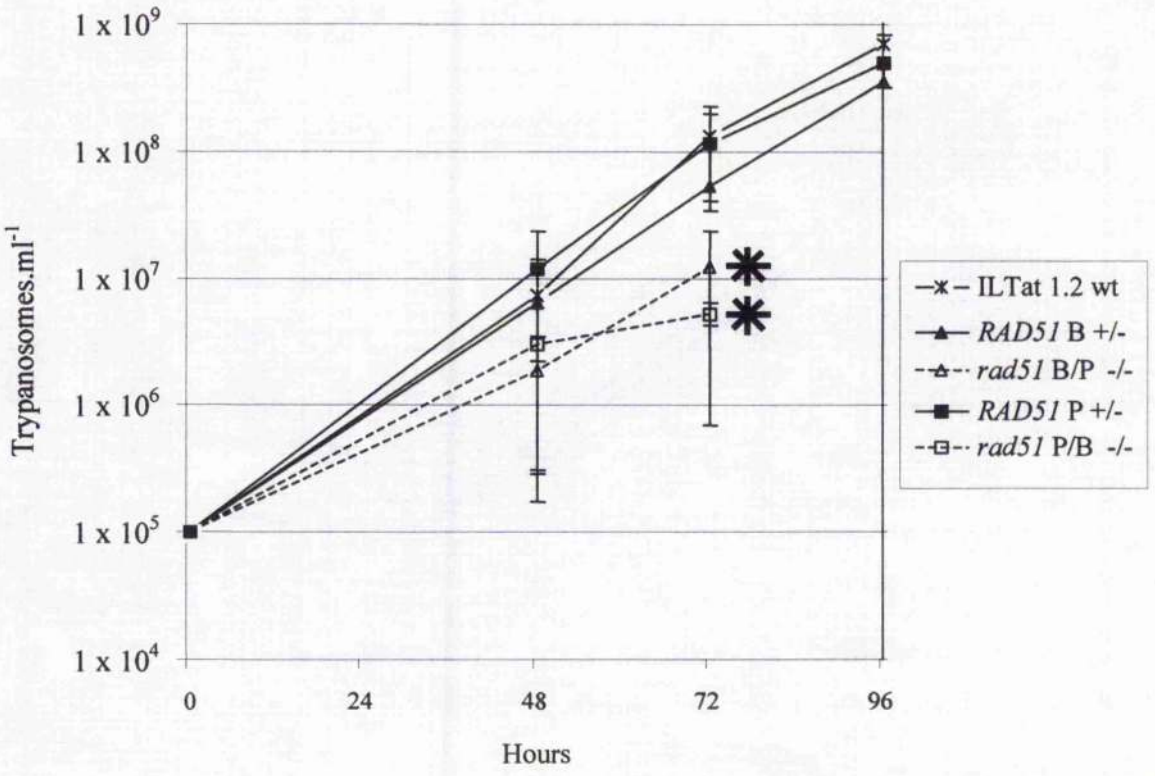


Figure 5.5 *in vivo* analysis of *rad51* mutant growth in mice. Two mice for each line were infected with 2 x 10^5 trypanosomes. Growth was measured every 24 h by extracting blood from the tail, diluting 100-fold in 0.85% ammonium chloride and counting the trypanosome density using a bright-line haemocytometer (Sigma). Data are the means of duplicate infections of each line. The large asterisks represent points in infections after which trypanosomes become undetectable. Vertical lines represent standard deviations and the strains used are given in the figure key.

5.4.3 Analysis of relapse of parasitaemia during trypanosome infection in the rat

In order to see whether the *rad51* null mutant parasites were capable of forming a relapse peak, and hence undergoing *VSG* switching, during infection, rats were individually infected with each line under study. Trypanosomes were initially grown in culture and then diluted to 40 trypanosomes.ml^{-1}. As in the study by Miller and Turner (1981) of trypanosome relapse during the infection of rats with the same ILTat line as used here, 10 trypanosomes from each independent line were used to infect individual rats, *via* interperitoneal injection. For the ILTat

1.2 wild-type, and each independent heterozygous *RAD51* mutant line, three rats were infected. For the homozygous mutant lines, 15 rats were infected, but only three rats showed signs of infection, all from the *rad51* -/- P1/B2 line, detectable by the method described by Herbert and Lumsden (1976). Note that the *rad51* -/- P1/B2 was the strain that grew more rapidly *in vitro*; whether or not this has any significance here is unknown. This indicates that the *rad51* null mutants have a reduced ability to yield infection when a small number of trypanosomes are inoculated. This is probably due to a reduced viability, a phenomenon associated with *in vivo* growth of these mutants in a mouse host (see section 5.4.2). All inoculations that showed no sign of infection were monitored over one month, and no parasites were detectable. For the other infections, parasitaemias were measured from a blood drop obtained from the tail of each rat, by counting the number of trypanosomes on a x 40 magnification field, as in Herbert and Lumsden (1976). Figure 5.6 displays the data for each individual rat and Figure 5.7 displays a summary of the data.

The three *rad51* -/- mutant parasitaemia profiles were distinct from those of the heterozygous mutants and the wild-type. In all three *rad51* -/- mutant infections, trypanosomes were not detected before the ninth day of infection (Figure 5.6). The wild-type and heterozygote mutants, in comparison, mostly gave rise to infections where trypanosomes were detectable by the seventh day, with only one infection having no detected trypanosomes until the ninth day. The growth seen on day nine of the *rad51* -/- mutant infections was transient, with trypanosome numbers falling below detection limits by the following day. In the heterozygous and wild-type infections, however, trypanosomes continued to grow to levels that would soon kill the animal so each rat was euthanised.

Two of the *rad51* -/- mutant infections relapsed again on days 16 and 17, indicating that trypanosomes were not wholly eradicated and were capable of evading the immune system. Taken together, these data (summarized in Figure 5.7), indicate a growth-related phenotype in the *rad51* -/- mutants that was not seen in the heterozygote mutants or wild-type cell line. Unfortunately, the difficulty in growing the *rad51* -/- mutant lines in rats made this method unfeasible for the study of *VSG* switching events in relapse populations, so another method was required (see section 5.5). Nevertheless, immunofluorescence was performed on blood smears from the last day of each infection using antibodies against the VSGs ILTat 1.21, ILTat 1.22, ILTat 1.25, ILTat 1.64, ILTat 1.64, ILTat 1.7, ILTat 1.71 and ILTat 1.73 (see

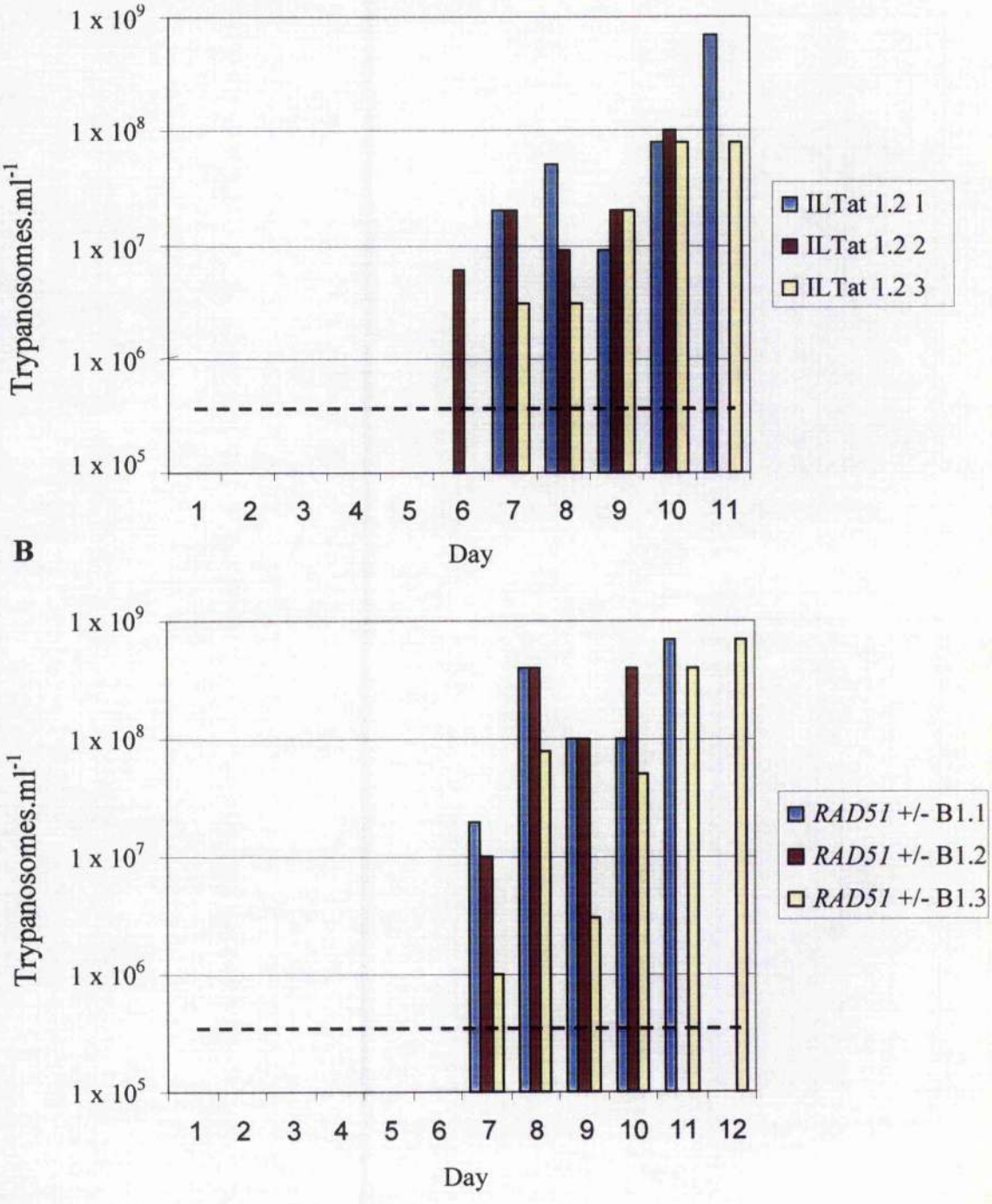
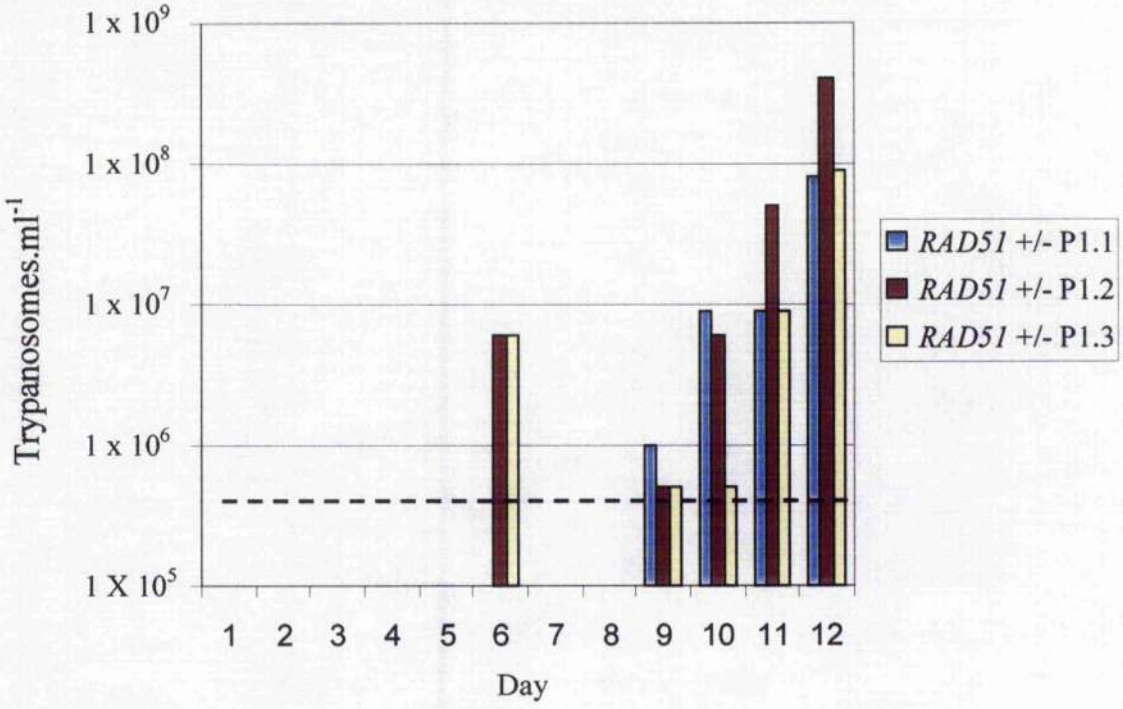


Figure 5.6 Analysis of trypanosome population relapse in rat infections. Ten trypanosomes were inoculated, via interperitoneal injection, into each rat and growth was measured daily using the Herbert and Lumdsen (1976) method. Three independent rats were infected for each line under study. **A.** Growth of the ILTat 1.2 wild-type strain. **B.** Growth of *RAD51* B1 +/- **C.** Growth of *RAD51* P1 +/-. **D.** Growth of *rad51* P1/B2 -/-. The dotted line in each graph represents the lowest detectable number of trypanosomes. See text for more details.

C



D

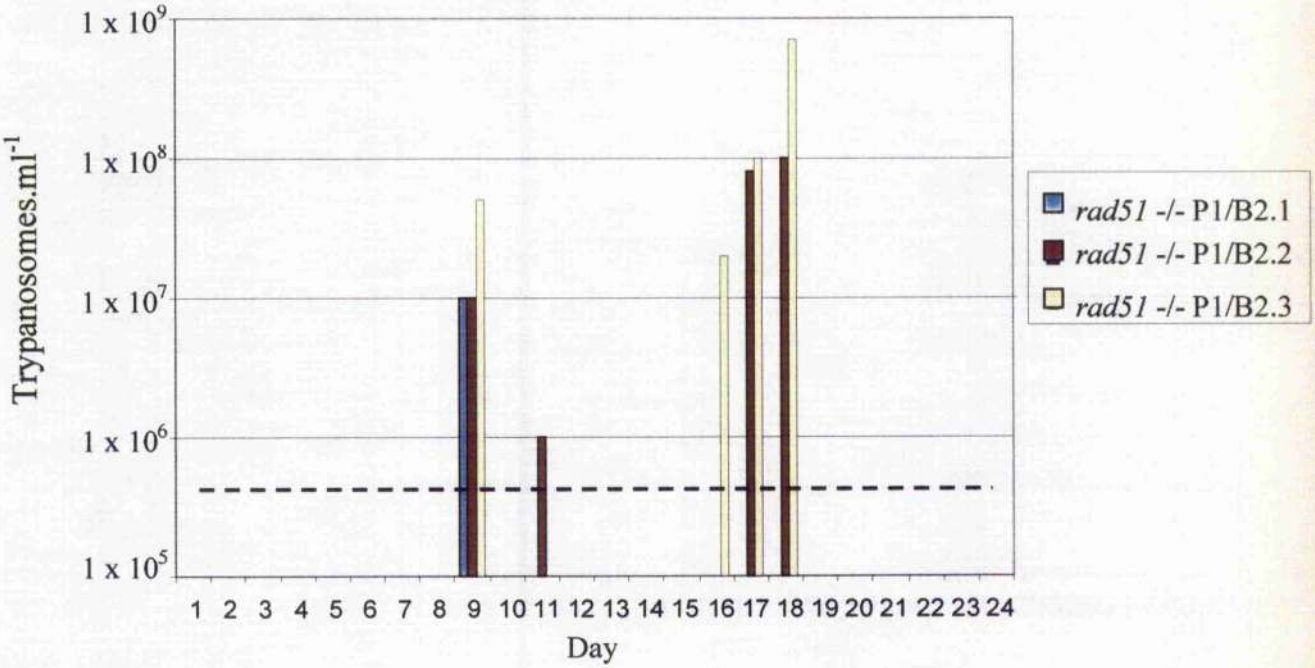


Figure 5.6 continued.

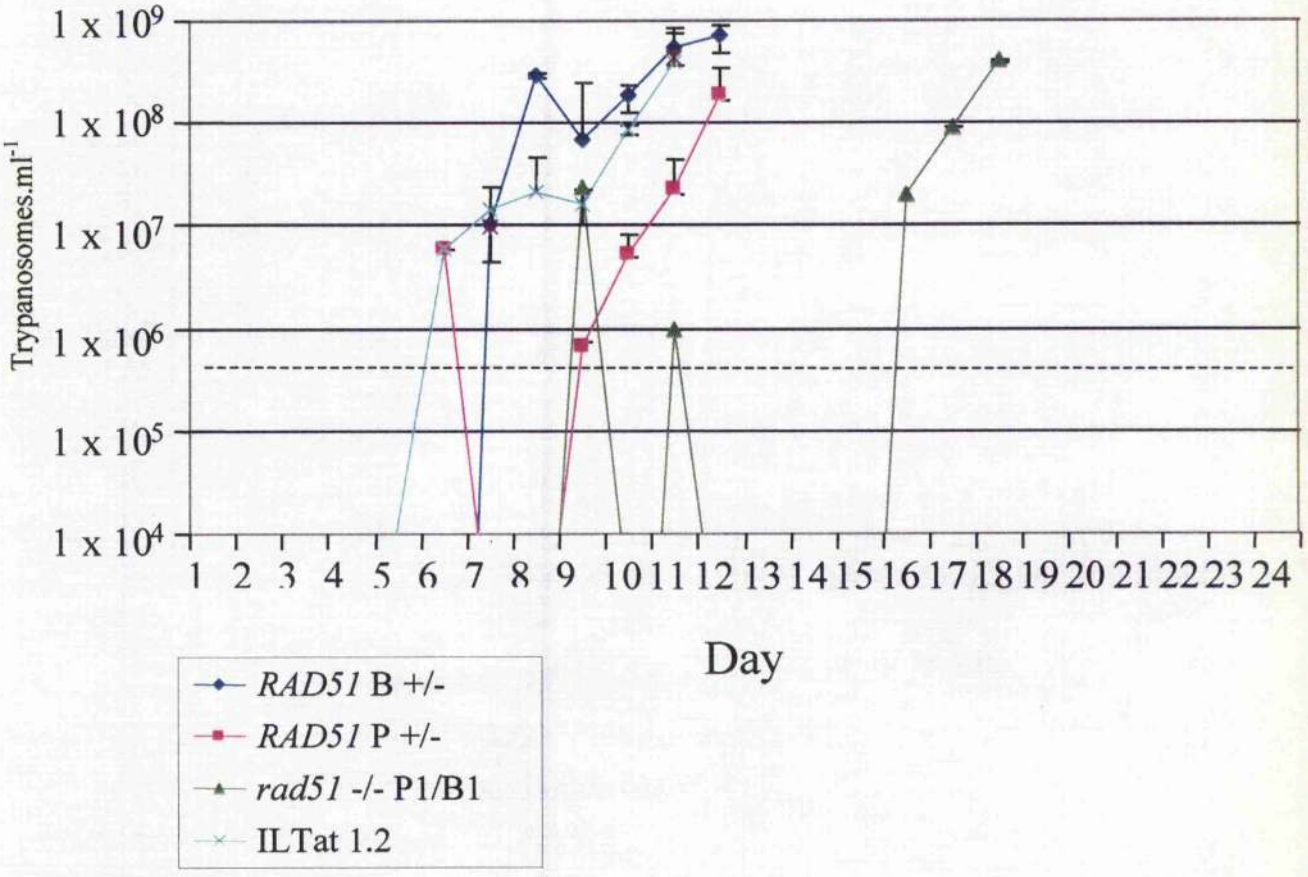


Figure 5.7 A summary of the growth of the ILTat 1.2 wild-type and *RAD51* heterozygous and homozygous mutants in rats inoculated with 10 trypanosomes. A lack of error bars on some data indicates that there was only one rat showing a parasitaemia on that day. The dashed line indicates the lowest number of trypanosomes detectable by this method (Herbert and Lumdsen, 1976)

Section 2.4). Despite the fact that these are known to be activated early in infection (Miller and Turner, 1981; Robinson *et al*, 1999), it was impossible to detect any trypanosomes with these VSGs on their surface.

5.5 Analysis of *VSG* switching

In order to determine whether or not the *VSG* switching frequency was reduced to the same extent in pleomorphic *rad51* null mutants as had been seen in the monomorphic mutants, each clonal line was grown in culture from stabilate. Approximately 2 x 10^5 trypanosomes of each line were then injected into independent mice; at least two mice were infected for each clonal line. Three days of parasite growth was allowed for the trypanosomes to reach a sufficient density, at which point the mice were cured by treatment with cymclarsen. Thus, mice were generated which were immune to the predominant VSG being expressed for each line.
Each clonal line was grown, again in culture, so that a total number of 4 x 10^7 trypanosomes could be injected into each immune mouse. Twenty-four hours after reinfection, the mice were exsanguinated by cardiac puncture, buffy coats were prepared from 0.4 ml of the blood, and the trypanosomes isolated from each 0.4 ml were plated over 96 wells in a total volume of 20 ml of HMI-9 medium, supplemented with 2.75% methylcellulose. Trypanosomes were then given up to one month to grow in the 96 well dishes. The number of wells showing growth were counted throughout and the *VSG* switching frequency was calculated as follows:
The number of wells containing living trypanosomes, representing the number of clonal switched variants surviving immune clearance, was multiplied by five, as the 0.4 ml of blood used for their isolation represents only a fifth of the total blood volume of the mouse. This figure was then multiplied by ten, as the plating efficiency of each strain is approximately 10% (see section 5.3). This figure then represents an estimate of the total amount of trypanosomes within the mouse at the time of bleeding, so this was altered to account for the number of population doublings that would have occurred during the 24 hours of growth. This figure was then converted into number of switched cells per 1 x 10^6 cells infected into the mouse. These numbers are shown in Table 5.3.
The switching levels appear to be generally equivalent for all the lines, which is in contrast to the 2-130 fold reduction seen in the monomorphic *rad51* -/- mutants (McCulloch and Barry, 1999). To support the finding that there was no difference in *VSG* switching rates in these pleomorphic *rad51* mutants, 2-paired t-tests were performed. These are summarized in Figure 5.8. Where the switching frequency of each mutant line was compared with the data for the ILTat 1.2 wild-type line using the 2-paired t-test there is no statistically significant difference

between any of the mutant lines and the ILTat 1.2 wild-type (P = 0.05). Also, not shown on the graph, the p-values generated upon comparison of the data for the homozygous null mutants with the data for their progenitor heterozygous mutants yielded no statistically significant difference in *VSG* switching levels (*rad51* -/- B1/P2: P = 0.169; *rad51* -/- P1/B2: P = 0.585).

Notably, the switching level of the *rad51* -/- B1/P2 null mutant was statistically very close to being significantly different from that of the ILTat 1.2 wild-type, however, any potential difference here represents an increase in the switching level of the mutant line, the opposite from what was predicted from the monomorphic mutant lines. It would be necessary to use a very large sample size if any small differences in *VSG* switching were to be detected by this method as the assay shows a high level of variation. It is clear, however, that there are no large differences between the *VSG* switching rates in the *rad51* ILTat 1.2 mutant lines and the wild-type line and this is in direct contrast to what was seen in the monomorphic lines (McCulloch and Barry, 1999).

5.6 Conclusion

In this chapter, independent *RAD51* heterozygous and homozygous mutants were generated in the pleomorphic ILTat 1.2 line. During *in vivo* growth, the null mutant trypanosomes displayed a strongly altered growth phenotype, where their population doubling time increased by approximately 4 h. This corresponded to the growth phenotype of *rad51* null mutants in the monomorphic Lister 427 line (McCulloch and Barry, 1999). It is possible that this *in vivo* growth deficiency also gave rise to a lengthened infection of rats, indicating that, although growth appeared to be attenuated, the trypanosomes were still capable of evading the immune system.

A major difference between the monomorphic lines and the pleomorphic lines was seen when *VSG* switching frequency was analysed. In the pleomorphic *rad51* -/- null mutant lines, a reduction in switching frequency was not detected as had previously been observed, by the same assay, in the *rad51* -/- null mutant monomorphic lines. Extensive differences in passaging history may potentially lead to differences between trypanosome lines in terms of growth and recombination, potentially allowing the *rad51* -/- mutants to up-regulate back-up

Trypanosome Line	Number of positive wells per 96 / mouse	Number of switched cells per 1 x 10^6 cells
ILTat1.2	32	5
	27	4.2
	7	1.1
	11	1.7
RAD51 +/- B1	52	8.1
	66	10.3
	14	2.2
	26	4.1
	3	0.5
RAD51 +/- P1	9	1.4
	4	0.6
	29	4.5
	36	5.6
rad51 -/- B1 P2	51	15.9
	44	13.8
	12	3.7
	26	8.1
rad51 -/- P1 B2	42	13.1
	11	3.4
	8	2.5

Table 5.3 The data generated for the analysis of *VSG* switching levels in the ILTat 1.2 *RAD51* mutants and their wild-type progenitor. See text for details on how switching rates were calculated.

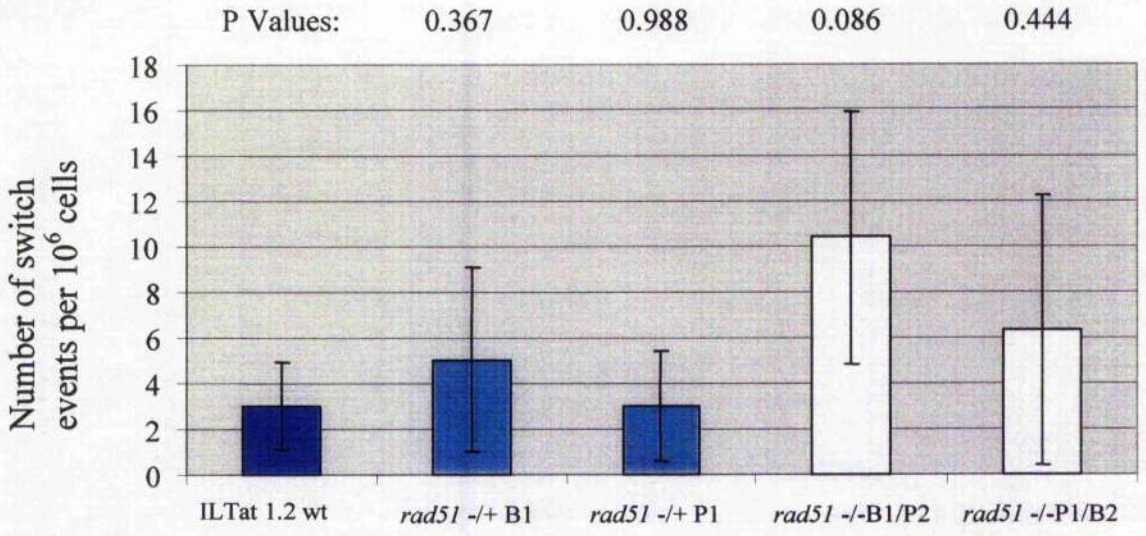


Figure 5.8 Analysis of *VSG* switching frequencies of the ILTat 1.2 *rad51* mutants. Switching frequencies were analysed as described in the text (see also Table 5.3). Above each column is a p-value that was generated by comparison of each data set with the ILTat 1.2 data set, using a 2-sample t-test. Neither the heterozygote or homozygote *rad51* mutants were statistically significantly different from the ILTat 1.2 wild-type (P = 0.05). Each data set is depicted as an average with standard deviations represented by the vertical error bars.

recombination pathways to increase the *VSG* switching frequency. It appears unlikely that this has happened in this case as differences in passage histories between the lines are minimal. Also, each line was cloned prior to analysis to reduce any possible discrepancies.

One way of confirming the presence of back-up recombination pathways is to test the mutants for sensitivity to DNA damaging agents. If pathways have been up-regulated that are capable of repairing induced DSBs in a *RAD51*-independent manner, then the cell lines would not appear sensitive to DNA damaging agents as the monomophic *rad51* -/- mutants were (McCulloch and Barry, 1999). This experiment would be of particular interest for the *rad51* P1/B2 -/- line which did not have a striking reduction of growth *in vitro*.

No re-introduction of the *RAD51* of the gene to the null mutants was performed to demonstrate that the growth phenotypes are due to the loss of *rad51*. In the time available this was deemed unfeasible due to the difficulty of transforming this line. As no difference in *VSG* switching frequency was demonstrated, reintegration of the gene into the null mutants also seems unnecessary, as there is no phenotype to alleviate. It appears, then, that the pleomorphic line may be utilising a pathway for *VSG* switching that is independent of *RAD51*, potentially

demonstrating the first genetic difference between *VSG* switching in a pleomorphic line and a monomorphic line.

CHAPTER 6

DISCUSSION

6.1 Introduction

The aim of this thesis was to study the mechanism of duplicative transposition of *VSG* genes in *Trypanosoma brucei*, with particular emphasis on elucidating a difference between the rapid switching pleomorphic trypanosomes and the infrequently switching monomorphic trypanosomes. The approaches taken were several fold: A detailed analysis of the conversion limits of duplicative activations in a pleomorphic line was performed, an assay to study the role of the 70-bp repeats in gene conversion events was created, attempts at identifying proteins interacting with the 70-bp repeats were made, and the role of RAD51 in *VSG* switching in a pleomorphic line was investigated.

6.2 *VSG* gene conversion and short sequence homology

The analyses of the independent activations of the telomeric ILTat 1.22 *VSG* gene demonstrated that the short (115 bp) 70-bp repeat region, located 1.4 kb upstream of the *VSG* gene, was utilized consistently as the upstream conversion limit of the duplicative transpositions. This confirmed the previous mapping analyses performed by Matthews *et al* (1990) for activations 1- 3, whereas the upstream conversion limits for activations 4 and 5 were first mapped, and then confirmed at the sequence level as for activations 1–3.
Upon examination of the sequence derived from the expression linked copy (ELC) of each activation it was clear that, despite having a short stretch of homology with which to recombine, the extent of the conversions varied. The reaction did appear to be regulated by homology, as duplication involved homologous sequence (*i.e.* 70-bp repeats from the basic copy (BC) and the active bloodstream expression site (BES)), and two independent activations appeared to be identical (activations 1 and 2). Differing lengths of the ILTat 1.22 70-bp repeat region, however, were duplicated into the expression site and the shortest stretch of homology used (*i.e.* the smallest amount of the ILTat 1.22 70-bp repeat DNA duplicated) was 42 bp.
The use of short stretches of homology involving 70-bp repeats to catalyse the duplicative transposition of *VSG* genes is not unique to ILTat 1.22 as, by restriction enzyme mapping, it appears that conversion during the activation of ILTat 1.64 utilises a 70-bp repeat region that is shorter than a single repeat unit. These data also agree with descriptions of single switch

events where as little as 24 bp (Shah *et al*, 1987) and 46 bp (Campbell *et al*, 1984) of perfect homology at the upstream conversion limit can be detected between the 70-bp repeats of the BC and the BES. Thus, this is not an uncommon feature of *VSG* gene duplication. Attempts were made to map the upstream conversion limit of an activation of the chromosome internal ILTat 1.71 gene to a sequence that appeared to be a degenerate 70-bp repeat, but this region was not the site of conversion. It is unknown, however, whether or not there is a 70-bp repeat region upstream of this site that is used for conversion. This would require cloning further upstream but, despite much effort, this was not achieved within the time scale of this study. In relation to the use of short 70-bp repeats as the upstream conversion limit of *VSG* duplicative transpositions, it is interesting to note that in a genomic survey of 56 internal *VSG* genes, of those that had 70-bp repeat regions upstream (~85 %), none of the repeat regions were greater than 266 bp or contained two or more full repeat units. In fact, the average length was 91 bp. If this is a common feature of *VSG* genes, and short repeats are utilised for telomeric *VSG* gene conversion, maybe they are also used for the activation of internal *VSG* genes. In fact, even when extensive numbers of 70-bp repeats are present in the donor sequence Shah *et al* (1987) demonstrated, when studying a single switch event, that only 24 bp of perfect homology between the recombining repeats was used. One important caveat of this hypothesis is that at least two-thirds of these internal *VSG* genes, by the criterion of open reading frame length, are incomplete. It could be that the arrays studied here are atypical, with there being arrays of intact *VSG* genes containing more extensive 70-bp repeats within the genome. As the incomplete genes may use mosaic gene formation for activation, their 70-bp repeats might now be functionally redundant and therefore have degenerated in number and sequence. This seems unlikely, as it would require some mechanism that would operate to select entire gene arrays to undergo, or be more prone to, mutation. It seems more likely that active *VSG* genes that mutate into pseudogenes would do so randomly and be accidentally selected by genetic drift or population bottlenecks (these *VSG* genes containing mutations, frame shifts or premature stop codons are possibly not classical pseudogenes, as it is questionable as to whether any of them have evolved from ancestrally active genes. The fact that part of these genes may be activated by mosaic gene formation in an active expression site also defies definition as a pseudogene). It is interesting to speculate that this lack of integrity

of *VSG* genes is the state of all internal *VSG* arrays and mosaic gene formation is extremely important in antigenic variation.

It could be argued that recombination using only short stretches of homology might be an uncommon, or even secondary mechanism, for activating *VSG* genes and the most frequent mechanism uses larger stretches of homology. In fact, a mechanism pairing a small portion of the expression site 70-bp repeat region over the entire ILTat 1.22 70-bp repeat region, but allowing mismatches that are corrected afterwards in favour of the expression site sequence (as there are differences in the ELC sequence to the ILTat 1.22 70-bp repeat region), might also be able to explain these activations. This, however, seems less likely, due to a bias in the positioning of the nucleotides that differ from the ILTat 1.22 70-bp region. If pairing occurred over the whole ILTat 1.22 70-bp region, which is a more likely length for efficient recombination reactions, then one might expect differences to occur over the length of the conversion, but what is observed is a higher proportion within the 5' end of the cloned ELC sequence. This is what is expected if invasion by the 70-bp repeat region from the expression site had occurred over a short, perfectly homologous, region of the repeat and DNA synthesis was primed in the 3' direction leading to a duplication of the ILTat 1.22 locus. Also, during MAT switching (McGill *et al*, 1989) and transformation with linear DNA in yeast (Leung *et al*, 1997), mismatch correction was in favour of the resident, unbroken donor strand. A gradient in the conversion of mismatches in favor of the donor sequence close to the DSB which decreases as a function of distance away from the break, however, has been detected in meiosis and mitosis (reviewed in Paques and Haber, 1999). But this distance is on the scale of kilobases and is not relevant to the ~100 bp seen to be involved in the ILTat 1.22 conversions. Alternative hypotheses leading to the same result include the short region of homology including mismatches but being corrected in favour of the ILTat 1.22 locus (leading to a model where not only are short regions of homology sufficient, but also heterology is tolerated – see below), or duplicative transposition being initiated downstream of the *VSG* and the reaction terminating using a short stretch of homology within the 70-bp repeat region. The latter of these hypotheses became less likely when the downstream conversion limits were identified (see below). One other possibility is that the presence of a base-pair difference between the ELC and BC does not reflect the end of a heteroduplex intermediate but rather the end of MMR in favor of the ILTat 1.22 sequence. Of course it is impossible, as we do not have

the sequence of the BES prior to conversion, to predict whether the duplicated ILTat 1.22 70-bp repeat sequence was created primarily by extension from an invading 3' strand or by conversion of mismatches generated by the heteroduplex created during strand invasion. It does not however detract from the fact that the level of homology must still have been very short, in fact it is possible that the level of homology used between the two sites was smaller than that duplicated. That is, those events which displayed longer stretches of duplication may have initiated in the upstream region of the repeat, using a short block of homology, and the downstream sequence has been converted via DNA synthesis as opposed to mismatch correction.

In *T. brucei*, it has recently been shown that including heterology in a fragment of DNA used for integration reduces transformation efficiency. If the targeting sequences are 11% divergent from the homologous region in the genome, there is a 93.4% reduction in transformation efficiency and just 1% divergence yields a 2.8-fold reduction (Bell and McCulloch, in press). These data are comparable with the results of similar experiments in other organisms, where increasing the number of mismatches results in an exponential decrease in recombination rates (Datta *et al*, 1997; Elliott and Jasin, 2001). This regulation of recombination in eukaryotes seems to be mediated by the mismatch-repair (MMR) system, and to some degree the recombination machinery itself (Datta *et al* 1997; Elliot and Jasin, 2001). As the percentage of non-homologous base-pairs between the GT and AT rich regions (Aline *et al*, 1985; Chapter 3) of different 70-bp repeats is almost always greater than 10% (this level of divergence is also seen in the $(TRR)_n$ repeat tract, however is less definable due to the variation in length), a *VSG* switching mechanism resembling that of DNA transformation would have reduced efficiency. The length of the 70-bp repeat unit is also variable, making homologous pairing over more than one repeat difficult unless some looping of the DNA occurs. It is interesting to note that mutation of genes involved in the MMR system does not influence *VSG* switching but does increase the transformation of heterologous sequences in comparison with wild-type (Bell and McCulloch, in press). These experiments, however, were conducted in a monomorphic line where there is not the same association of conversion with the 70-bp repeats (although it does occur), so it is possible that in monomophic lines the initiating event of switching regulates the rate and later processing by the MMR system has little detectable impact upon this frequency, as the sequences involved may be less divergent than the 70-bp repeats. It is unknown what

role the MMR system may have in pleomorphic lines, where mismatches may be tolerated to a greater extent in order to utilise the 70-bp repeats and also produce an increased *VSG* switching rate, though it seems that the increased switching rate in the pleomorphic lines is not solely due to a deregulation or suppression of the MMR system. It is important to note, however, that although the MMR mutants still displayed a drop in transformation efficiency even with only a 1% sequence divergence, the decrease was not as extreme as in the wild-type cells. Thus a system that does not rely on these MMR factors can also regulate the success of recombination reactions based upon sequence. In yeast, it has been shown that, between sequences showing greater than 10% divergence, recombination is severely depressed in a manner independent of the MMR system (Datta *et al*, 1996; Datta *et al*, 1997). This is most likely due to the inability to form a stable base-paired recombination intermediate, as there are too many mismatches between the sequences. Maybe it is this undetermined route of regulation, or no mechanism other than the recombination machinery itself, that regulates *VSG* switching.

One proviso should be placed on our interpretation of the effect of mutations on VSG switching. The assay used to detect the frequency of *VSG* switches may not have been sensitive enough to identify any differences between the wild-type cells and the mutants. If *VSG* switching is occurring at a background level then any increase may be too subtle to detect by this assay.

In some pleomorphic lines the switching rate can be as high as 10^{-2} to 10^{-3} switch/ trypanosome/ generation. It is therefore unlikely that the MMR system which regulates DNA transformation is the same as that which regulates *VSG* switching involving the 70-bp repeats. If the initiating event of *VSG* switching was a DSB, then repair presumably must occur by utilising another *VSG* locus (sub-telomeres often interact ectopically; for a review, see Barry *et al*, 2003), unless sister chromatids are present during S-phase of the cell cycle. If recombination is aborted frequently, due to heterology, there may be sufficient loci for repair to occur eventually. DNA transformation occurs approximately 100-fold less efficiently with 11% sequence divergence. Thus, if 100% of theoretical DSBs occurring during *VSG* switching could be repaired (in yeast up to 90% of a HO endonuclease-induced DSB can be repaired (Inbar and Kupiec, 1999)) with perfectly matched sequences, then if reactions instead involved 70-bp repeats (sequences divergent by over 10%) and MMR operated upon them,

then at least 100 heteroduplex intermediate formations, on average, would need to be initiated to allow for a single successful *VSG* duplicative transposition. Failure to repair the DSBs would lead to a large number of cell deaths or a large variation in the length of individual cell cycles. *VSG* switching therefore does either not utilise the 70-bp repeats and/or the MMR system in its regulation, or a DSB is not the initiating factor and homologous recombination is not the driving force behind antigenic variation. As we know that 70-bp repeats are involved in the duplicative transposition of *VSG* genes (in the ILTat 1.2 line) and MMR seems to have no influence upon *VSG* switching (in monomorphic lines), it seems more likely that a recombination pathway which is not constrained by heteroduplex rejection by the MMR system is catalyzing the switching. Porter *et al* (1996) claimed that, when perfectly matched homologous sequences recombine, they generate long stable heteroduplexes, as seen by long conversion tracts with single mismatches being corrected. Divergence in the sequences leads to shorter conversion tracts and reduced levels of recombination. They suggest that stable recombination intermediates between divergent sequences occur as a result of extension of the heteroduplex intermediate by replication following strand invasion, rather than branch migration in the opposite direction. That is, during homeologous recombination, an invading 3'-end pairs with only a short region of perfect homology which is enough to prime DNA synthesis, thus leading to an extension of perfectly homologous paired DNA, stabilising the intermediate. Evidence for this is derived from those authors' plasmid-chromosome intermolecular spontaneous recombination assay, which revealed short regions of homology (3 – 50 bp) between the donor and recipient at the 5' and 3' ends of the conversion tracts. This pattern was seen also in an experiment involving intrachromosomal long inverted repeat stimulated recombination (Tran *et al,* 1997). Tran *et al* (1997) also found that defects in the MMR system had little or no effect upon the previously mentioned assay, which is predicted as most of the heteroduplex is formed from newly synthesized DNA and would not contain any mismatches (other than those incorporated during the DNA synthesis). If this was the case for *VSG* switching, then there would be no role for the MMR system in terms of assessing the recombination intermediates for heterology.

As different studies employ different assays to assess the effect of divergent sequences and MMR upon recombination frequencies, it is perhaps not surprising to find that the influence these factors have varies from assay to assay. The differences seen between assays might

reflect MMR influencing only certain pathways of recombination, allowing others to escape its anti-recombinogenic mechanisms. Perhaps the likelihood that diverged sequences will undergo recombination depends on the initiating lesion and the DNA organisation. Datta *et al* (1997) found that different sequences, even when donor and recipient are perfectly identical, can recombine at much different frequencies to each other (approximately a range of an order of magnitude). Thus, it appears that factors other than sequence identity are critical in determining the rate at which recombination takes place. One postulated role for the anti-recombinogenic action of MMR on homeologous sequence is to reduce recombination between diverged repeats. Recombination between such regions in the genome could lead to chromosomal reaarangements such as inversion, deletion or translocations. In fact, malignancies arising in MMR (*msh2*) deficient mice were hypothesised to occur through such rearrangements (Abuin *et al*, 2000). If the 70-bp repeats are used for the gene conversion events involved in the duplication of *VSG* genes, then it is likely that if MMR was involved it would have a similar anti-recombinogenic effect as the 70-bp repeats are divergent. It is again hard to imagine that such a negative effect could be occurring upon cells experiencing a high rate of *VSG* switching. Perhaps the 70-bp repeats are divergent to reduce the risk of spurious recombination between them, allowing only the directed process of *VSG* switching.

6.3 *RAD51*-independent recombination in *VSG* switching

It seems that certain homologous recombination reactions initiated by only short stretches of homology can escape the anti-recombinogenic influence of MMR (Tran *et al*, 1997). These reactions are strikingly similar to those seen for the ILTat 1.22 duplicative transpositions and it is interesting to contemplate that the recombination pathway driving the conversion of the ILTat 1.22 *VSG* gene is not the same as that which drives normal double-strand break repair and integration of DNA into the genome during transformation. In the yeast *S. cerevisiae*, a pathway of DSB repair termed break-induced replicaton (BIR) is available. Repair by this pathway is initiated by a DSB, followed by the invasion of one of the ends of this break into a homologous sequence elsewhere in the genome. Extensive DNA replication then occurs, often to the end of the chromosome, resulting in the conversion of all of the DNA downstream of the break to the sequence which synthesis was initiated upon. It became apparent, during the

course of this investigation, that the duplicative transpositions of the ILTat 1.22 *VSG* gene appeared very similar in nature to the described BIR reactions. Through cloning of the sub-telomeric region (*i.e.* the region between the *VSG* gene and the telomere tract) of the ILTat 1.22 BC gene, and subsequent restriction enzyme analysis, the length of conversion of each activation extended at least as far as the telomere tract, and could have continued until the very end of the chromosome. It is therefore likely that conversion initiated in the 70-bp repeats and continued to the chromosome end as in BIR.

Conversion to the chromosome end is not the only similarity to BIR: the upstream homology requirements are also very similar. In *S. cerevisiae*, BIR with as little as 33 bp of homology on either side of the DSB can occur 60% as efficiently as a reaction involving over 1 kb of homology (Ira and Haber, 2002). Interestingly, yeast *RAD*52 is essential for both classical gene conversion and BIR, whereas genetic inactivation of *rad51*, *rad54*, *rad55* or *rad57* can abolish gene conversion but not BIR (Signon *et al*, 2001). Mutations in *rad50*, *rad59* and *tid1* however, did not significantly affect gene conversion or BIR, except when coupled with *rad51* or *rad54* mutations, where both gene conversion and BIR are eliminated. It appears that a *RAD51*- and *RAD54*- independent pathway of BIR is present that relies upon *RAD59*, *TID1* and *RAD50*. In fact Rad51 appears to be detrimental to these reactions, because in wild-type cells the incidence of successful recombination events occurring at short stretches of homology is lower than that in a *rad51*Δ strain by approximately 50% (Ira and Haber, 2002). The Rad51-independent pathway allows successful strand invasion leading to new DNA synthesis 10% of the time with a stretch of homology as short as 29 bp on either side of the break, whereas mutants of this pathway (i.e. *rad59* null mutants) require approximately 70 bp of homology on either side of the break to allow the same level of recombination. Thus, in S. *cerevisiae*, recombination events operating upon short stretches of homology tend to be performed without Rad51 and occur via BIR. As these recombination events appear to be identical to those utilised to perform the duplicative transpositions of the ILTat 1.22 gene, it can be predicted that the mutation of Rad51 would have no effect on, or would raise the frequency of, *VSG* switching via this pathway. However, in *rad51* mutants generated in the monomorphic Lister 427 strain a 2-130 fold decrease in *VSG* switching was detected (McCulloch and Barry, 1999) showing that the *RAD51*-dependent pathway of *VSG* switching is critical in that strain. In the present study, mutation of *rad51* in the ILTat 1.2 line revealed

that a reaction similar to that described in yeast as *RAD51*-independent is perhaps involved in the duplicative transposition of a least some *VSG* genes.

Both independent *rad51* null mutants displayed an increased population doubling time in comparison with the ILTat 1.2 wild-type and the independent *rad51* heterozygote mutants *in vivo*, as expected from the *rad51* mutant doubling times in the monomorphic line. Only one *rad51* mutant, however, showed an increased population doubling time in comparison with the wild-type *in vitro*. This could be due to a general lack of adaptation of the ILTat 1.2 line to growth in culture. Alternatively, the independent *rad51* null mutant which doubles at approximately the same rate as the wild-type may have up-regulated an alternative pathway that allows it to deal with the genotoxic stresses encountered in culture medium, but not those suffered in the mammalian host. It is known that, in trypanosomes, there is a *RAD51*-independent pathway capable of performing homologous recombination and therefore potentially the repair of DSBs (Conway *et al*, 2002).

The *in vivo* growth phenotype observed in the mice was unusual because not only did the null mutant trypanosomes increase in number at a slower rate, but they also became undetectable (via the Herbert and Lumsden (1976) detection method) after 72 hours, whereas the wild-type and heterozygous mutants continued to expand. This phenomenon was also seen in the Lister 427 mutants, but the sudden disappearance of the trypanosomes was alleviated by immunosupression of the mice prior to infection (Richard McCulloch, personal communication). Thus, it seems likely that the decreased growth rate caused by the mutation of *rad51* allows the host's immune system to eradicate the trypanosome population more rapidly, as the threshold for antibody-mediated killing may be reached earlier during null mutant growth due to less trypanosomes.

The ILTat 1.2 *rad51* null mutants also exhibited a reduced ability to establish a detectable infection in rats inoculated with only ten trypanosomes. Of the 3 out of 15 attempts that did establish an infection, however, 2 did form a relapse peak, so presumably the null mutants are capable of *VSG* switching. Due to the low frequency of infections displayed by the null mutants, examination of the *VSG* variants from relapse populations in rats was not deemed a viable means to analyse *VSG* switching.

The frequency of *VSG* switch events in the ILTat 1.2 wild-type line and *RAD51* mutants was measured in a manner similar to that used to measure the frequency of switch events in Lister

427 *RAD51* mutants (McCulloch and Barry, 1999). The only difference was that the medium was supplemented with 2.75% methyl-cellulose and the calculated switching frequency was adjusted to accommodate a 10% plating efficiency. No statistically significant difference in switch frequency could be detected between any of the lines ($P > 0.01$). This is in sharp contrast to the Lister 427 strain, where mutation of *rad51* caused a 2-130 fold decrease in *VSG* switching. Thus it appears, as predicted by the DNA homology requirements of the ILTat 1.22 and ILTat 1.64 duplicative transpositions, that the ILTat 1.2 line predominantly uses a RAD51-independent mechanism for *VSG* switching. It is possible that there is an effect on *VSG* switching that is too subtle to be detected by this rather insensitive assay. It can be said, however, that there is a difference between the ILTat 1.2 strain and the Lister 427 strain. It is not known what effect these mutations and growth in culture have had on the pleomorphicism, but by visual assessment stumpy-form trypanosomes did appear to be present at high population densities, although this was not formally demonstrated. Also, the *VSG* switching rate was in the range previously determined for this strain and was higher than that measured by the same assay for Lister 427.

It appears then, that there are at least two mechanisms of switching *VSG* genes by recombinational mechanisms, as predicted from the analysis of general recombination pathways in Lister 427 *rad51* mutants. These are *RAD51*-dependent and *RAD51*-independent, and appear to have different roles in *VSG* switching depending on the strain being analysed. The Lister 427 line predominantly uses the *RAD51*-dependent pathway, whereas the ILTat 1.2 line uses the *RAD51*-independent pathway.

As the available evidence suggests that the ILTat 1.2 line uses the 70-bp repeat region as the upstream conversion limit for the duplicative transposition of *VSG* genes, it is interesting to speculate that the *RAD51*-independent pathway is more applicable to catalyzing these recombination events (see above). In fact, a *RAD51*-independent pathway would explain the preponderance of BIR-like reactions seen catalyzing the ILTat 1.22 switches. If extrapolated, this model provides a simple explanation of the prevalence of telomeric *VSG* genes being activated more frequently than internal genes in the ILTat 1.2 line (Robinson *et al*, 1999). A reaction of this sort would require only an interaction of one end of the recipient recombining duplex, unlike the internal genes, which would require both. Thus, the BIR reactions catalysed by the *RAD51*-independent mechanism may be more parsimonious and therefore more likely

to occur. One possibility is that the 70-bp repeats are not commonly used for switching in the ILTat 1.2 strain and those analysed in this and other studies are a biased or unrepresentative sample. This can only be determined by further analysis of *VSG* switching events in this, and other, strains.

Although the *RAD51*-independent BIR pathway model provides a means of explaining the activation of the telomeric *VSG* genes in the ILTat 1.2 line, can it be applied also to the activation of internal genes? A simple one-ended invasion in this case would presumably lead to a duplication of a large region of another chromosome, leading to the disruption of the expression site. Therefore, homologous regions within or downstream of the *VSG* genes involved would have to interact to complete a recombination event. In yeast, a pathway known as single-strand annealing (SSA) can join two homologous sequences flanking a unique sequence resulting in the deletion of this unique sequence (Paques and Haber, 1999). This reaction has been shown to require as little as 29 bp of homology and the authors suggest that the lower limit is somewhere between 5 and 29 bp (Sugawara *et al*, 2000). Moreover, *RAD51* is not required for this reaction either (Ivanov *et al*, 1996; Sugawara *et al*, 2000), although it is dependent on the strand annealing proteins Rad52 and Rad59, with the evidence suggesting that Rad59 enhances Rad52 strand annealing (Davies and Symington, 2001). In fact, a study involving intraplasmid recombination between inverted repeats led to the conclusion that the most likely explanation for the detected *RAD51*-independent events was a coupling of BIR with SSA (Ira and Haber, 2002). Another means of completing a reaction initiated by BIR is a reaction that uses only 1 to 3 bp of homology and is reminiscent of nonhomologous end-joining (Kraus *et al*, 2001). It seems, then, that *RAD51*-independent BIR initiated events can be coupled to other homologous or non-homologous recombination pathways with little homology requirements. Thus, if the activation of an internal *VSG* gene was initiated by a *RAD51*-independent event it is likely that it could be completed even if the extent of homology between the downstream regions of the genes involved was low. Homologies in the region of *VSG* genes encoding the C-termini of the proteins have been demonstrated, but they are not extensive (Rice-Ficht *et al*, 1981).

RAD51-independent homologous recombination reactions in trypanosomes seem to involve short stretches of homology, sometimes interspersed with mismatches (Conway *et al*, 2002).

Although that study assayed merely the integration of exogenous DNA, it is possible that the same machinery is available for DSB repair and *VSG* switching.

In the present study, attempts to integrate a construct (pPB1005) in Lister 427 *rad51* null mutants were unsuccessful, with none of the resulting nine transformants having integrated the DNA into the targeted intergenic region between the two copies of the aldolase gene. Aberrant integrations are seen in *rad51* null mutants (Conway *et al*, 2002). However the level of non-targeted integrations, but not the level of transformation efficiency, was greater for this construct than for one designed to target the tubulin array. This could be due to the differences in target copy number, *i.e.* the construct designed to integrate in the tubulin array had more opportunity to integrate. Less loci exist for the construct used in this study, although sequences homologous to the active BES were present within it and these appeared to be used at a relatively high rate, approximately 50% of the time. This could have been for one of several reasons, with either the sequences (the 70-bp repeats or the co-transposed region) being more recombinogenic or the active BES being more susceptible to recombination. An alternative possibility is that there is a lot more 70-bp repeat sequence present in the expression site than other sequences within the genome that are homologous to sequences within the construct. An increased number of targets may lead to an increased probability of recombination, as it is known that sequences that are not terminal within the transforming DNA can be used to catalyse recombination in yeast (Inbar and Kupiec, 1999).

6.4 Models of *VSG* gene duplicative activation

Compiling the *VSG* switching data from this study and others, and from studies on recombination in yeast and other organisms, it possible to construct models for *VSG* switching (Figures 6.1 and 6.2). Here, a break in the DNA would occur within the 70-bp repeats upstream of the *VSG* gene in the active BES. This break is currently conjectural and problems with this hypothesis abound. One is that 70-bp repeats are in many places in the genome, so how might a system operate to ensure the break occurs in only the active expression site (or perhaps any expression site, as *VSG* gene conversions do occur into silent BESs (Myler *et al*, 1988; Aline *et al*, 1989)). Navarro and Gull (2002) recently demonstrated that the active BES is located in a sub- nuclear body containing at least RNA Pol I, and is

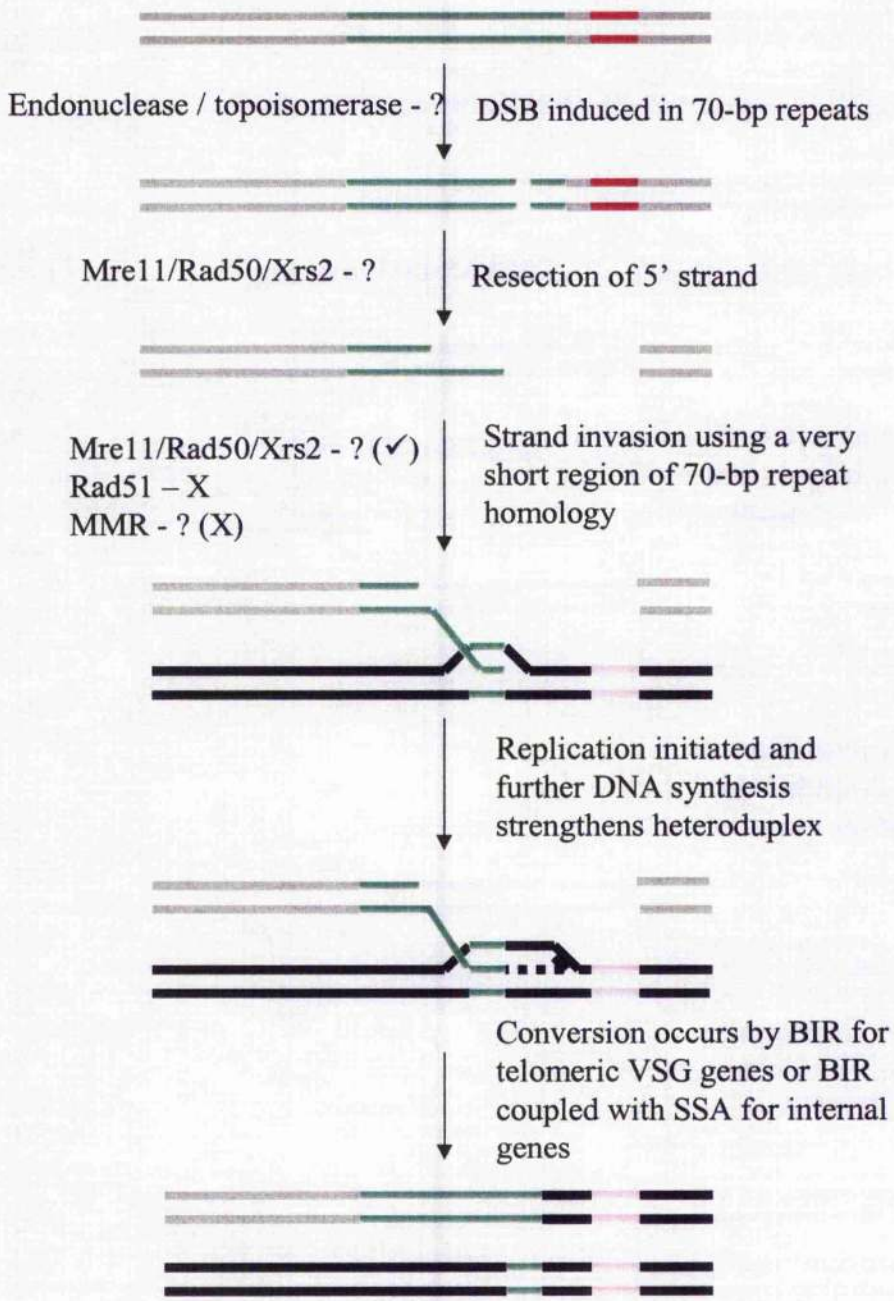


Figure 6.1 A model for *RAD51*-independent *VSG* gene duplicative activation. The factors postulated to be involved in the process are listed at each stage on the left. A tick represents a role for the protein and cross indicates no role, those marks in brackets indicate a prediction. The *VSG* genes involved are represented in red and lilac, the 70-bp repeats are in green, the gray duplex is the expression site and the black duplex is the donor sequence.

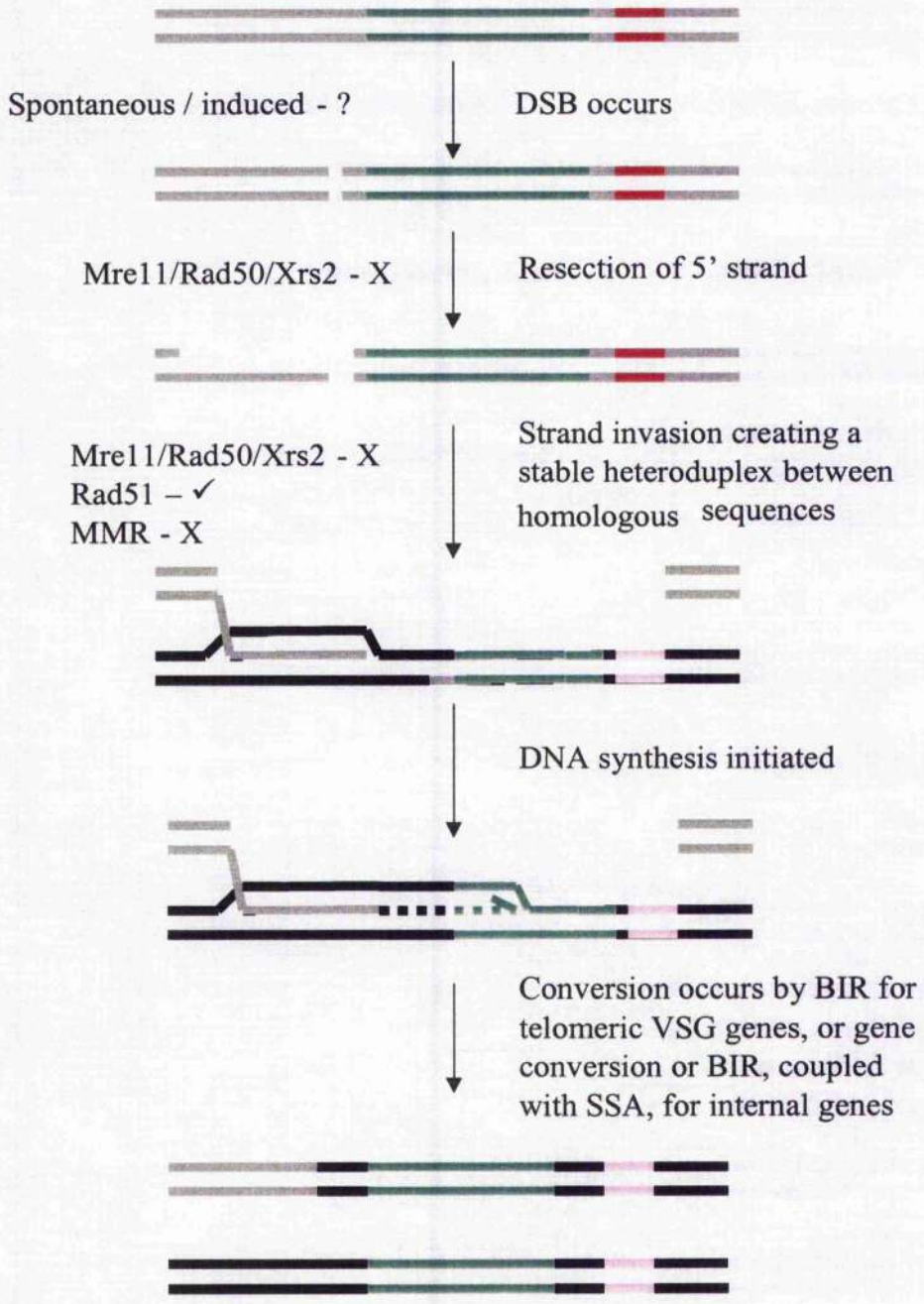


Figure 6.1 A model for *RAD51*-dependent *VSG* gene duplicative activation. The factors postulated to be involved in the process are listed at each stage on the left. A tick represents a role for the protein and cross indicates no role. The *VSG* genes involved are represented in red and lilac, the 70-bp repeats are in green, the gray duplex is the expression site and the black duplex is the donor sequence.

separate from the nucleolus. It is possible that this body provides a means of isolating the active expression site for cleavage of the 70-bp repeats, thus initiating a *VSG* switch. It is therefore critical to identify other factors that are held within this expression site body. One other possibility is that all of the 70-bp repeat sequences other than those in the expression site are held in some form of chromatin that does not allow access of the *VSG* switching machinery as the active *VSG* chromatin is preferentially degraded by single-strand endonucleases (Greaves and Borst, 1987). It is difficult to imagine how an enzyme system could operate to generate single cleavages in order to initiate a *VSG* switch when acting upon a repetitive structure. That is, surely any potential cleavage sites would be present throughout the array, making it difficult to create a single break. It is possible that such an enzyme could be of very low abundance which, along with isolation of the active BES, could lead to only a single break being made in the repeat tract. A low abundance of the enzyme(s) could also explain the difficulty in obtaining 70-bp repeat specific binding proteins via affinity purification. One other means of creating a DNA break in the 70-bp repeats is by an intrinsic (*e.g.* DNA secondary structure) or extrinsic (*e.g.* RNA polymerase or 70-bp repeat specific binding protein) block to a replication fork. Subsequent collapse of stalled forks can lead to a recombinogenic structure that could be used to catalyse *VSG* switching. From this study it appears that 70-bp repeat containing DNA is less stable than non-70-bp repeat containing DNA when incubated with a trypanosome nuclear extract, although the data are highly tentative. If this is the case, then it is possible that the sequence forms a secondary structure or is more recombinogenic, which could possibly influence *VSG* switching. Such intrinsic initiation of *VSG* switching, however raises more problems. Firstly, it encounters the same problems as the idea of an endonuclease, namely how does such a system differentiate between the active BES and other long tracts of 70-bp repeats? Also, such an initiation event could not on its own explain the differences in the switching rates seen between strains, and such an explanation would have to be found in downstream events.

Of course, it is possible that a DSB is not the initiating event of a *VSG* switch, but it is the most likely candidate, as gene conversion reactions, of which these appear to be a form, require such a break (or other similar lesion). It is also possible that the break does not occur within the 70-bp repeat region, and instead occurs inside the *VSG* gene or a region downstream of it. If that is the case, it is difficult to imagine how the ILTat 1.22 conversions

dissected in this study occurred. Here conversion occurred up to and possibly including the telomere, whereas if a break had occurred downstream of the *VSG*, then resection of the duplex must have continued to the telomere tract itself which must then have initiated the event. Such a telomere interaction would be unusual and almost excludes the activation of internal genes, unless the homologies seen in the downstream regions of the *VSG* genes are used (Rice-Ficht *et al*, 1981). The idea of the 70-bp repeat being the site of initiation fits with the data generated for the ILTat 1.2 line, but not for the Lister 427 strain. In the Lister 427 strain, *VSG* switches most likely using the 70-bp repeats account for approximately only a third of the recombination-based switches (McCulloch and Barry, 1999). Also, removal of the 70-bp repeat region in the Lister 427 line made no difference to the frequency of *VSG* switching (McCulloch *et al*, 1997). This indicates that, in Lister 427, the 70-bp repeats are not important for switching. It is tempting to speculate that the reduction of a rapid wild-type switching rate to background spontaneous mutation frequency via laboratory adaptation of the Lister 427 strain is due to the loss of a *VSG* switch initiation event that acts upon the 70-bp repeat. This, however, requires much more extensive research in the pleomorphic lines to determine whether the 70-bp repeats are truly associated with *VSG* switching and that this association is linked to the rapid switching rate.

After the initiating event, which in this model is a DSB, the DNA must be processed to provide a recombinogenic substrate. The complex proposed to be involved in this process in yeast is Mre11/Rad50/Xrs2 (see Chapter 1). When the *MRE11* gene was deleted from the monomorphic Lister 427 line, no difference was seen in the *VSG* switching frequency, so it is unlikely that it is involved in the processing of the switching events. Null mutants of this gene, however, do show a reduction in the transformation efficiency of linear DNA, so the role of this protein in other pathways of homologous recombination is unclear (Robinson *et al*, 2002; Tan *et al*, 2002). Mutants in the pleomorphic line ILTat 1.2 have been made, but data on *VSG* switching have yet to be generated.

For a gene conversion to occur, the duplex receiving the break needs to invade a homologous duplex to initiate DNA synthesis. In the majority of vegetative eukaryotic cells, invasion is catalysed by the recombinase Rad51. As described above, in the monomorphic line Lister 427 this enzyme is critical for the catalysis of most *VSG* switching (McCulloch and Barry, 1999), while the experiments performed in this study show it is not required in the pleomorphic ILTat

1.2 line. It is possible that the difference in the use of these pathways reflects the difference in the use of DNA substrates to drive the conversion reactions. Also discussed in detail above is the fact that the *RAD51*-dependent pathway tends to be less capable at catalyzing recombination reactions involving short stretches of homology than the *RAD51*-independent pathway. Perhaps then the ILTat 1.2 line, along with other pleomorphic lines, utilises the 70-bp repeats as a means of catalyzing recombinational *VSG* switching and, in doing so, has exploited the *RAD51*-independent recombination pathway. In the Lister 427 line, however, larger stretches of homology (between expression sites for example) may be used, in a *RAD51*-dependent manner. If this is the case, then the specialized use of a *RAD51*-independent, 70-bp repeat dependent, pathway could explain the high switching rates seen in the non-laboratory strains. The loss of such a directed pathway during laboratory adaptation could explain the RAD51-dependence of *VSG* switching. This dependence is exemplified by the fact that the VSG switching frequency of the RAD51-dependent pathway is equivalent to the spontaneous background mutation rate. This hypothesis makes a prediction, that the ILTat 1.2 *mre11* null mutants should have a reduced *VSG* switching rate, as the protein complex it is an integral part of (Mre11/Rad50/Xrs2) is essential for recombination in the absence of *RAD51* (Signon *et al*, 2001).

How could a system driving recombination via degenerate repeats have evolved? Repetitive DNA can often be a source of instability leading to genomic rearrangements, so degeneracy in the repeats may reduce the risk of spontaneous detrimental rearrangements by the *RAD51*-dependent mechanism. This implies that, in the ILTat 1.2 line, the presence of the usually background *RAD51*-independent pathway is not increased, but perhaps is specifically channeled to *VSG* switching events. Recombination reactions using short stretches of homology have also been shown to decrease the frequency of crossing over in yeast (Inbar *et al*, 2000), which in trypanosomes could be a preferential property in terms of the success of internal *VSG* gene activation.

It has also been shown in yeast that, during the homology searching step following a DSB, sequences distant from the DSB can often be used more efficiently than sequences close to the break to initiate recombination (Inbar and Kupiec, 1999). A large array of 70-bp repeats within an expression site may reduce recombination between homologous sequences on either side of

the array, that is, reduce the frequency of switching events involving gene conversions with other expression sites.

The studies of *RAD51* in trypanosomes indicate that the difference in switching rates seen between the pleomorphic and monomorphic lines may be ascribed to differences in the utilisation of different homologous recombination pathways. This is opposed to the idea that the reduced switching rates seen in the monomorphic lines are due to loss of a factor during laboratory adaptation. The most likely explanation, in my opinion, is that *VSG* switching is a directed process, yielding the extremely high levels of switching seen in the pleomorphic lines, and that during laboratory adaptation there is no selective pressure to maintain such a potentially demanding and genotoxic process. Thus, over time those trypanosomes that switch at a lower rate are selected and eventually the ability to perform this activity is lost, leading to switching at an undirected, background level. It is not necessary, in this case, for the phenotypes of a reduced switching rate and inability to differentiate naturally to be linked. One other possibility is that, in the absence of recombination factors having been identified to be involved in switching in pleomorphic cell lines, then perhaps an entirely novel mechanism is catalyzing their *VSG* switching. Unlike mating-type switching in yeast, where an initiating event occurs and then and subsequent gene conversion is performed by the homologous recombination machinery, the trypanosome may have a dedicated system used for no other cellular process.

In conclusion, it appears that the ILTat 1.2 line performs recombinational *VSG* switching in a manner different from that of Lister 427, not only in terms of the differences in the level of *in situ* versus gene conversion (Liu *et al*, 1983; Robinson *et al*, 1999) but also mechanistically. Whether the 70-bp repeats are used exclusively for switching in the ILTat 1.2 line requires further experimentation, but the evidence presented here confirms previous work (Matthews *et al*, 1990) and further substantiates this claim. The observations on *RAD51* also provide the first genetic evidence for a difference between monomorphic and pleomorphic lines in relation to *VSG* switching. It is impossible to say which pathway would drive switching in trypanosomes in the wild, as both Lister 427 and ILTat 1.2 have undergone some laboratory adaptation. The ILTat 1.2 line, however, has undergone much less serial passaging in mice and is therefore considerably closer to a field isolate, so it is tempting to speculate that this line is a more realistic model for analyzing *VSG* switching. It should be noted that it is likely that both

pathways are available to each cell line but a difference in how they are regulated is present between the strains.

Thus, this work provides stimulus for further research upon the both the *cis*- and *trans*- acting factors involved in the catalysis of *VSG* switching.

6.5 Future work

One experiment that is perhaps crucial for some of the claims made in this discussion is to determine the percentage of switching events that utilise the 70-bp repeats in the ILTat 1.2 line and other pleomorphic lines. This could be achieved by isolating the sequences surrounding the basic copy of *VSG* genes activated by duplication, using extensive PCR based genomic-walking or the generation of a library. Following this, DNA could be prepared from clones known to have activated the gene and restriction enzyme mapping could determine the limit of conversion. Ideally, such experiments would be performed in the strain used for the genome project (TREU 927/4), thus removing the need to identify surrounding sequences of most *VSG* genes. However, problems arise when using strains with a high switching rate, as it is difficult to obtain clonal populations expressing one *VSG*. This reason perhaps makes the ILTat 1.2 line perfect for study. It is pleomorphic, but displays a switching rate that is at the lower end of such lines.

One aspect that is critical for understanding *VSG* switching is identifying how such events are initiated. One way to identify the potential DSB is to use the linker based PCR utilised by the TOPO-walker (Invitrogen) method. Here, DNA could be extracted from a large number of bloodstream cells, some of which are presumably undergoing a recombinational *VSG* switch. A single primer specific to the active *VSG* gene could be used to extend either towards or away from the 70-bp repeats. If a DSB was present in a number of the cells from which the DNA was prepared, then the primer extension by Taq polymerase would generate the adenosine overhang required for the linker to be attached by the conjugated topoisomerase from the kit. PCR, and further nested PCRs, could then be performed using primers specific for the attached linker and for the *VSG* in the active site. Sequencing of this product would identify the break point, or sequence being processed shortly after the break. Of course, this could also pick up random DNA breaks, so by performing the same experiment on genes in

inactive expression sites or on other telomeric or internal genes, any differences between a catalysed break in the active expression site and random breaks elsewhere could be determined.

Two-dimensional gel electrophoresis can be used to identify recombination intermediates and replication fork pausing (Kobayashi *et al*, 1992; Courcelle *et al*, 2003). It may be possible to apply this technology to trypanosomes in order to see whether the 70-bp repeats have the propensity to pause replications forks and subsequently collapse, creating a recombinogenic intermediate.

One other way of identifying factors involved in the initiation of *VSG* switching is to extend the attempts made in this study to find proteins that bind to sequences potentially involved in *VSG* switching. Here, the short region of 70-bp repeats derived from the ILTat 1.22 locus was used in an experiment designed to isolate proteins that bound this sequence specifically. It would be interesting to see if a longer array of these sequences, equivalent to those present in the active expression site, would bind specific proteins. It is also possible that the *VSG* sequence, particularly the region encoding the C-terminus of the protein, or the region downstream of the *VSG* in the active expression site, could be bound by any specific proteins. Binding of any factors to these regions may require conditions different from those used in this study, so a rigorous optimization would be required. Also, an extract from a line where switching is very rapid may contain a higher abundance of any *VSG* switching specific factors.

The assay created here in order to analyse the role of the 70-bp repeats in duplicating sequence from a chromosomal location into the active expression site showed that it was possible to generate such duplications, at a low frequency, in the Lister 427 line but not in the ILTat 1.2 line. This could be for many reasons, so improvement of the assay is required. One is by using a selectable marker that is not transcribed in the chromosomal internal location, this would be activated solely by duplication into the expression site instead of selecting upon an increased drug concentration. Such a high drug concentration may have been detrimental to the selection of growth in the ILTat 1.2, line which appears less stable in culture than the Lister 427 line. If this method could be made sensitive enough, or the assay undertaken in a rapid switching strain (*i.e.* EATRO 795), the selectable marker in this new assay could be replaced by a fluorescent marker and the switching events could be assayed by fluorescence-associated cell analysis. By creating an assay that is high-throughput and readily detects switching events,

modifications of the sequences given to recombine the marker could be analysed in detail. In fact this study is already underway in our laboratory.

The absence of a detectable role for Rad51 in the ILTat 1.2 lines poses many interesting questions. Firstly, is this protein required for switching in other pleomorphic and monomorphic lines? It would be very interesting to make *rad51* null mutants in high switching strains, as this would allow detection of more subtle changes and allow us to determine whether Rad51 truly has a role in *VSG* switching. By making *rad51* null mutants in several monomorphic and pleomorphic strains of varying switching rates, it could be determined whether switching really is catalysed differently between monomorphic and pleomorphic lines and whether this difference is related to the frequency of switching, as opposed to simply a specific difference between the Lister 427 line and the ILTat 1.2 line.

It would be interesting to analyse how switching events are catalysed in the ILTat 1.2 *rad51* null mutants. Is there a change in the frequency of *in situ* events compared to the recombinational events? Are the 70-bp repeats still involved? Ideally, in this case, ILTat 1.22 expressers could be isolated and subjected to the same detailed analysis of the *cis*-acting elements as performed in this study. Are telomeric genes still activated more frequently than internal genes? Can internal genes be activated at all, considering that the predominant pathway for two-ended gene conversion is absent?

Most interestingly of all, perhaps, would be to determine if the ILTat 1.2 *mre11* null mutants displayed a reduction in the frequency of *VSG* switching, as predicted by the model described above and in Figures 6.1 and 6.2.

REFERENCE LIST

Aboussekhra, A., Chanet, R., Adjiri, A., and Fabre, F. (1992). Semidominant suppressors of Srs2 helicase mutations of *Saccharomyces cerevisiae* map in the *RAD51* gene, whose sequence predicts a protein with similarities to procaryotic RecA proteins. *Mol.Cell Biol.* **12**, 3224-3234.

Abuin, A., Zhang, H., and Bradley, A. (2000). Genetic analysis of mouse embryonic stem cells bearing Msh3 and Msh2 single and compound mutations. *Mol.Cell Biol.* **20**, 149-157.

Adzuma, K. (1998). No sliding during homology search by RecA protein. *J.Biol.Chem.* **273**, 31565-31573.

Alarcon, C. M., Son, H. J., Hall, T., and Donelson, J. E. (1994). A monocistronic transcript for a trypanosome variant surface glycoprotein. *Mol.Cell Biol.* **14**, 5579-5591.

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

Alexandre, S., Paindavoine, P., Hanocq-Quertier, J., Paturiaux-Hanocq, F., Tebabi, P., and Pays, E. (1996). Families of adenylate cyclase genes in *Trypanosoma brucei*. *Mol.Biochem.Parasitol.* **77**, 173-182.

Alexeev, A., Mazin, A., and Kowalczykowski, S. C. (2003). Rad54 protein possesses chromatin-remodeling activity stimulated by the Rad51-ssDNA nucleoprotein filament. *Nat.Struct.Biol.* **10**, 182-186.

Aline, R., Jr., MacDonald, G., Brown, E., Allison, J., Myler, P., Rothwell, V., and Stuart, K. (1985). (TAA)n within sequences flanking several intrachromosomal variant surface glycoprotein genes in *Trypanosoma brucei*. *Nucleic Acids Res.* **13**, 3161-3177.

Aline, R. F., Jr. and Stuart, K. (1989). *Trypanosoma brucei*: conserved sequence organization 3' to telomeric variant surface glycoprotein genes. *Exp.Parasitol.* **68**, 57-66.

Aline, R. F., Jr., Myler, P. J., and Stuart, K. D. (1989). *Trypanosoma brucei*: frequent loss of a telomeric variant surface glycoprotein gene. *Exp.Parasitol.* **68**, 8-16.

Baltz, T., Giroud, C., Baltz, D., Roth, C., Raibaud, A., and Eisen, H. (1986). Stable expression of two variable surface glycoproteins by cloned *Trypanosoma equiperdum*. *Nature* **319**, 602-604.

Barry, J. D. (1986). Antigenic variation during *Trypanosoma vivax* infections of different host species. *Parasitology* **92**, 51-65.

Barry, J. D. (1997). The relative significance of mechanisms of antigenic variation in African trypanosomes. *Parasitol. Today* **13**, 212-218.

Barry, J. D., Graham, S. V., Fotheringham, M., Graham, V. S., Kobryn, K., and Wymer, B. (1998). *VSG* gene control and infectivity strategy of metacyclic stage *Trypanosoma brucei*. *Mol.Biochem.Parasitol.* **91**, 93-105.

Barry, J. D. and McCulloch, R. (2001). Antigenic variation in trypanosomes: enhanced phenotypic variation in a eukaryotic parasite. *Adv.Parasitol.* **49**, 1-70.

Barry, J. D., Ginger, M. L., Burton, P., and McCulloch, R. (2003). Why are parasite contingency genes often associated with telomeres? *Int.J.Parasitol.* **33**, 29-45.

Baumann, P. and West, S. C. (1997). The human Rad51 protein: polarity of strand transfer and stimulation by hRP-A. *EMBO J.* **16**, 5198-5206.

Baumann, P. and West, S. C. (1998). Role of the human RAD51 protein in homologous recombination and double-stranded-break repair. *Trends Biochem.Sci.* **23**, 247-251.

Bell, J.S., Harvey, T.I., Sims, A.M., and McCulloch, R. (2003). Characterisation of the mismatch repair machinery in *Trypanosoma brucei*. *Mol. Microbiol.*, in press.

Bell, J.S. and McCulloch, R. (2003). Mismatch repair regulates mismatch repair, but has little influence on antigenic variation, in *Trypanosoma brucei*. *J. Biol.Chem.,* in press.

Bell, S. D. and Barry, J. D. (1995). Trypanosome nuclear factors which bind to internal promoter elements of tRNA genes. *Nucleic Acids Res.* **23**, 3103-3110.

Benson, F. E., Baumann, P., and West, S. C. (1998). Synergistic actions of Rad51 and Rad52 in recombination and DNA repair. *Nature* **391**, 401-404.

Bernards, A., Kooter, J. M., and Borst, P. (1985). Structure and transcription of a telomeric surface antigen gene of *Trypanosoma brucei*. *Mol.Cell Biol.* **5**, 545-553.

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

Biet, E., Sun, J., and Dutreix, M. (1999). Conserved sequence preference in DNA binding among recombination proteins: an effect of ssDNA secondary structure. *Nucleic Acids Res.* **27**, 596-600.

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

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

Boddy, M. N., Gaillard, P. H., McDonald, W. H., Shanahan, P., Yates, J. R., III, and Russell, P. (2001). Mus81-Eme1 are essential components of a Holliday junction resolvase. *Cell* **107**, 537-548.

Borst, P. (1991). Transferrin receptor, antigenic variation and the prospect of a trypanosome vaccine. *Trends Genet.* **7**, 307-309.

Borst, P. and van Leeuwen, F. (1997). beta-D-glucosyl-hydroxymethyluracil, a novel base in African trypanosomes and other Kinetoplastida. *Mol.Biochem.Parasitol.* **90**, 1-8.

Borst, P. and Fairlamb, A. H. (1998). Surface receptors and transporters of *Trypanosoma brucei*. *Annu.Rev.Microbiol.* **52**, 745-778.

Borst,P. (2002) In Craig,N.L. (ed.) *Mobile DNA II*. ASM Press, Washington, DC, pp. 953–971.

Borst, P. (2002). Antigenic variation and allelic exclusion. *Cell* **109**, 5-8.

Bosco, G. and Haber, J. E. (1998). Chromosome break-induced DNA replication leads to nonreciprocal translocations and telomere capture. *Genetics* **150**, 1037-1047.

Boulton, S. J. and Jackson, S. P. (1998). Components of the Ku-dependent non-homologous end-joining pathway are involved in telomeric length maintenance and telomeric silencing. *EMBO J.* **17**, 1819-1828.

Breidbach, T., Ngazoa, E., and Steverding, D. (2002). *Trypanosoma brucei*: *in vitro* slender-to-stumpy differentiation of culture-adapted, monomorphic bloodstream forms. *Exp.Parasitol.* **101**, 223-230.

Bressan, D. A., Baxter, B. K., and Petrini, J. H. (1999). The Mre11-Rad50-Xrs2 protein complex facilitates homologous recombination-based double-strand break repair in Saccharomyces cerevisiae. *Mol.Cell Biol.* **19**, 7681-7687.

Bringaud, F., Biteau, N., Melville, S. E., Hez, S., El Sayed, N. M., Leech, V., Berriman, M., Hall, N., Donelson, J. E., and Baltz, T. (2002). A new, expressed multigene family containing a hot spot for insertion of retroelements is associated with polymorphic subtelomeric regions of *Trypanosoma brucei*. *Eukaryot.Cell* **1**, 137-151.

Butikofer, P., Ruepp, S., Boschung, M., and Roditi, I. (1997). 'GPEET' procyclin is the major surface protein of procyclic culture forms of *Trypanosoma brucei* brucei strain 427. *Biochem.J.* **326 (Pt 2)**, 415-423.

Capbern, A., Giroud, C., Baltz, T. and Mattern, P. (1997) *Trypanosoma equiperdum*: étude des variation antigéniques au cours de la trypanosomose expérimentale du lapin. *Exp. Parasitol.* **42**, 6-13.

Carrington, M., Miller, N., Blum, M., Roditi, I., Wiley, D., and Turner, M. (1991). Variant specific glycoprotein of *Trypanosoma brucei* consists of two domains each having an independently conserved pattern of cysteine residues. *J.Mol.Biol.* **221**, 823-835.

Carrington, M. and Boothroyd, J. (1996). Implications of conserved structural motifs in disparate trypanosome surface proteins. *Mol.Biochem.Parasitol.* **81**, 119-126.

Chanet, R., Heude, M., Adjiri, A., Maloisel, L., and Fabre, F. (1996). Semidominant mutations in the yeast Rad51 protein and their relationships with the Srs2 helicase. *Mol.Cell Biol.* **16**, 4782-4789.

Chaves, I., Zomerdijk, J., Dirks-Mulder, A., Dirks, R. W., Raap, A. K., and Borst, P. (1998). Subnuclear localization of the active variant surface glycoprotein gene expression site in *Trypanosoma brucei*. *Proc.Natl.Acad.Sci.U.S.A* **95**, 12328-12333.

Chaves, I., Rudenko, G., Dirks-Mulder, A., Cross, M., and Borst, P. (1999). Control of variant surface glycoprotein gene-expression sites in *Trypanosoma brucei*. *EMBO J.* **18**, 4846-4855.

Chen, C. and Kolodner, R. D. (1999). Gross chromosomal rearrangements in *Saccharomyces cerevisiae* replication and recombination defective mutants. *Nat.Genet.* **23**, 81-85.

Chen, X. B., Melchionna, R., Denis, C. M., Gaillard, P. H., Blasina, A., Van, de Weyer, I, Boddy, M. N., Russell, P., Vialard, J., and McGowan, C. H. (2001). Human Mus81-associated endonuclease cleaves Holliday junctions *in vitro*. *Mol.Cell* **8**, 1117-1127.

Connelly, J. C. and Leach, D. R. (2002). Tethering on the brink: the evolutionarily conserved Mre11-Rad50 complex. *Trends Biochem.Sci.* **27**, 410-418.

Constantinou, A., Davies, A. A., and West, S. C. (2001). Branch migration and Holliday junction resolution catalyzed by activities from mammalian cells. *Cell* **104**, 259-268.

Constantinou, A., Chen, X. B., McGowan, C. H., and West, S. C. (2002). Holliday junction resolution in human cells: two junction endonucleases with distinct substrate specificities. *EMBO J.* **21**, 5577-5585.

Conway, C., McCulloch, R., Ginger, M. L., Robinson, N. P., Browitt, A., and Barry, J. D. (2002). Ku is important for telomere maintenance, but not for differential expression of telomeric VSG genes, in African trypanosomes. *J.Biol.Chem.* **277**, 21269-21277.

Conway, C., Proudfoot, C., Burton, P., Barry, J. D., and McCulloch, R. (2002a). Two pathways of homologous recombination in *Trypanosoma brucei*. *Mol.Microbiol.* **45**, 1687-1700.

Courcelle, J., Donaldson, J. R., Chow, K. H., and Courcelle, C. T. (2003). DNA damage-induced replication fork regression and processing in *Escherichia coli*. *Science* **299**, 1064-1067.

Cox, M. M. (2002). The nonmutagenic repair of broken replication forks via recombination. *Mutat.Res.* **510**, 107-120.

Craig, A. and Scherf, A. (ed.) *Antigenic Variation*. Academic Press.

Cromie, G. A., Connelly, J. C., and Leach, D. R. (2001). Recombination at double-strand breaks and DNA ends: conserved mechanisms from phage to humans. *Mol.Cell* **8**, 1163-1174.

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

Cross, M., Kieft, R., Sabatini, R., Wilm, M., de Kort, M., van der Marel, G. A., van Boom, J. H., van Leeuwen, F., and Borst, P. (1999). The modified base J is the target for a novel DNA-binding protein in kinetoplastid protozoans. *EMBO J.* **18**, 6573-6581.

Cross, M., Kieft, R., Sabatini, R., Dirks-Mulder, A., Chaves, I., and Borst, P. (2002). J-binding protein increases the level and retention of the unusual base J in trypanosome DNA. *Mol.Microbiol.* **46**, 37-47.

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

D'Amours, D. and Jackson, S. P. (2002). The Mre11 complex: at the crossroads of DNA repair and checkpoint signalling. *Nat.Rev.Mol.Cell Biol.* **3**, 317-327.

Datta, A., Adjiri, A., New, L., Crouse, G. F., and Jinks, Robertson S. (1996). Mitotic crossovers between diverged sequences are regulated by mismatch repair proteins in Saccaromyces cerevisiae. *Mol.Cell Biol.* **16**, 1085-1093.

Datta, A., Hendrix, M., Lipsitch, M., and Jinks-Robertson, S. (1997). Dual roles for DNA sequence identity and the mismatch repair system in the regulation of mitotic crossing-over in yeast. *Proc.Natl.Acad.Sci.U.S.A* **94**, 9757-9762.

Davis, A. P. and Symington, L. S. (2001). The yeast recombinational repair protein Rad59 interacts with Rad52 and stimulates single-strand annealing. *Genetics* **159**, 515-525.

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

de Jager, M., Dronkert, M. L., Modesti, M., Beerens, C. E., Kanaar, R., and van Gent, D. C. (2001). DNA-binding and strand-annealing activities of human Mre11: implications for its roles in DNA double-strand break repair pathways. *Nucleic Acids Res.* **29**, 1317-1325.

De Lange, T., Kooter, J. M., Michels, P. A., and Borst, P. (1983). Telomere conversion in trypanosomes. *Nucleic Acids Res.* **11**, 8149-8165.

De Lange, T., Kooter, J. M., Luirink, J., and Borst, P. (1985). Transcription of a transposed trypanosome surface antigen gene starts upstream of the transposed segment. *EMBO J.* **4**, 3299-3306.

Doe, C. L., Ahn, J. S., Dixon, J., and Whitby, M. C. (2002). Mus81-Eme1 and Rqh1 involvement in processing stalled and collapsed replication forks. *J.Biol.Chem.* **277**, 32753-32759.

Donelson, J. E., Murphy, W. J., Brentano, S. T., Rice-Ficht, A. C., and Cain, G. D. (1983). Comparison of the expression-linked extra copy (ELC) and basic copy (BC) genes of a trypanosome surface antigen. *J.Cell Biochem.* **23**, 1-12.

Donelson, J. E. (2003). Antigenic variation and the African trypanosome genome. *Acta Trop.* **85**, 391-404.

Doyle, J. J., Hirumi, H., Hirumi, K., Lupton, E. N., and Cross, G. A. (1980). Antigenic variation in clones of animal-infective *Trypanosoma brucei* derived and maintained *in vitro*. *Parasitology* **80**, 359-369.

Eggleston, A. K. and West, S. C. (1997). Recombination initiation: easy as A, B, C, D... chi? *Curr.Biol.* **7**, R745-R749.

El Sayed, N. M., Hegde, P., Quackenbush, J., Melville, S. E., and Donelson, J. E. (2000). The African trypanosome genome. *Int.J.Parasitol.* **30**, 329-345.

El Sayed, N. M., Ghedin, E., Song, J., MacLeod, A., Bringaud, F., Larkin, C., Wanless, D., Peterson, J., Hou, L., Taylor, S., Tweedie, A., Biteau, N., Khalak, H. G., Lin, X., Mason, T., Hannick, L., Caler, E., Blandin, G., Bartholomeu, D., Simpson, A. J., Kaul, S., Zhao, H., Pai, G., Van Aken, S., Utterback, T., Haas, B., Koo, H. L., Umayam, L., Suh, B., Gerrard, C., Leech, V., Qi, R., Zhou, S., Schwartz, D., Feldblyum, T., Salzberg, S., Tait, A., Turner, C. M., Ullu, E., White, O., Melville, S., Adams, M. D., Fraser, C. M., and Donelson, J. E. (2003). The sequence and analysis of *Trypanosoma brucei* chromosome II. *Nucleic Acids Res.* **31**, 4856-4863.

Elliott, B. and Jasin, M. (2001). Repair of double-strand breaks by homologous recombination in mismatch repair-defective mammalian cells. *Mol.Cell Biol.* **21**, 2671-2682.

Esser, K. M., Schoenbechler, M. J., and Gingrich, J. B. (1982). *Trypanosoma rhodesiense* blood forms express all antigen specificities relevant to protection against metacyclic (insect form) challenge. *J.Immunol.* **129**, 1715-1718.

Evans, E. and Alani, E. (2000). Roles for mismatch repair factors in regulating genetic recombination. *Mol.Cell Biol.* **20**, 7839-7844.

Ferguson, D. O. and Holloman, W. K. (1996). Recombinational repair of gaps in DNA is asymmetric in *Ustilago maydis* and can be explained by a migrating D-loop model. *Proc.Natl.Acad.Sci.U.S.A* **93**, 5419-5424.

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

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

Florent, I., Baltz, T., Raibaud, A., and Eisen, H. (1987). On the role of repeated sequences 5' to variant surface glycoprotein genes in African trypanosomes. *Gene* **53**, 55-62.

Frank, S. A. (1999). A model for the sequential dominance of antigenic variants in African trypanosome infections. *Proc.R.Soc.Lond B Biol.Sci.* **266**, 1397-1401.

Gabrielsen, O. S., Homes, E., Korsnes, L., Ruet, A., and Oyen, T. B. (1989). Magnetic DNA affinity purification of yeast transcription factor tau--a new purification principle for the ultra rapid isolation of near homogeneous factor. *Nucleic Acids Res.* **17**, 6253-6267.

Gasior, S. L., Olivares, H., Ear, U., Hari, D. M., Weichselbaum, R., and Bishop, D. K. (2001). Assembly of RecA-like recombinases: distinct roles for mediator proteins in mitosis and meiosis. *Proc.Natl.Acad.Sci.U.S.A* **98**, 8411-8418.

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

Gilbertson, L. A. and Stahl, F. W. (1996). A test of the double-strand break repair model for meiotic recombination in *Saccharomyces cerevisiae*. *Genetics* **144**, 27-41.

Ginger, M. L., Blundell, P. A., Lewis, A. M., Browitt, A., Gunzl, A., and Barry, J. D. (2002). *Ex vivo* and *in vitro* identification of a consensus promoter for *VSG* genes expressed by metacyclic-stage trypanosomes in the tsetse fly. *Eukaryot.Cell* **1**, 1000-1009.

Gottesdiener, K., Garcia-Anoveros, J., Lee, M. G., and Van der Ploeg, L. H. (1990). Chromosome organization of the protozoan *Trypanosoma brucei*. *Mol.Cell Biol.* **10**, 6079-6083.

Gottesdiener, K., Chung, H. M., Brown, S. D., Lee, M. G., and Van der Ploeg, L. H. (1991). Characterization of *VSG* gene expression site promoters and promoter-associated DNA rearrangement events. *Mol.Cell Biol.* **11**, 2467-2480.

Gottesdiener, K. M., Goriparthi, L., Masucci, J. P., and Van der Ploeg, L. H. (1992). A proposed mechanism for promoter-associated DNA rearrangement events at a variant surface glycoprotein gene expression site. *Mol.Cell Biol.* **12**, 4784-4795.

Grab, D. J. and Bwayo, J. J. (1982). Isopycnic isolation of African trypanosomes on Percoll gradients formed in situ. *Acta Trop.* **39**, 363-366.

Gray, A. R. (1965). Antigenic variation in a strain of *Trypanosoma brucei* transmitted by Glossina morsitans and G. palpalis. *J.Gen.Microbiol.* **41**, 195-214.

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

Haag, J., O'hUigin, C., and Overath, P. (1998). The molecular phylogeny of trypanosomes: evidence for an early divergence of the Salivaria. *Mol.Biochem.Parasitol.* **91**, 37-49.

Haber, J. E. (1998). The many interfaces of Mre11. *Cell* **95**, 583-586.

Haber, J. E. and Heyer, W. D. (2001). The fuss about Mus81. *Cell* **107**, 551-554.

Haber, J.E.. (2002) In Craig,N.L. (ed.) *Mobile DNA II*. ASM Press, Washington, DC, pp. 927-952.

Hall, N., Berriman, M., Lennard, N. J., Harris, B. R., Hertz-Fowler, C., Bart-Delabesse, E. N., Gerrard, C. S., Atkin, R. J., Barron, A. J., Bowman, S., Bray-Allen, S. P., Bringaud, F., Clark, L. N., Corton, C. H., Cronin, A., Davies, R., Doggett, J., Fraser, A., Gruter, E., Hall, S., Harper, A. D., Kay, M. P., Leech, V., Mayes, R., Price, C., Quail, M. A., Rabbinowitsch, E., Reitter, C., Rutherford, K., Sasse, J., Sharp, S., Shownkeen, R., MacLeod, A., Taylor, S., Tweedie, A., Turner, C. M., Tait, A., Gull, K., Barrell, B., and Melville, S. E. (2003). The DNA sequence of chromosome I of an African trypanosome: gene content, chromosome organisation, recombination and polymorphism. *Nucleic Acids Res.* **31**, 4864-4873.

Hendriks, E., van Deursen, F. J., Wilson, J., Sarkar, M., Timms, M., and Matthews, K. R. (2000). Life-cycle differentiation in *Trypanosoma brucei*: molecules and mutants. *Biochem.Soc.Trans.* **28**, 531-536.

Hendriks, E. F., Robinson, D. R., Hinkins, M., and Matthews, K. R. (2001). A novel CCCH protein which modulates differentiation of *Trypanosoma brucei* to its procyclic form. *EMBO J.* **20**, 6700-6711.

Herbert, W. J. and Lumsden, W. H. (1976). *Trypanosoma brucei*: a rapid "matching" method for estimating the host's parasitemia. *Exp.Parasitol.* **40**, 427-431.

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

Hobbs, M. R. and Boothroyd, J. C. (1990). An expression-site-associated gene family of trypanosomes is expressed *in vivo* and shows homology to a variant surface glycoprotein gene. *Mol.Biochem.Parasitol.* **43**, 1-16.

Hoek, M., Engstler, M., and Cross, G. A. (2000). Expression-site-associated gene 8 (*ESAG8*) of *Trypanosoma brucei* is apparently essential and accumulates in the nucleolus. *J.Cell Sci.* **113 (Pt 22)**, 3959-3968.

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

Holbeck, S. L. and Strathern, J. N. (1997). A role for REV3 in mutagenesis during double-strand break repair in Saccharomyces cerevisiae. *Genetics* **147**, 1017-1024.

Holliday, R. (1964). A mechanism for gene conversion in fungi. *Genet. Res.* **5**, 282-304.

Holmes, A. M. and Haber, J. E. (1999). Double-strand break repair in yeast requires both leading and lagging strand DNA polymerases. *Cell* **96**, 415-424.

Holmes, V. F., Benjamin, K. R., Crisona, N. J., and Cozzarelli, N. R. (2001). Bypass of heterology during strand transfer by *Saccharomyces cerevisiae* Rad51 protein. *Nucleic Acids Res.* **29**, 5052-5057.

Hopfner, K. P., Putnam, C. D., and Tainer, J. A. (2002). DNA double-strand break repair from head to tail. *Curr.Opin.Struct.Biol.* **12**, 115-122.

Horn, D. and Cross, G. A. (1997). Analysis of *Trypanosoma brucei VSG* expression site switching *in vitro*. *Mol.Biochem.Parasitol.* **84**, 189-201.

Hsieh, P. (2001). Molecular mechanisms of DNA mismatch repair. *Mutat.Res.* **486**, 71-87.

Inbar, O. and Kupiec, M. (1999). Homology search and choice of homologous partner during mitotic recombination. *Mol.Cell Biol.* **19**, 4134-4142.

Inbar, O., Liefshitz, B., Bitan, G., and Kupiec, M. (2000). The relationship between homology length and crossing over during the repair of a broken chromosome. *J.Biol.Chem.* **275**, 30833-30838.

Ira, G. and Haber, J. E. (2002). Characterization of *RAD51*-independent break-induced replication that acts preferentially with short homologous sequences. *Mol.Cell Biol.* **22**, 6384-6392.

Isobe, T., Holmes, E. C., and Rudenko, G. (2003). The Transferrin Receptor Genes of *Trypanosoma equiperdum* Are Less Diverse in Their Transferrin Binding Site than Those of the Broad-Host Range *Trypanosoma brucei*. *J.Mol.Evol.* **56**, 377-386.

Ivanov, E. L., Sugawara, N., White, C. I., Fabre, F., and Haber, J. E. (1994). Mutations in XRS2 and RAD50 delay but do not prevent mating-type switching in Saccharomyces cerevisiae. *Mol.Cell Biol.* **14**, 3414-3425.

Ivanov, E. L., Sugawara, N., Fishman-Lobell, J., and Haber, J. E. (1996). Genetic requirements for the single-strand annealing pathway of double-strand break repair in *Saccharomyces cerevisiae*. *Genetics* **142**, 693-704.

Jiang, H., Xie, Y., Houston, P., Stemke-Hale, K., Mortensen, U. H., Rothstein, R., and Kodadek, T. (1996). Direct association between the yeast Rad51 and Rad54 recombination proteins. *J.Biol.Chem.* **271**, 33181-33186.

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

Johnson, R. D. and Symington, L. S. (1995). Functional differences and interactions among the putative RecA homologs Rad51, Rad55, and Rad57. *Mol.Cell Biol.* **15**, 4843-4850.

Kaliraman, V., Mullen, J. R., Fricke, W. M., Bastin-Shanower, S. A., and Brill, S. J. (2001). Functional overlap between Sgs1-Top3 and the Mms4-Mus81 endonuclease. *Genes Dev.* **15**, 2730-2740.

Kamper, S. M. and Barbet, A. F. (1992). Surface epitope variation via mosaic gene formation is potential key to long-term survival of *Trypanosoma brucei*. *Mol.Biochem.Parasitol.* **53**, 33-44.

Kimmel, B. E., ole-MoiYoi, O. K., and Young, J. R. (1987). Ingi, a 5.2-kb dispersed sequence element from *Trypanosoma brucei* that carries half of a smaller mobile element at either end and has homology with mammalian LINEs. *Mol.Cell Biol.* **7**, 1465-1475.

Klar, A. J. and Strathern, J. N. (1984). Resolution of recombination intermediates generated during yeast mating type switching. *Nature* **310**, 744-748.

Kobayashi, T., Hidaka, M., Nishizawa, M., and Horiuchi, T. (1992). Identification of a site required for DNA replication fork blocking activity in the rRNA gene cluster in *Saccharomyces cerevisiae*. *Mol.Gen.Genet.* **233**, 355-362.

Kogoma, T. (1997). Stable DNA replication: interplay between DNA replication, homologous recombination, and transcription. *Microbiol.Mol.Biol.Rev.* **61**, 212-238.

Kolodner, R. D. and Marsischky, G. T. (1999). Eukaryotic DNA mismatch repair. *Curr.Opin.Genet.Dev.* **9**, 89-96.

Kooter, J. M., van der Spek, H. J., Wagter, R., d'Oliveira, C. E., van der, Hoeven F., Johnson, P. J., and Borst, P. (1987). The anatomy and transcription of a telomeric expression site for variant-specific surface antigens in *T. brucei*. *Cell* **51**, 261-272.

Kooter, J. M., Winter, A. J., de Oliveira, C., Wagter, R., and Borst, P. (1988). Boundaries of telomere conversion in *Trypanosoma brucei*. *Gene* **69**, 1-11.

Kraus, E., Leung, W. Y., and Haber, J. E. (2001). Break-induced replication: a review and an example in budding yeast. *Proc.Natl.Acad.Sci.U.S.A* **98**, 8255-8262.

LaCount, D. J., El Sayed, N. M., Kaul, S., Wanless, D., Turner, C. M., and Donelson, J. E. (2001). Analysis of a donor gene region for a variant surface glycoprotein and its expression site in African trypanosomes. *Nucleic Acids Res.* **29**, 2012-2019.

Lamont, G. S., Tucker, R. S., and Cross, G. A. (1986). Analysis of antigen switching rates in *Trypanosoma brucei*. *Parasitology* **92 (Pt 2)**, 355-367.

Landfear, S. M. and Ignatushchenko, M. (2001). The flagellum and flagellar pocket of trypanosomatids. *Mol.Biochem.Parasitol.* **115**, 1-17.

Laurent, M., Pays, E., Delinte, K., Magnus, E., Van Meirvenne, N., and Steinert, M. (1984). Evolution of a trypanosome surface antigen gene repertoire linked to non-duplicative gene activation. *Nature* **308**, 370-373.

Lee, M. G. and Van der Ploeg, L. H. (1987). Frequent independent duplicative transpositions activate a single *VSG* gene. *Mol.Cell Biol.* **7**, 357-364.

Lee, S. E., Moore, J. K., Holmes, A., Umezu, K., Kolodner, R. D., and Haber, J. E. (1998). Saccharomyces Ku70, mre11/rad50 and RPA proteins regulate adaptation to G2/M arrest after DNA damage. *Cell* **94**, 399-409.

Leung, W., Malkova, A., and Haber, J. E. (1997). Gene targeting by linear duplex DNA frequently occurs by assimilation of a single strand that is subject to preferential mismatch correction. *Proc.Natl.Acad.Sci.U.S.A* **94**, 6851-6856.

Lewis, L. K. and Resnick, M. A. (2000). Tying up loose ends: nonhomologous end-joining in *Saccharomyces cerevisiae*. *Mutat.Res.* **451**, 71-89.

Liu, A. Y., Van der Ploeg, L. H., Rijsewijk, F. A., and Borst, P. (1983). The transposition unit of variant surface glycoprotein gene 118 of *Trypanosoma brucei*. Presence of repeated elements at its border and absence of promoter-associated sequences. *J.Mol.Biol.* **167**, 57-75.

Liu, A. Y., Michels, P. A., Bernards, A., and Borst, P. (1985). Trypanosome variant surface glycoprotein genes expressed early in infection. *J.Mol.Biol.* **182**, 383-396.

Longacre, S., Raibaud, A., Hibner, U., Buck, G., Eisen, H., Baltz, T., Giroud, C., and Baltz, D. (1983). DNA rearrangements and antigenic variation in *Trypanosoma equiperdum*: expression-independent DNA rearrangements in the basic copy of a variant surface glycoprotein gene. *Mol.Cell Biol.* **3**, 410-414.

Longacre, S., Hibner, U., Raibaud, A., Eisen, H., Baltz, T., Giroud, C., and Baltz, D. (1983a). DNA rearrangements and antigenic variation in *Trypanosoma equiperdum*: multiple expression-linked sites in independent isolates of trypanosomes expressing the same antigen. *Mol.Cell Biol.* **3**, 399-409.

Longacre, S. and Eisen, H. (1986). Expression of whole and hybrid genes in *Trypanosoma equiperdum* antigenic variation. *EMBO J.* **5**, 1057-1063.

Louis, E. J. and Haber, J. E. (1990). Mitotic recombination among subtelomeric Y' repeats in Saccharomyces cerevisiae. *Genetics* **124**, 547-559.

Lukes, J., Guilbride, D. L., Votypka, J., Zikova, A., Benne, R., and Englund, P. T. (2002). Kinetoplast DNA network: evolution of an improbable structure. *Eukaryot.Cell* **1**, 495-502.

Malkova, A., Ivanov, E. L., and Haber, J. E. (1996). Double-strand break repair in the absence of RAD51 in yeast: a possible role for break-induced DNA replication. *Proc.Natl.Acad.Sci.U.S.A* **93**, 7131-7136.

Malkova, A., Signon, L., Schaefer, C. B., Naylor, M. L., Theis, J. F., Newlon, C. S., and Haber, J. E. (2001). *RAD51*-independent break-induced replication to repair a broken chromosome depends on a distant enhancer site. *Genes Dev.* **15**, 1055-1060.

Masson, J. Y., Tarsounas, M. C., Stasiak, A. Z., Stasiak, A., Shah, R., McIlwraith, M. J., Benson, F. E., and West, S. C. (2001). Identification and purification of two distinct complexes containing the five RAD51 paralogs. *Genes Dev.* **15**, 3296-3307.

Matthews, K. R., Shiels, P. G., Graham, S. V., Cowan, C., and Barry, J. D. (1990). Duplicative activation mechanisms of two trypanosome telomeric *VSG* genes with structurally simple 5' flanks. *Nucleic Acids Res.* **18**, 7219-7227.

Matthews, K. R. and Gull, K. (1994). Evidence for an interplay between cell cycle progression and the initiation of differentiation between life cycle forms of African trypanosomes. *J.Cell Biol.* **125**, 1147-1156.

Matthews, K. R. and Gull, K. (1997). Commitment to differentiation and cell cycle re-entry are coincident but separable events in the transformation of African trypanosomes from their bloodstream to their insect form. *J.Cell Sci.* **110**, 2609-2618.

Matthews, K. R. (1999). Developments in the differentiation of *Trypanosoma brucei*. *Parasitol.Today* **15**, 76-80.

Mazin, A. V., Bornarth, C. J., Solinger, J. A., Heyer, W. D., and Kowalczykowski, S. C. (2000). Rad54 protein is targeted to pairing loci by the Rad51 nucleoprotein filament. *Mol.Cell* **6**, 583-592.

McCulloch, R., Rudenko, G., and Borst, P. (1997). Gene conversions mediating antigenic variation in *Trypanosoma brucei* can occur in variant surface glycoprotein expression sites lacking 70-base-pair repeat sequences. *Mol.Cell Biol.* **17**, 833-843.

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

McGill, C. B., Holbeck, S. L., and Strathern, J. N. (1998). The chromosome bias of misincorporations during double-strand break repair is not altered in mismatch repair-defective strains of Saccharomyces cerevisiae. *Genetics* **148**, 1525-1533.

Mehlert, A., Zitzmann, N., Richardson, J. M., Treumann, A., and Ferguson, M. A. (1998). The glycosylation of the variant surface glycoproteins and procyclic acidic repetitive proteins of *Trypanosoma brucei*. *Mol.Biochem.Parasitol.* **91**, 145-152.

Melville, S. E., Leech, V., Gerrard, C. S., Tait, A., and Blackwell, J. M. (1998). The molecular karyotype of the megabase chromosomes of *Trypanosoma brucei* and the assignment of chromosome markers. *Mol.Biochem.Parasitol.* **94**, 155-173.

Melville, S. E., Gerrard, C. S., and Blackwell, J. M. (1999). Multiple causes of size variation in the diploid megabase chromosomes of African tyrpanosomes. *Chromosome.Res.* **7**, 191-203.

Michels, P. A., Liu, A. Y., Bernards, A., Sloof, P., Van der Bijl, M. M., Schinkel, A. H., Menke, H. H., Borst, P., Veeneman, G. H., Tromp, M. C., and van Boom, J. H. (1983). Activation of the genes for variant surface glycoproteins 117 and 118 in *Trypanosoma brucei*. *J.Mol.Biol.* **166**, 537-556.

Michiels, F., Matthyssens, G., Kronenberger, P., Pays, E., Dero, B., Van Assel, S., Darville, M., Carvador, A., Steinert, M., and Hamers, R. (1983). Gene activation and re-expression of a *Trypanosoma brucei* variant surface glycoprotein. *EMBO J.* **2**, 1185-1192.

Miller, E. N. and Turner, M. J. (1981). Analysis of antigenic types appearing in first relapse populations of clones of *Trypanosoma brucei*. *Parasitology* **82**, 63-80.

Moreau, S., Ferguson, J. R., and Symington, L. S. (1999). The nuclease activity of Mre11 is required for meiosis but not for mating type switching, end joining, or telomere maintenance. *Mol.Cell Biol.* **19**, 556-566.

Mortensen, U. H., Bendixen, C., Sunjevaric, I., and Rothstein, R. (1996). DNA strand annealing is promoted by the yeast Rad52 protein. *Proc.Natl.Acad.Sci.U.S.A* **93**, 10729-10734.

Moxon, E. R., Rainey, P. B., Nowak, M. A., and Lenski, R. E. (1994). Adaptive evolution of highly mutable loci in pathogenic bacteria. *Curr.Biol.* **4**, 24-33.

Myler, P. J., Aline, R. F., Jr., Scholler, J. K., and Stuart, K. D. (1988). Multiple events associated with antigenic switching in *Trypanosoma brucei*. *Mol.Biochem.Parasitol.* **29**, 227-241.

Myung, K. and Kolodner, R. D. (2002). Suppression of genome instability by redundant S-phase checkpoint pathways in Saccharomyces cerevisiae. *Proc.Natl.Acad.Sci.U.S.A* **99**, 4500-4507.

Nagoshi, Y. L., Alarcon, C. M., and Donelson, J. E. (1995). The putative promoter for a metacyclic VSG gene in African trypanosomes. *Mol.Biochem.Parasitol.* **72**, 33-45.

Namsaraev, E. and Berg, P. (1997). Characterization of strand exchange activity of yeast Rad51 protein. *Mol.Cell Biol.* **17**, 5359-5368.

Namsaraev, E. A. and Berg, P. (2000). Rad51 uses one mechanism to drive DNA strand exchange in both directions. *J.Biol.Chem.* **275**, 3970-3976.

Nantulya, V. M., Musoke, A. J., Rurangirwa, F. R., and Moloo, S. K. (1984). Resistance of cattle to tsetse-transmitted challenge with *Trypanosoma brucei* or Trypanosoma congolense after spontaneous recovery from syringe-passaged infections. *Infect.Immun.* **43**, 735-738.

Nassif, N., Penney, J., Pal, S., Engels, W. R., and Gloor, G. B. (1994). Efficient copying of nonhomologous sequences from ectopic sites via P-element-induced gap repair. *Mol.Cell Biol.* **14**, 1613-1625.

Navarro, M. and Cross, G. A. (1996). DNA rearrangements associated with multiple consecutive directed antigenic switches in *Trypanosoma brucei*. *Mol.Cell Biol.* **16**, 3615-3625.

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

New, J. H., Sugiyama, T., Zaitseva, E., and Kowalczykowski, S. C. (1998). Rad52 protein stimulates DNA strand exchange by Rad51 and replication protein A. *Nature* **391**, 407-410.

Ogawa, T., Yu, X., Shinohara, A., and Egelman, E. H. (1993). Similarity of the yeast Rad51 filament to the bacterial RecA filament. *Science* **259**, 1896-1899.

Ohshima, K., Kang, S., Larson, J. E., and Wells, R. D. (1996). TTA.TAA triplet repeats in plasmids form a non-H bonded structure. *J.Biol.Chem.* **271**, 16784-16791.

Orr-Weaver, T. L. and Szostak, J. W. (1983). Yeast recombination: the association between double-strand gap repair and crossing-over. *Proc.Natl.Acad.Sci.U.S.A* **80**, 4417-4421.

Overath, P., Chaudhri, M., Steverding, D. and Ziegelbauer, K. (1994). Invariant surface proteins in bloodstream forms of *Trypanosoma brucei*. *Parasitol. Today* **10**, 53-58.

Overath, P., Haag, J., Lischke, A., and O'hUigin, C. (2001). The surface structure of trypanosomes in relation to their molecular phylogeny. *Int.J.Parasitol.* **31**, 468-471.

Paindavoine, P., Rolin, S., Van Assel, S., Geuskens, M., Jauniaux, J. C., Dinsart, C., Huet, G., and Pays, E. (1992). A gene from the variant surface glycoprotein expression site encodes one of several transmembrane adenylate cyclases located on the flagellum of *Trypanosoma brucei*. *Mol.Cell Biol.* **12**, 1218-1225.

Paques, F., Leung, W. Y., and Haber, J. E. (1998). Expansions and contractions in a tandem repeat induced by double-strand break repair. *Mol.Cell Biol.* **18**, 2045-2054.

Paques, F. and Haber, J. E. (1999). Multiple pathways of recombination induced by double-strand breaks in *Saccharomyces cerevisiae*. *Microbiol.Mol.Biol.Rev.* **63**, 349-404.

Paull, T. T. and Gellert, M. (1998). The 3' to 5' exonuclease activity of Mre 11 facilitates repair of DNA double-strand breaks. *Mol.Cell* **1**, 969-979.

Pays, E., Houard, S., Pays, A., Van Assel, S., Dupont, F., Aerts, D., Huet-Duvillier, G., Gomes, V., Richet, C., Degand, P., and . (1985). *Trypanosoma brucei*: the extent of conversion in antigen genes may be related to the DNA coding specificity. *Cell* **42**, 821-829.

Pays, E., Tebabi, P., Pays, A., Coquelet, H., Revelard, P., Salmon, D., and Steinert, M. (1989). The genes and transcripts of an antigen gene expression site from *T. brucei*. *Cell* **57**, 835-845.

Pays, E. (1989). Pseudogenes, chimaeric genes and the timing of antigen variation in African trypanosomes. *Trends Genet.* **5**, 389-391.

Pays, E., Lips, S., Nolan, D., Vanhamme, L., and Perez-Morga, D. (2001). The VSG expression sites of *Trypanosoma brucei*: multipurpose tools for the adaptation of the parasite to mammalian hosts. *Mol.Biochem.Parasitol.* **114**, 1-16.

Pazin, M. J. and Kadonaga, J. T. (1997). SWI2/SNF2 and related proteins: ATP-driven motors that disrupt protein-DNA interactions? *Cell* **88**, 737-740.

Petukhova, G., Stratton, S., and Sung, P. (1998). Catalysis of homologous DNA pairing by yeast Rad51 and Rad54 proteins. *Nature* **393**, 91-94.

Petukhova, G., Van Komen, S., Vergano, S., Klein, H., and Sung, P. (1999). Yeast Rad54 promotes Rad51-dependent homologous DNA pairing via ATP hydrolysis-driven change in DNA double helix conformation. *J.Biol.Chem.* **274**, 29453-29462.

Petukhova, G., Stratton, S. A., and Sung, P. (1999a). Single strand DNA binding and annealing activities in the yeast recombination factor Rad59. *J.Biol.Chem.* **274**, 33839-33842.

Porter, G., Westmoreland, J., Priebe, S., and Resnick, M. A. (1996). Homologous and homeologous intermolecular gene conversion are not differentially affected by mutations in the DNA damage or the mismatch repair genes *RAD1, RAD50, RAD51, RAD52, RAD54, PMS1 and MSH2. Genetics* **143**, 755-767.

Rao, B. J. and Radding, C. M. (1995). RecA protein mediates homologous recognition via non-Watson-Crick bonds in base triplets. *Philos.Trans.R.Soc.Lond B Biol.Sci.* **347**, 5-12.

Rattray, A. J. and Symington, L. S. (1995). Multiple pathways for homologous recombination in *Saccharomyces cerevisiae. Genetics* **139**, 45-56.

Reuner, B., Vassella, E., Yutzy, B., and Boshart, M. (1997). Cell density triggers slender to stumpy differentiation of *Trypanosoma brucei* bloodstream forms in culture. *Mol.Biochem.Parasitol.* **90**, 269-280.

Rice-Ficht, A.C.,Chen,K.K. and Donelson, J.E. (1981) Sequence homologies near the C-termini of the variable surface glycoproteins of *Trypanosoma brucei. Nature* **294**, 53-57.

Rijkers, T., Van Den, Ouweland J., Morolli, B., Rolink, A. G., Baarends, W. M., Van Sloun, P. P., Lohman, P. H., and Pastink, A. (1998). Targeted inactivation of mouse RAD52 reduces homologous recombination but not resistance to ionizing radiation. *Mol.Cell Biol.* **18**, 6423-6429.

Robinson, N. P., Burman, N., Melville, S. E., and Barry, J. D. (1999). Predominance of duplicative *VSG* gene conversion in antigenic variation in African trypanosomes. *Mol.Cell Biol.* **19**, 5839-5846.

Robinson, N. P., McCulloch, R., Conway, C., Browitt, A., and Barry, J. D. (2002). Inactivation of Mre11 does not affect *VSG* gene duplication mediated by homologous recombination in *Trypanosoma brucei. J.Biol.Chem.* **277**, 26185-26193.

Roditi, I., Furger, A., Ruepp, S., Schurch, N., and Butikofer, P. (1998). Unravelling the procyclin coat of *Trypanosoma brucei. Mol.Biochem.Parasitol.* **91**, 117-130.

Roth, C. W., Longacre, S., Raibaud, A., Baltz, T., and Eisen, H. (1986). The use of incomplete genes for the construction of a *Trypanosoma equiperdum* variant surface glycoprotein gene. *EMBO J.* **5**, 1065-1070.

Rudenko, G., Bishop, D., Gottesdiener, K., and Van der Ploeg, L. H. (1989). Alpha-amanitin resistant transcription of protein coding genes in insect and bloodstream form *Trypanosoma brucei*. *EMBO J.* **8**, 4259-4263.

Rudenko, G., McCulloch, R., Dirks-Mulder, A., and Borst, P. (1996). Telomere exchange can be an important mechanism of variant surface glycoprotein gene switching in *Trypanosoma brucei*. *Mol.Biochem.Parasitol.* **80**, 65-75.

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

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular cloning: a laboratory manual. 2edn, Cold Spring harbor Press.

Scholler, J. K., Myler, P. J., and Stuart, K. D. (1989). A novel telomeric gene conversion in *Trypanosoma brucei*. *Mol.Biochem.Parasitol.* **35**, 11-19.

Shah, J. S., Young, J. R., Kimmel, B. E., Iams, K. P., and Williams, R. O. (1987). The 5' flanking sequence of a *Trypanosoma brucei* variable surface glycoprotein gene. *Mol.Biochem.Parasitol.* **24**, 163-174.

Sheader, K., Berberof, M., Isobe, T., Borst, P., and Rudenko, G. (2003). Delineation of the regulated Variant Surface Glycoprotein gene expression site domain of *Trypanosoma brucei*. *Mol.Biochem.Parasitol.* **128**, 147-156.

Shinohara, A., Ogawa, H., and Ogawa, T. (1992). Rad51 protein involved in repair and recombination in *S. cerevisiae* is a RecA-like protein. *Cell* **69**, 457-470.

Shinohara, A. and Ogawa, T. (1998). Stimulation by Rad52 of yeast Rad51-mediated recombination. *Nature* **391**, 404-407.

Shinohara, A., Shinohara, M., Ohta, T., Matsuda, S., and Ogawa, T. (1998). Rad52 forms ring structures and co-operates with RPA in single-strand DNA annealing. *Genes Cells* **3**, 145-156.

Signon, L., Malkova, A., Naylor, M. L., Klein, H., and Haber, J. E. (2001). Genetic Requirements for *RAD51*- and *RAD54*-Independent Break-Induced Replication Repair of a Chromosomal Double-Strand Break. *Mol.Cell Biol.* **21**, 2048-2056.

Sloof, P., Menke, H. H., Caspers, M. P., and Borst, P. (1983). Size fractionation of *Trypanosoma brucei* DNA: localization of the 177-bp repeat satellite DNA and a variant surface glycoprotein gene in a mini-chromosomal DNA fraction. *Nucleic Acids Res.* **11**, 3889-3901.

Sonoda, E., Sasaki, M. S., Buerstedde, J. M., Bezzubova, O., Shinohara, A., Ogawa, H., Takata, M., Yamaguchi-Iwai, Y., and Takeda, S. (1998). Rad51-deficient vertebrate cells accumulate chromosomal breaks prior to cell death. *EMBO J.* **17**, 598-608.

Steverding, D., Stierhof, Y. D., Fuchs, H., Tauber, R., and Overath, P. (1995). Transferrin-binding protein complex is the receptor for transferrin uptake in *Trypanosoma brucei*. *J.Cell Biol.* **131**, 1173-1182.

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

Sugawara, N., Ivanov, E. L., Fishman-Lobell, J., Ray, B. L., Wu, X., and Haber, J. E. (1995). DNA structure-dependent requirements for yeast *RAD* genes in gene conversion. *Nature* **373**, 84-86.

Sugawara, N., Paques, F., Colaiacovo, M., and Haber, J. E. (1997). Role of *Saccharomyces cerevisiae* Msh2 and Msh3 repair proteins in double-strand break-induced recombination. *Proc.Natl.Acad.Sci.U.S.A* **94**, 9214-9219.

Sugawara, N., Wang, X., and Haber, J. E. (2003). *In vivo* roles of Rad52, Rad54, and Rad55 proteins in Rad51-mediated recombination. *Mol.Cell* **12**, 209-219.

Sugiyama, T., Zaitseva, E. M., and Kowalczykowski, S. C. (1997). A single-stranded DNA-binding protein is needed for efficient presynaptic complex formation by the *Saccharomyces cerevisiae* Rad51 protein. *J.Biol.Chem.* **272**, 7940-7945.

Sung, P. (1994). Catalysis of ATP-dependent homologous DNA pairing and strand exchange by yeast RAD51 protein. *Science* **265**, 1241-1243.

Sung, P. and Robberson, D. L. (1995). DNA strand exchange mediated by a RAD51-ssDNA nucleoprotein filament with polarity opposite to that of RecA. *Cell* **82**, 453-461.

Sung, P. (1997). Function of yeast Rad52 protein as a mediator between replication protein A and the Rad51 recombinase. *J.Biol.Chem.* **272**, 28194-28197.

Szostak, J. W., Orr-Weaver, T. L., Rothstein, R. J., and Stahl, F. W. (1983). The double-strand-break repair model for recombination. *Cell* **33**, 25-35.

Takata, M., Sasaki, M. S., Tachiiri, S., Fukushima, T., Sonoda, E., Schild, D., Thompson, L. H., and Takeda, S. (2001). Chromosome instability and defective recombinational repair in knockout mutants of the five Rad51 paralogs. *Mol.Cell Biol.* **21**, 2858-2866.

Tan, K. S., Leal, S. T., and Cross, G. A. (2002). *Trypanosoma brucei MRE11* is non-essential but influences growth, homologous recombination and DNA double-strand break repair. *Mol.Biochem.Parasitol.* **125**, 11-21.

Tan, T. L., Essers, J., Citterio, E., Swagemakers, S. M., de Wit, J., Benson, F. E., Hoeijmakers, J. H., and Kanaar, R. (1999). Mouse Rad54 affects DNA conformation and DNA-damage-induced Rad51 foci formation. *Curr.Biol.* **9**, 325-328.

Thacker, J. (1999). A surfeit of *RAD51*-like genes? *Trends Genet.* **15**, 166-168.

Thon, G., Baltz, T., and Eisen, H. (1989). Antigenic diversity by the recombination of pseudogenes. *Genes Dev.* **3**, 1247-1254.

Thon, G., Baltz, T., Giroud, C., and Eisen, H. (1990). Trypanosome variable surface glycoproteins: composite genes and order of expression. *Genes Dev.* **4**, 1374-1383.

Timmers, H. T., De Lange, T., Kooter, J. M., and Borst, P. (1987). Coincident multiple activations of the same surface antigen gene in *Trypanosoma brucei*. *J.Mol.Biol.* **194**, 81-90.

Tracy, R. B. and Kowalczykowski, S. C. (1996). *In vitro* selection of preferred DNA pairing sequences by the *Escherichia coli* RecA protein. *Genes Dev.* **10**, 1890-1903.

Tracy, R. B., Baumohl, J. K., and Kowalczykowski, S. C. (1997). The preference for GT-rich DNA by the yeast Rad51 protein defines a set of universal pairing sequences. *Genes Dev.* **11**, 3423-3431.

Tran, H., Degtyareva, N., Gordenin, D., and Resnick, M. A. (1997). Altered replication and inverted repeats induce mismatch repair-independent recombination between highly diverged DNAs in yeast. *Mol.Cell Biol.* **17**, 1027-1036.

Treumann, A., Zitzmann, N., Hulsmeier, A., Prescott, A. R., Almond, A., Sheehan, J., and Ferguson, M. A. (1997). Structural characterisation of two forms of procyclic acidic repetitive protein expressed by procyclic forms of *Trypanosoma brucei*. *J.Mol.Biol.* **269**, 529-547.

Trujillo, K. M., Yuan, S. S., Lee, E. Y., and Sung, P. (1998). Nuclease activities in a complex of human recombination and DNA repair factors Rad50, Mre11, and p95. *J.Biol.Chem.* **273**, 21447-21450.

Tsubouchi, H. and Ogawa, H. (1998). A novel *mre11* mutation impairs processing of double-strand breaks of DNA during both mitosis and meiosis. *Mol.Cell Biol.* **18**, 260-268.

Turner, C. M., Barry, J. D., Maudlin, I., and Vickerman, K. (1988). An estimate of the size of the metacyclic variable antigen repertoire of *Trypanosoma brucei rhodesiense*. *Parasitology* **97 (Pt 2)**, 269-276.

Turner, C. M. and Barry, J. D. (1989). High frequency of antigenic variation in *Trypanosoma brucei rhodesiense* infections. *Parasitology* **99 Pt 1**, 67-75.

Turner, C. M. (1997). The rate of antigenic variation in fly-transmitted and syringe-passaged infections of *Trypanosoma brucei*. *FEMS Microbiol.Lett.* **153**, 227-231.

Tyler, K. M., Higgs, P. G., Matthews, K. R., and Gull, K. (2001). Limitation of *Trypanosoma brucei* parasitaemia results from density-dependent parasite differentiation and parasite killing by the host immune response. *Proc.R.Soc.Lond B Biol.Sci.* **268**, 2235-2243.

Ulbert, S., Chaves, I., and Borst, P. (2002). Expression site activation in *Trypanosoma brucei* with three marked variant surface glycoprotein gene expression sites. *Mol.Biochem.Parasitol.* **120**, 225-235.

Usui, T., Ohta, T., Oshiumi, H., Tomizawa, J., Ogawa, H., and Ogawa, T. (1998). Complex formation and functional versatility of Mre11 of budding yeast in recombination. *Cell* **95**, 705-716.

Van Den, Abbeele J., Claes, Y., van Bockstaele, D., Le Ray, D., and Coosemans, M. (1999). *Trypanosoma brucei* spp. development in the tsetse fly: characterization of the post-mesocyclic stages in the foregut and proboscis. *Parasitology* **118**, 469-478.

Van der Ploeg, L. H., Valerio, D., De Lange, T., Bernards, A., Borst, P., and Grosveld, F. G. (1982). An analysis of cosmid clones of nuclear DNA from *Trypanosoma brucei* shows that the genes for variant surface glycoproteins are clustered in the genome. *Nucleic Acids Res.* **10**, 5905-5923.

Van der Ploeg, L. H., Liu, A. Y., and Borst, P. (1984). Structure of the growing telomeres of Trypanosomes. *Cell* **36**, 459-468.

Van Komen, S., Petukhova, G., Sigurdsson, S., Stratton, S., and Sung, P. (2000). Superhelicity-driven homologous DNA pairing by yeast recombination factors Rad51 and Rad54. *Mol.Cell* **6**, 563-572.

van Leeuwen, F., Wijsman, E. R., Kuyl-Yeheskiely, E., van der Marel, G. A., van Boom, J. H., and Borst, P. (1996). The telomeric GGGTTA repeats of *Trypanosoma brucei* contain the hypermodified base J in both strands. *Nucleic Acids Res.* **24**, 2476-2482.

van Leeuwen, F., Wijsman, E. R., Kieft, R., van der Marel, G. A., van Boom, J. H., and Borst, P. (1997). Localization of the modified base J in telomeric *VSG* gene expression sites of *Trypanosoma brucei*. *Genes Dev.* **11**, 3232-3241.

van Leeuwen, F., Taylor, M. C., Mondragon, A., Moreau, H., Gibson, W., Kieft, R., and Borst, P. (1998). beta-D-glucosyl-hydroxymethyluracil is a conserved DNA modification in kinetoplastid protozoans and is abundant in their telomeres. *Proc.Natl.Acad.Sci.U.S.A* **95**, 2366-2371.

Van Meirvenne,N., Janssens,P.G. and Magnus, E. (1975) Antigenic variation in syringe passaged populations of T*rypanosoma* (*Trypanozoon*) *brucei*. I. Rationalization of the experimental approach. *Annales De La Societe Belge De Medecine Tropicale* **55**, 1-23.

Vanhamme, L., Poelvoorde, P., Pays, A., Tebabi, P., Van Xong, H., and Pays, E. (2000). Differential RNA elongation controls the variant surface glycoprotein gene expression sites of *Trypanosoma brucei*. *Mol.Microbiol.* **36**, 328-340.

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

Vassella, E., Reuner, B., Yutzy, B., and Boshart, M. (1997). Differentiation of African trypanosomes is controlled by a density sensing mechanism which signals cell cycle arrest via the cAMP pathway. *J.Cell Sci.* **110 (Pt 21)**, 2661-2671.

Vickerman, K. (1965). Polymorphism and mitochondrial activity in sleeping sickness trypanosomes. *Nature* **208**, 762-766.

Vickerman, K. (1969). On the surface coat and flagellar adhesion in trypanosomes. *J.Cell Sci.* **5**, 163-193.

Vickerman, K. (1985). Developmental cycles and biology of pathogenic trypanosomes. *Br.Med.Bull.* **41** , 105-114.

Vijayasarathy, S., Ernest, I., Itzhaki, J. E., Sherman, D., Mowatt, M. R., Michels, P. A., and Clayton, C. E. (1990). The genes encoding fructose bisphosphate aldolase in *Trypanosoma brucei* are interspersed with unrelated genes. *Nucleic Acids Res.* **18**, 2967-2975.

Wang, J., Bohme, U., and Cross, G. A. (2003). Structural features affecting variant surface glycoprotein expression in *Trypanosoma brucei*. *Mol.Biochem.Parasitol.* **128**, 135-145.

Weiden, M., Osheim, Y. N., Beyer, A. L., and Van der Ploeg, L. H. (1991). Chromosome structure: DNA nucleotide sequence elements of a subset of the minichromosomes of the protozoan *Trypanosoma brucei*. *Mol.Cell Biol.* **11**, 3823-3834.

Welburn, S. C., Fevre, E. M., Coleman, P. G., Odiit, M., and Maudlin, I. (2001). Sleeping sickness: a tale of two diseases. *Trends Parasitol.* **17**, 19-24.

West, S. C. (1997). Processing of recombination intermediates by the RuvABC proteins. *Annu.Rev.Genet.* **31**, 213-244.

White, C. I. and Haber, J. E. (1990). Intermediates of recombination during mating type switching in Saccharomyces cerevisiae. *EMBO J.* **9**, 663-673.

Yamaguchi-Iwai, Y., Sonoda, E., Buerstedde, J. M., Bezzubova, O., Morrison, C., Takata, M., Shinohara, A., and Takeda, S. (1998). Homologous recombination, but not DNA repair, is reduced in vertebrate cells deficient in RAD52. *Mol.Cell Biol.* **18**, 6430-6435.

Ziegelbauer, K., Quinten, M., Schwarz, H., Pearson, T. W., and Overath, P. (1990). Synchronous differentiation of *Trypanosoma brucei* from bloodstream to procyclic forms *in vitro*. *Eur.J.Biochem.* **192**, 373-378.

Zomerdijk, J. C., Ouellette, M., ten Asbroek, A. L., Kieft, R., Bommer, A. M., Clayton, C. E., and Borst, P. (1990). The promoter for a variant surface glycoprotein gene expression site in *Trypanosoma brucei*. *EMBO J.* **9**, 2791-2801.

Zomerdijk, J. C., Kieft, R., Shiels, P. G., and Borst, P. (1991). Alpha-amanitin-resistant transcription units in trypanosomes: a comparison of promoter sequences for a *VSG* gene expression site and for the ribosomal RNA genes. *Nucleic Acids Res.* **19**, 5153-5158.