MISMATCH REPAIR, RECOMBINATION AND GENETIC VARIABILITY IN *TRYPANOSOMA BRUCEI*

REBECCA LOUISE BARNES

WELLCOME CENTRE FOR MOLECULAR PARASITOLOGY
GLASGOW BIOMEDICAL RESEARCH CENTRE
UNIVERSITY OF GLASGOW

SUPERVISOR: RICHARD MCCULLOCH

Submitted for the degree of Doctor of Philosophy
June 2006

Abstract

The protozoan parasite *Trypanosoma brucei* has a complex life cycle with stages in mammalian hosts, where it is the causative agent of sleeping sickness in humans and nagana in cattle, and in the tsetse fly vector. In order to evade the host immune system, *T. brucei* undergoes a process called antigenic variation in the mammalian bloodstream. In this process, a single Variant Surface Glycoprotein (VSG) is expressed on the cell surface, acting as a protective coat. The molecular identity of the VSG coat is periodically and spontaneously changed by a number of different switching mechanisms. *T. brucei* is known to have conserved DNA repair pathways, including homologous recombination (HR) and mismatch repair (MMR). While the central recombination factor RAD51 and its paralogue RAD51-3 have been shown to be important, but not essential, in *VSG* switching, a number of other factors, including the MMR proteins MSH2 and MLH1, seem not to be involved. Work in this thesis sought to examine several aspects of MMR function in *T. brucei*, and concentrated on homologues of the bacterial MutS protein.

The requirements for substrate length and sequence homology in *T. brucei* HR were studied using a DNA transformation assay. It was shown that reduction in either the length or the sequence identity of recombination substrates causes a significant reduction in the transformation efficiency of linear DNA, at least at an interstitial site. Genetic disruption of the *MSH2* gene only seemed to affect HR using substrates over 100 bp in length and with 5% divergence from the target sequence; shorter sequences and sequences with either 0% or 11% mismatches apparently remained unaffected. A number of transformants from all classes of transformation retained an undisturbed copy of the target locus, hypothesised to be due to low-level trisomy within the population. In addition, and at a very low rate, distinct recombination events, resulting in observable changes in the *T. brucei* chromosomes, were observed. This work reveals some of the factors which influence the pathways of recombination used by *T. brucei*.

A potential role for *T. brucei* homologues of the meiosis-specific MutS homologues MSH4 and MSH5 was also examined. Sequence comparisons show that these genes are present in *T. brucei* and the related kinetoplastids, *T. cruzi* and *L. major*. Like their orthologues in other organisms, *T. brucei* MSH4 and MSH5 lack a detectably functional mismatch interaction domain. Although MSH4 and MSH5 would only be expected to be required at the epimastigote life cycle stage, expression of *MSH5* can be detected by northern blot in procyclic form and bloodstream stage cells. Although creating genetic knockouts of these

genes was not successful, attempts were made to force expression of *MSH4* and *MSH5* ORFs from an ectopic locus, though this did not disrupt MMR function, nor reveal other observable phenotypes.

Finally, potential variation in MMR gene sequence and MMR functions in different *T. brucei* strains and subspecies was investigated. Many bacterial strains, known as mutators, have mutations in MMR genes, causing impaired MMR function and therefore increased variability in the population. It has recently been reported that this phenomenon is also observed in *T. cruzi*. *MSH2* and *RAD51* nucleotide and protein sequences were compared between nine *T. brucei* strains, and showed extremely low levels of polymorphism. However, four *T. brucei* strains were found to vary in their tolerance to the DNA damaging agents MNNG, H_2O_2 and MMS; whether this is due to differences in MMR, another DNA repair pathway, or drug uptake, is yet to be determined.

Table of Contents

List of Tables

List of Figures

Acknowledgements

A massive thank you to Richard, a great supervisor and a great guy. We really stuck it to the Man.

Thanks to the Barry and McCulloch groups past and present for all their help and suggestions.

Thanks to Annette MacLeod for supplying genomic DNA samples and microsatellite primers for chapter 5.

Thanks to Alice da Silva Machado for technical help and discussion.

Thanks to Derek Pike for amazing help with the statistics in Chapter 3.

Thanks to the Aragon, Byres Road for supplying 5% ethanol, and to the perverts and extroverts at the Cecilian Society for supplying scandal. Thanks to Monkey for being my cheerleader for the last few months. Special recognition also goes to the BG crew support group.

And of course, thank you to Mum, Dad and Emma for a lifetime's logistical, emotional and financial support.

Author's Declaration

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

Rebecca Barnes

Abbreviations

A	adenine
aa	amino acid
ADP	adenosine diphosphate
AE	axial element of the synaptonemal complex
ATP	adenosine triphosphate
BER	base excision repair
BES	bloodstream expression site
bp	base pair
BIR	break-induced replication
BSD	blasticidin S deaminase
C	cytosine
cAMP	cyclic adenosine monophosphate
cisplatin	cis-diamminedichloroplatinum(II)
CO	crossover
COI	crossover interference
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleotide triphosphate
DSB	double-strand break
EATRO	East African Trypanosomiasis Research Organisation
EDTA	ethylenediaminetetraacetic acid
ELC formation	expression-linked copy formation
ESAG	expression-site-associated gene
G	guanine
GPI	glycophospatidylinositol
HJ	Holliday junction
HNPCC	human nonpolyposis colorectal carcinoma
HR	homologous recombination
HTUB	hygromycin-tubulin
Hyg^R	hygromycin resistant
Hyg^S	hygromycin sensitive
kb	kilobase pair
LB broth	Luria-Bertani broth
LE	lateral element of the synaptonemal complex
MBSU	Molecular Biology Sequencing Unit

MEPS	minimal efficient processing segment
MES	metacyclic expression site
MMEJ	microhomology-mediated end-joining
MMS	methyl methane sulfonate
MMR	mismatch repair
MNE buffer	MOPS/sodium acetate/EDTA buffer; 1x: 0.024 M MOPS, 5mM NaOAC, 1mM EDTA, pH adjusted to 7 with NaOH
MNNG	N-methyl-N'-nitroN-nitrosoguanidine
mRNA	messenger RNA
MRX complex	Mre11, Rad50, Xrs2 complex
MTT	3[4,5-dimethylthiazol-2-yl]-2,5-diphenyltetrazolium bromide
NER	nucleotide excision repair
NCO	noncrossover
NHEJ	non-homologous end-joining
nt	nucleotide
ORF	open reading frame
PARP	procyclic acidic repetitive protein
PCNA	proliferating cell nuclear antigen
PCR	polymerase chain reaction
PUR	puromycin N acetyltransferase
RNA	ribonucleic acid
RNAi	RNA interference
RPA	replication protein A
rRNA	ribosomal RNA
RT-PCR	reverse transcriptase polymerase chain reaction
SC	synaptonemal complex
SDS	sodium dodecyl sulphate
SDSA	synthesis-dependent strand annealing
SNP	single-nucleotide polymorphism
SSA	single-strand annealing
ssDNA	single-stranded DNA
STIB	Swiss Tropical Institute, Basel
T	thymine
TAE	TRIS/acetate/EDTA buffer; 1x: 0.04 M TRIS base, 0.04 M glacial acetic acid, 1mM EDTA

TB1/10E	TRIS/borate/1/10 EDTA buffer; 1x: 0.089 M TRIS base, 0.089 M ortho-boric acid, 0.2 mM EDTA
TBE	TRIS/borate/EDTA buffer; 1x: 0.089 M TRIS base, 0.089 M ortho-boric acid, 2 mM EDTA
TREU	Trypanosomiasis Research, Edinburgh University
UTR	untranslated region
UV	ultraviolet
VSG	variant surface glycoprotein
Wt	wild-type

CHAPTER 1

GENERAL INTRODUCTION

1 General introduction

1.1 *Trypanosoma brucei*

The protozoan parasite *Trypanosoma brucei* is the causative agent of trypanosomiasis in humans (Human African Trypanosomiasis, or sleeping sickness), as well as in cattle (nagana) and in a range of wild animals, which provide a reservoir of human infection. Trypanosomiasis is endemic to 36 countries of sub-Saharan Africa, with 60 million people at risk of disease; in 1998, 45,000 cases of sleeping sickness were reported, although the World Health Organisation has estimated that the actual prevalence of the disease could be up to ten times higher (www.who.int/mediacentre/factsheets/fs259/en), although more recently the numbers have fallen to 70,000 (Barrett, *et al.*, 2003). *T. brucei* is spread by the tsetse fly (*Glossina* spp.), with a chancre or local inflammatory reaction visible at the site of the bite, after which parasites pass into the bloodstream and lymphatic tissues. The symptoms of early human infection include a general malaise, fever, weakness and weight loss, whereas late stage infection, when the parasites have crossed the blood-brain barrier into the central nervous system, causes psychiatric, motor and sensory disorders and sleep disturbances, followed by seizures, somnolence and coma. Sleeping sickness is fatal in many cases if left untreated, with disease severity sometimes depending on host genotype (Sternberg, 2004). Two different subspecies of *T. brucei* are infective to humans, with different disease outcomes: *T. b. gambiense*, endemic to central and western Africa, causes a chronic infection where progression to late stage infection can take years, whereas *T. b. rhodesiense*, found in southern and eastern parts of the continent, causes an acute form of the disease which can progress to late stage within months (Sternberg, 2004). The few drugs available for the treatment of sleeping sickness are all associated with major problems, including unacceptable side effects and an increasing rate of treatment failure (Barrett *et al.*, 2003; Kennedy, 2004). A third *T. brucei* subspecies, *T. b. brucei*, is not infective to humans but can still infect cattle and wildlife (Gibson, 2002). The reasons for this difference in host range are discussed in section 5.2.1.1.

The *T. brucei* genome, the near-complete sequence of which has been published (Berriman *et al.*, 2005), contains 11 diploid chromosomes, known as megabase chromosomes. These range in size from 0.9 to over 6 Mb, with a total haploid genomic content of 26 Mb, and are predicted to contain 9068 genes. In addition (El Sayed *et al.*, 2000), there are several intermediate chromosomes of between 200 and 900 kb, and about 100 minichromosomes measuring between 50 and 150 kb. Minichromosomes are composed mainly of repetitive,

palindromic sequences (El Sayed *et al.*, 2000; Wickstead *et al.*, 2004) that are present also on the intermediate chromosomes. To date, only *VSG* and expression site-associated genes (see section 1.1.4) have been found on these chromosomes (Melville *et al.*, 1998; Wickstead *et al.*, 2004). The published *T. brucei* genome sequence was derived from the TREU 927/4 strain; however, genomic content can differ by up to 25% between strains, and the sizes of individual chromosomes can vary by a similar amount between the allelic copies of a chromosome in the same strain (Melville *et al.*, 2000; El Sayed *et al.*, 2000). Much of this fluctuation is thought to be due to telomeric and subtelomeric rearrangements possibly associated with antigenic variation, as described below (Melville *et al.*, 2000).

T. brucei belongs to the family Trypanosomatidae, of the protist order Kinetoplastida and the order Eugelenozoa. The order Kinetoplastida is defined by the presence of the kinetoplast, a single large mitochondrion made up of a complex network of circular DNA molecules (Liu *et al.*, 2005). One peculiarity of the Kinetoplastida is the organisation of their genes into polycistronic transcription units, as discussed in more detail in section 4.2.5.1. Phylogenetic analysis of the eukaryotes based on a number of protein sequences estimated divergence of this order from the eukaryotic lineage at approximately 1.98 billion years ago (Hedges *et al.*, 2004; see figure 1.1). The genomes of two other human-infective members of the Trypanosomatidae family, *Trypanosoma cruzi* and *Leishmania major*, which cause Chagas disease and leishmaniasis, respectively, have also been sequenced (El Sayed *et al.*, 2005; Ivens *et al.*, 2005). Although these three parasites are grouped in different branches of the trypanosomatid phylogeny (Hughes and Piontkivska, 2003), there are high levels of synteny between *T. brucei*, *T. cruzi* and *L. major* (Ghedin *et al.*, 2004).

1.1.1 The T. brucei *life cycle*

T. brucei has a complex life cycle, with stages in both mammalian hosts (including humans and cattle) and the tsetse fly *Glossina* spp (Barry and McCulloch, 2001; see figure 1.2). Life cycle stages, all of which are extracellular, were identified by virtue of differences in morphology, for example movement of the kinetoplast, from the posterior end of the cell in the bloodstream stage to a more central location in the procyclic form (Hendriks *et al.*, 2000). The replicative stages are the long slender bloodstream form, procyclic form, and epimastigote form, which undergo constant mitotic division and cause infection in the mammalian bloodstream and in the midgut and the salivary glands of the tsetse, respectively. Transmission stages, on the other hand, do not undergo mitosis and appear to

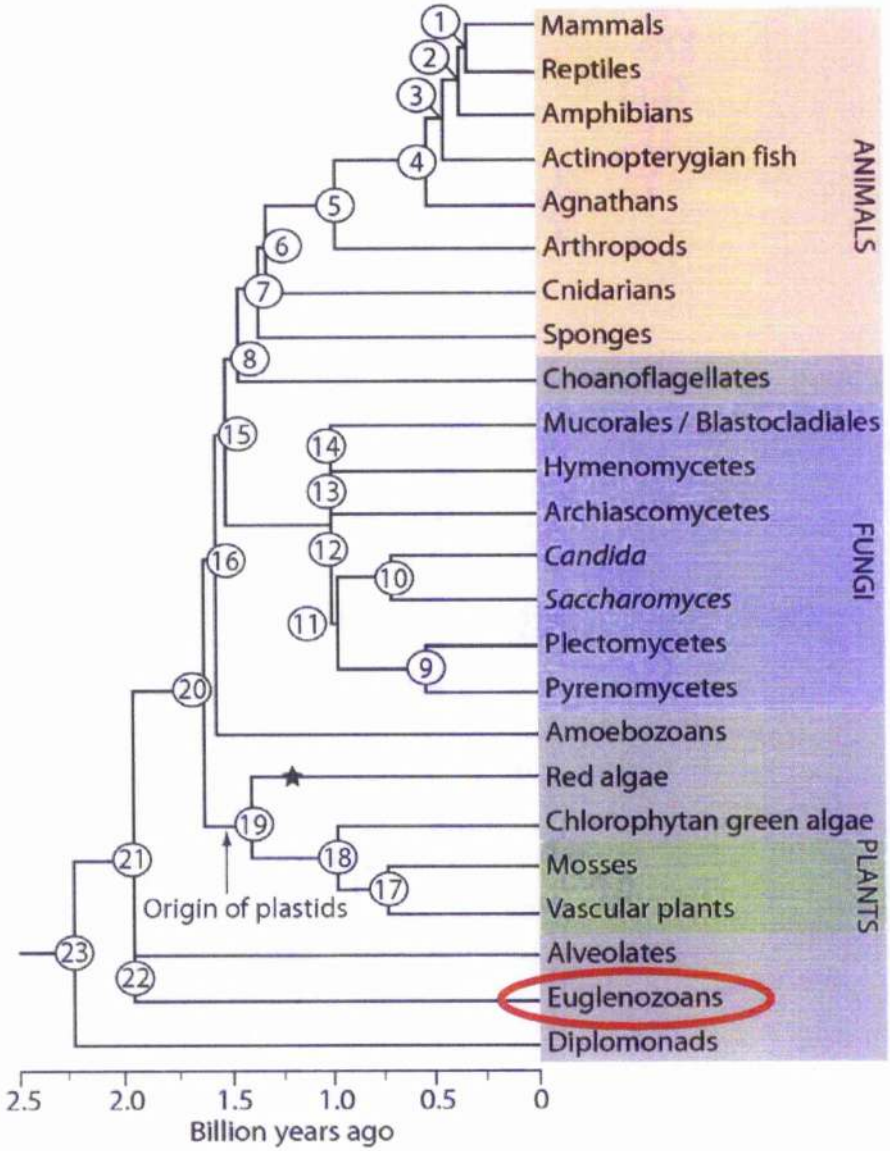


Figure 1.1. A phylogenetic tree of the eukaryotes. A phylogenetic tree of a number of eukaryotic organisms was constructed based on a number of protein sequences, and used to calculate times of divergence of various organisms from the tree. *T. brucei* belongs to the Euglenozoan order, circled in red. Taken from Hedges *et al.*, 2004.

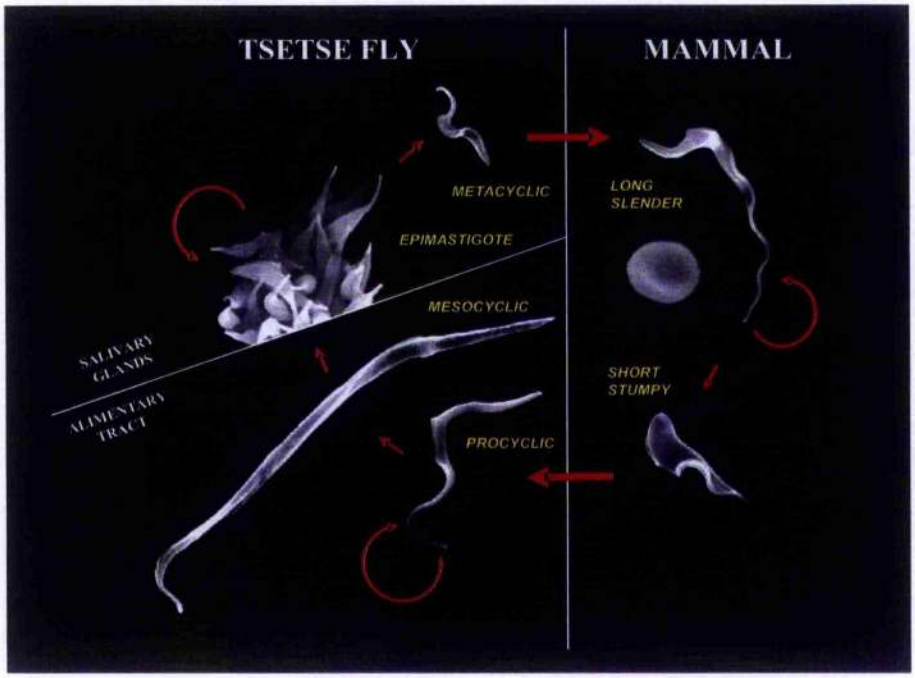


Figure 1.2. The *T. brucei* life cycle. *T. brucei* life cycle stages are shown as scanning electron micrographs, shown to scale; an erythtrocyte is shown next to the long slender bloodstream stage for comparison. The host organism and the name of the life cycle stage are indicated. Circular arrows represent replicative stages, whereas straight arrows represent differentiation and progression through the life cycle. Taken from Barry and McCulloch, 2001.

be cell cycle arrested. These are the bloodstream short stumpy stage, capable of transmission from the mammalian bloodstream to the tsetse fly, the metacyclic stage, capable of transmission from the tsetse salivary gland to the mammalian host (both as part of an infected bloodmeal), and the mesocyclic form, capable of transmission from the tsetse alimentary tract to the salivary gland. Transmission stage parasite forms appear to be a dead end in the life cycle if not passed on to the next host; for example, it has been observed that short stumpy cells can only survive for a matter of days if not transmitted to the tsetse fly (Turner *et al.*, 1995). Short stumpy form trypanosomes also display partial adaptations to survival in the tsetse fly; for example, they display increased resistance to acidity and proteolytic stress compared to long slender cells (Nolan *et al.*, 2000). A similar pre-adaptation is seen in tsetse metacyclic form cells, which express a *VSG* coat (see below) in preparation for transition into the mammalian bloodstream (Ginger *et al.*, 2002). Differentiation from the long slender to short stumpy bloodstream forms is dependent on cell density in the host bloodstream (Reuner *et al.*, 1997), communicated by a compound known as "stumpy induction factor", which triggers cell cycle arrest in G_1/G_0 phase through a cAMP signalling pathway (Vassella *et al.*, 1997).

Bloodstream stage *T. brucei* are covered by a protective coat of Variant Surface Glycoprotein (VSG), which shields invariant surface molecules from the host immune system, protects the cell against complement-mediated lysis, and undergoes antigenic variation (as discussed in section 1.1.5). In contrast, tsetse fly-infective cycle stages are covered by procyclin, also known as procyclic acidic repetitive protein (PARP; not to be confused with poly(ADP-ribose) polymerase, an enzyme involved in DNA damage signalling). Both VSGs and procyclins are anchored to the cell surface by glycophosphatidylinositol (GPI) anchors (Boothroyd *et al.*, 1980; Ferguson, 1999; Matthews *et al.*, 2004). Once the short stumpy form has been ingested as part of a tsetse bloodmeal, loss of VSG and their replacement by procyclin on the surface of the parasite is completed in a matter of hours (Hendriks *et al.*, 2000; Matthews *et al.*, 2004). Different procyclins are characterised by different internal repeats: EP forms contain internal Glu-Pro repeats, whereas GPEET forms contain the Gly-Pro-Glu-Glu-Thr repeat (Roditi *et al.*, 1998). These variants are expressed differentially, in different isoforms, throughout tsetse infection (Acosta-Serrano *et al.*, 2001; Roditi and Liniger, 2002). Some doubt remains as to the function of the procyclin coat, but roles have been suggested in protection against trypanocidal factors found in the midgut, and detection of tsetse-specific factors stimulating further differentiation (Roditi *et al.*, 1998).

1.1.2 Antigenic variation is a conserved mechanism for the evasion of host immunity

A challenge facing all pathogenic organisms is survival in the face of host immunity. To this end, many pathogens employ a system of highly mutable "contingency genes" (Moxon *et al.*, 1994), which make up only a small subset of the genome, yet allow a much higher rate of phenotypic variation than would be possible using only spontaneous mutations. In fact, contingency genes' functions are not limited to the avoidance of host immunity, but allow enhanced propagation, including differential tissue binding, cell invasion or avoidance of phagocytosis. This strategy has been observed in viral, bacterial, fungal and protozoal pathogens (Barry *et al.*, 2003), and can be classified either as phase variation or antigenic variation, as described below. In most of these systems the rate of variation is surprisingly similar (about 10^{-2} to 10^{-3} switch/cell/population doubling time), even though the considerable diversity in mechanisms used indicates convergent evolution (Deitsch *et al.*, 1997).

Many bacterial pathogens undergo phase variation, where expression of a given factor is controlled by a simple "on/off" switch (Henderson *et al.*, 1999). The most-cited examples of phase variation are to be found in Gram-negative bacteria, where changes in surface structures lead to differences in observable phenotypes such as cellular aggregation and colony morphology. Some mechanisms of phase variation involve recombination-mediated genetic rearrangements: either site-specific recombination such as involving the fimA gene in *Escherichia coli* (Kulasekara and Blomfield, 1999), or RecA-dependent homologous recombination, such as involving the type IV pilin in *Neisseria gonorrhoea* (Mehr and Seifert, 1998). Alternatively, gene expression can be modulated by slipped-strand mispairing, where highly mutable short sequence repeats can be used as a switch either by affecting a gene's promoter or by altering its reading frame; examples of this are seen in a number of *Helicobacter pylori* genes (Saunders *et al.*, 1998).

Antigenic variation is distinguished from phase variation by the fact that it is strictly concerned with evasion of acquired immunity through the expansion of alternative surface molecules. As such, it is more complex, since it involves switches between multiple surface antigens, rather than switches between two different states, as is seen in phase variation. In intracellular life cycle stages of *Plasmodium falciparum*, the apicomplexan parasite that is the causative agent of malaria, highly polymorphic parasite-encoded antigens are expressed on the surface of parasitized erythrocytes. These are encoded by

three families of genes, the *var* genes (which encode PfEMP1, or *P. falciparum* erythrocyte membrane protein 1), the *rif* (repetitive interspersed family) genes, and the *stevor* (subtelomeric variant open reading frame) genes, which are expressed one (or a very few) at a time (Deitsch and Hviid, 2004). For the *var* genes, this is controlled by transcriptional silencing agents (Deitsch *et al.*, 2001) which mediate changes in chromatin structure (Deitsch, 2005); for the other gene families, the method of control is not known. As well as acting in immune evasion, some of these factors also have roles in virulence and pathogenesis, as they affect the ability of the red cells to adhere to vascular endothelium, causing accumulation of parasites in certain organs, in particular the brain (Kyes *et al.*, 2001). Another example of antigenic variation can be found in the protozoan intestinal parasite, *Giardia lamblia* (Nash, 2002). In this organism, expression of a single 20-200 aa, cysteine-rich Variant-specific Surface Protein (VSP) from a repertoire of approximately 150 is controlled by transcriptional mechanisms (Mowatt *et al.*, 1991). *T. brucei* also uses antigenic variation in the evasion of host immunity; this mechanism involves both transcriptional and recombinational processes, as outlined below.

1.1.3 T. brucei *Variant Surface Glycoproteins*

VSG molecules are the antigens that undergo switching in *T. brucei* antigenic variation. They cover the surface of bloodstream form cells in a densely-packed monolayer (see figure 1.3), shielding necessarily invariant conserved surface molecules (Borst and Fairlamb, 1998) from the mammalian immune system (Overath *et al.*, 1994). This protective function causes an immune response to be mounted against the expressed VSG, necessitating periodic changing of the coat as described in section 1.1.5. There are approximately 5.5×10^6 VSG homodimers on the surface of each *T. brucei* cell (Cross, 1975; Auffret and Turner, 1981), and these are endocytosed and recycled at a very high rate *via* the flagellar pocket (Overath and Engstler, 2004).

VSGs are generally between 400 and 500 aa long, and are made up of two domains: the C-terminal domain of 40-80 aa, closest to the plasma membrane of the cell; and the N-terminal domain of 350-400 aa, which extends beyond the invariant surface molecules and contains the exposed epitopes (Pays, Salmon, Morrison, Marcello and Barry, in press). While only minimal primary sequence conservation within these domains can be detected, 3D crystallographic structures of both the N-terminal (Blum *et al.*, 1993) and C-terminal (Chattopadhyay *et al.*, 2005) domains demonstrate that secondary and tertiary structure are well conserved. The C-terminal domain has an elongated structure, which is proposed to

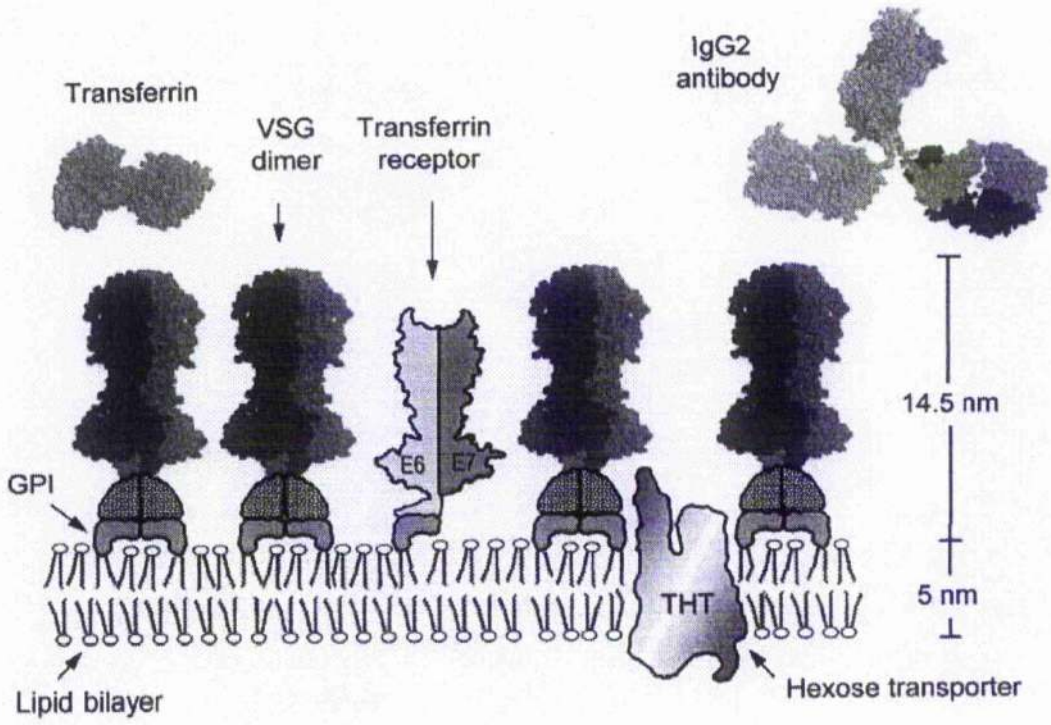


Figure 1.3. A schematic representation of the cell surface of bloodstream form *T. brucei*. VSG dimers (attached to the GPI anchor via the C-terminal domain), a transferrin receptor and an hexose transporter are shown associated with the plasma membrane. A transferrin and immunoglobulin G (IgG2) molecule are also shown for size comparison. Taken from Borst and Fairlamb, 1998

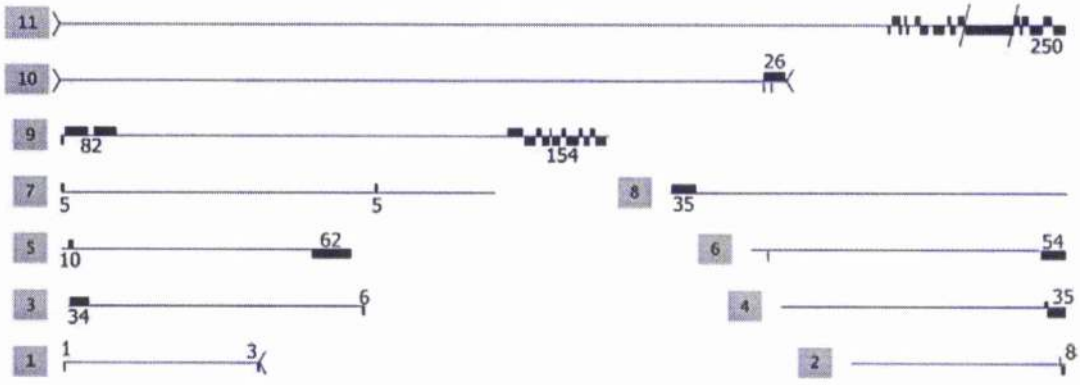


Figure 1.4. Schematic representation of the silent *VSG* repertoire on the megabase chromosomes of *T. brucei* strain TREU 927/4. Chromosomes are represented by horizontal lines, with the chromosome number in a grey box to the left of each chromosome. Arrays of *VSG*s are depicted by black blocks; the orientation of sets of *VSG*s is shown by the position of the box above or below the line. The provisional number of *VSG*s in each array is shown. Breaks in contiguation are represented by oblique lines. Taken from Barry *et al.*, 2005.

allow the exposed N-terminal domain to extend out substantially from the plasma membrane (Chattopadhyay *et al.*, 2005). The N-terminal domain is made up of a "VSG fold" of two antiparallel alpha helices, which interact with the corresponding helices of the partner VSG of the homodimer creating a coiled-coil structure. Both domains can be classified into different groups based on the number and organisation of cysteine residues (Carrington *et al.*, 1991), and although certain combinations of N and C domains seem to be preferred, all are possible (Carrington *et al.*, 1991; Hutchinson *et al.*, 2003).

The basis of antigenic variation is that *VSG*s are expressed one at a time, with a single *VSG* being selected from a large repertoire of silent *VSG*s. The original estimate of the number of *VSG*s, calculated by probing cosmid libraries at low stringency with 70-bp repeat (see section 1.1.4) sequence, was approximately 1,000 (Van der Ploeg *et al.*, 1982), and analysis of the genome sequence has confirmed this figure to be roughly correct: the total number of *VSG*s on the megabase chromosomes has been estimated at 1,600, with 940 annotated to date (L. Marcello and J.D. Barry, pers. comm.). The total number will, however, be higher, as approximately 100-200 *VSG*s are found at subtelomeric sites within the minichromosomes (Wickstead *et al.*, 2004), which have not been sequenced. A number of surprising facts about the *VSG* archive have been uncovered thanks to genomic sequence analysis (Barry *et al.*, 2005). Firstly, silent *VSG* genes on the megabase chromosomes are found overwhelmingly in subtelomeric arrays (see figure 1.4), rather than in interstitial arrays as previously thought (Barry and McCulloch, 2001), with only 5 *VSG*s being found in interstitial locations (Barry *et al.*, 2005). This occupation of subtelomeres by genes involved in antigenic variation is also found in other pathogens, including *P. falciparum* (Scherf *et al.*, 2001). One explanation for why this may be advantageous is that it allows for high levels of ectopic recombination (Barry *et al.*, 2003), perhaps facilitating gene diversification with minimal risk to essential chromosome-internal genes. Secondly, only 5% of genes in the *VSG* archive are fully functional, encoding all known features of known expressed *VSG*s; of the remaining genes, 9% are classified as atypical and are predicted not to be adequately folded or modified due to defective GPI signals and/or cysteine pattern, 62% are pseudogenes with frameshifts and/or stop codons, and 19% are gene fragments, in many cases due to *ingi* insertion (Barry *et al.*, 2005; L. Marcello, pers.comm.).

The strategy of placing surface antigen genes in subtelomeric locations is also employed by a number of other organisms (Barry *et al.*, 2003), including the parasites *T. cruzi* (Chiurillo *et al.*, 1999) and *P. falciparum* (Scherf *et al.*, 2001), the eubacterium *Borrelia hermsii* (Kitten and Barbour, 1990), and the fungal parasite *Pneumocystis carnii* (Wada

and Nakamura, 1996). A notable exception to this trend is the protist *Giardia lamblia*, all of whose *vsp* genes are interstitial (Adam, 2000). The effectively aneuploid nature of the subtelomeres means that they can recombine ectopically with each other rather than being limited to homologous chromosomes (Horn and Barry, 2005). In many other organisms such as *H. sapiens*, the subtelomeres is where recombination and expansion of genomic repeat sequences is common (Linardopoulou *et al.*, 2005).

Metacyclic form trypanosomes encode a relatively specific set of *VSG*s, which are expressed for a few days after infection of the mammalian host before being replaced by bloodstream-specific *VSG* (Ginger *et al.*, 2002). The repertoire of metacyclic *VSG*s is much smaller than that of bloodstream stage *VSG*s, calculated by immunological methods to contain only 27 genes (Turner *et al.*, 1988), not all of which are known to be functional. However, gradual turnover of this small repertoire prevents the development of herd immunity (Barry *et al.*, 1983).

1.1.4 Expression sites

*VSG*s are expressed from specialised polycistronic transcription units called expression sites (see figure 1.5). There are approximately 20 of these in the *T. brucei* genome (Becker *et al.*, 2004) that are used in the bloodstream, which are referred to as bloodstream expression sites (BESs). In addition, metacyclic *VSG*s are expressed from metacyclic stage-specific expression sites (MESs). BESs and MESs are located exclusively in subtelomeric locations, with a single one being expressed at once. BESs vary between ~40 and 100 kb in size (Becker *et al.*, 2004), and have a variable but conserved structure (Berriman *et al.*, 2002). The *VSG* gene is located closest to the telomere, within approximately 5 kb of the telomeric repeats (Horn and Barry, 2005). Upstream of the *VSG* is a stretch of up to 20 kb (McCulloch *et al.*, 1997) of short, divergent repeat sequences known as the 70-bp repeats (Liu *et al.*, 1983), shorter stretches of which are also found upstream of most silent *VSG* genes in subtelomeric arrays (Barry *et al.*, 2005). Between 8 and 10 expression-site-associated genes (*ESAG*s) can be found upstream of the 70-bp repeats. Some of their functions are as yet unknown and in many cases they have degenerated to pseudogenes. In addition, several are members of multigene families with paralogues in interstitial locations (Pays *et al.*, 2001; Donelson, 2003). None of the *ESAG*s have been found to be directly involved in antigenic variation, although some have roles in host-parasite interactions. The human serum resistance-associated gene which confers human infectivity on certain *T. brucei* strains (see section 5.2.1.1), is expressed as an *ESAG* (Xong *et al.*, 1998). *ESAG6* and *ESAG7*, the only *ESAG*s identified in every BES

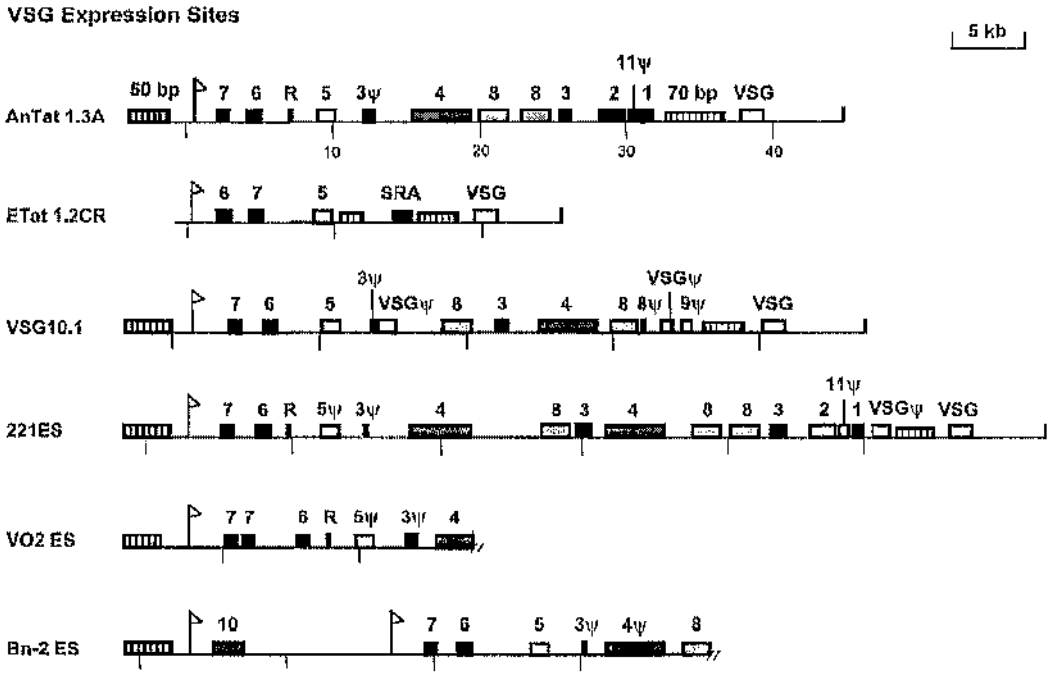


Figure 1.5. Some bloodstream expression sites of *T. brucei*. A schematic representation of several BESs (AnTat 1.3A, ETat 1.2CR, VSG10.1, 221ES, VO2 ES and Bn-2 ES). The *VSG* is indicated by a white box, the 50-bp and 70-bp repeat arrays (not to scale) by striped boxes, and the promoters by flags. The *ESAG*s and pseudo-*ESAG*s (indicated by ψ) are represented by numbered grey and black boxes. RHS genes are annotated R, and SRA genes SRA. Taken from Berriman *et al.*, 2002.

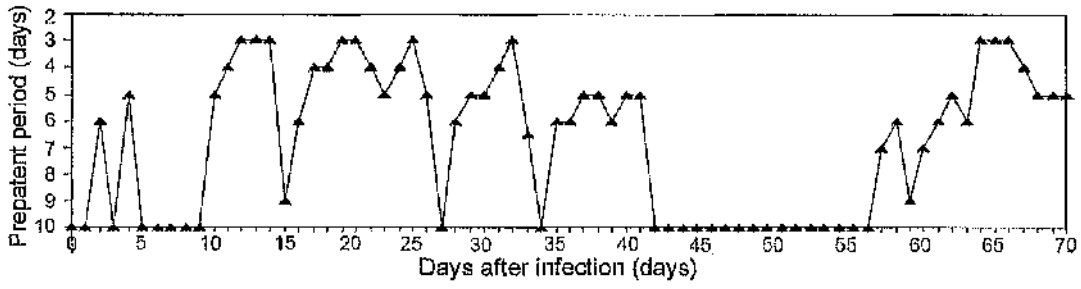


Figure 1.6. Chronic infection of a cow with *T. brucei* strain ILTat 1.2. The timeline of the 70-day infection is shown on the x-axis, whereas parasite density in the cow is shown as prepatent period on the y-axis; the number of days for parasitaemia to reach a certain level following inoculation of an immunosuppressed mouse with cattle blood. Taken from Morrison *et al.*, 2005.

studied, encode the two subunits of the *T. brucei* transferrin receptor (Schell *et al.*, 1991; Ligtenberg *et al.*, 1994), and are presumably essential, as they provide the parasite's sole means of iron uptake from the host bloodstream (Berriman *et al.*, 2002). Different alleles of *ESAG6* and *ESAG7* are found in different BESs, with differing affinities for transferrins from different hosts (Bitter *et al.*, 1998). This led to the theory that the use of different BESs helps to establish host specificity, though this has been cast into doubt when it was shown recently that differences in binding do not correspond to differences in growth in the sera of different mammalian species (Salmon *et al.*, 2005). Transcription of the active BES (and of procyclin genes) is carried out by RNA polymerase I, usually reserved in eukaryotes for the transcription of ribosomal RNA (Lee and Van der Ploeg, 1997; Laufer *et al.*, 1999; Gunzl *et al.*, 2003), although differences between BES and rRNA promoters have been identified (Zomerdijk *et al.*, 1990; Horn and Cross, 1997b). These genes all require transcription at very high rates. However, another possible reason for the use of RNA polymerase I is that it allows differential control of expression, rather than the unregulated transcription provided by RNA polymerase II; it is known that these polymerases interact with the RNA processing machinery in different ways (Vanhamme *et al.*, 2000). The BES is insulated from the subtelomeric *VSG* arrays by large arrays of repetitive sequences. Directly upstream of the BES promoter are arrays of 50-bp repeats which can stretch 40-50 kb upstream (Zomerdijk *et al.*, 1990; Zomerdijk *et al.*, 1991), marking the boundary of BES-specific transcriptional control (Sheader *et al.*, 2003). Upstream of each 50-bp repeat array is an "island" consisting of tens to hundreds of kb of repetitive elements, such as the RIME and *ingi* transposons (Berriman *et al.*, 2002; Bringaud *et al.*, 2002), further separating the BES from chromosome-internal sequences.

Five MESs for the expression of metacyclic VSGs (Graham *et al.*, 1999) have been annotated to date. These are also found in subtelomeric locations and transcribed by RNA polymerase I (Barry and McCulloch, 2001; Berriman *et al.*, 2002). However, their structure is much simpler than that of the BESs, consisting simply of a promoter, a small stretch of 70-bp repeats and a single *VSG*. In fact, the MESs are the only *T. brucei* genes known to be transcribed monocistronically (Alarcon *et al.*, 1994; Nagoshi *et al.*, 1995). It is possible that MESs are derived from BESs, as they are flanked upstream by divergent *ESAG* sequence (Bringaud *et al.*, 2001).

A novel nucleotide, known as β-D-Glucosyl-hydroxymethyluracil or J (Gommers-Ampt *et al.*, 1993), has been discovered in *T. brucei*, and linked by some to the control of antigenic variation. This modified version of uracil is incorporated in place of a subset of thymine residues, and accounts for 0.2 mol% of nuclear DNA in *T. brucei* (van Leeuwen *et al.*,

1997). The presence of J at inactive, but not active, BESs (van Leeuwen *et al.*, 1997) suggested a role in antigenic variation. However, J has also been found in other kinetoplastids, even those incapable of antigenic variation (van Leeuwen *et al.*, 1998). In addition, it is not exclusive to telomeric structures but also found in other tandemly repeated sequences (van Leeuwen *et al.*, 2000). J is bound by JBP1 (J-binding protein 1), which promotes propagation of this unusual base (Cross *et al.*, 1999), and by the chromatin remodelling protein JBP2 (J-binding protein 2), a member of the SWI2/SNF2 family (Dipaolo *et al.*, 2005), and has been proposed to be an epigenetic marker of heterochromatin and/or repetitive DNA, rather than controlling antigenic variation (Borst and Ulbert, 2001; Pays *et al.*, 2004). The chromatin structure of the active BES is more open than that of silent ones (Pays *et al.*, 1981), and the active BES appears to contain increased amounts of single-stranded DNA, presumably associated with the high levels of transcription (Greaves and Borst, 1987). As for J, whether or not these characteristics are a cause rather than a consequence of BES expression is unclear, however.

1.1.5 VSG *switching in* T. brucei

Antigenic variation in *T. brucei* consists of the parasite periodically changing, or "switching", its VSG coat; reviews are available on this subject (Barry, 1997; Barry and McCulloch, 2001). Trypanosome infection of the mammalian host follows a pattern of recurring peaks of parasitaemia, as shown in figure 1.6 (Barry, 1986; Morrison *et al.*, 2005), caused by the host's anti-VSG antibody responses, and by density-dependent differentiation of parasites to the non-dividing short stumpy life cycle stage (Tyler *et al.*, 2001). *VSG* switching is spontaneous and is not triggered by this immune response, and can indeed occur *in vitro*, in the absence of antibodies (Doyle *et al.*, 1980), but allows the few cells that have switched their coat to survive immune destruction. A single peak of infection can be a mixture of descendents of several different *VSG* switching events (Miller and Turner, 1981; Robinson *et al.*, 1999), and even cells expressing the same VSG within an infection peak can arise from separate switching events (Timmers *et al.*, 1987). The genetic factors influencing antigenic variation are discussed in section 1.4.

The rate of *VSG* switching has been demonstrated to occur at rates as high as 10^{-2} to 10^{-3} switch/cell/population doubling time (Turner and Barry, 1989; Turner, 1997). However, in lab-adapted cell lines derived by serial syringe passage through rodents, the switching rate is much lower at only 10^{-6} to 10^{-7} switch/cell/population doubling time (Lamont *et al.*, 1986). Such lab adaptation causes the parasite to grow to high parasitaemia in rodents, and to the loss of the cells' ability to differentiate from the long slender to the short stumpy

form and hence to transmit through the tsetse fly. *T. brucei* strains that undergo this differentiation are referred to as "pleomorphic", whilst strains that have lost the capacity are called "monomorphic". The basis for this deficiency is unknown, as is its potential relationship with a reduction in *VSG* switching rate.

1.1.5.1 Transcriptional "*in situ*" switching

Two different *VSG*s can be expressed simultaneously if they are found in the same BES (Munoz-Jordan *et al.*, 1996), whereas simultaneous, maximal expression from two different BESs at once is impossible (Chaves *et al.*, 1999). Together, this implies that control of *VSG* expression must be acting at the level of BES, rather than *VSG*, selection. How is this phenomenon, known as allelic exclusion, regulated? An important breakthrough in understanding this was the discovery of a putative specialised nuclear compartment outside the nucleolus, known as the Expression Site Body (Navarro and Gull, 2001). This complex is associated with the active BES, but none of the silent ones. In addition, it is detectable in bloodstream, but not procyclic, stage cells. Apart from RNA polymerase I, the components of this structure are not known. Small amounts of transcript from silent BESs can be detected for *ESAG*s and BES sequence close to the promoter, whereas transcription of more than one *VSG* can never be detected (Ansorge *et al.*, 1999; Vanhamme *et al.*, 2000; Borst and Ulbert, 2001; Amiguet-Vercher *et al.*, 2004), showing that it is RNA processing and elongation, rather than differential initiation of transcription, that underlies the control of BES expression. The elements that underpin this reaction, either as DNA sequence or potential factors, are unknown. Telomere position effect could also contribute to this process. Telomere position effect was first observed in *Saccharomyces cerevisiae* (Gottschling *et al.*, 1990), and is a repression of transcription in telomere-proximal regions of the genome, dependent on the peripheral position of telomeres within the nucleus (Feuerbach *et al.*, 2002). This reaction has also been observed in *H. sapiens* (Baur *et al.*, 2001) and in *T. brucei* (Horn and Cross, 1995; Rudenko *et al.*, 1995), and it remains possible that it could contribute to the regulation of *VSG* expression.

One form of *VSG* switching is known as transcriptional or *in situ* switching, where transcription is simply moved from the promoter of one BES to another (see fig 1.7). No clear examples of genetic rearrangements have been shown to be involved in this form of switching reaction. Although some sequence alterations are seen in *in situ* switch reactions in one study (Navarro and Cross, 1996), and a significant minority of cases have been

described where the active BES is deleted (Cross *et al.*, 1998), other reactions have been identified without observable sequence changes (Horn and Cross, 1997a; Rudenko *et al.*, 1998). It seems likely, therefore, that *in situ* switches do not involve recombination of the BES sequences. Nevertheless, the mechanistic basis of the reaction, the factors involved and its relationship with BES silencing are not yet known. Recent work has shown that derepression of silent BESs, as well as other PolI transcription units, can be triggered by blocking transcription or subjecting the cell's genome to damage (Sheader *et al.*, 2004) so it is possible that the reaction is linked to DNA repair in as yet unknown ways.

1.1.5.2 Recombinational switching pathways

A number of *VSG* switching mechanisms have also been detected that clearly rely on rearrangement of the parasite's genetic content by homologous recombination (see fig 1.7). These recombinational mechanisms mainly involve duplicative transposition by gene conversion, *i.e.* a one-way transfer of genetic information from a silent locus (or loci) into the active BES, deleting the resident *VSG* locus but leaving the silent copy (or copies) intact for activation in subsequent infections. A number of distinct reactions fall in this class.

The first type of duplicative transposition, named expression-linked copy (ELC) formation, involves gene conversion of a silent *VSG* gene, from a tandem array in a megabase chromosome, a silent BES or the subtelomere of a minichromosome, into the active BES, with associated loss of the previously-expressed *VSG*. The gene conversion tract generally extends from the 70-bp repeats upstream (Liu *et al.*, 1983; Matthews *et al.*, 1990), and the downstream limit in this pathway of *VSG* switching is normally at homologies at the 3' end of the *VSG* coding sequence (Michels *et al.*, 1983), although it can also extend beyond the ORF (Michels *et al.*, 1983; Timmers *et al.*, 1987). In monomorphic lines the situation is slightly different: initiation of recombination can be observed up to 6 kb upstream of the *VSG* (Lee and Van der Ploeg, 1987), and indeed the 70-bp repeats can be deleted from the active BES of a monomorphic cell line without any change in switching frequencies (McCulloch *et al.*, 1997). It has been proposed that a specific endonuclease might create double-strand breaks in the 70-bp repeats to initiate recombination (Pays *et al.*, 1994; Robinson *et al.*, 1999), but none has yet been identified, and breaks could be created by other means. The fact that switching events do not share specific endpoints argues against a site-specific recombination mechanism for *VSG* switching, but rather in favour of homologous recombination (Rudenko *et al.*, 1998; Barry and McCulloch, 2001).

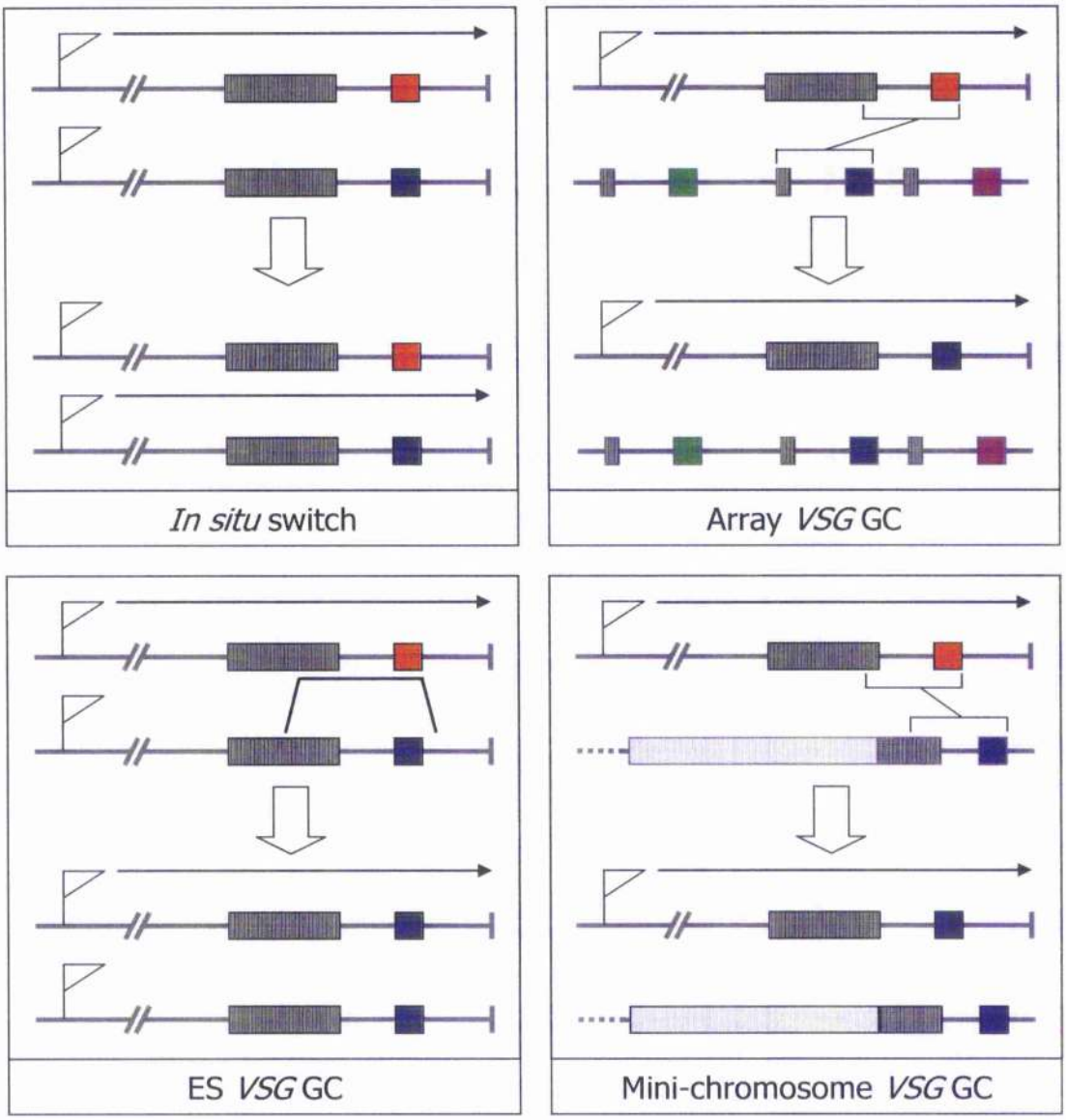


Figure 1.7. The *VSG* switching mechanisms of *T. brucei*. A schematic representation of the mechanisms of *VSG* switching (see text). Horizontal grey lines represent the expression site or *VSG* array chromosome; vertical grey lines represent the end of a telomere; triangles represent expression site promoters; black arrows represent transcription of an expression site; black and white striped boxes represent the 70-bp repeats; grey and white striped boxes represent the 177-bp repeats found upstream of metacyclic expression sites; coloured squares represent distinct *VSG*s; black lines represent the extent of sequence copied into the active expression site. Taken from C. Proudfoot, PhD thesis, University of Glasgow.

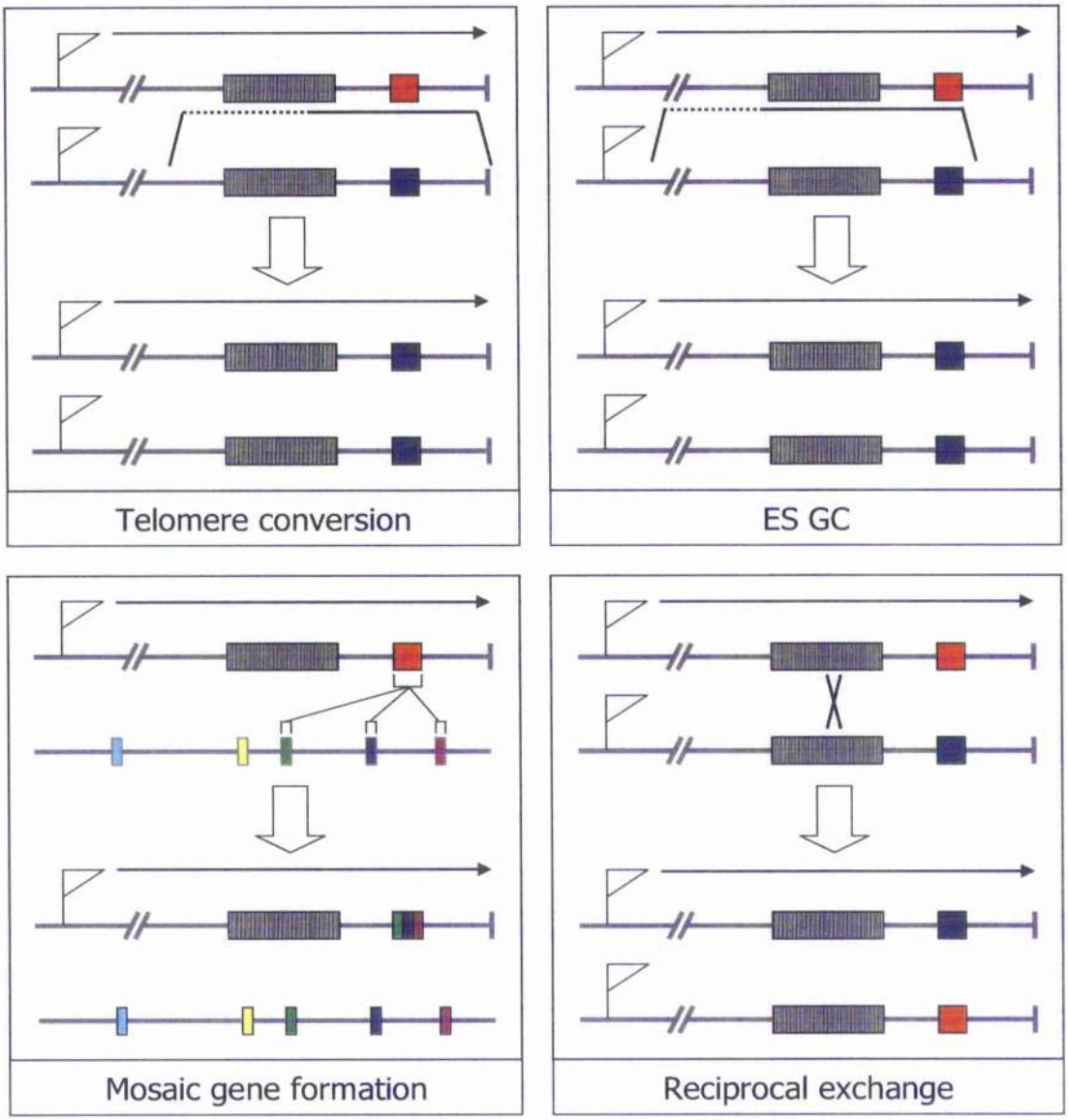


Figure 1.7 (cont). For legend, see previous page.

In a second type of duplicative transposition, named telomere conversion, the upstream limit of the transferred sequence is also usually the 70-bp repeats (Shah *et al.*, 1987), but in this case transfer continues up to the telomere (de Lange *et al.*, 1983). It is possible that this is the same reaction as duplicative transposition, simply using the telomere tract as downstream homology. However, it is also possible that this pathway is distinct, and occurs by a break-induced replication mechanism, as described in section 1.2.2.

Another mechanism of recombination-driven *VSG* switching is also found, and referred to as telomere reciprocal recombination (Rudenko *et al.*, 1996; Barry, 1997). Unlike gene conversion reactions, this involves a simple crossover event between telomeric *VSG*s, with both *VSG*s remaining intact at telomeric locations. The site of crossover can either be the 70-bp repeats (Pays *et al.*, 1985) or further upstream (Shea *et al.*, 1986) in the BES. Although this has been documented on many occasions (Rudenko *et al.*, 1998), it seems likely that it is a minor reaction, being limited only to the telomeric *VSG* repertoire.

The third form of *VSG* switching appears also to involve gene conversion, and hence be related to duplicative transposition. This is mosaic gene formation (Thon *et al.*, 1990; Barbet and Kamper, 1993), where novel, composite *VSG*s are created by segmental gene conversion using sections of two or more silent *VSG*s. This is the only pathway that allows the *VSG* pseudogenes to contribute productively to antigenic variation, and differs from the duplicative transposition events described above as it does not rely on flanking homology, but instead on short regions of homology within the *VSG* ORFs, and it is not clear whether the same mechanism is employed for both switching pathways. Mosaic gene formation is thought to be employed to prolong infection once the intact *VSG*s have been recognised by the immune system (Barbet and Kamper, 1993). Although it has been suggested that mosaic gene formation takes place within the active BES (Pays, 1989), it has also been pointed out that any incomplete products formed at this locus would be lethal to the cell (Barry and McCulloch, 2001).

1.1.5.3 Importance of different pathways

The above mechanisms of *VSG* switching contribute to antigenic variation to different extents in pleomorphic and monomorphic *T. brucei*. In monomorphic cells, *in situ* switching predominates (Liu *et al.*, 1983; Aitcheson *et al.*, 2005), and the 70-bp repeats have a less important role in switching (as discussed above), perhaps indicating that monomorphic cells have lost the ability to carry out ELC formation (Barry, 1997). In pleomorphic *T. brucei*, most *VSG* switching takes place by duplicative transposition

(Robinson *et al.*, 1999). The presence of a very high number of *VSG* pseudogenes in the *T. brucei* genome suggests that mosaic gene formation must have a greater importance in antigenic variation than previously thought, although whether or not it occurs equivalently in pleomorphic and monomorphic strains is unclear.

The order of expression of different *VSG*s from the archive has often been described as "semi-predictable", *i.e.* certain *VSG*s tend to be used at different stages of infection in a hierarchical order of switching (Miller and Turner, 1981; Robinson *et al.*, 1999; Aitcheson *et al.*, 2005; Morrison *et al.*, 2005). The semi-ordered pattern of expression allows the prolongation of infection within the host, as it allows "new" parts of the archive to be unveiled to the host slowly. Order is determined by a probability of activation of each *VSG*; *VSG*s with a lower probability of activation are used later in infection, once VSGs with higher probability have already prompted an immune response and so become unusable (Morrison *et al.*, 2005). The locus inhabited by the donor *VSG*, but not sequence homology to the previously expressed *VSG*, was also found to have an effect on determination of the order of expression (Morrison *et al.*, 2005); genes from subtelomeric, minichromosomal sites tend to be preferentially activated earlier in infection (Robinson *et al.*, 1999).

1.1.6 Meiosis in trypanosomes

Trypanosomes, and other kinetoplastids, undergo genetic exchange in the wild (discussed in section 4.1.1). Classical genetic analysis has been the main approach employed so far to study meiosis in *T. brucei*. Research based on laboratory crosses showed that *T. brucei* genetic markers show allelic segregation and independent assortment in the ratios expected in a Mendelian system (Macleod *et al.*, 2005). This is consistent with the *T. brucei* pathway of genetic exchange being standard meiotic recombination, probably involving the production and fusion of haploid gametes in the salivary gland of the tsetse fly. The fact that meiotic crossing over occurs in this Mendelian manner means that it has been possible to construct a genetic map (MacLeod *et al.*, 2005). This opens up the opportunity of employing a forward genetics approach to identify factors involved in phenotypes such as drug resistance. Bioinformatic analysis of the kinetoplastid genomes has revealed that they contain homologues of a number of meiosis-specific genes (discussed in section 4.1.3). In conjunction with the genetic evidence, the presence of these genes strongly points to the existence of meiosis in *T. brucei*. This is the basis of the experiments in Chapter 4.

1.2 DNA double strand break repair

The most dangerous form of DNA damage that a cell can suffer is a double-strand break (DSB), which can occur due to ionising and UV radiation or treatment with radiomimetic drugs, attack by reactive oxygen species, the collapse of replication forks (Jackson, 2001), as well as being induced for specific purposes. The danger of DSBs is reflected by the fact that a variety of strategies have evolved to deal with them (see figure 1.8). If left unrepaired, DSBs have disastrous consequences including chromosomal fragmentation and translocation, which can lead to cancer in multicellular organisms due to inactivation or loss of a tumour suppression gene or ectopic activation of an oncogene (Khanna and Jackson, 2001). Unrepaired DSBs can also lead to the induction of apoptosis in higher eukaryotes (Khanna and Jackson, 2001). Two different pathways for the repair of DSBs have been detailed. Homologous recombination (HR) repairs DSBs using undamaged, homologous DNA as a template for error-free repair. In contrast, non-homologous end joining (NHEJ) rejoins the broken ends of a DSB with no homology requirements, which can lead to sequence changes. Although both systems are conserved throughout the eukaryotes, their relative use varies considerably. In general, it appears that "higher" eukaryotes favour NHEJ, whereas HR is the pathway predominantly used in lower eukaryotes (Liang *et al.*, 1998). Beyond this predisposition, a range of factors can determine which pathway is used at a given DSB. These include the cell cycle stage at which the damage occurs, with NHEJ predominating at G_1-early S phase but HR being favoured at late S-G_2 phase in chicken cells (Takata *et al.*, 1998); the position of the break along the chromosome, with NHEJ being used more frequently to repair telomere-proximal DSBs in *S. cerevisiae* (Ricchetti *et al.*, 2003); and the nature of the DNA ends, with HR appearing to prefer to act on substrates with long ssDNA ends (Ristic *et al.*, 2003). In addition, homologous recombination has relatively recently been shown to be important in the vital ability to bypass replication fork stalling (Michel *et al.*, 2004). These findings all argue against the suggestion that HR and NHEJ compete against each other at each DSB, as has been suggested (Van Dyck *et al.*, 1999). As well as dealing with dangerous spontaneous DSBs, HR and NHEJ are involved in processes such as V(D)J recombination (Xu *et al.*, 2005) and meiosis (see chapter 4), where DSBs are created as part of specific cellular processes by the RAG1 and RAG2, and SPO11 proteins, respectively. Here, general DNA repair systems are used in specific circumstances to generate diversity.

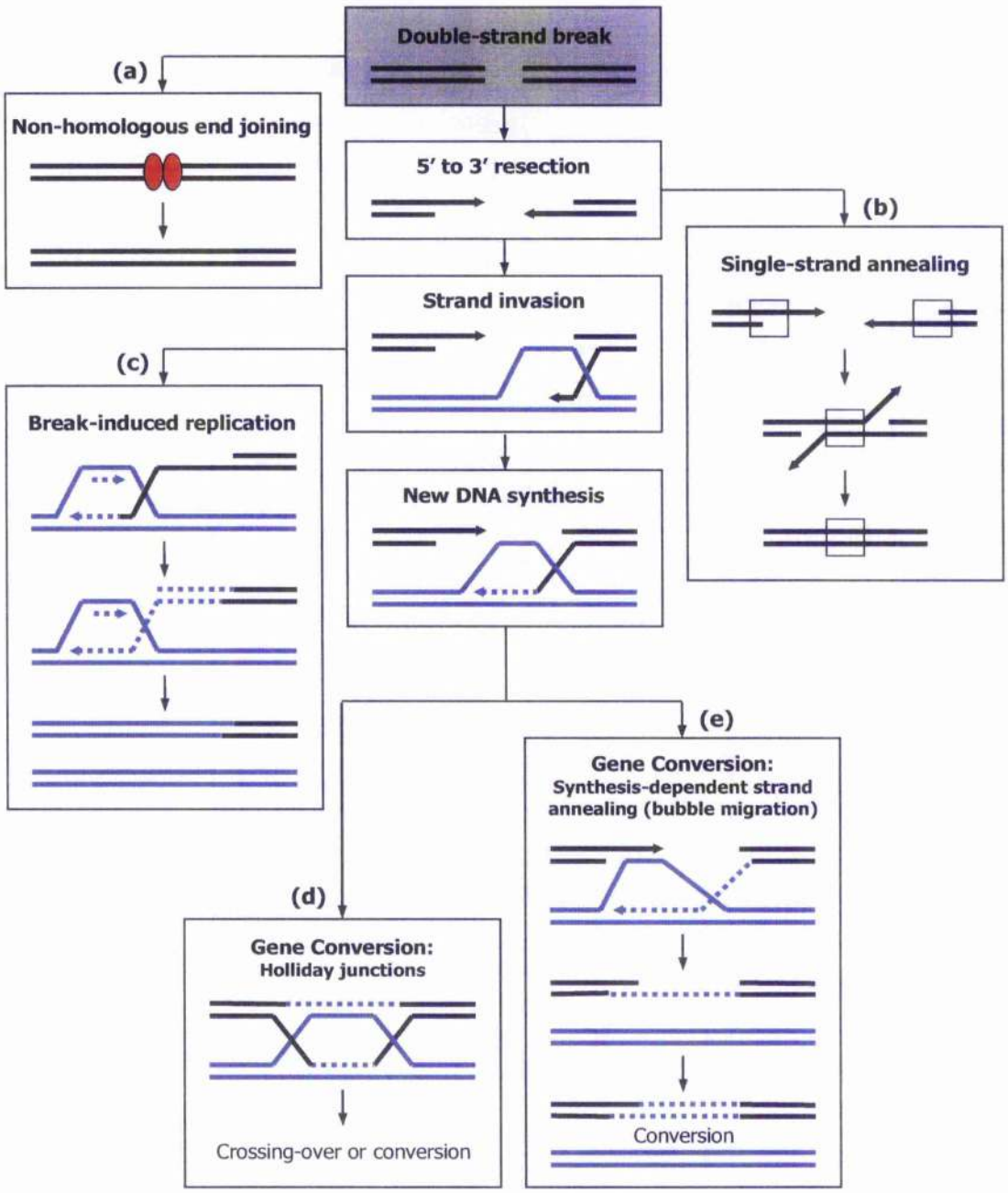


Figure 1.8. Pathways of eukaryotic double strand break repair. A schematic representation of DSB repair mechanisms in eukaryotic cells. DNA containing a DSB is represented by black lines, intact duplex DNA by blue lines, newly synthesised DNA by dashed lines, and NHEJ machinery by red circles. Taken from J.S. Bell, PhD thesis 2002.

1.2.1 Homologous recombination

HR allows accurate repair of a DSB by using homologous sequence as a template. The basic stages of the process, and the factors involved, are conserved from bacteria to humans, and the eukaryotic mechanism is described below. Eukaryotic HR is carried out by a group of proteins defined in yeast as the Rad52 epistasis group, which includes Rad52, Rad50, Rad51, Rad54, Rad55, Rad57, Rad59, Mre11 and Xrs2 (Symington, 2002).

The first task performed by the HR machinery upon detection of a DSB is resection of each side of the break to form invasive 3' ssDNA overhangs (White and Haber, 1990), which can be thousands of bp long (Sun *et al.*, 1991). In eukaryotes, the MRX complex, made up of Mre11, Rad50 and Xrs2 (or its functional analogue Nbs1 in humans, (Carney *et al.*, 1998)), may be responsible for this reaction (Trujillo *et al.*, 1998). Mre11, the central binding partner of the complex (Usui *et al.*, 1998) which contains a nuclease activity (Paull and Gellert, 1998; Tsubouchi and Ogawa, 1998) is stabilised by Rad50 (Paull and Gellert, 1998), and the nuclease reaction is potentiated by Xrs2/Nbs1 binding (Paull and Gellert, 1999). However, the nucleolytic activity required to create 3' overhangs at DSBs is 5'-3', whereas Mre11 acts in the opposite polarity, leading to the suggestion that Mre11 is important to "clean up" adducts to the ends of the DSB, allowing other, redundant nucleases to create the overhangs (Krogh and Symington, 2004). Mre11 and Rad50 form heterotetramers, with Mre11 binding next to the Rad50 ATPase domain and probably controlled by ATP-dependent conformational changes of this region (Hopfner *et al.*, 2001). The MRX complex also has roles in NHEJ, DNA damage signalling, meiosis and the maintenance of telomeres (Symington, 2002; Assenmacher and Hopfner, 2004).

Following resection of the DSB ends, the single-stranded tails are bound by Rad51, forming a characterisitic nucleoprotein filament (Shinohara *et al.*, 1992). Rad51 (Shinohara *et al.*, 1992) is homologous to the bacterial RecA protein and to RadA in the Archaea (Brendel *et al.*, 1997; Seitz *et al.*, 1998). The preferred substrate for Rad51 being dsDNA with ssDNA tails (Mazin *et al.*, 2000b). Indeed, the proteins have similar structures (Conway *et al.*, 2004), despite limited sequence similarity (Brendel *et al.*, 1997), illustrating that it is universally conserved. Each forms a helical nucleoprotein filament with very similar structure, though Rad51 and RecA filaments have opposite polarities (Radding, 1991; Sung and Robberson, 1995; McIlwraith *et al.*, 2001). One Rad51 monomer binds to 3 DNA nucleotides, in an extended conformation within the filament (Sung and Robberson, 1995). Although ATP hydrolysis by Rad51 is important for its

function (Stark *et al.*, 2002), it is ATP binding rather than hydrolysis that is needed for Rad51 filament formation (Chi *et al.*, 2006).

Repair proteins are grouped into foci, which contain a number of repair factors and, apparently, multiple repair events (Lisby *et al.*, 2003). Rad51 foci are formed as the protein is recruited to DSBs following DNA damage (Tarsounas *et al.*, 2003). This DNA damage response is mediated by the Brca2 protein, which binds to a number of other proteins involved in DNA metabolism (Zhang *et al.*, 1998; Liu and West, 2002). Brca2 contains a number of BRC repeats, which bind Rad51 (Davies *et al.*, 2001; Galkin *et al.*, 2005) and sequester it, allowing selective mobilisation upon DNA damage (Yu *et al.*, 2003) and recruitment to the dsDNA-ssDNA boundary at the DSB (Yang *et al.*, 2005). Brca2 can be phosphorylated at its C-terminus by cyclin-dependent kinases, blocking its interaction with Rad51; DNA damage decreases this phosphorylation, making a sort of molecular switch (Esashi *et al.*, 2005). However, this appears not to be universal, as Brca2 is found in mammals, plants, nematodes and the fungus *Ustilage maydis*, but not in *S. cerevisiae* (Kowalczykowski, 2002).

In *S. cerevisiae*, ssDNA created by DSB resection is coated by Replication Factor A (RPA), a homologue of the bacterial single-strand binding protein SSB (Sugiyama *et al.*, 1997). RPA protects the DNA from nucleases and removes secondary structures from the 3' overhangs (Sugiyama *et al.*, 1997), as well as being needed later in the repair process (Wang and Haber, 2004). As has been shown *in vitro*, Rad51 can bind efficiently to ssDNA only in the absence of RPA (Sung, 1997a), and therefore a number of proteins are needed to mediate the formation of the Rad51 nucleoprotein filament. Firstly, Rad52 binds to DNA ends (Mortensen *et al.*, 1996) and facilitates removal of RPA from the ssDNA and its replacement by Rad51 in both yeast and humans (Sung, 1997a; Benson *et al.*, 1998). Nevertheless, Rad52 is not present in all organisms, being absent in *D. melanogaster*, *C. elegans* and *T. brucei*. Rad51-related proteins also aid Rad51 function. In *S. cerevisiae*, the Rad51 homologues Rad55 and Rad57 form a heterodimer which also aids nucleoprotein filament formation (Sung, 1997b), using a separate mechanism to Rad52 (Gasior *et al.*, 1998); however, an absence of Rad55 or Rad57 can be compensated for *in vivo* by overexpression of either Rad51 (Johnson and Symington, 1995) or Rad52 (Hays *et al.*, 1995). In mammalian cells, the situation is more complex as five Rad51 paralogues, Rad51B, Rad51C, Rad51D, Xrcc2 and Xrcc3, as well as Rad52 (Symington, 2002). The roles of the mammalian Rad51 paralogues are not yet clear, though there is evidence that at least some seem to act in the same way as Rad55-Rad57 in yeast (Sigurdsson *et al.*, 2001).

The mature Rad51 nucleoprotein filament is capable of interacting with intact homologous DNA and initiating strand exchange, or synapsis, where the Rad51 filament invades the homoduplex, aligning and base-pairing with homologous sequence. Homology searching, at least in bacteria, occurs thanks to random collisions with homologous sequence, rather than a sliding mechanism along the DNA (Adzuma, 1998). The polarity of Rad51-mediated strand exchange is 5' to 3' with respect to the complementary strand of the DNA duplex (Sung and Robberson, 1995), in contrast to bacterial RecA (Kahn *et al.*, 1981 PNAS). Strand invasion leads to the formation of a structure called a D-loop, a bubble of unwound DNA where the complementary strand has been displaced from the duplex. Rad54 (Emery *et al.*, 1991), a member of the Swi2/Snf2 family of chromatin remodelling proteins (Eisen *et al.*, 1995), is also important for strand invasion. Rad51 and Rad54 interact in an equimolar ratio (Mazin *et al.*, 2000a), in an interaction which strongly stimulates strand invasion (Petukhova *et al.*, 1998). Rad51 and Rad54 act cooperatively in chromatin remodelling (Alexiadis and Kadonaga, 2002), using Rad54 ATP hydrolysis function to supercoil and separate the strands of the homologous DNA (Van Komen *et al.*, 2000; Sigurdsson *et al.*, 2002). Rad54 also assists Rad51 binding to ssDNA in an ATP-independent fashion (Wolner and Peterson, 2005).

According to the DSB repair model of HR developed by Szostak *et al.* (Paques and Haber, 1999), following strand invasion and D-loop formation, DNA synthesis is carried out within the D-loop to recover any lost sequence, starting from the 3' OH of the invading strand and requiring both leading and lagging strand DNA synthesis (Holmes and Haber, 1999). The other 3' end of the broken DNA also invades the D-loop, providing another primer for DNA synthesis and creating a structure with two Holliday Junctions. This process is called gene conversion, with the intact sister chromatid donating the genetic information to repair the break. Two putative Holliday Junction resolvases have been described in human cells. One is composed of Mus81-Eme1 (Boddy *et al.*, 2001; Chen *et al.*, 2001), an endonuclease capable of acting on recombination intermediates including Holliday Junctions (Smith *et al.*, 2003) and perhaps for replication fork rescue (Constantinou *et al.*, 2002). The second, the RAD51C-XRCC3 heterodimer, may be mammal-specific and is capable of *in vitro* Holliday Junction resolution (Liu *et al.*, 2004). In addition, mammalian RAD51B has been shown to bind Holliday Junctions, suggesting a possible role in junction resolution or branch migration (Yokoyama *et al.*, 2003). Resolution of each HJ can occur in two orientations, leading to either crossover or non-crossover outcomes; each of these would be expected to occur in 50% of cases (Krogh and Symington, 2004). Another model of HR, called synthesis-dependent strand annealing (SDSA; Nassif *et al.*, 1994; Paques and Haber, 1999), has been proposed to account for the

fact that the actual amount of crossovers observed is much less than the 50% predicted by the DSBR model. In the SDSA model, no HJs are formed. Instead, conservative replication occurs inside a D-loop, which migrates along the intact chromosome until second end capture allows the D-loop to be dismantled. This model can also be adapted to allow for crossovers, with second end capture annealing and forming a single HJ, resolved with or without crossing over (Ferguson and Holloman, 1996).

The reaction described above is the type of recombination that occurs in meiotic recombination (see section 4.1.2): DSBs created by SPO11 (Keeney *et al.*, 1997) promote HR catalysed by RAD51 and its meiosis-specific homologue DMC1 (see section 4.1.2.2).

1.2.2 Rad51-independent homologous recombination

In the absence of Rad51 protein, HR is reduced but not removed altogether (Rattray and Symington, 1994). Two proteins responsible for Rad51-independent recombination are Rad52 (Mezard and Nicolas, 1994) and its paralogue Rad59 (Bai and Symington, 1996; Bai *et al.*, 1999), which is found in some yeast species but not in mammalian genomes (Wu *et al.*, 2006). Rad59, like Rad51, also has a role in Rad51-dependent recombination, which is to help Rad52 form complexes with RPA and Rad51 (Davis and Symington, 2001; Davis and Symington, 2003). It should be noted that Rad52 has markedly different roles in fungal and mammalian cells; mutation of Rad52 causes a 3000-fold reduction in *S. cerevisiae* recombination, whereas the effect is considerably less pronounced in mammalian cells (Rijkers *et al.*, 1998; Yamaguchi-Iwai *et al.*, 1998a). Two pathways of homologous recombination have been described allowing DSBs to be repaired in the absence of Rad51: single-strand annealing and break-induced replication; in fact, in some circumstances one reaction may be followed by the other (Kang and Symington, 2000; Malagon and Aguilera, 2001; Ira and Haber, 2002).

Break-induced replication (BIR; McEachern and Haber, 2006) is a one-ended invasion process, which can occur by Rad51-dependent or Rad51-independent mechanisms, although both are usually masked in wild-type cells by the high amounts of gene conversion that take place (Davis and Symington, 2004). After strand invasion of a homologous chromosome by sequences centromere-proximal to the DSB, a replication fork is established and the chromosome is copied for up to 100 kb, normally up to the chromosome end (Malkova *et al.*, 1996). This feature of the mechanism makes BIR a useful tool in telomere maintenance (McEachern and Haber, 2006). Rad51-independent BIR requires Rad52 (Bosco and Haber, 1998), Rad50 and Rad59 (Signon *et al.*, 2001), and

is in fact hindered by the presence of Rad51 (Ira and Haber, 2002). It requires shorter lengths of homologous substrate than Rad51-mediated recombination (Ira and Haber, 2002), and formation of the replication fork is promoted by a distant enhancer element (Malkova *et al.*, 2001). Rad51-dependent BIR requires the same co-factors as Rad51-mediated gene conversion (Davis and Symington, 2004), and, in fact, is much more efficient than its Rad51-independent equivalent (Malkova *et al.*, 2005). Regardless of the factors involved, three different mechanisms for BIR have been suggested following strand invasion, involving either conservative or semi-conservative replication of the large tract of DNA (Kraus *et al.*, 2001; McEachern and Haber, 2006).

Single-strand annealing (SSA) is a Rad51-independent repair pathway acting specifically on direct repeat sequences and leading to the loss of some genetic material is known as single strand annealing (Paques and Haber, 1999). Resection of the DSB ends to 3' overhangs causes exposure of complementary sequences, which can then anneal to each other with no need for a strand invasion step. This reaction does not require Rad51, Rad54, Rad55 or Rad57, although Rad52 (Ivanov *et al.*, 1996) and Rad59 (Sugawara *et al.*, 2000) are needed, Rad59 in particular for SSA between shorter direct repeats. Removal of the nonhomologous tails, requiring the Rad1-Rad10 and Msh2-Msh3 heterodimers (Sugawara *et al.*, 2003), and ligation of the nicks completes the reaction.

1.2.3 NHEJ

NHEJ catalyses the ligation of DSB ends, requiring no homology or as little as 2-4 bp of homologous sequence, and often leading to changes in nucleotide sequence at the DSB site. The two main components of the human NHEJ machinery (Weterings and van Gent, 2004) are DNA-dependent protein kinase (although this protein is not conserved in other organisms) and the DNA ligase IV – XRCC4 complex. Conservation of these factors is retained in bacteria (Della *et al.*, 2004), demonstrating that this is an ancient process. The MRX complex is involved in the early stages of NHEJ, removing proteins already bound to the DNA ends (Connelly and Leach, 2002) and bridging the DNA ends together (Stracker *et al.*, 2004). The DNA-dependent protein kinase holoenzyme is composed of the DNA-PKcs catalytic subunit and the Ku70/80 heterodimer (Walker *et al.*, 2001). Ku binds to DNA ends before translocating along the DNA in an ATP-independent manner, stabilising the binding of DNA-PKcs to broken DNA (Smith and Jackson, 1999) and forming a bridge between broken DNA ends (Ramsden and Gellert, 1998). The Ku complex also has roles in telomere maintenance (Tsukamoto *et al.*, 2005). The role of the kinase activity of DNA-PK, which belongs to the PIKK family that contains the DNA

damage signalling proteins ATM and ATR, is not clear: though a number of its targets have been identified, including p53, Ku, XRCC4 and itself (Smith and Jackson, 1999), the relevance of phosphorylation in the NHEJ mechanism remains unknown. It is notable that DNA-PKcs is absent from yeast, suggesting that such phosphorylation may not be essential. The second component of the NHEJ apparatus is the ATP-dependent DNA ligase IV (Wilson *et al.*, 1997), acting in complex with XRCC4 (Critchlow and Jackson, 1998), which stabilises and activates the ligation activity (Grawunder *et al.*, 1997). Some types of breaks require processing before end-joining can take place, for example by the addition of phosphate groups (Chappell *et al.*, 2002) or by DNA re-synthesis by repair polymerases (Fan and Wu, 2004).

1.3 Mismatch repair

The mismatch repair (MMR) system has three main roles. Firstly, it monitors replication, where errors can lead not only to simple mismatches caused by misincorporation of a base, but also (especially in microsatellite tracts and mononucleotide runs) to small insertion-deletion loops created by replication slippage. Secondly, it removes chemical damage to DNA, including that caused by alkylating agents such as *N*-methyl-*N'*-nitro-*N*-nitrososguanidine and cisplatin (Jiricny, 2006). Such damage can be perceived as a mismatch, but in most cases MMR does not effect repair of these lesions and can, in fact, hinder cell survival (see section 4.2.7.4). Finally, MMR proteins are involved in recombination. One component of this is in the prevention of homologous recombination between similar but non-identical sequences (see section 3.1). In addition, MMR components also act in the regulation of crossover formation during meiotic recombination (see section 4.1.2) and can also promote some forms of mitotic recombination. The human cancer hereditary nonpolyposis colorectal carcinoma (HNPCC), an autosomal dominant cancer predisposition syndrome accounting for approximately 5% of colorectal cancer, is caused by mutations in some MMR genes, especially Msh2 and Mlh1 (Fishel *et al.*, 1994; Buermeyer *et al.*, 1999). A number of reviews are available on the molecular mechanisms underlying this repair pathway (Modrich, 1991; Modrich and Lahue, 1996; Buermeyer *et al.*, 1999; Kolodner and Marsischky, 1999; Harfe and Jinks-Robertson, 2000; Kunkel and Erie, 2005; Jiricny, 2006).

1.3.1 The E. coli *MMR paradigm*

The MMR system has been best characterised in bacteria (Modrich, 1991), and the bacterial system can be thought of as a basis for understanding the reaction in eukaryotes (see figure 1.9). Bacterial MMR is initiated when the first of the "mutator" proteins, MutS, binds a mismatch in heteroduplex DNA (Su and Modrich, 1986), as a homodimer (Allen *et al.*, 1997), with much higher affinity for basepair mismatches and small insertion-deletion loops (IDLs) of up to 4 nucleotides (Parker and Marinus, 1992) compared with perfectly matched DNA (Gradia *et al.*, 2000). The MutS dimer is then bound by a MutL homodimer (Grilley *et al.*, 1989; Galio *et al.*, 1999), which catalyses the ATPase activity of MutS, promoting translocation of the complex along the DNA away from the mismatch, as well as downstream events of the MMR process as described below. MutH endonuclease provides strand specificity to the bacterial post-replicative MMR system (Lu *et al.*, 1983), by creating nicks in DNA at hemimethylated GATC sites on either side of the mismatch (Au *et al.*, 1992). These nicks allow the section of the newly replicated strand containing the mismatch to be excised in a reaction requiring the UvrD helicase (also known as MutU) (Dao and Modrich, 1998), which is loaded onto the MutH-induced nick in a MutL-dependent manner. The strand containing the mismatch is removed by redundant exonucleases (Viswanathan *et al.*, 2001), resynthesis of the strand is catalysed by DNA polymerase III, and the ends of the repaired section are joined together by DNA ligase (Kunkel and Eric, 2005). Eukaryotic MMR is orthologous to the system described in *E. coli*, although more complex; however, MutS and MutL homologues are central to the detection and repair of mismatches.

1.3.2 *MutS family proteins*

Eukaryotic genomes contain between five and seven MutS Homologues (MSH proteins), all with different functions (see figure 1.9). The MSH proteins will be described here in some detail as they are central both to MMR and to the experimental work in this thesis. Msh2, Msh3 and Msh6 are involved in MMR of nuclear DNA. Msh2 binds to either Msh3 or Msh6 to form heterodimers with distinct but overlapping mismatch specificities (Acharya *et al.*, 1996; Marsischky *et al.*, 1996; Johnson *et al.*, 1996b; Bowers *et al.*, 1999; Edelmann *et al.*, 2000). The Msh2-Msh6 dimer, known as MutSα in mammals and plants, binds base-base mismatches and short IDLs, whereas Msh2-Msh3, or MutSβ, binds IDLs of various sizes (McCulloch *et al.*, 2003). Msh3 and Msh6 cannot interact with each other, and are not capable of self-interaction (Acharya *et al.*, 1996). In plants, an additional

Msh6-like MutS homologue, Msh7, has been identified (Culligan and Hays, 2000; Dong *et al.*, 2002; Horwath *et al.*, 2002). Msh7 also forms heterodimers with Msh2 (MutSγ), although this complex is only capable of binding base-base mismatches with high affinity. This difference in substrate affinity is possibly due to Msh7's inability to bind non-specifically to the DNA backbone (Wu *et al.*, 2003).

The two partners in an *E. coli* MutS homodimer adopt different conformations, leading to the formation of a functional heterodimer and echoing the heterodimeric relationships of the eukaryotic MSH proteins (Lamers *et al.*, 2000). The MutS or MSH dimer forms two channels, one of which contains the mismatched DNA, kinked by 60° (Junop *et al.*, 2003). In addition, at the interface of the dimer are two composite ATPase sites (Obmolova *et al.*, 2000), which bind ATP and ADP with different affinities (Antony and Hingorani, 2004). ATP hydrolysis is essential in the function of MutS homologues (Kunkel and Erie, 2005). The MutS/MSH dimer has been described as a sliding clamp, bound to DNA until binding is released by ATP hydrolysis (Gradia *et al.*, 1997), allowing translocation along the DNA. An α-loop of DNA is thus formed, usually containing the mismatch (Allen *et al.*, 1997). Diffusion away from the mismatch is bidirectional and allows the binding of numerous MutS/MSH dimers (Acharya *et al.*, 2003). At physiological concentrations, MutS also forms tetramers, where two dimers join together adopting asymmetrical conformations (Su and Modrich, 1986; Biswas *et al.*, 1999; Bjornson *et al.*, 2003). Tetramerisation, stimulated by ATP and DNA binding (Bjornson *et al.*, 2003) is dependent on the extreme N-terminal of the protein, which is also essential for MMR function. However, the relevance of this higher order structure needs to be investigated in more detail.

The Msh4 and Msh5 proteins (see figure 1.10), described in more detail in section 4.1.2, are not involved in MMR but instead in the regulation of meiotic recombination (Ross-Macdonald and Roeder, 1994; Hollingsworth *et al.*, 1995). Msh4 and Msh5 are not required in all meiotic systems, and are not found in all eukaryotic genomes. Another MutS homologue with divergent properties is Msh1, found in some organisms, including yeast and plants. Exceptionally, this eukaryotic MutS homologue acts as a homodimer (Nakagawa *et al.*, 1999) with roles in the repair of mitochondrial DNA, both in post-replicative repair and the prevention of ectopic recombination (Reenan and Kolodner, 1992; Chi and Kolodner, 1994a; Chi and Kolodner, 1994b; Koprowski *et al.*, 2002; Mookerjee *et al.*, 2005).

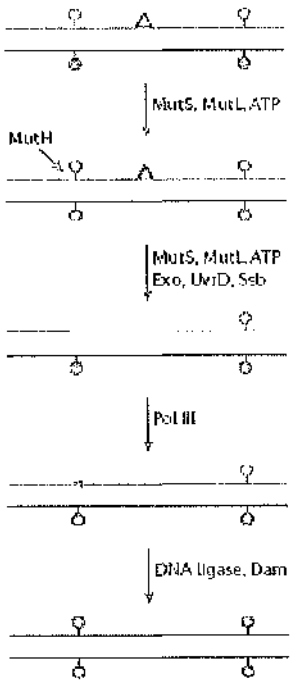


Figure 1.9. *E. coli* post-replicative MMR. A diagram summarising the stages of bacterial MMR, as described in the text. Factors needed for each step are named next to the arrow. Newly-synthesised DNA is shown in green, whereas the template strand is shown in blue. Unmethylated GATC sites are represented by open circles, and methylated sites by closed circles. The mismatched base is represented by a red triangle and DNA added by PolIII by a red arrow. Taken from Jiricny, 2006.

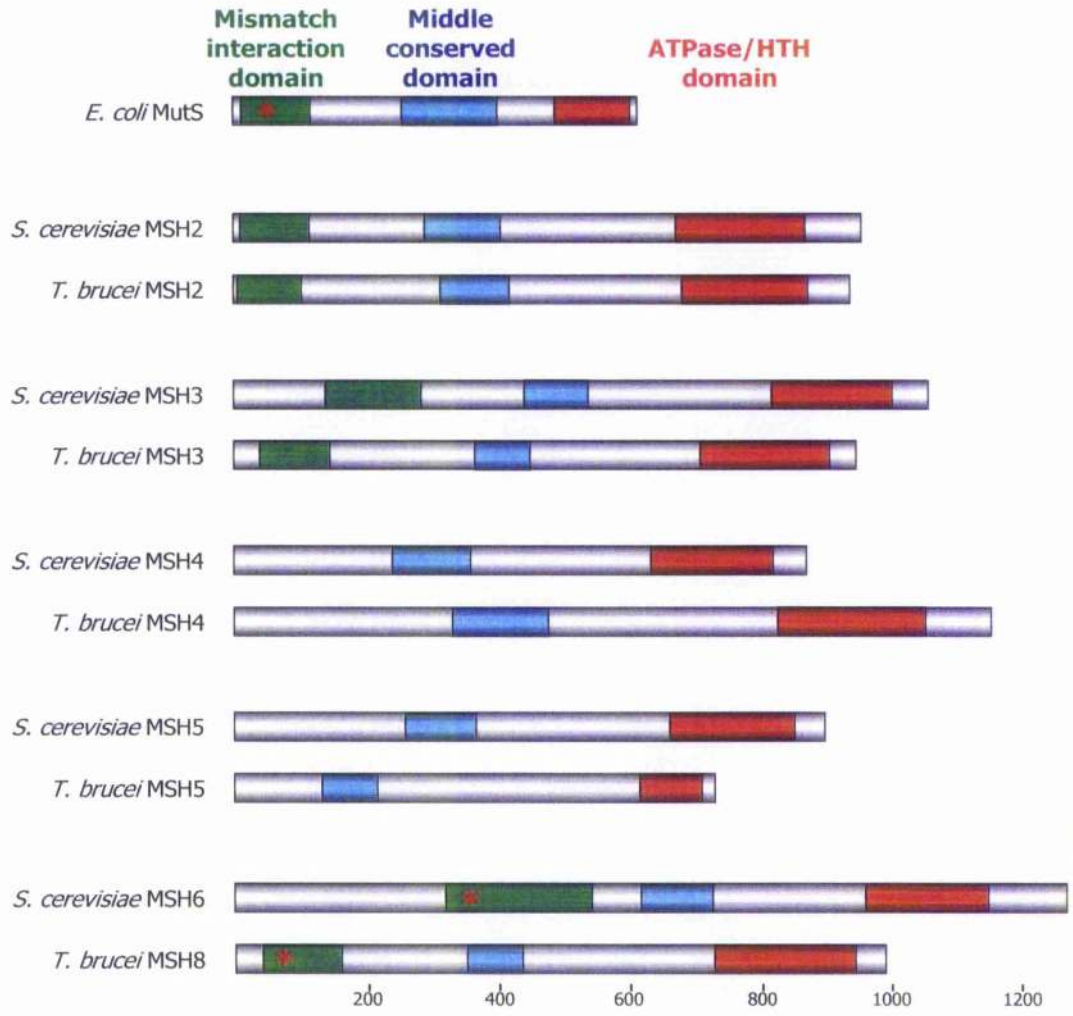


Figure 1.10. MutS homologues in *S. cerevisiae* and *T. brucei*. A schematic representation of MutS from *E. coli* and MSH2, MSH3, MSH4, MSH5 and MSH6/8 from *S. cerevisiae* and *T. brucei*. Proteins are depicted by grey boxes, and the conserved mismatch interaction, middle conserved and ATPase/helix-turn-helix domains by green, blue and orange boxes respectively. A red asterisk represents the conserved Phe-X-Glu residues of the mismatch interaction domain. A scale of the protein lengths (in amino acids) is shown at the bottom.

Early analysis of bacterial MutS showed that the protein's DNA binding ability requires the N-terminal portion of the protein, whereas dimerisation and interactions with MutL are dependent on C-terminal sequences (Wu and Marinus, 1994). The structure of MutS and its eukaryotic homologues has since been analysed in more detail, and four main conserved domains have been identified (see figure 1.10). The conserved domain nearest to the N-terminal is the mismatch interaction domain, containing a six-stranded mixed β-sheet surrounded by three α-helices (Lamers *et al.*, 2000). Within this domain, the Phe-39 and Glu-41 residues of the *E. coli* and *T. aquaticus* proteins (Malkov *et al.*, 1997; Schofield *et al.*, 2001) are needed for mismatch binding, by aromatic ring stack and hydrogen binding, respectively (Obmolova *et al.*, 2000). These residues are conserved in Msh6 but not Msh2 or Msh3 (Bowers *et al.*, 1999; Drotschmann *et al.*, 2001), despite Msh3 retaining general conservation in the mismatch interaction domain (Culligan *et al.*, 2000; Lamers *et al.*, 2000), implying that the mechanism of binding to base-base mispairs and to IDLs must be different (Kunkel and Erie, 2005). In keeping with their roles in recombination but not MMR, this domain is lacking altogether in Msh4 and Msh5 (see section 4.2.3.3). The middle conserved domain was identified by alignment of complete protein sequences (Culligan *et al.*, 2000), but its role in MMR is not clear. It is found on the surface of the molecule (las Alas *et al.*, 1998), suggesting interactions with other members of the MMR machinery, but does not seem to be needed for dimerisation. One mutation in this domain has been shown to have a dominant negative mutator phenotype (Wu and Marinus, 1994), but the structural basis for this is not known. Structural integrity of the middle section of MutS is needed for non-specific binding of the dimer to DNA (Obmolova *et al.*, 2000). The C-terminal is the best-conserved part of the protein (Eisen, 1998; Culligan *et al.*, 2000), and most *msh2* mutations causing cancer are situated in this region (las Alas *et al.*, 1998). It contains an ATPase domain (whose function is described above), containing Walker A and Walker B motifs, that is classified with the ABC ATPase superfamily (Obmolova *et al.*, 2000). As mentioned above, a functional ATPase domain is composed of parts of both proteins of the dimer (Obmolova *et al.*, 2000), with the nucleotide binding sites situated at the dimer interface (Lamers *et al.*, 2000). Nearest to the C-terminal of all MutS homologues is a helix-turn-helix domain, mediating dimerisation and essential for MMR (Alani *et al.*, 1997; Biswas *et al.*, 2001).

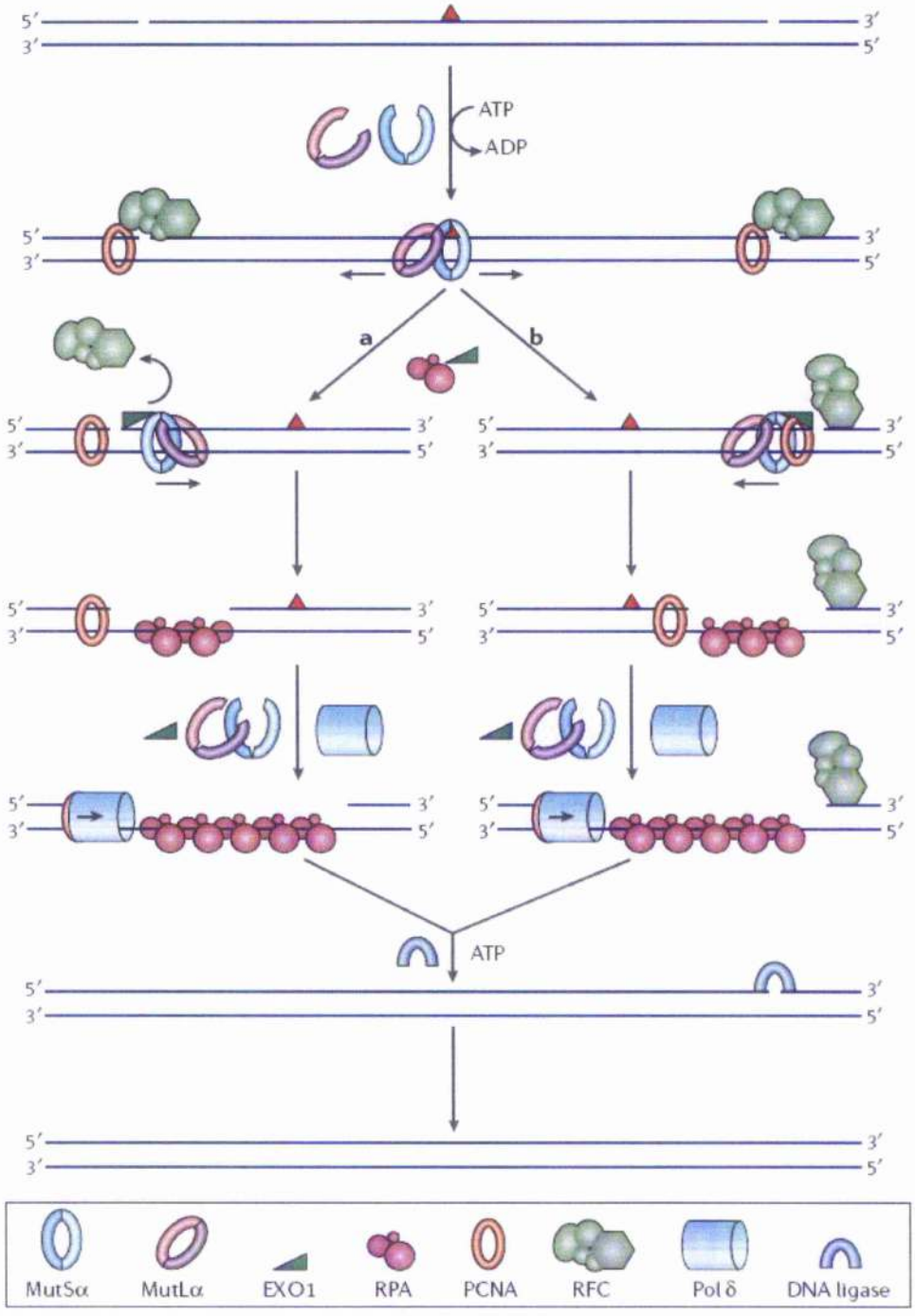


Figure 1.11. The human MMR system. A diagram showing the essential elements of the human MMR system (as described in the text), the components of which are shown in the legend. DNA strands are represented by horizontal blue lines and the mismatched base by a red triangle. Taken from Jiricny, 2006.

1.3.3 Downstream factors in MMR

After recognition of mismatched DNA by MutS homologues, binding of the MutS-DNA complex by MutL homologues occurs. Eukaryotic MutL homologues (MLH proteins) also act as heterodimers, some of which were identified in screens for defects in post-meiotic segregation (hence their description as PMS genes). Mlh1 interacts either with a protein known as Pms1 in yeast and PMS2 in human cells, or with Mlh2 or Mlh3 (Wang *et al.*, 1999), all of which compete with each other for Mlh1 binding (Kondo *et al.*, 2001). Dimerisation occurs via a conserved domain at the C-termini of each protein (Pang *et al.*, 1997; Guerrette *et al.*, 1999; Wang *et al.*, 1999; Kondo *et al.*, 2001). In addition, this family of proteins have an ATPase domain of the GHKL superfamily (Dutta and Inouye, 2000) at their N-termini. Most mutations causing mutator phenotypes and human cancer are found in the ATPase domain (Ban and Yang, 1998), which is essential for MMR (Spampinato and Modrich, 2000). ATP hydrolysis of MutL causes changes in this domain that can modulate interactions with other MMR proteins (Ban and Yang, 1998), suggesting an action as a molecular switch (Ban *et al.*, 1999). A non-specific DNA binding role important in MMR is probably associated with the dimerisation groove (Junop *et al.*, 2000; Kunkel and Erie, 2005). MutL heterodimers interact with MutS heterodimers with differing specificities (Kolodner and Marsischky, 1999): Mlh1-Pms1 can bind to both Mutsα and MutSβ (Greene and Jinks-Robertson, 1997; Blackwell *et al.*, 2001; Bowers *et al.*, 2001), enhancing both mismatch-binding properties (Habraken *et al.*, 1998) and stability of the mismatch-bound complex (Bowers *et al.*, 2001), whereas Mh1-Mlh3 only interacts with MutSα (Flores-Rozas and Kolodner, 1998). The role of Mlh2 has not been studied in detail; however, it is capable of repairing mutational intermediates *in vivo* (Harfe *et al.*, 2000), and weak tolerance of *Mlh2* mutants to certain drugs implies a role in MMR (Durant *et al.*, 1999).

The eukaryotic MMR machinery lacks a MutH homologue, and so there is some doubt surrounding the strand discrimination mechanism for post-replicative MMR. A methyl-CpG-binding endonuclease, MED1, has been proposed to act in this role (Bellacosa *et al.*, 1999), but the variable levels of methylation in the genomes of the higher eukaryotes make this unlikely (Jiricny, 1998). The essential replication factor PCNA (proliferating cell nuclear antigen) is a more likely candidate, providing a link between the replication machinery and post-replicative MMR (Buermeyer *et al.*, 1999). This protein has been shown to interact with both MutS and MutL homologues (Umar *et al.*, 1996; Gu *et al.*, 1998; Clark *et al.*, 2000; Flores-Rozas *et al.*, 2000; Kleczkowska *et al.*, 2001), and *PCNA*

mutants are defective in MMR (Johnson *et al.*, 1996a; Kokoska *et al.*, 1999). PCNA is required for MMR both at the early stages of MMR and at the resynthesis step (Gu *et al.*, 1998; Flores-Rozas *et al.*, 2000; Bowers *et al.*, 2001), and it has been suggested that it acts as a scaffold, allowing proteins from different stages of MMR to be brought together (Lee and Alani, 2006). Preferential repair of the lagging replication strand (Pavlov *et al.*, 2002; Pavlov *et al.*, 2003) is consistent with the higher concentration of PCNA on this strand.

The exonuclease Exo1 is also involved in MMR, interacting with Msh2 in a yeast two-hybrid screen (Tishkoff *et al.*, 1997), and in human cells (Schmutte *et al.*, 1998), and forming higher order structures with most known MMR factors (Amin *et al.*, 2001). Exo1 catalyses excision of the mismatched strand whether the nick is 5' or 3' of the mismatch (Genschel *et al.*, 2002). The 3'-5' exonuclease activities of DNA polymerases Polδ and Polε are also involved in MMR, however (Longley *et al.*, 1997; Tran *et al.*, 1999). The 3'-5' exonuclease activity of Mre11, which interacts with Mlh1, might also be involved *in vivo* (Vo *et al.*, 2005).

The factors essential to MMR have been identified in *in vitro* reconstitutions of MMR activity from human and yeast cell extracts (Dzantiev *et al.*, 2004; Constantin *et al.*, 2005; Zhang *et al.*, 2005; Jiricny, 2006). In addition to the proteins mentioned above, RPA (Lin *et al.*, 1998; Umezu *et al.*, 1998) and the non-histone chromatin component HMBG1 (high mobility group box 1) and RFC (replication factor C) were identified. RFC binds to the 5' end of nicked DNA, preventing degradation in the 5'-3' direction, a function needed when the nick is created 5' of the mismatch. Bidirectional MMR could be reconstituted *in vitro* using only the human factors MSH2, MSH6, MLH1, MLH3, RPA, EXO1, PCNA, RFC, HMGB1 (for optimal efficiency), DNA polymerase δ and DNA ligase I.

1.3.4 Mismatch repair, genetic variability, and selective pressure

The role of the MMR system is to reduce the amount of mutation in populations at the time of damage or miscopying (Modrich and Lahue, 1996; Jiricny, 2006). However, DNA repair is not perfect and some variation can be found between individuals of all species; indeed, some diversity is desirable as it provides the raw material permitting evolution. Whether a mutation is eventually maintained in a population as a polymorphism, or is lost because the organisms that contain it die or fail to produce viable offspring, depends on the selective pressure on the region of the genome containing the mutation. A mutation in an intergenic region is unlikely to be detrimental to the organism harbouring it, whereas in an

important motif of an open reading frame or gene control elements, a mutation could well cause a reduction in fitness and selective disadvantage.

To take the well-studied human genome as an example, nearly 12 million single nucleotide polymorphisms (SNPs) have been reported to the NCBI SNP database, dbSNP (www.ncbi.nlm.nih.gov/projects/SNP), of which almost half are found within genes. In a study looking at SNPs across the American population, the majority (96%) of single nucleotide polymorphisms (SNPs) within genes are found in noncoding regions of the gene (reviewed in Crawford, Akey and Nickerson, 2005). A separate investigation (Zhao *et al.*, 2003) looking at the different categories of SNPs in the human genome found that within genic regions, less SNPs were situated in exons compared to intronic and untranslated regions (5.28, 8.21 and 7.51 SNPs per 10 kb, respectively). In addition, a higher rate of SNPs was found in intergenic compared to genic regions (8.44 compared to 8.09 SNPs per 10 kb). In a third study, it was shown that SNPs resulting in an amino acid change or within the conserved 5' untranslated region of genes are less common than those in intergenic regions (Hughes *et al.*, 2003). These are some of the examples showing that selection is acting against the retention of polymorphism in important genomic regions.

Microsatellites are short, tandemly repeated sequences, found throughout genomes, that are highly prone to replication slippage. Clusters of these therefore constitute highly polymorphic regions of the genome. Disruption of the MMR system leads to a defect in repair following this slippage and therefore to microsatellite instability, and microsatellite instability is a hallmark of HNPCC (reviewed in Oda, Zhao and Maehara, 2005). Microsatellites are useful as tools for studying genetic and phylogenetic relationships between related organisms.

1.4 DNA repair and antigenic variation in *T. brucei*

The DNA repair machinery in *T. brucei* is especially interesting because of the genetic rearrangements taking place during antigenic variation. A number of important proteins and repair pathways have been identified and studied in some detail in this parasite, as summarised in the following section.

The first *T. brucei* protein shown to be involved in DNA repair was RAD51 (McCulloch and Barry, 1999). Disruption of the *RAD51* ORF in bloodstream stage and procyclic form *T. brucei* causes a significant increase in population doubling time and increased sensitivity to DNA damaging agents, as well as a decrease in, but not complete removal of,

recombination efficiency (measured by transformation of linear constructs). *Rad51* mutants also suffer a decreased frequency of *VSG* switching, although surprisingly both transcriptional and recombinational switching mechanisms are affected, implying that both RAD51-dependent and RAD51-independent mechanisms are involved in all types of switching, at least in the monomorphic cell line used for these experiments. *RAD51* is also found in the genomes of *L. major* (McKean *et al.*, 2001) and *P. falciparum* (Bhattacharyya and Kumar, 2003; Bhattacharyya *et al.*, 2004), and in both these protozoans it retains ATPase activity and is upregulated in response to DNA damage.

Six *RAD51* homologues have been identified in the *T. brucei* genome. One of these proteins, an orthologue of DMC1, has no known role in recombination, DNA repair or antigenic variation in bloodstream stage cells, perhaps consistent with a more specific role in meiotic recombination (Proudfoot and McCulloch, 2006). Of the others, only one protein, RAD51-3, has been implicated in antigenic variation (like RAD51, mutation of this paralogue leads to a reduction in all switching pathways), although both RAD51-3 and RAD51-5 are involved in HR (Proudfoot and McCulloch, 2005). This could be equivalent to the situation in mammalian cells, where RAD51D has special roles at the telomere as well as in DSB repair (Tarsounas *et al.*, 2004).

Study of transformation in *RAD51-/- T. brucei* has allowed some analysis of RAD51-independent recombination to be conducted (Conway *et al.*, 2002b). In the absence of this central recombination protein, the rate of integration of linear constructs is reduced approximately ten-fold, and of the events that are able to take place, only some are faithful HR reactions. Others integrate aberrantly into a number of genomic locations, using short stretches of homology and leading to some DNA loss at the integration sites. The factors involved in this reaction, and its detailed mechanism in relation to what has been described in other eukaryotes, is not yet clear. Nevertheless, this analysis suggests that RAD51-independent HR is able to use much shorter substrates than RAD51-dependent reactions, as seen in other organisms (see section 3.1.1).

T. brucei retains both components of the KU complex (Conway *et al.*, 2002a), and mutation of either gene shows that it is important for telomere length maintenance (Janzen *et al.*, 2004). The same mutations have no effect on *VSG* switching, cell growth or on sensitivity to DNA damaging agents. Indeed, the lack of NHEJ-specific ligases in the *T. brucei* genome suggests that this parasite is lacking a canonical eukaryotic NHEJ system (Burton, McBride, Wilkes, Barry and McCulloch, in preparation). If correct, this suggests that the integrations described in *rad51* mutant cells are not NHEJ-catalysed, and may be

explained by a microhomology-mediated end joining (MMEJ) pathway. Such reactions have been seen in other organisms, but the evolutionary conservation is unclear due to a lack of understanding of the factors involved; however, a recent study of the *T. brucei* ligases (J. Wilkes, University of Glasgow, pers. comm) may shed some light on this.

Genetic inactivation of *T. brucei* MRE11 (Robinson *et al.*, 2002; Tan *et al.*, 2002) leads to dramatically decreased cell viability reflected by increased population doubling time, similar to that seen in *rad51* mutants. *T. brucei mre11* mutants also accumulate gross chromosomal rearrangements, associated with shortening and rearrangement of the megabase chromosomes, but without altering telomere length. In addition, mutant cells are defective in HR, and show increased sensitivity to phleomycin though not MMS-induced DNA damage. Surprisingly, and similar to the phenotype described for the RAD51 paralogue RAD51-5, *mre11* mutants show no deficiency in *VSG* switching.

T. brucei has been shown to possess a functional MMR system, with the genome encoding homologues of MSH2, MSH3, MSH4, MSH5, MSH6 (see figure 1.9), MLH1 and PMS1, although not MLH3 or MSH1 (Bell *et al.*, 2004). Genetic disruption of *MLH1* or *MLH3* does not lead to changes in growth rate, but does result in increased methylation tolerance (see section 4.2.7.4) and microsatellite instability, both phenotypes associated with MMR deficiency in other organisms (Bell *et al.*, 2004). In addition, it has been demonstrated that MSH2 acts to regulate *T. brucei* recombination between divergent sequences, as described in more detail in section 3.1.2 (Bell and McCulloch, 2003). However, the MMR system appears to have no effect on antigenic variation, at least in monomorphic cells; genetic disruption of *MSH2* and *MLH1* has no effect on the frequency or spectrum of *VSG* switching events (Bell and McCulloch, 2003).

1.5 Aims of this thesis

The work in this thesis has sought to address three main questions surrounding the functions provided by the *T. brucei* MMR system, and by the MutS-related genes in this parasite. Each of these questions will be outlined in more detail in individual introductions to each chapter, but are summarised here.

i) What are the sequence requirements for homologous recombination in *T. brucei*, and does MMR regulate recombination using all types of substrate?

ii) Do the *T. brucei* homologues of MSH4 and MSH5 act in a way consistent with meiosis-specific regulators of recombination?

iii) Are there differences in the sequence of MMR genes from different trypanosome subspecies, and do any such differences correspond to alterations in MMR efficiency, as have been observed in bacteria and in *T. cruzi*?

CHAPTER 2

MATERIALS AND METHODS

2 Materials and Methods

2.1 Trypanosome culture

2.1.1 Trypanosome strains and their growth

2.1.1.1 Bloodstream stage cells

The *Trypanosoma brucei* MITat1.2a strain used for the bloodstream-stage work in this thesis is derived from the monomorphic Lister 427 stock and expresses VSG221. The 427 stock was derived from many syringe passages through rodents over a number of years, although its exact derivation is uncertain (Melville *et al.*, 2000). *In vitro* growth of *T. brucei* bloodstream forms was carried out using HMI-9 growth medium (Hirumi and Hirumi, 1989) at 37 °C in a humidified 5% CO_2 incubator. The population doubling time of this strain is ~8 h (Proudfoot and McCulloch, 2006). To keep a working culture of *T. brucei* bloodstream-stage cell lines, cells were passaged twice weekly by addition of ~100 µl of a log-phase culture (at a density of ~4 x 10^6 cells.ml^{-1}) to 1.5 ml HMI-9 medium in a 24-well plate. Bloodstream-stage *T. brucei* were grown in petri dishes in volumes of 25 ml to obtain large numbers of cells for experiments.

2.1.1.2 Procyclic form cells

The procyclic form *Trypanosoma brucei* strains used in this study are Swiss Tropical Institute, Basel (STIB) 247, STIB 386, East African Trypanosomiasis Research Organisation (EATRO) 795 and Trypanosomiasis Research, Edinburgh University (TREU) 927, whose derivations are described in section 5.2.1.1. *in vitro* growth of procyclic forms was carried out using SDM-79 growth medium (Brun and Schonenberger, 1979) at 27 °C. To keep a working culture of *T. brucei* procyclic form cell lines, cells were passaged twice weekly by addition of ~500 µl of a log-phase culture (at a density of ~8 x 10^6 cells.ml^{-1}) to 3 ml SDM-79 medium in a 25 cm^2 tissue culture flask. Procyclic form *T. brucei* were grown in 25 cm^2 tissue culture flasks in volumes of up to 25 ml to obtain large numbers of cells for experiments.

2.1.2 Stabilate preparation and retrieval

Stabilates for long-term storage of *T. brucei* were prepared by mixing 900 µl of *T. brucei* culture at a density of ~2 x 10^6 cells.ml^{-1} (bloodstream stage cells) or ~7 x 10^6 cells.ml^{-1} (procyclic cells) with 100 µl sterile 100% glycerol in cryotubes (Nunc), by inverting the tube. Stabilates were initially frozen at -80 °C overnight before being moved to liquid nitrogen.

To begin growing parasites, stabilates were taken out of liquid nitrogen, defrosted at 37 °C (bloodstream form cells) or 27 °C (procyclic cells), and placed in 10 ml HMI-9 growth medium (bloodstream stage cells) or 3 ml SDM-79 growth medium (procyclic form cells) overnight; the cells were then passaged normally as described above.

2.1.3 *Transformations of 427 strain bloodstream form* T. brucei

2.1.3.1 Transformations of gene knockout/ectopic expression constructs

T. brucei bloodstream stage cultures were grown to a density of ~1.5 x 10^6 cells.ml^{-1} and centrifuged at room temperature for 10 min at 583 x g. The cells were resuspended in Zimmerman post-fusion medium (5 M NaCl, 1 M KCl, 1 M Na_2HPO_4, 1 M KH_2HPO_4, 1 M MgOAc, 0.2 M $CaCl_2$) supplemented with 1 M D-glucose (ZMG), at a concentration of 1 x 10^8 cells.ml^{-1}. 5 x 10^7 cells per transformation were electroporated in 0.5 ml ZMG at 1.5 kV and 25 µF capacitance using a BioRad Gene Pulser II normally using ~5 µg of purified plasmid DNA that had been restriction digested, phenol-chloroform extracted and ethanol precipitated. After electroporation, cells were placed in 10 ml of HMI-9 for three population doubling times (normally 24 h) before being subjected to antibiotic selection. To do this, all recovered cells were centrifuged at room temperature for 10 min at 583 x g and resuspended in 50 ml of HMI-9 containing the appropriate antibiotic at concentrations described in the text, typically to a concentration of 1 x 10^6 cells.ml^{-1}. ~1 ml of this solution was pipetted into each well of two 24-well plates. Transformants were counted after 7-10 days by looking at the plates under a light microscope (Leitz) and counting the number of wells that contained growing cells.

2.1.3.2 Transformation efficiency assay

Transformations to assay recombination efficiency (section 3.2.1) were carried out as described above but with a number of modifications. 2.5 x 10^7 cells were used per

transformation and resuspended in ZMG to a concentration of 5 x 10^7 cells.ml^{-1}. To select for transformants, cells were counted and put on antibiotic selection at specific densities as described in the text (section 3.2.1). Approximately 3 µg of DNA that had been PCR-amplified using Herculase high-fidelity DNA polymerase (see section 2.7; Stratagene) was used for each transformation. For each transformation construct, 24 PCR reactions were performed, the products pooled, and purified using the Qiagen PCR purification kit according to manufacturer's instructions. Five volumes of Buffer PB was added to one volume of pooled PCR samples, and mixed. 750 µl of sample was applied to a QIAquick spin column in a 2 ml collection tube, centrifuged in a microcentrifuge at 16 x g for 1 min, and the flow-through was discarded. This step was repeated four times; so one QIAquick column was used per six PCR reactions. 750 µl of Buffer PE was added to the column and centrifuged in a microcentrifuge at 16 x g for 1 min, and the flow-through was discarded and the column centrifuged in a microcentrifuge at 16 x g for and additional minute. 50 µl of dH_2O was applied to the column and centrifuged in a microcentrifuge at 16 x g for 1 min. This elution step was repeated with a second volume of distilled water (dH_2O) for maximal elution of DNA from the column. All centrifugation steps were carried out at room temperature.

2.2 Isolation of material from trypanosomes

2.2.1 Isolation of genomic DNA

2.2.1.1 Lysis

15 ml of a bloodstream-stage *T. brucei* culture grown to a density of ~4 x 10^6 cells.ml^{-1} were harvested by centrifugation at 1620 x g for 10 min at room temperature, and resuspended in 500 µl of buffer (1 mM ethylenediaminetetraacetic acid (EDTA), 100 mM NaCl, 50 mM Tris-HCl pH 8). To lyse the cells, 50 µl of 10% sodium dodecyl sulphate (SDS) and 2.5 µl of a 20 µg.µl^{-1} proteinase K solution were added, and the solution incubated at 37 °C overnight. DNA was recovered from the lysis reaction by phenol/chloroform extraction and ethanol precipitation.

2.2.1.2 Phenol/chloroform extraction and ethanol precipitation of genomic DNA

An equal volume of a 1:1 mixture of phenol/chloroform (Sigma) was added to the lysis reaction and mixed by inverting the eppendorf tube several times. The phenol and aqueous

phases were then separated by centrifugation at 16 x g in a microcentrifuge for 10 min at room temperature, and the upper, aqueous phase containing the DNA was transferred to a new eppendorf tube. 2 volumes of 100% ethanol and 1/10 volume 3 M sodium acetate (pH 5.2) were added, the solution mixed by inverting the tube several times, and incubated at -20 °C for 30 min to overnight. DNA was pelleted by centrifugation at maximum speed in a microcentrifuge for 30 min at 4 °C. The 100% ethanol was removed and the nucleic acid pellet washed by addition of 1 ml 70% ethanol, followed by centrifugation at 16 x g in a microcentrifuge for 2 min at room temperature. The 70% ethanol was removed by aspiration and the pellet air-dried. Genomic DNA was resuspended in a typical volume of 30 µl of sterile dH_2O or TE (100 mM Tris, 10 mM EDTA, pH 7.4), to a final concentration of approximately 1 µg/µl.

2.2.2 Isolation of total RNA

50 ml of a bloodstream stage *T. brucei* culture grown to a density of ~4 x 10^6 cells.ml^{-1}, or 20 ml of a procyclic form T. brucei culture grown to a density of ~8 x 10^6 cells.ml^{-1}, were harvested by centrifugation at 1620 x g for 10 min at room temperature and removing the supernatant. Total RNA was isolated using the RNEasy mini kit (Qiagen) following manufacturer's instructions. 600 µl of Buffer RLT (containing the appropriate amount of 2-mercaptoethanol) was added and pelleted cells were mixed by pipetting, and the sample was homogenised by passing the lysate 5 times through a 25 G needle fitted to an RNase-free 1-ml syringe. 600 µl of 70% ethanol was added to the sample and mixed by pipetting. 700 µl of this solution was applied to an RNeasy column placed in a 2-ml collection tube, and centrifuged for 15 sec at 16 x g in a microentrifuge, and the flow-through was discarded. The column was washed by applying 700 µl of Buffer RW1 and centrifuging for 15 sec at 16 x g in a microcentrifuge, then by transferring the column to a new collection tube, applying 500 µl of Buffer RPE , centrifuging for 15 sec at 16 x g in a microcentrifuge. The flow-through was discarded and a final wash step carried out by applying another 500 µl of Buffer RPE to the column and contrifuging for 2 min at 16 x g in a microcentrifuge. RNA was eluted from the column by placing it in a clean eppendorf tube, adding 30 µl of RNase-free dH_2O and centrifuging for 1 min at 16 x g in a microcentrifuge. RNA was concentrated when necessary by ethanol precipitation as described in section 2.2.1.2.

2.2.3 Preparation of genomic agarose plugs

Each agarose plug contained ~4 x 10^7 bloodstream stage trypanosomes. Cells were grown in HMI-9 to a density of ~2 x 10^6 cells.ml^{-1}, centrifuged at 583 x g for 10 min at room temperature, washed by resuspending the pelleted cells in 10 ml PSG (1x PBS, 1% w/v glucose) and centrifuging them as before, then resuspended in PSG at a concentration of 1.6 x 10^6 cells.ml^{-1}. The cells were then warmed at 37 °C for 1 min and an equal volume of 1.4% Microsieve low-melt agarose (Flowgen) made with dH_2O was added and mixed. Disposable plug moulds (BioRad) were filled with ~50 µl agarose and placed at 4 °C for ~4 h to set. The agarose plugs were then removed from the moulds, incubated in NDS buffer (0.5 M EDTA, 1 mM Tris base and 34.1 mM lauroyl sarcosine) pH 9.0 containing 1mg.ml^{-1} proteinase K at 50 °C for ~24 h, transferred into NDS buffer pH 8.0 containing 1mg.ml^{-1} proteinase K at 50 °C for ~24 h, and finally transferred into NDS buffer pH8 for storage at 4°C.

2.3 Electrophoresis

2.3.1 DNA electrophoresis

Plasmid and genomic DNA samples were routinely separated by electrophoresis on 0.8% agarose gels (Seakem LE agarose, BioWhittaker Molecular Applications) made with 1 x TAE buffer (40 mM Tris, 19mM acetic acid, 1mM EDTA) and containing 0.2 µg.ml^{-1} ethidium bromide (Sigma), using apparatus supplied by Gibco BRL. A commercial 1kb DNA ladder was used as a size marker (Invitrogen). Typically, PCR products and plasmid DNA were electrophoresed in 1 x TAE at ~110 V for ~1 – 3 h, whereas genomic DNA for Southern blotting analysis was electrophoresed in 1 x TAE at ~30 V overnight.

2.3.2 RNA electrophoresis

RNA molecules were separated by electrophoresis on 1% agarose gels (Seakem LE agarose, BioWhittaker Molecular Applications) made with 0.4 x MNE buffer and containing 2.46M formaldehyde. Gels were typically run for ~16 h at 30 V in 1x MNE buffer, using a commercial 500 – 9000 b (New England Biolabs) or 281-6583 b (Promega) ladder as a size marker. RNA samples (typically 10-20 µg) were added to 20 µl RNA loading buffer (7.38 M formaldehyde, 20% v/v formamide, in 1 x MNE buffer) and 1 µl ethidium bromide at 0.2 µg.ml^{-1}, and incubated at 65 °C for 5 min before loading. To

faciliatate this analysis, MNE buffer was prepared as a 5 x stock (0.12 M MOPS, 25 mM sodium acetate, 5 mM EDTA, adjusted to pH 7.0), and diluted in RNAase-free H_2O to the appropriate concentration.

DNA and RNA electrophoresis gels were visualised using a trans-UV illuminator and Gel Doc software (BioRad).

2.3.3 Pulse field gel electrophoresis

Prior to electrophoretic separation, the pulsed field gel electrophoresis (PFGE) apparatus (CHEF-DR III, BioRad) was cleaned by the circulation of 2 l of 0.1% SDS overnight at 20 °C. The tank was then rinsed twice by circulating dH_2O for ~1 h at 15 °C, and once by circulating the appropriate electrophoresis buffer for ~1 h at 15 °C (1 x TB1/10E (90 mM Tris base, 90 mM boric acid, 2 mM EDTA) was used for the separation of megabase chromosomes, whereas 0.5 x TBE (45 mM Tris base, 45 mM boric acid, 10 mM EDTA) was used for the separation of mini-chromosomes). Gels were electrophoresed in 2 l buffer, which was circulated in the tank for at least 30 min at 15 °C before the gel was run.

All separations were conducted using 1.2% agarose (Seakem LE, BioWhittaker Molecular Applications). To do this, agarose was dissolved to the correct concentration in 150 ml of the appropriate electrophoresis buffer, and 140 ml used to prepare a gel using the tray provided with the PFGE system; the remainder was kept at 37 °C. After the agarose gel had set, the comb was removed, agarose genomic plugs placed into the wells, and the wells sealed with the agarose at 37 °C. Agarose genomic plugs had been prepared by dialysis overnight in 3 ml of electrophoresis buffer on a shaking table, followed by three rounds of further dialysis each of 1 h in fresh electrophoresis buffer. Gels were electrophoresed at 15 °C, either at 2.5 $V.cm^{-1}$ for 144 h with an initial switch time of 1400 sec and final switch time of 700 sec for the separation of megabase chromosomes, or at 5.8 $V.cm^{-1}$ for 24h with initial and final switch times of 20 sec for the separation of mini-chromosomes. Chromosomes were visualised by placing agarose gels in 200 ml electrophoresis buffer containing 4 µl ethidium bromide at 10 $\mu g.\mu l^{-1}$ on a rocking table for ~30 min. They were then destained in dH_2O for ~30 min, or until they could be visualised clearly by UV illumination.

2.4 Blotting

2.4.1 Southern blotting

Agarose gels to be Southern blotted were photographed on a UV transilluminator, alongside a ruler to allow calculation of the sizes of fragments hybridised by radioactively labelled DNA (see section 2.5). The gels were then soaked in 125 mM HCl for 15 min, denaturation solution (0.5 M NaOH, 1.5 M NaCl) for 30 min and neutralisation solution (1 M Tris base, 1.5 M NaCl, 186 mM HCl) for 30min. DNA was transferred to a nylon membrane (Hybond XL, Amersham Biosciences) by overnight capillary blotting (Sambrook *et al.*, 1989) using 20 x SSC (3 M NaCl, 300 mM NaOAc) as transfer buffer. Following transfer, the DNA was crosslinked to the membrane using the auto-crosslink setting on a UV Stratalinker (Stratagene).

Pulse field gels were blotted essentially as described above, though with slightly different treatments. After ethidium bromide staining, gels were soaked in 125 mM HCl for two periods of 7 min, transferred to denaturation solution for two periods of 15 min and to neutralisation solution for 30 min; in addition, they were rinsed with dH_20 between each treatment. Finally, the gels were rinsed in 20 x SSC before overnight capillary blotting and crosslinking performed as above.

2.4.2 Northern blotting

Agarose gels to be northern blotted were photographed on a UV transilluminator, alongside a ruler to allow calculation of the sizes of fragments hybridised by radioactively labelled DNA (see section 2.5). Gels were soaked in sodium phosphate (10 mM Na_2HPO_4,/NaH_2PO_4) for 15 min before the transfer of RNA to a nylon membrane (Hybond XL, Amersham Biosciences) by overnight capillary blotting (Sambrook *et al.*, 1989) using sodium phosphate as transfer buffer. RNA was crosslinked to the membrane using the auto-crosslink setting on a UV Stratalinker (Stratagene).

2.5 Radiolabelling and hybridisation of DNA probes

2.5.1 Probe manufacture by random hexamer labelling of DNA

DNA fragments used for probe manufacture were specific PCR products amplified as described in section 2.7, separated on an agarose gel and purified using the Qiagen gel extraction kit, following the manufacturer's protocol. DNA fragments to be purified were excised from the agarose gel using a scalpel, and dissolved in 3 volumes of Buffer QG (*e.g.*, a gel fragment weighing 100 mg was dissolved in 300 µl of buffer) by incubation at 50 °C for 10 min. 1 gel volume of isopropanol was added to the solution and mixed, then the sample was applied to a QIAquick column in a collection tube, centrifuged at 16 x g for 1 min in a microcentrifuge, and the flow-through was discarded. The column was washed by addition of 750 µl of Buffer PE and centrifugation at 16 x g for 1 min in a microcentrifuge. The flow-through was discarded and the column was centrifuged at top speed for a further 1 min in a microcentrifuge. The column was placed in a clean eppendorf tube and DNA was eluted by applying 30 µl of sterile dH_2O to the column, letting it stand for 1 min and centrifuging the column at 16 x g for 1 min in a microcentrifuge.

Radiolabelling of these fragments was carried out using the Prime It II kit (Stratagene). ~25 ng DNA was mixed with 10 µl random hexameric oligonucleotides at 27 OD units.ml^{-1} and dH_20 in a final reaction volume of 36 µl, and the DNA denatured by incubation at 95 °C for 5 min. 10 µl of 5 x dATP or dCTP primer buffer, 2 µl of α^{32}P-labelled dATP or dCTP (~0.74 MBq) and 1 µl Klenow DNA polymerase (5 U.ml^{-1}) were added and the reaction incubated at 37 °C for 4-10 min. Probes made in this way were purified from unincorporated nucleotides by size exclusion chromatography using Microspin columns (Amersham Biosciences) according to the manufacturer's protocol, and were denatured at 95 °C for 5 min before hybridisation.

2.5.2 Hybridisation of radiolabelled DNA probes

Nylon filters blotted with DNA or RNA (section 2.4) were placed in hybridisation tubes (Hybaid) with ~50 ml pre-warmed 0.5 M Church Gilbert solution (342 mM Na_2HPO_4, 158 mM NaH_2PO_4, 7% SDS, 1 mM EDTA) and pre-hybridised for ~1 h – overnight at 65 °C in a rotating hybridisation oven. Denatured, radiolabelled probe was added to the Church Gilbert solution in the tube and allowed to hybridise to the blot overnight at 65 °C in a

rotating hybridisation oven. Filters were then washed in a rotating hybridisation oven with 50 ml of 2 x SSC, 0.1% SDS for 30 min at 65 °C and 50 ml of 0.2 x SSC, 0.1% SDS for 30 min at 65 °C. Followig washing, filters were sealed in plastic and exposed to a phosphorimaging screen (Fuji) at room temperature for 4-72 h (depending on the strength of the signal) and visualised using a Typhoon 8600 phosphorimager (Amersham Biosciences).

2.5.3 Stripping of hybridised nylon membranes

To strip nylon membranes of hybridised probe DNA, membranes were placed in a heat-proof container, then boiling 0.1% SDS was poured onto them and allowed to cool to room temperature. The SDS solution was then poured off and the procedure repeated. Stripping was checked by exposure to a phosphorimage screen (Fuji) for ~24 h and visualisation using a Typhoon 8600 (Amersham Biosciences).

2.6 Restriction enzyme digestion of DNA

Routinely, restriction digestions were carried out in a final reaction volume of 30µl, containing ~10 µg of DNA, 3 µl of restriction enzyme (NEB at 10 or 20 $U.\mu l^{-1}$) and 3 µl of 10 x buffer (NEB) as recommended by the manufacturer. Digests were incubated at the appropriate temperature for the enzyme(s) for ~2 h for plasmid DNA, or overnight for genomic DNA.

2.7 Polymerase chain reaction (PCR)

2.7.1 Standard PCR

PCRs were normally set up in volumes of 25µl for diagnostic reactions and 50µl for reactions intended to amplify DNA fragments for cloning or transformation. The amounts of reagents used in 25µl reactions were simply half those used in the 50µl reactions, which contained either 1µl of either Taq (ABGene, at $5U.ml^{-1}$) or Herculase (Stratagene, at $5U.ml^{-1}$) DNA polymerase. 5µl of the manufacturer's 10 x reaction buffer, 2 µl of 10 mM dNTPs and 2 µl of forward and reverse oligonucleotide primers (5 mM). For Taq-based PCRs, $MgCl_2$ was typically added to a final concentration of 1.5 mM, although this was occasionally increased to improve efficiency. For Herculase-based PCRs, the reaction buffer provides 2.0 mM Mg^{++}, although this was also occasionally increased by addition of

$MgCl_2$, up to a maximum concentration of 6 mM. In both reactions, dH2O was added to a final volume of 50 µl. PCR was conducted either in Robocycler (Stratagene) or PCRSprint (Hybaid) machines. Reaction conditions were 95 °C for 5 min, followed by 30 cycles of 95 °C for 1 min, 50-60 °C for 1 min, and 72 °C for 1 min per kb of expected product, and a final cycle of 72 °C for 10 min. PCR products were routinely purified using the Qiagen PCR Purification and Gel Extraction kits, following manufacturer's instructions as described in sections 2.1.3.2 and 2.5.1. Appendix 1 contains a list of oligonucleotides used for PCRs, and specific primers are referred to in the text.

The PCRs for amplification of microsatellites were performed using a cocktail provided by Annette MacLeod: PCR reactions were performed in 10 µl reaction volume in 45 mM Tris–HCl pH 8.8, 11 mM $(NH_4)_2SO_4$, 4.5 mM $MgCl_2$, 6.7 mM 2-mercaptoethanol, 4.4 µM EDTA pH 8.0, 113 µg ml^{-1} BSA, 1 mM each of the four deoxyribonucleotide triphosphates, 1 µM of each oligonucleotide primer, 1 unit of Taq DNA polymerase (ABGene, at $5U.ml^{-1}$). These reactions were carried out in a Robocycler machine (Stratagene) under the following conditions: 96°C for 50 s, 66°C for 50 s and 70°C for 90 s, for a total of 32 cycles.

2.7.2 Reverse transcription PCR (RT-PCR)

Total RNA was treated with DNAaseI to remove genomic DNA prior to cDNA preparation. To do this, 2 µg of RNA was incubated with 1 µl of DNAaseI (Invitrogen, $1U.ml^{-1}$) and 1 µl of 10 x DNAaseI buffer in a final reaction volume of 10 µl. The reaction was terminated by the addition of 1 µl 0.25 mM EDTA pH 8.0 and incubation at 65 °C for 20 min.

cDNA was prepared from DNase-treated RNA using the Superscript First-Strand Synthesis System for RT-PCR kit (Invitrogen), according to manufacturer's instructions. 50 ng of random hexamers and 1 µl of dNTPs were added to 8 µl of DNAase-treated RNA and the mixture incubated at 65 °C for 5 min and on ice for 1 min. 4 µl of 25 mM $MgCl_2$, 2 µl of 0.1 M DTT, 2 µl of 10 x RT buffer and 1 µl of RNaseOUT recombinant ribonuclease inhibitor were added, and incubated for 2 min at 25 °C. 1 µl of Superscript II reverse transcriptase (RT; 200 $U.ml^{-1}$) was then added, and the reaction incubated at 25 °C for 10 min, followed by 42 °C for 50 min. For each RT reaction, a duplicate reaction was set up using the same RNA but water instead of RT, thereby acting as a control for DNA contamination in downstream experiments. Following cDNA generation, RT was heat-inactivated at 70 °C for 15 min. Finally, 1 µl RNaseH ($3.8\ U.ml^{-1}$) was added and the

reaction incubated at 37 °C for 20 min. cDNA prepared in this way was used directly in PCR reactions, with 1 µl of undiluted cDNA routinely acting as a substrate in 25 µl reaction volumes.

2.8 Cloning of DNA fragments

2.8.1 Cloning using T4 DNA ligase

DNA fragments for cloning were prepared either by PCR-amplification and purification (section 2.7), or by restriction digestion (section 2.6). When vectors were restriction digested using a single enzyme, self-ligation was prevented by treatment with calf intestinal phosphatase (CIP; Roche). To do this, 0.5 µl of CIP (10 $U.ml^{-1}$) was added to the restriction digestion reaction and incubated at 37 °C for 10 min. After CIP treatment, vectors were purified by agarose gel extraction following electrophoresis using the Qiagen gel extraction kit according to manufacturer's instructions as described in section 2.5.1. Inserts for cloning, either derived by PCR or by plasmid digestion, were also purified by gel extraction following agarose gel electrophoresis.

Ligation of DNA fragments into a plasmid vector were carried out in a 10 µl reaction volume, containing 1 µl of T4 DNA ligase (400 $U.ml^{-1}$, New England Biolabs) and 1 µl of ligase buffer (New England Biolabs), and were incubated at room temperature for ~6 h or 16 °C overnight. 1 µl of the 10 µl ligation reaction was used to transform 60-120 µl of *E. coli* XL-1 blue MRF' cells (see section 2.9).

2.8.2 Cloning into the TOPO vector

Cloning DNA fragments into the TOPO TA vector (Invitrogen) occurs using the 3' single adenosine overhang that is present on all PCR products generated by Taq DNA polymerase. Therefore, PCR products generated by Herculase DNA polymerase were treated by the addition of 1 µl of Taq DNA polymerase per 50 µl reaction and incubation at 72 °C for 20 min prior to cloning. For either Taq or Herculase PCRs, 0.5-4 µl of PCR product was incubated with 1µl of salt solution (provided with the vector) and 1 µl TOPO TA vector, made up to a final reaction volume of 6 µl with dH_2O and incubated for 5 min at room temperature. 2 µl of this reaction was then used to transform 25 µl TOP10 F' *E. coli* cells (Invitrogen) (see section 2.9).

2.9 Transformation of *E. coli* and plasmid retrieval

Transformation of both XL-1 blue MRF' (Stratagene) and TOP10 F' (Invitrogen) *E. coli* cells was carried out using heat shock. Ligations (see section 2.8) and cells were mixed and left on ice for 20 min. The cells were then heat-shocked at 42 °C for 45 sec and transferred to ice for ~2 min. All plasmids used in this study encode ampicillin resistance; to select for transformants, cells were therefore spread on L-agar plates containing ampicillin at a final concentration of 100 $\mu g.ml^{-1}$ (Sigma) and incubated overnight at 37 °C.

Single colonies from bacterial plates were picked and used to inoculate 3-200 ml of L-broth containing ampicillin (Sigma) at a final concentration of 100 $\mu g.ml^{-1}$ and grown up overnight at 37 °C in a shaking incubator. Plasmids were purified from 1.5 ml of the overnight culture using the Qiagen miniprep kit, following manufacturer's instructions. Cells were pelleted by centrifugation at 16 x g in a microcentrifuge for 1 min, and the supernatant was discarded. Cells were then resuspended in 250 µl of resuspension buffer P1 (containing RNase A added following manufacturer's instructions) by pipetting, 250 µl of alkaline lysis buffer P2 was added and the solution mixed by inverting the eppendorf tube 4-6 times. 350 µl of high-salt neutralising buffer N3 was added and the solution mixed as before, then centrifuged at 16 x g in a microcentrifuge for 10 min. The supernatant was applied to a QIAprep column in a collection tube, centrifuged at top speed in a microcentrifuge for 1 min, and the flow-through was discarded. 750 µl of Buffer PE was added to the column and centrifuged at 16 x g in a microcentrifuge for 1 min, the flow-through discarded, and the column centrifuged for an additional 1 min to remove residual buffer. Plasmid DNA was eluted from the column by addition of 50 µl of dH_2O to the column (placed in a clean eppendorf tube) and centrifugation 16 x g in a microcentrifuge for 1 min.

When larger amounts of DNA were required, plasmids were purified from 200 ml of the overnight culture using the Sigma maxiprep kit, according to manufacturer's instructions. Cells were pelleted by centrifugation at 5000 x g for 10 min, then resuspended in 12 ml of Resuspension Solution by pipetting. Cells were then lysed by addition of 12 ml of Lysis Solution and mixture by inversion of the tube several times, and neutralised by addition of 12 ml of neutralisation solution. 9 ml of binding solution was added, the tube of cells was inverted 1 or 2 times, applied to the barrel of a filter syringe, and left for 5 min. During this incubation step, the binding column was prepared: 12 ml of Column Preparation Solution was added to the column, which was then centrifuged at 3000 x g for 2 min. Half

of the cleared lysate was expelled through the filter syringe into the binding column, and centrifuged at 3000 x g for 2 min; the eluate was discarded, and this step was repeated with the other half of the lysate. The column was washed twice: first by addition of 12 ml of Wash Solution 1, centrifugation at 3000 x g for 2 min and discarding of the eluate, and secondly by addition of 12 ml of Wash Solution 2 and centrifugation at 3000 x g for 5 min. The binding column was transferred to a clean 50 ml collection tube and 3 ml of Elution Solution was added to the column; plasmid DNA was eluted by centrifugation at 3000 x g for 5 min.

CHAPTER 3

MISMATCH REPAIR AND THE CONTROL OF HOMOLOGOUS RECOMBINATION IN *T. BRUCEI*

3 Mismatch repair and the control of homologous recombination in *T. brucei*

3.1 Introduction

One of the roles of the MMR system is the prevention of homologous recombination (HR) between similar yet divergent (homeologous) sequences. MMR-dependent prevention of recombination between divergent substrates is important in the maintenance of speciation in bacteria (Zahrt *et al.*, 1994; Matic *et al.*, 1995), yeast (Hunter *et al.*, 1996; Chen and Jinks-Robertson, 1999) and mice (te Riele *et al.*, 1992; Deng and Capecchi, 1992). In addition, MMR prevents recombination between divergent repeats in genomes, which can lead to gross chromosomal rearrangements (Petit *et al.*, 1991).

As well as preventing HR as described in more detail below, MMR can promote recombination in certain specialised circumstances. For example, in *S. cerevisiae*, once a mismatched heteroduplex is formed, MMR can act within it (Priebe *et al.*, 1994) to catalyse gene conversion with more than 85% of mismatches corrected in favour of the resident, unbroken strand (Leung *et al.*, 1997). In addition, the Msh2-Msh3 heterodimer is involved in binding specifically to (Evans and Alani, 2000; Evans *et al.*, 2000), and catalysing the removal of, nonhomologous single-strand tails created during recombinational processes, in conjunction with the Rad1-Rad10 heterodimeric endonuclease (Saparbaev *et al.*, 1996; Sugawara *et al.*, 1997; Colaiacovo *et al.*, 1999). Similarly, the Msh2-Msh3 dimer stimulates the two-end invasion pathway of targeted gene replacement in *S. cerevisiae* (Langston and Symington, 2005).

3.1.1 The importance of sequence length and homology in homologous recombination

HR is exquisitely sensitive to the presence of sequence differences between recombining substrates. When heterologous sequences undergo strand exchange, divergence between them results in base mismatches during strand transfer, and even very low levels of divergence can have a striking effect on recombination efficiency. A single base difference causes a significant reduction (up to four-fold) in recombination efficiency in yeast (Datta *et al.*, 1997), mice (Lukacsovich and Waldman, 1999) and plants (Opperman *et al.*, 2004). At higher levels of divergence, the reduction is even stronger (Bailis and

Rothstein, 1990; Harris *et al.*, 1993; Priebe *et al.*, 1994), with, for example, 83% sequence homology (17% divergence) causing a 180-fold reduction in yeast recombination compared to perfectly matched substrates (Selva *et al.*, 1995). There is an exponential reduction in recombination efficiency with increasing substrate heterology for substrates of a fixed length in *Streptococcus pneumoniae* over the range 0.6 to 27% (Majewski *et al.*, 2000), *S. cerevisiae* over the range 0-26% (Datta *et al.*, 1997) and higher eukaryotes, over the range 0.5-9% in *A. thaliana* (Li *et al.*, 2006). An exception to this is the specialised transformation-associated recombination (TAR) cloning, where sequence length has little effect (Kouprina et al 1998 PNAS). In at least one case, at higher levels of sequence divergence, recombination rates plateau: in one experiment in *Bacillus subtilis*, no additional decrease in recombination was observed once divergence rose above 8% (Zawadzki *et al.*, 1995), although whether this can be extrapolated beyond this bacterium is not known.

The MMR system contributes to the partial prevention of homeologous recombination described above. Genetic inactivation of either MutS or MutL in bacterial species *E. coli*, *S. pneumoniae* and *S. typhimurium* has been observed to partially alleviate the restraint on recombination between DNA sequences in the range 1% to 18% sequence divergence (Rayssiguier *et al.*, 1989; Abdulkarim and Hughes, 1996; Zahrt and Maloy, 1997; Westmoreland *et al.*, 1997; Worth, Jr. *et al.*, 1998; Majewski *et al.*, 2000). Mutation of MMR factors also improves the rate of recombination between homeologous substrates in eukaryotic organisms, although all components of the machinery do not have an equal effect. Genetic disruption of *MSH2* causes an increase in homeologous recombination in the range 0.3-9% in *S. cerevisiae* (Selva *et al.*, 1995; Datta *et al.*, 1997; Negritto *et al.*, 1997; Elliott and Jasin, 2001; Li *et al.*, 2006), in the range 0.5-9% in *A. thaliana* (Li *et al.*, 2006) and for substrates with 1.5% divergence in murine cells (Elliott and Jasin, 2001). In the moss *Physcomitrella patens*, the 22-fold reduction in recombination efficiency caused by 3% divergence in the substrates is completely alleviated by genetic disruption of *MSH2* (Trouiller *et al.*, 2006). Msh2's partners Msh3 and Msh6 are partially redundant to each other in their roles in recombination surveillance (Spell and Jinks-Robertson, 2003), and *msh3* mutants show increased recombination between substrates containing base-base mismatches, even though Msh3 does not recognise this substrate in post-replicative MMR (Marsischky *et al.*, 1996; Earley and Crouse, 1998). The effects of genetic disruption of both Msh2 and Msh3 are more extreme than genetic disruption of Msh2 alone (Selva *et al.*, 1995). Disruption of MutS homologues has a greater effect on the regulation of recombination than removal of the MutL homologues (Chen and Jinks-Robertson, 1999; Nicholson *et al.*, 2000; Welz-Voegele *et al.*, 2002). Separation-of-function alleles of both

Pms1 (Welz-Voegele *et al.*, 2002) and bacterial MutS (Calmann *et al.*, 2005) have been isolated, showing that ATPase activity and the extreme C-terminal sections of these proteins, respectively, are needed for their anti-recombination functions, although not mismatch recognition function. The absence of MMR has also been observed in some cases to have a stimulatory effect on recombination between perfectly matched substrates: a lack of proofreading activity presumably allows the recombination reaction to proceed more rapidly (Datta *et al.*, 1996; Negritto *et al.*, 1997; Chen and Jinks-Robertson, 1998).

The recombinational machinery also has MMR-independent roles in the prevention of homeologous recombination (Chen and Jinks-Robertson, 1999), although this only comes into play at higher levels of divergence. *E. coli* RecA is capable of catalysing branch migration *in vitro* between sequences that are 3% divergent, although at a reduced rate compared to homologous sequences; this branch migration is partially prevented by addition of MutS (DasGupta and Radding, 1982; Worth, Jr. *et al.*, 1994). On the other hand, substrates with more than ~10% sequence non-homology (Bazemore *et al.*, 1997) or containing insertions or deletions (Bucka and Stasiak, 2001) cannot undergo RecA-catalysed branch migration *in vivo*; whether this is also true for eukaryotic Rad51 is not known. *in vivo* analysis also indicates that the HR machinery is sensitive to base mismatches in ways that can be independent of MMR. In *E. coli*, upregulation of RecA can increase the amount of recombination at all levels of sequence divergence, whereas elimination of MutS increases recombination proportionally to divergence (Vulic *et al.*, 1997). Similarly, in yeast, recombination between substrates with 23% divergence is unaffected by MMR, whereas the reduction in MMR caused by the presence of 9% divergence in the recombination substrates is almost completely removed in MMR-deficient cells, with recombination rates approaching those seen in perfectly matched substrates (Datta *et al.*, 1996). In *A. thaliana*, on the other hand, the increase in recombination frequency in inactivation of *MSH2* is consistently roughly 3-fold in the range 0.5-9% (Li *et al.*, 2006). In fact, the relative importance of the RecA/Rad51-mediated and MMR-mediated controls of homeologous recombination differs between species. In bacteria, it has been observed that MMR is more important in the control of homeologous recombination in *Escherichia* and *Salmonella* species than in *Bacillus* and *Streptococcus* species (Humbert *et al.*, 1995; Majewski *et al.*, 2000; Prunier and Leclercq, 2005), whereas *Mycobacteria smegmatis* lack a MMR system altogether while still maintaining stringent requirements for homology in RecA-mediated HR (Springer *et al.*, 2004). In eukaryotic organisms, MMR has been calculated to act more stringently in the control of HR in yeast than in plants (Li *et al.*, 2006; see section 3.3).

Yeast Rad51-dependent and Rad59-dependent, Rad51-independent HR (see section 1.2) differ both in their requirements for substrate homology and in their interactions with the MMR machinery (Spell and Jinks-Robertson, 2003). The Rad51-dependent pathway is regulated by Msh2, since mutation causes alleviation of homeologous recombination suppression, whereas mutations of Mlh1 have only a limited effect. In contrast, the Rad59-dependent pathway, which appears to be more tolerant of sequence divergence and to act on shorter substrates (Ira and Haber, 2002), is regulated equally by Msh2 and Mlh1, and mutations of either MMR gene causes a smaller increase in homeologous recombination relative to Rad51-dependent recombination. Genetic disruption of Mlh1 nevertheless causes a greater increase in homeologous recombination in wild-type cells than in either *rad51* or *rad59* mutants, suggesting that the two pathways co-operate in the cell.

Another factor influencing the rate of HR is the length of the homologous sequence catalysing recombination. Although the exact substrate lengths vary between studies, it can generally be stated that HR rates reduce with substrate length. Exponential reductions in recombination efficiency have been described between 405 bp and 27 bp in *E. coli* (Shen and Huang, 1986) and 960 bp and 80 bp in *S. cerevisiae* (Jinks-Robertson *et al.*, 1993). In a more recent study in *E. coli* (Lovett *et al.*, 2002), a linear relationship was only observed between 200 bp and 50 bp, with a plateau in recombination efficiency observed above this range. In *S. cerevisiae*, an exponential reduction in recombination efficiency was seen between 2 kb and 26 bp (Ahn *et al.*, 1988). Longer substrates seem to be required for efficient mammalian HR, with recombination observed as much less efficient with substrates shorter than 163 bp in one assay (Rubnitz and Subramani, 1984), whereas in another study, a threshold of 1.7 kb in another (Hasty *et al.*, 1991). A plateau in mammalian recombination frequency was observed for substrates longer than 14 kb in another study (Deng and Capecchi, 1992). However, small but significant amounts of recombination have been seen with substrates measuring only 25 bp in bacterial and human cells (Ayares *et al.*, 1986; Lovett *et al.*, 2002). These reactions are thought to be RecA/Rad51-independent: yeast Rad59-dependent recombination is known to require shorter substrates than Rad51-dependent reactions (Bai and Symington, 1996; Sugawara *et al.*, 2000), and recombination using the shortest substrates in a bacterial experiment is independent of RecA, with RecA-independent reactions having much less length dependence than RecA-dependent ones (Lovett *et al.*, 2002).

The Minimal Efficient Processing Segment (MEPS), which can be calculated by linear regression, is defined as the minimum substrate length permitting efficient recombination, below which the rate of recombination decreases sharply (Shen and Huang, 1986). A

longer substrate can be considered as a series of overlapping MEPS units, explaining the log-linear relationship between substrate length and recombination efficiency. In homeologous substrates, the MEPS can also be thought of as a mismatch-free block of sequence in which the local mismatch concentration is low enough to escape the MMR system and initiate recombination: the length of this block has been calculated as 15 bp in MMR-deficient, but 62 bp in wild-type bacteria (Vulic *et al.*, 1997), and 18 bp in MMR-deficient, but 28 bp in wild-type *S. cerevisiae* cells (Datta *et al.*, 1997). The location of mismatches within a recombination substrate is important, with branch migration able to continue through a small number of mismatches as long as recombination is initiated within a perfectly matched region (Waldman and Liskay, 1988; Majewski and Cohan, 1999). In addition to this role in the initiation of recombination, MMR monitors the nascent heteroduplex. The heteroduplex rejection model (Rayssiguier *et al.*, 1989)) proposes that MMR enzymes, in conjunction with helicases, can abort recombination reactions producing mismatched heteroduplexes; Datta *et al* (Datta *et al.*, 1997) calculated that 610 bp of uninterrupted homology is needed for escape of the MMR system. Although the alternative heteroduplex destruction model (Rayssiguier *et al.*, 1989), involving nicking and destruction of mismatched heteroduplex, is also proposed, the heteroduplex rejection model is supported by the fact that many interactions between recombination and MMR proteins and helicases of the RecQ family have been identified (Pedrazzi *et al.*, 2001; Wu and Hickson, 2001; Langland *et al.*, 2001; Spell and Jinks-Robertson, 2004). For this unwinding to occur, the mismatched heteroduplex must be recognised immediately at its formation: MMR has no effect on preformed heteroduplexes (Worth, Jr. *et al.*, 1994; Westmoreland *et al.*, 1997; Worth, Jr. *et al.*, 1998).

3.1.2 A system to study the effects of sequence divergence on recombination efficiency of long substrates in T. brucei

Methods for stable transformation of linear DNA by integration into the genomes of kinetoplastids following electroporation of cultured cells were developed a number of years ago (Ten Asbrock *et al.*, 1990; Cruz and Beverley, 1990; Cruz *et al.*, 1991; Cruz *et al.*, 1993; Cooper *et al.*, 1993; Hariharan *et al.*, 1993; Otsu *et al.*, 1993), and this technique is now widely used for genetic manipulation of these species. Integration of transformed DNA into the correct locus occurs in effectively 100% of cases in wild-type cells (Eid and Sollner-Webb, 1987; Cooper *et al.*, 1993; Otsu *et al.*, 1993; Conway *et al.*, 2002b). These experiments have yielded some data on transformation efficiency in the kinetoplastids using different substrates. It has been observed in *T. brucei* that linear DNA recombines

with higher efficiency than circular plasmids and, secondly, that transformation efficiency is improved when 5 kb rather than 900 bp substrates are used (Ten Asbroek *et al.*, 1990). In contrast, no effect of substrate length on transformation efficiency was observed in *T. brucei* in two separate studies: Wickstead *et al* (Wickstead *et al.*, 2003) found no differences in rates comparing constructs with integration flanks between 80 and 400 bp, and Shen *et al* (Shen *et al.*, 2001) saw no difference in transformation using targeting sequences of 50 – 90 bp in length. Despite these variations in findings, more broadly the substrates needed for transformation in *T. brucei* and *T. cruzi* (Hariharan *et al.*, 1993) seem to be shorter than those of at least 180 bp that appear to be needed by *L. major* (Papadopoulou and Dumas, 1997); perhaps the lower threshold for efficient recombinationis different between these organisms. Transformation in procyclic form *T. brucei* was suggested not to be affected by the number of potential target loci present in the genome (Wickstead *et al.*, 2003). This was interpreted as suggesting that it is the recombination reaction itself, and not the search for sequence homology, that is rate-limiting in *T. brucei*. This contrasts with the situation in *S. cerevisiae*, where the amount of transformants obtained is proportional to the number of targets (Wilson *et al.*, 1994).

Transformation of any organism is likely to be influenced by a number of factors, including DNA concentration, transformation conditions, and growth of the cells undergoing electroporation. Despite these potential problems, an assay, summarised in figure 3.1, has been developed to measure the recombination efficiency of both homologous and homeologous DNA sequences in *T. brucei*, in the presence and absence of a functional MMR system (Bell and McCulloch, 2003). As they will form the basis of the work in this chapter, the original study will be described in detail here. Although transformation assays are not an ideal experimental system with which to measure recombination, assay systems not involving electroporation such as have been developed in *S. cerevisiae* (*e.g.* the inverted intron system used in Datta *et al.*, 1996) are not available in *T. brucei*. Transformation has been used successfully to measure recombination efficiency in the absence of *T. brucei* RAD51, MRE11, RAD51-3 and RAD51-5, in addition to the results described below, validating this experimental approach.

The first component of the assay is a series of cell lines containing unique sites for recombination. The hygromycin phosphotransferase ORF (*HYG*) was integrated into the tubulin array (Seebeck *et al.*, 1983) on chromosome I of Lister 427 bloodstream stage cells, creating a unique site for recombination referred to as the *HTUB* locus. An integrated foreign sequence as a target site for recombination offers a number of advantages. Firstly, it provides a defined, single site for recombination from which a

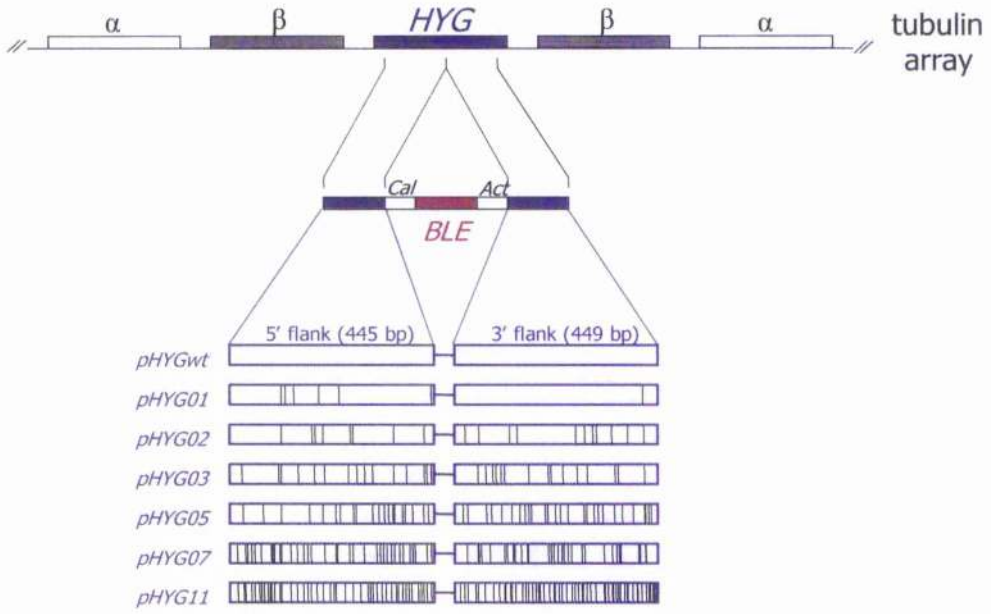


Figure 3.1. A schematic representation of an assay to measure the effect of MMR on *T. brucei* homologous recombination (adapted from Bell and McCulloch, 2003). *HYG*, encoding hygromycin resistance, was integrated into the repetitive tubulin array of Lister 427 bloodstream stage cells, replacing one of the copies of α tubulin. *MSH2* mutants were made in this genetic background. Recombination into this locus using a range of constructs was measured. All constructs contained *BLE*, the bleomycin resistance gene, together with processing signals derived from the calmodulin (*Cal*) and actin (*Act*) loci, and integration flanks corresponding to the two halves of *HYG*, with differing levels of divergence from the wild-type sequence. Sites of base mismatches are represented by vertical lines.

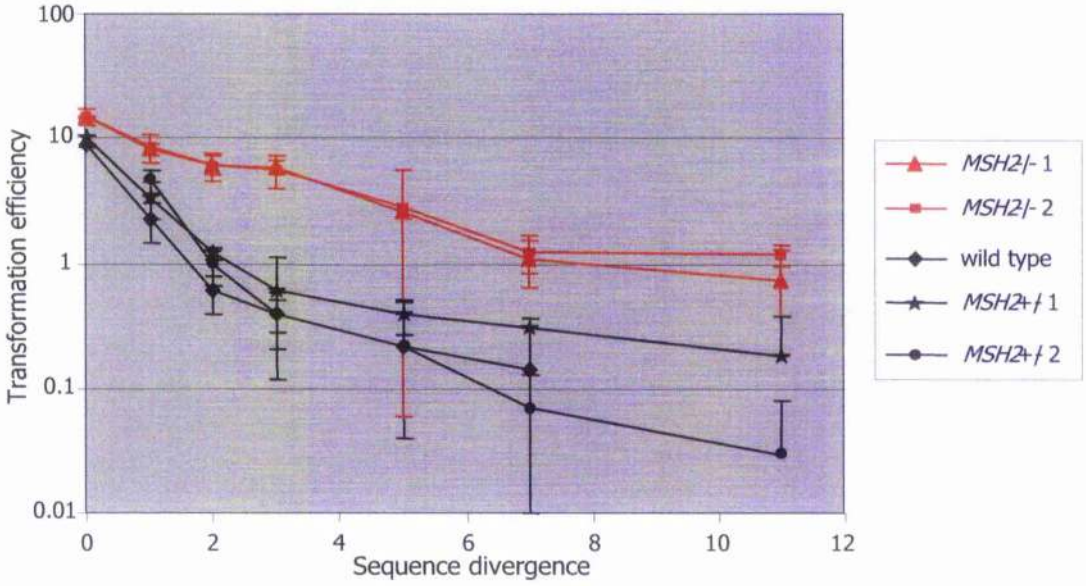


Figure 3.2. A graph showing transformation efficiency (in transformants x10^{-6} cells) against sequence divergence (as a percentage) of the 450-bp recombination substrates. Adapted from Bell and McCulloch, 2003.

number of parameters (sequence length, similarity) can be varied. Secondly, the potential for integrations into related sequences elsewhere in the genome when endogenous *T. brucei* targeting flanks are used is reduced. Thirdly, this strategy could be adapted to allow

comparison of recombination in different parts of the genome, as the *HYG* target can subsequently be integrated into different genomic locations. To examine the effect of MMR in *T. brucei* recombination, *MSH2* knockout cell lines were made in the *HTUB* genetic background by eliminating the *MSH2* ORF from the genome using constructs *ΔMSH2::BSD* and *ΔMSH2::PUR*; the cell lines are referred to as *HTUB wt*, *HTUB MSH2+/-* and *HTUB MSH2-/-*. To assay recombination substrate requirements, a series of constructs based on a bleomycin resistance cassette were generated. The ORF encoding the bleomycin resistance protein (*BLE*), surrounded by intergenic regions from the actin and calmodulin loci, was flanked by 445 bp and 449 bp *HYG*-derived sequences, allowing recombination with the *HTUB* locus following transformation. The constructs differed from each other in having increasing numbers of base mismatches (relative to the integrated *HYG*) that had been generated by PCR mutagenesis. Constructs with 0%, 1%, 2%, 3%, 5%, 7%, 9% and 11% divergence from the original *HYG* sequence were selected and named *pHYGwt*, *pHYG01*, *pHYG02*, *pHYG03*, *pHYG05*, *pHYG07*, *pHYG09* and *pHYG11*, respectively. To assay transformation, each construct was excised from the plasmid backbone by restriction digestion, phenol chloroform extracted and ethanol precipitated, and a constant amount of DNA was transformed into *HTUB wt*, *HTUB MSH2+/-* and *HTUB MSH2-/-* cell lines by electroporation.

The recombination efficiencies observed from this assay are summarised in figure 3.2. The highest recombination efficiency observed (as calculated by dividing the number of positive wells counted, by the number of cells put on selection), after electroporation of the *pHYGwt* construct, was 14.7 +/- SE 0.573 transformants x 10^{-6} cells in *HTUB MSH2-/-* cells and 8.8 +/- SE 0.146 transformants x 10^{-6} cells in *HTUB wt* cells; this slight increase in recombination rate between perfectly matched substrates in MMR-deficient cells has been seen in other organisms (see section 3.1.1). A reduction in the level of sequence similarity between the recombination substrates significantly reduced recombination efficiency. For instance, in the *HTUB wt* cell line, the transformation rate of the *pHYG01* construct (which has 1% sequence divergence from the genomic *HYG* sequence) was 3-fold lower than that of *pHYGwt*. Moreover, increasing sequence divergence had an exponentially detrimental effect on recombination, such that transformation of the *pHYG11* construct (11% divergence) was nearly 100-fold lower than *pHYGwt*. This constraint on recombination between divergent sequences was alleviated, although not altogether

removed, in the *HTUB MSH2-/-* cell line, indicating that MMR regulates HR as in other organisms. However, even in the absence of MSH2, recombination became less efficient with increasing divergence, indicating that the *T. brucei* HR machinery is itself sensitive to base mismatches. Overall, the levels of transformation in the *HTUB MSH2-/-* cells were approximately 6-12-fold higher than in *HTUB wt*, in the range 2-11% divergence.

These experiments have demonstrated that the MMR system in *T. brucei* is functional and acts to limit recombination between homeologous substrates. The work in the following chapter is based on an extended version of this assay, which was used to assess the requirements for substrate length and sequence homology in HR, and to investigate whether different substrates are monitored in different ways by the MMR machinery.

3.2 Results

3.2.1 An assay to study the requirements for substrate length and homology levels in MMR+ and MMR- T. brucei

The plasmids *pHYGwt*, *pHYG05* and *pHYG11* described above (see section 3.1.2) were used as substrates for high-fidelity PCR to create a series of new transformation constructs (see figure 3.3), using the primers shown in figure 3.4. These PCR reactions yielded a series of constructs with integration flanks of approximately 25 bp, 50 bp, 100 bp, 150 bp and 200 bp, which were named *0%-25bp*, *0%-50bp*, *etc.*, as shown in table 3.1. Although they will be referred to as 5% and 11% divergent, the targeting flanks amplified from the *pHYG05* and *pHYG11* plasmids did not contain precisely 5% and 11% mismatches compared to the wild-type *HYG* locus, depending on the section of the flank that was amplified; the exact number of mismatches is shown in table 3.1. Linear DNA was prepared for these transformations by PCR amplification, rather than restriction digestion of constructs from plasmids as previously (Bell and McCulloch, 2003); this change in protocol was necessary because non-homologous overhangs at the DNA ends after digestion, while insignificant for an integration flank of 450 bp, could have larger effects on recombination efficiencies mediated by the shorter substrates. PCR products were purified and ethanol precipitated (see figure 3.5), and a constant amount of DNA (3 μg) was transformed into 2.5 x 10^7 *HTUB wt*, *HTUB MSH2+/-* or *HTUB MSH2-/-* cells by electroporation. Transformants were selected in 24-well plates, using phleomycin at 2.5 $\mu g.ml^{-1}$, and the number of transformants counted after 7-10 days. Differing amounts of cells were put on selection, depending on the expected differences in transformation

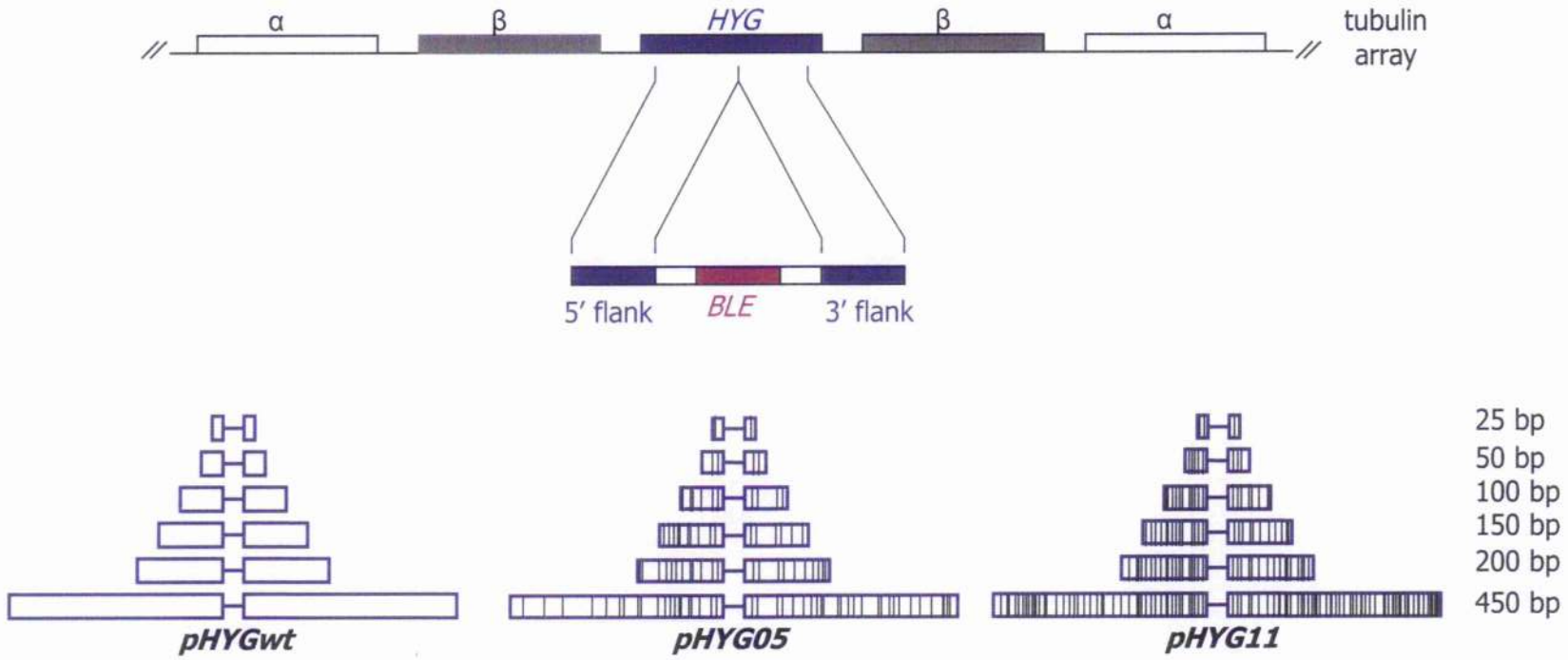
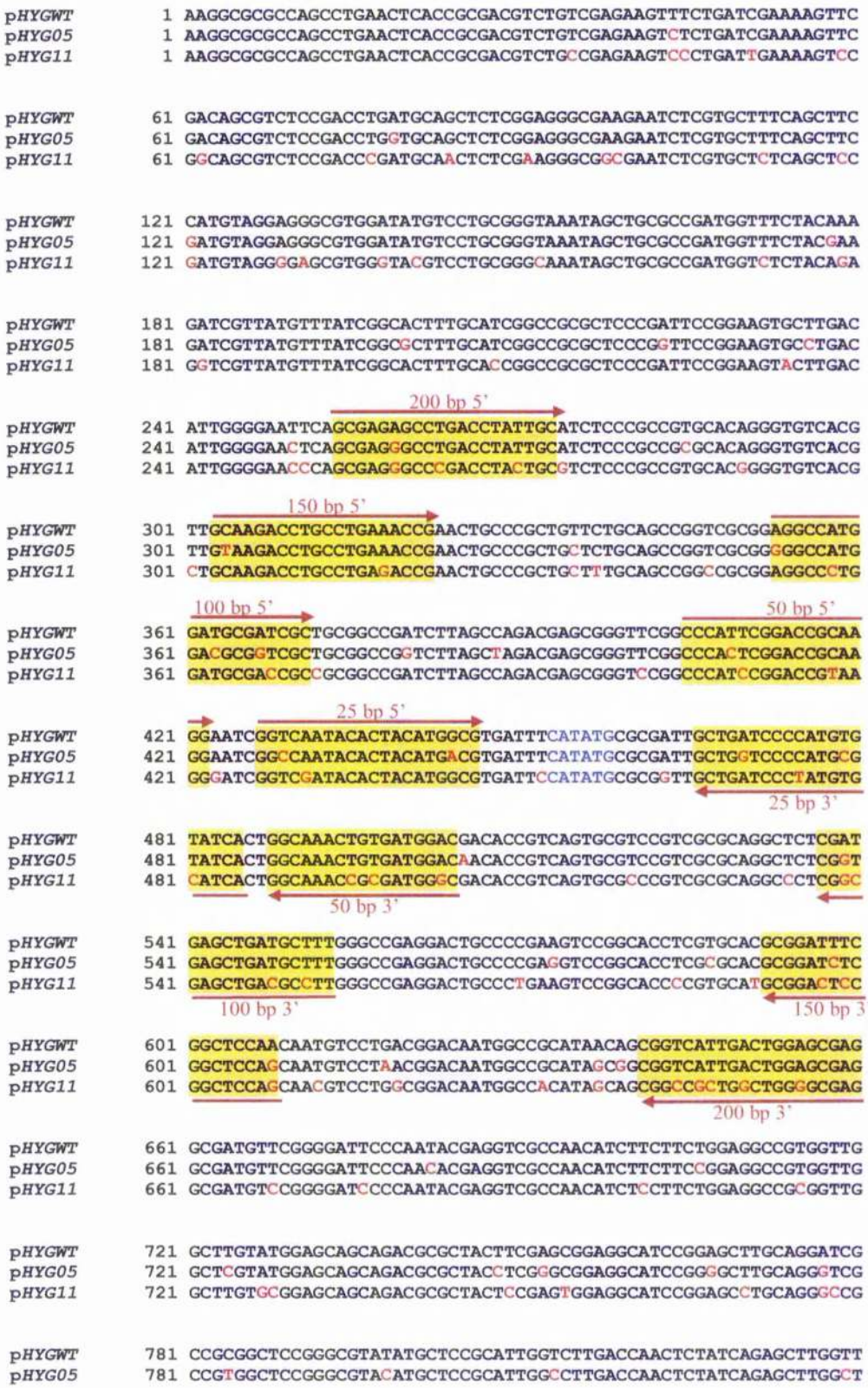


Figure 3.3. A schematic representation of an assay to measure the influence of mismatch repair on recombination using substrates of varying length and sequence homology. The *HYG* ORF was integrated into the tubulin array of Lister 427 bloodstream stage cells, and recombination of a number of constructs was measured by selecting for phleomycin resistance. These were made by PCR amplification of sections of increasing size using *pHYGwt*, *pHYG05* and *pHYG11* as templates; they had integration flanks of approximately 25, 50, 100, 150, 200 and 450 bp as indicated to the right of the diagram.


pHYGWT 1 AAGGCGCGCCAGCCTGAACTCACCGCGACGTCTGTCGAGAAGTTTCTGATCGAAAAGTTC
pHYG05 1 AAGGCGCGCCAGCCTGAACTCACCGCGACGTCTGTCGAGAAGTCTCTGATCGAAAAGTTC
pHYG11 1 AAGGCGCGCCAGCCTGAACTCACCGCGACGTCTGCCGAGAAGTCCCTGATTGAAAAGTCC
pHYGWT 61 GACAGCGTCTCCGACCTGATGCAGCTCTCGGAGGGCGAAGAATCTCGTGCTTTCAGCTTC
pHYG05 61 GACAGCGTCTCCGACCTGGTGCAGCTCTCGGAGGGCGAAGAATCTCGTGCTTTCAGCTTC
pHYG11 61 GGCAGCGTCTCCGACCCGATGCAACTCTCGAAGGGCGGCGAATCTCGTGCTCTCAGCTCC
pHYGWT 121 CATGTAGGAGGGCGTGGATATGTCCTGCGGGTAAATAGCTGCGCCGATGGTTTCTACAAA
pHYG05 121 GATGTAGGAGGGCGTGGATATGTCCTGCGGGTAAATAGCTGCGCCGATGGTTTCTACGAA
pHYG11 121 GATGTAGGGGAGCGTGGGTACGTCCTGCGGGCAAATAGCTGCGCCGATGGTCTCTACAGA
200 bp 5'
pHYGWT 241 ATTGGGGAATTCAGCGAGAGCCTGACCTATTGCATCTCCCGCCGTGCACAGGGTGTCACG
pHYG05 241 ATTGGGGAACTCAGCGAGGGCCTGACCTATTGCATCTCCCGCCGCGCACAGGGTGTCACG
pHYG11 241 ATTGGGGAACCCAGCGAGGGCCCGACCTACTGCGTCTCCCGCCGTGCACGGGGTGTCACG
150 bp 5'
pHYGWT 301 TTGCAAGACCTGCCTGAAACCGAACTGCCCGCTGTTCTGCAGCCGGTCGCGGAGGCCATG
pHYG05 301 TTGTAAGACCTGCCTGAAACCGAACTGCCCGCTGCTCTGCAGCCGGTCGCGGGGGCCATG
pHYG11 301 CTGCAAGACCTGCCTGAGACCGAACTGCCCGCTGCTTTGCAGCCGGCCGCGGAGGCCCTG
100 bp 5'
50 bp 5'
pHYGWT 361 GATGCGATCGCTGCGGCCGATCTTAGCCAGACGAGCGGGTTCGGCCCATTCGGACCGCAA
pHYG05 361 GACGCGGTCGCTGCGGCCGGTCTTAGCTAGACGAGCGGGTTCGGCCCACTCGGACCGCAA
pHYG11 361 GATGCGACCGCCGCGGCCGATCTTAGCCAGACGAGCGGGTCCGGCCCATCCGGACCGTAA
25 bp 5'
pHYGWT 421 GGAATCGGTCAATACACTACATGGCGTGATTTCATATGCGCGATTGCTGATCCCCATGTG
pHYG05 421 GGAATCGGCCAATACACTACATGACGTGATTTCATATGCGCGATTGCTGGTCCCCATGCG
pHYG11 421 GGGATCGGTCGATACACTACATGGCGTGATTCCATATGCGCGGTTGCTGATCCCTATGTG
25 bp 3'
pHYGWT 481 TATCACTGGCAAACTGTGATGGACGACACCGTCAGTGCGTCCGTCGCGCAGGCTCTCGAT
pHYG05 481 TATCACTGGCAAACTGTGATGGACAACACCGTCAGTGCGTCCGTCGCGCAGGCTCTCGGT
pHYG11 481 CATCACTGGCAAACCGCGATGGGCGACACCGTCAGTGCGCCCGTCGCGCAGGCCCTCGGC
50 bp 3'
pHYGWT 541 GAGCTGATGCTTTGGGCCGAGGACTGCCCCGAAGTCCGGCACCTCGTGCACGCGGATTTC
pHYG05 541 GAGCTGATGCTTTGGGCCGAGGACTGCCCCGAGGTCCGGCACCTCGCGCACGCGGATCTC
pHYG11 541 GAGCTGACGCCTTGGGCCGAGGACTGCCCTGAAGTCCGGCACCCCGTGCATGCGGACTCC
100 bp 3'
150 bp 3'
pHYGWT 601 GGCTCCAACAATGTCCTGACGGACAATGGCCGCATAACAGCGGTCATTGACTGGAGCGAG
pHYG05 601 GGCTCCAGCAATGTCCTAACGGACAATGGCCGCATAGCGGCGGTCATTGACTGGAGCGAG
pHYG11 601 GGCTCCAGCAACGTCCTGGCGGACAATGGCCACATAGCAGCGGCCGCTGGCTGGGGCGAG
200 bp 3'
pHYGWT 661 GCGATGTTCGGGGATTCCCAATACGAGGTCGCCAACATCTTCTTCTGGAGGCCGTGGTTG
pHYG05 661 GCGATGTTCGGGGATTCCCAACACGAGGTCGCCAACATCTTCTTCCGGAGGCCGTGGTTG
pHYG11 661 GCGATGTCCGGGGATCCCCAATACGAGGTCGCCAACATCTCCTTCTGGAGGCCGCGGTTG
pHYGWT 721 GCTTGTATGGAGCAGCAGACGCGCTACTTCGAGCGGAGGCATCCGGAGCTTGCAGGATCG
pHYG05 721 GCTCGTATGGAGCAGCAGACGCGCTACCTCGGGCGGAGGCATCCGGGGCTTGCAGGGTCG
pHYG11 721 GCTTGTGCGGAGCAGCAGACGCGCTACTCCGAGTGGAGGCATCCGGAGCCTGCAGGGCCG
pHYGWT 781 CCGCGGCTCCGGGCGTATATGCTCCGCATTGGTCTTGACCAACTCTATCAGAGCTTGGTT
pHYG05 781 CCGTGGCTCCGGGCGTACATGCTCCGCATTGGCCTTGACCAACTCTATCAGAGCTTGGCT

```
pHYGWT      841 GACGGCAATTTCAATGATGCAGCTTGGGCGCAGGGTCGATGCGACGCAATCGTCCGATCC
pHYG05      841 GACGGCAATTTCGGTGATACAGCCTGGGCGCAGGGTCGATGCGACGCAATCGTCCGATCC
pHYG11      841 GGCGGCAATTTCGACGGTGCAGCCTGGGCGCAGGGTCGGTACGACGCAATCGTCCGATCC

pHYGWT      901 GGAGGGCGCGCCAT
pHYG05      901 GGAGGGCGCGCCAT
pHYG11      901 GGAGGGCGCGCCAT
```

Figure 3.4. An alignment of the integration constructs used in this chapter. Multiple alignment of the sequences of the *pHYGwt*, *pHYG05* and *pHYG11* plasmids, aligned using Multalin (http://prodes.toulouse.inra.fr/multalin/multalin.html; Corpet, 1988). Identical residues are shown in black, residues differing from the genomic *HYG* sequence are shown in red, and the restriction site used to insert the bleomycin resistance cassette is shown in blue. Primers are indicated by a purple arrow and yellow shading. Adapted from J. Bell, PhD thesis, University of Glasgow

		Primers		Integration flank size		Number of mismatches		Percentage mismatches			Uninterrupted homology		
PCR product	**Substrate**	**5'**	**3'**	**5'**	**3'**	**5'**	**3'**	**5'**	**3'**	**Average**	**5'**	**3'**	**Selection**
0%-25bp	*pHYGwt*	25bp-5'-0%	25bp-3'	26 bp	27 bp	0	0	0	0	0	26 bp	27 bp	10
0%-50bp	*pHYGwt*	50bp-5'-0%	50bp-3'-0%	48 bp	46 bp	0	0	0	0	0	48 bp	46 bp	8
0%-100bp	*pHYGwt*	100bp-5'-0%	100bp-3'-0%	100 bp	95 bp	0	0	0	0	0	100 bp	95 bp	6
0%-150bp	*pHYGwt*	150bp-5'-0%	150bp-3'-0%	150 bp	149 bp	0	0	0	0	0	150 bp	149 bp	4
0%-200bp	*pHYGwt*	200bp-5'-0%	200bp-3'-0%	199 bp	202 bp	0	0	0	0	0	199 bp	202 bp	2
5%-25bp	*pHYG05*	25bp-5'-0%	25bp-3'-0%	26 bp	27 bp	2	2	7.7	7.4	7.55	14 bp	11 bp	10
5%-50bp	*pHYG05*	50bp-5'-0%	50bp-3'-0%	48 bp	46 bp	3	2	6.3	4.3	5.3	19 bp	25 bp	10
5%-100bp	*pHYG05*	100bp-5'-0%	100bp-3'-0%	100 bp	95 bp	8	4	8	4.2	6.1	20 bp	33 bp	8
5%-150bp	*pHYG05*	150bp-5'-0%	150bp-3'-0%	150 bp	149 bp	10	7	6.6	4.7	5.65	30 bp	33 bp	6
5%-200bp	*pHYG05*	200bp-5'-0%	200bp-3'-0%	199 bp	202 bp	12	11	6	5.4	5.7	30 bp	33 bp	4
11%-25bp	*pHYG11*	25bp-5'	25bp-3'	26 bp	27 bp	2	3	7.7	11.1	9.4	20 bp	11 bp	10
11%-50bp	*pHYG11*	50bp-5'	50bp-3'	48 bp	46 bp	5	6	10.4	13	11.7	20 bp	13 bp	10
11%-100bp	*pHYG11*	100bp-5'	100bp-3'	100 bp	95 bp	9	12	9	12.6	10.8	28 bp	16 bp	10
11%-150bp	*pHYG11*	150bp-5'	150bp-3'	150 bp	149 bp	13	17	8.7	11.4	10.05	28 bp	18 bp	8
11%-200bp	*pHYG11*	200bp-5'	200bp-3'	199 bp	202 bp	19	27	9.5	13.3	11.4	28 bp	18 bp	6

Table 3.1. PCR products used to measure transformation efficiency. A table showing the PCR products used in the transformation efficiency assay (PCR product), the plasmids used as substrates for the PCR (Substrate), the names of the primers used (Primers), the actual sizes of the integration flanks amplified (Integration flank size), the number of mismatches in each flank (Number of mismatches), the corresponding percentage of mismatches (Percentage mismatches) in each flank and as an average of both flanks, the longest uninterrupted stretch of perfect homology with the genomic *HYG* locus in each flank (Uninterrupted homology), and the number of cells ($\times 10^6$) that were placed on selection per 24-well plate following electroporation (Selection).

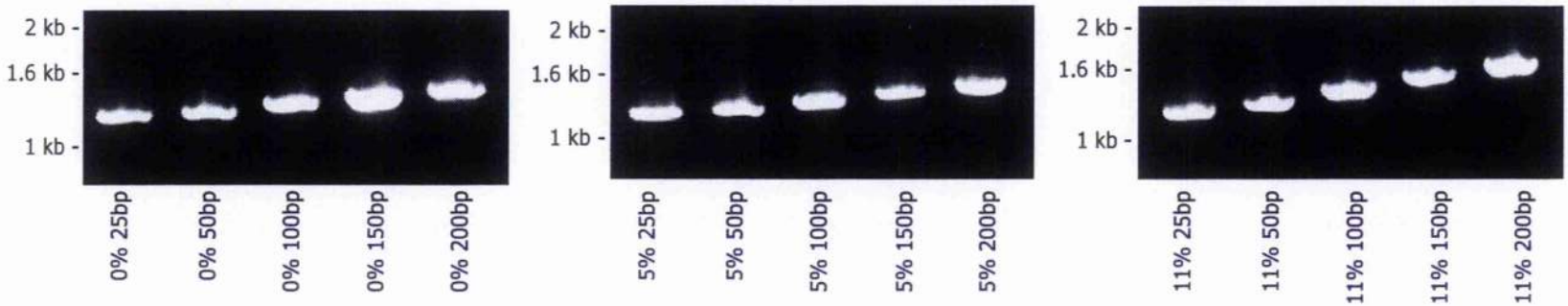


Figure 3.5. PCR products used to measure transformation efficiency. Constructs were prepared by PCR amplification using *pHYGwt*, *pHYG05* and *pHYG11* as templates followed by purification and ethanol precipitation. 1 µl of each construct was run on an ethidium bromide-stained agarose gel.

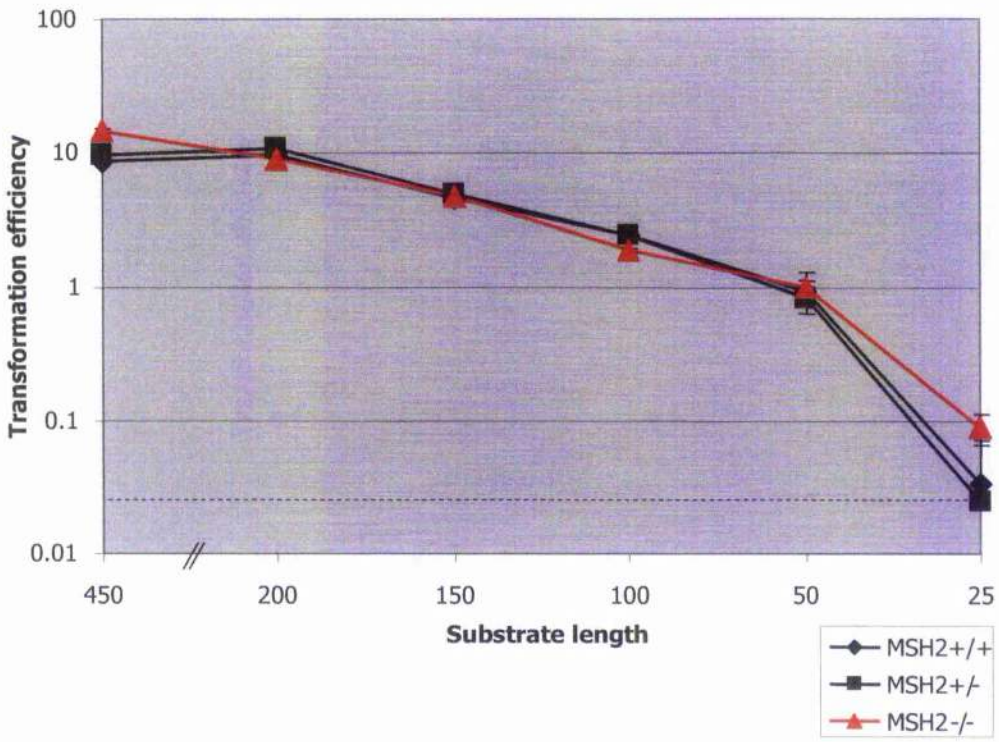
efficiency of the different recombination substrates (see table 3.1). This was done in order to ensure near-clonal growth of transformants in the more efficient reactions, while detecting the few transformants obtained in the less efficient reactions. Transformations for each cell line were typically carried out on three separate occasions, with two different selection plates counted from each transformation. The threshold of transformant detection for each plate is 1 transformant x 10^{-7} cells. Where no transformants were detected for any of the transformations, this frequency can be divided by the total number of plates analysed (maximum 6), providing a minimum detectable transformation efficiency of 0.17 x 10^{-7}. For clarity, this value was plotted in the analysis below when no transformants were found.

Calculations of transformation efficiency were first carried out by the method used in previous studies by the McCulloch and Barry groups: the number of positive wells scored after 7 – 10 days was divided by the number of cells put on selection. These calculations are shown in sections 3.2.1.1-3.2.1.3. However, this simple method has the drawback of not taking into account the possibility that more than one transformant could be found in each well, leading to an underestimation of the transformation rates for the more efficient reactions. Therefore, more in-depth statistical analysis was carried out in addition to the original simple calculations, and the revised transformation efficiencies are described and discussed in section 3.2.2.

3.2.1.1 0% divergence

Transformation constructs derived from PCR-amplification of sections of the *pHYGwt* construct, and therefore with integration flanks perfectly matched to the genomic *HYG* locus, were transformed into *HTUB wt*, *HTUB MSH2+/-* and *HTUB MSH2-/-* cell lines as described above. The transformation efficiencies obtained, along with results previously obtained using the *pHYGwt* full-length sequence (450 bp; Bell and McCulloch, 2003), are shown in graphical form in figure 3.6. The presence of an active MMR system does not seem to have any effect on the recombination of perfectly matched substrates, as striking similarities were observed between the results of the wild-type, *MSH2+/-* and *MSH2-/-* cell lines. In the original experiments, recombination of the full-length, 450 bp construct was slightly more efficient in the *HTUB MSH2-/-* cell line compared with the *HTUB wt* cell line (Bell and McCulloch, 2003); this difference was not observed when shorter substrates were used. However, the transformation efficiencies obtained using the 200 bp integration flanks were highly similar to those obtained using the 450 bp integration flank in *HTUB wt* cells: the *0%-200bp* construct generated BLE-resistant clones in the *HTUB wt* and *HTUB MSH2-/-* cell lines at average frequencies of 9.83 +/- SE 0.423 and 9.13 +/- SE 0.754

a)



b)

	MSH2+/+		MSH2-/-		
Length (bp)	Transformants/10^6 cells	Reduction	Transformants/10^6 cells	Reduction	MMR-/MMR+
200	9.833		9.125		0.928
150	4.667	2.107	4.813	1.896	1.031
100	2.417	4.069	1.854	4.921	0.767
50	0.917	10.727	0.975	9.359	1.064
25	0.033	295	0.088	104.286	2.625

	MSH2+/-		MSH2-/-		
Length (bp)	Transformants/10^6 cells	Reduction	Transformants/10^6 cells	Reduction	MMR-/MMR+
200	10.833		9.125		0.842
150	4.875	2.222	4.813	1.896	0.987
100	2.444	4.432	1.854	4.921	0.759
50	0.813	13.333	0.975	9.359	1.2
25	0.025	433.333	0.088	104.286	3.5

Figure 3.6. Recombination mediated by substrates perfectly matched to the target sequence. a) a log-linear plot of transformation efficiency (in transformations x 10-6 cells) versus substrate length (in bp). b) mean transformation efficiencies of wild-type (MSH2+/+; top), *MSH2+/-* (bottom) and *MSH2-/-* (both) for each substrate length. The column marked Reduction shows the fold reduction relative to substrates with 200 bp targeting flanks, whereas the MMR-/MMR+ column shows the fold increase in transformation efficiency of *MSH2-/-* cells relative to *MSH2+/+* or *MSH2+/-* cells.

transformants x 10^{-6} cells, respectively. This may suggest that *T. brucei* recombination efficiency plateaus at 200 bp, and that increasing the substrate length beyond this does not improve recombination. It is also possible, however, that this represents the maximum achievable transformation efficiency in these cells, at least with the parameters used here. Shortening the integration flanks below 200 bp caused a log-linear decrease in recombination efficiency, for all cell lines studied, down to 50 bp substrate length. The average transformation rates for the 50 bp construct were 0.912 +/- SE 0.175 and 0.975 +/- SE 0.281 transformants x 10^{-6} cells in *HTUB wt* and *HTUB MSH2-/-* cells, respectively, corresponding to an approximately 10-fold reduction in efficiency compared to *0%-200bp* (see figure 3.6 b). At the shortest substrate length of 25 bp, the reactions were extremely inefficient with average transformation rates of only 0.033 +/- SE 0.037 and 0.088 +/- SE 0.023 transformants x 10^{-6} cells in *HTUB wt* and *HTUB MSH2-/-* cells, respectively. This represents a reduction of 295-fold and 104-fold compared with *0%-200bp*, and 28-fold and 11-fold compared with *0%-50bp*, respectively. Together, this appears to indicate a break in the log-linear relationship at around 25 bp, perhaps indicating that the recombination machinery in *T. brucei* becomes very inefficient around this substrate length. Comparing the transformation efficiencies for all constructs between *HTUB wt* or *HTUB MSH2+/-* and *HTUB MSH2-/-* cells (figure 3.6 b) shows that impairment of MMR had no discernible effect on recombination when the substrates are sequence matched.

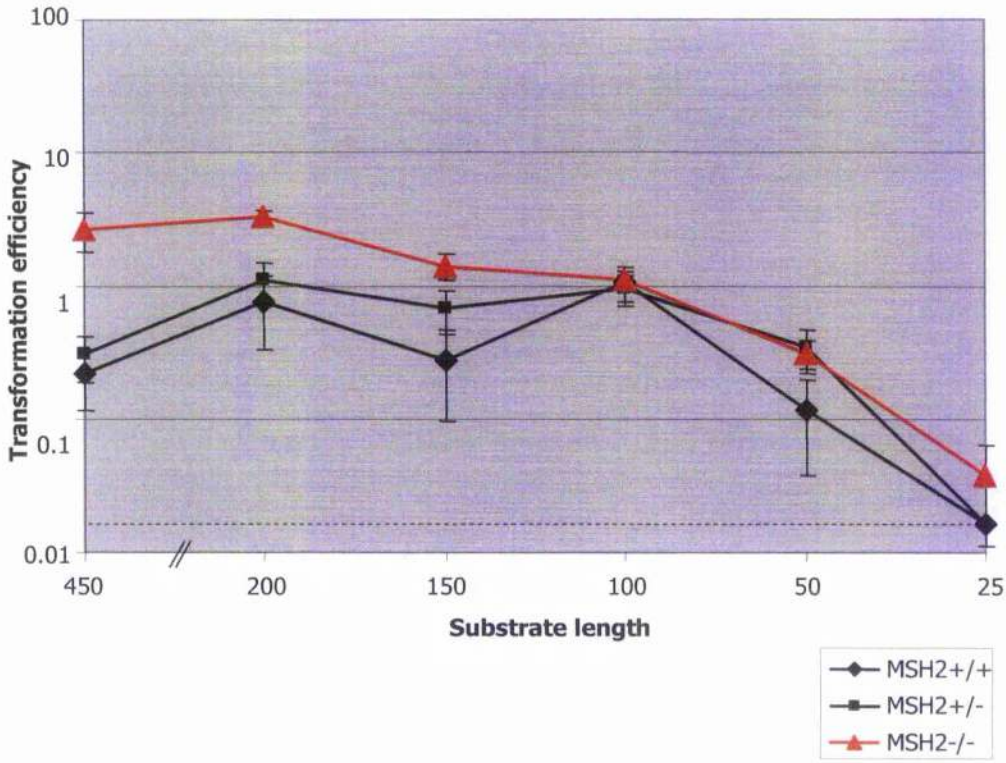
3.2.1.2 5% divergence

Transformation constructs derived from PCR-amplification of sections of the *pHYG05* construct, and therefore with integration flanks ~5% diverged from the genomic *HYG* locus, were next transformed into *HTUB wt*, *HTUB MSH2+/-* or *HTUB MSH2-/-* cell lines as before. These results, along with those obtained using the *pHYG05* full-length sequence, are shown in graphical form in figure 3.7. Two primary conclusions can be drawn from these results. Firstly, at all substrate lengths, there is a considerable reduction in recombination efficiency compared to substrates perfectly matched to the genomic *HYG* sequence. For example, *5%-200bp* had a transformation efficiency of 0.75 x 10^{-6} compared with 9.8 x 10^{-6} for *0%-200bp* (a 13-fold reduction). The second conclusion is that, at this level of sequence divergence, mutation of *MSH2*, and therefore MMR impairment, influences recombination. The transformation frequencies obtained in *HTUB MSH2+/-* cells are perhaps slightly higher than those in *HTUB wt* cells, but both follow the same basic pattern; although large fluctuations were seen, there appeared to be a plateau in efficiency with all substrates longer than 100 bp generating transformants at approximately the same frequency (between 0.278-1.071 x 10^{-6} transformants for wild-type, and 0.667-

1.083 x 10^{-6} transformants for *MSH2*+/-). Below this threshold length, there was a steep drop-off in recombination efficiency, with *5%-50bp* and *5%-25bp* showing 6.43-fold and 30-fold reductions in recombination efficiency respectively compared with *5%-200bp* in *HTUB wt* cells; in *HTUB MSH2*+/- cells, 3-fold and 6-fold reductions were observed (see figure 3.7 b).

The *HTUB MSH2*-/- cell line displayed a different pattern of efficiency, however. Here, the results obtained with the full-length *pHYG05* construct were similar to those obtained with *5%-200bp* (average recombination frequencies of 2.63 +/- SE 0.849 and 3.22 +/- SE 0.425 transformants x 10^{-6} cells, respectively; see figure 3.7 b). 5% diverged constructs with integration flanks between 200 bp and 100 bp showed a gradual decrease in recombination efficiency, with reductions of 2.29-fold and 2.93-fold when the recombination efficiencies obtained with *5%-150bp* and *5%-100bp*, respectively, are compared with *5%-200bp*. With substrates shorter than 100 bp, a more pronounced reduction in recombination efficiency was observed, with 10.36-fold and 85.83-fold reductions in efficiency when rates obtained with the *5%-50bp* and *5%-25bp* constructs, respectively are compared with *5%-200bp*. It appears, therefore, that *MSH2* mutation allows the recombination machinery to act with increasing efficiency in constructs above 100 bp, which is not seen in MMR-proficient cells. Below ~100 bp, as for wild-type or *MSH2*+/- cells, recombination becomes highly inefficient, analogous to the sharp decline seen at 50 bp with perfectly matched substrates. Mutation of *MSH2* increases the rate of recombination on 5% diverged substrates. This is most obvious above 100 bp, as *pHYG05*, *5%-200bp* and *5%-150bp* constructs recombined at 2.6, 3.2 and 1.4 x 10^{-6} cells respectively, representing a 2-5-fold increase relative to *HTUB wt* or *MSH2*+/- cells.

a)



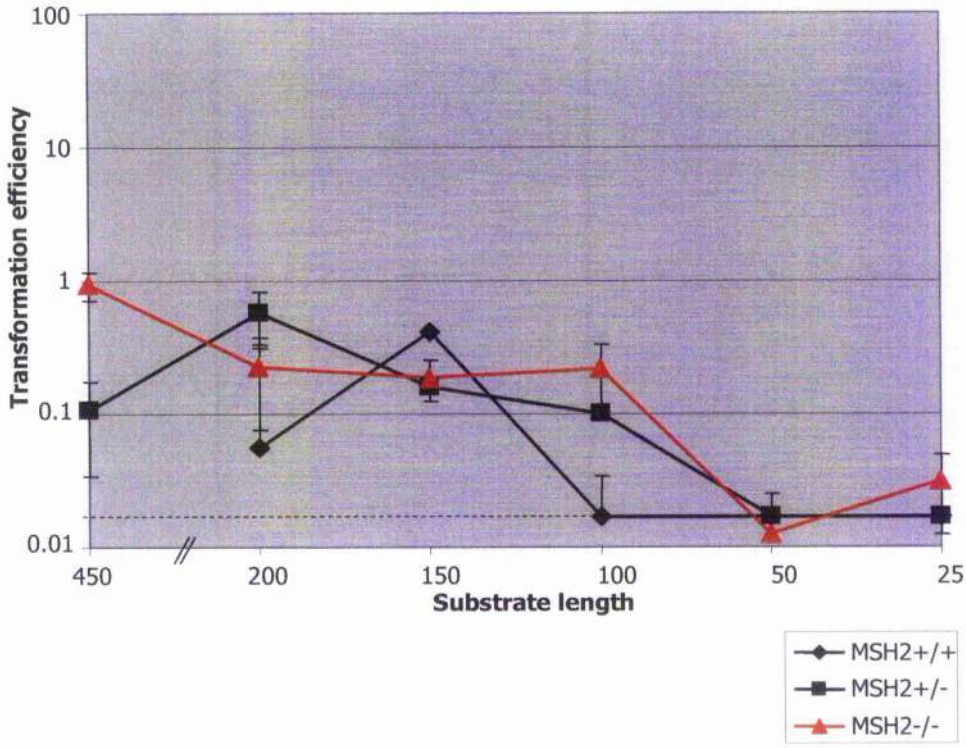
b)

	MSH2+/+		MSH2-/-		
Length (bp)	Transformants/10^6 cells	Reduction	Transformants/10^6 cells	Reduction	MMR-/MMR+
200	0.750		3.219		4.292
150	0.278	2.7	1.4	2.299	5.04
100	1.071	0.7	1.099	2.930	1.025
50	0.117	6.429	0.311	10.364	2.662
25	0.016	47	0.038	85.833	1.5

	MSH2+/-		MSH2-/-		
Length (bp)	Transformants/10^6 cells	Reduction	Transformants/10^6 cells	Reduction	MMR-/MMR+
200	1.083		3.219		2.971
150	0.667	1.625	1.4	2.299	2.100
100	0.938	1.156	1.099	2.930	1.172
50	0.343	3.158	0.311	10.364	0.905
25	0.016	67.708	0.038	85.833	2.344

Figure 3.7. Recombination mediated by substrates 5% divergent from the target sequence. a) a log-linear plot of transformation efficiency (in transformations x 10-6 cells) versus substrate length (in bp). b) mean transformation efficiencies of wild-type (MSH2+/+; top), *MSH2+/-* (bottom) and *MSH2-/-* (both) for each substrate length. The column marked Reduction shows the fold reduction relative to substrates with 200 bp targeting flanks, whereas the MMR-/MMR+ column shows the fold increase in transformation efficiency of *MSH2-/-* cells relative to *MSH2+/+* or *MSH2+/-* cells.

a)



b)

	MSH2+/+		MSH2-/-		
Length (bp)	Transformants/10^6 cells	Reduction	Transformants/10^6 cells	Reduction	MMR-/MMR+
200	0.056		0.222		4.000
150	0.417	0.133	0.188	1.185	0.450
100	0.017	3	0.217	1.026	12.745
50	0.017	3	0.013	17.778	0.750
25	0.017	3	0.031	7.196	1.817

	MSH2+/-		MSH2-/-		
Length (bp)	Transformants/10^6 cells	Reduction	Transformants/10^6 cells	Reduction	MMR-/MMR+
200	0.583		0.222		0.381
150	0.156	3.733	0.188	1.185	1.200
100	0.100	5.833	0.217	1.026	2.167
50	0.017	34.314	0.013	17.778	0.735
25	0.017	34.314	0.031	7.196	1.817

Figure 3.8. Recombination mediated by substrates 11% divergent from the target sequence. **a**) a log-linear plot of transformation efficiency (in transformations x 10-6 cells) versus substrate length (in bp). b) mean transformation efficiencies of wild-type (MSH2+/+; top), *MSH2+/-* (bottom) and *MSH2-/-* (both) for each substrate length. The column marked Reduction shows the fold reduction relative to substrates with 200 bp targeting flanks, whereas the MMR-/MMR+ column shows the fold increase in transformation efficiency of *MSH2-/-* cells relative to *MSH2+/+* or *MSH2+/-* cells.

3.2.1.3 11% divergence

Transformation constructs derived from PCR-amplification of sections of the *pHYG11* construct, and therefore with integration flanks ~11% diverged from the genomic *HYG* locus, were transformed into *HTUB wt*, *HTUB MSH2+/-* and *HTUB MSH2-/-* cell lines as described above. The recombination efficiencies obtained, along with results previously obtained using the *pHYG11* full-length sequence (apart from in *HTUB wt* cells: these data were not available), are shown in graphical form in figure 3.8. Recombination at this high level of sequence divergence is uniformly very inefficient, but does not appear to be affected by MMR. Even at this high level of sequence divergence, transformation efficiency appears to be dependent on substrate length, although there are large amounts of fluctuation between experiments. The maximum transformation efficiencies obtained, with the *11%-200bp* construct, were 0.056 +/- SE 0.056 transformants x 10^{-6} cells in *HTUB wt* cells, 0.583 +/- SE 0.25 transformants x 10^{-6} cells in *HTUB MSH2+/-* cells, and 0.222 +/- SE 0.147 transformants x 10^{-6} cells in *HTUB MSH2-/-* cells. A roughly log-linear reduction in transformation efficiency was observed down to 50 bp substrate length; with substrates measuring 50 bp and 25 bp, very few transformants were recovered.

3.2.2 Revised transformation efficiency calculations

More complex statistical analysis was needed to get a more accurate estimate of *T. brucei* transformation efficiencies. This was done with the invaluable mathematical help of Derek Pike. In similar cases to this one in the literature (for example, Wickstead *et al.*, 2003), the Poisson distribution has been used to calculate transformation efficiency. Here, λ, the mean number of positive transformants per well, is calculated with the equation $\lambda = \ln [1 - (n+/N)]$, where n+ is the number of positive wells and N is the total number of wells. This was considered a good measure of λ as long as less than 80% of wells were positive. This concurs with a long known fact in the statistical world that "the Poisson distribution is the limiting form of the binomial distribution when there is a large number of trials but only a small probability of success at each of them" (Bulmer, 1967). Unfortunately, in the experiments discussed here, in many of the more efficient reactions up to 23/24 wells were positive, and an alternative strategy was necessary.

The Poisson distribution is appropriate for the study of rare events in a certain area or time. The binomial distribution is similar to the Poisson distribution when the probabilities involved are small. It has the advantages of firstly, being valid over a wider range of values, whether the event is very rare or relatively frequent, and secondly, of allowing the

calculation of error based on the probability of multiple events occurring in the same interval (in this case, multiple cells being transformed in the same well), as described below. Therefore, this is the approach we used to calculate better estimations of transformation efficiencies in this case.

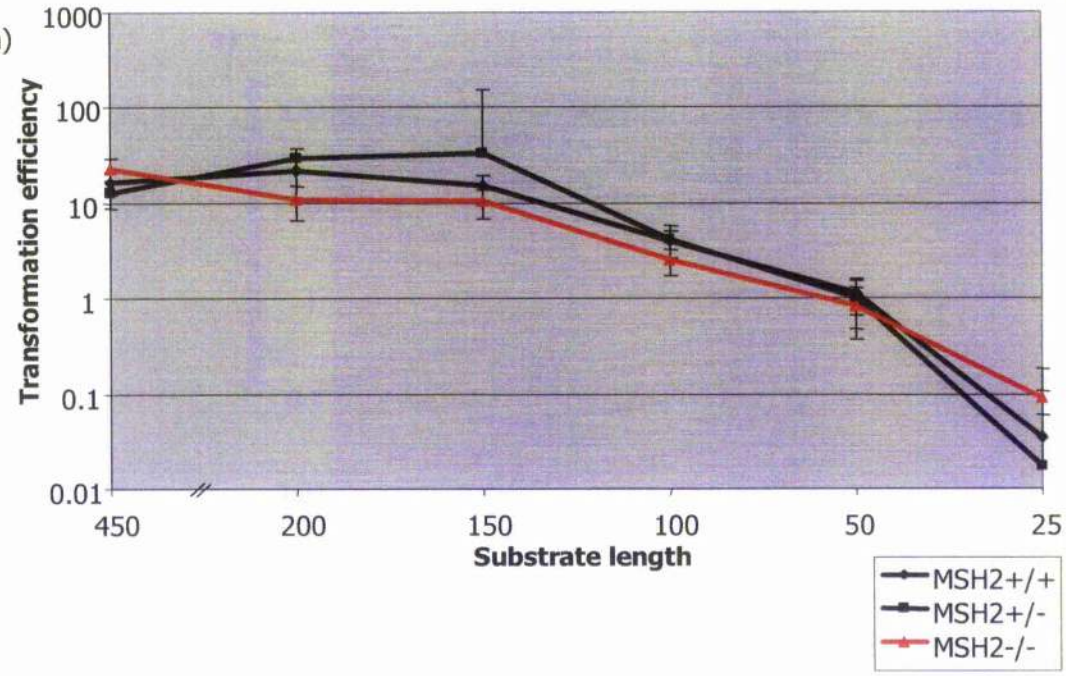
Firstly, we calculated the probability p of a cell being successfully transformed. By the binomial distribution, the probability q of no cells in a well being successfully transformed: $(1-p)^n = 1-q$. The probability of at least one cell in a well being successfully transformed: $q = 1 - (1-p)^n$. The probability of r out of 24 wells showing growth: ${}^{24}C_r\, q^r\, (1-q)^{24-r}$. From this, the maximum likelihood estimate of q is r/24. Therefore, $(1-p)^n = [1-(r/24)]$ and so $p = 1 - [1 - (r/24)]^{(1/n)}$.

The next problem was to calculate the standard error of the average transformation efficiencies calculated in this way. This needs to take into account both the error coming from not knowing how many cells were actually transformed in each positive well, but also the error caused by variation between repititions of the experiment. The first part of the total error, the variance of this estimate of probabililty, was calculated using a first order approximation based on a Taylor Series expansion of the desired function. Using this method, the "within-replicate variance", variance of the estimate of $p = (\{[1-q]^{[(1/n)-1]}\}/n)^2 * \mathrm{Var}(q)$. The second part of the error, the "data variance" or variation between replicates of the experiment, was calculated from the variation between all the transformation efficiencies as described above. These two sorts of variation were combined in the following way. The "unit variance" is the average of the within-replicate variances. The "replicate variance" = (data variance – unit variance) / 2. We can add this to the unit variance to get a total calculation of variance, and the square root of this number gives us the total standard error.

3.2.2.1 0% divergence

Transformation efficiencies were calculated using data obtained by transformation of constructs derived from the *pHYGwt* construct, using the binomial distribution as described above (shown in graphical form in figure 3.9). These revised transformation efficiencies differ from those generated originally by in two ways. The first difference is that the

a)



b)

	MSH2+/+		MSH2-/-		
Length (bp)	Transformants/10^6 cells	Reduction	Transformants/10^6 cells	Reduction	MMR-/MMR+
200	22.205		10.696		0.482
150	14.733	1.507	10.275	1.041	0.697
100	3.922	5.661	2.512	4.259	0.640
50	1.138	19.505	0.820	13.039	0.721
25	0.035	637.999	0.090	118.923	2.584

	MSH2+/-		MSH2-/-		
Length (bp)	Transformants/10^6 cells	Reduction	Transformants/10^6 cells	Reduction	MMR-/MMR+
200	29.347		10.696		0.364
150	33.275	0.882	10.275	1.041	0.309
100	4.184	7.014	2.512	4.259	0.600
50	0.998	29.394	0.820	13.039	0.822
25	0.017	1723.894	0.090	118.923	5.283

Figure 3.9. Recombination mediated by substrates perfectly matched to the target sequence; transformation efficiencies calculated using the binomial distribution. a) a log-linear plot of transformation efficiency (in transformations x 10^{-6} cells) vs substrate length (in bp). b) mean transformation efficiencies of wild-type (MSH2+/+; top), MSH2+/- (bottom) and MSH2-/- (both) for each substrate length. The column marked Reduction shows the fold reduction relative to substrates with 200 bp

standard errors are larger using the second method (especially for the results obtained using the *0%-150bp* construct). This is because the added error due to not knowing the number of transformants in each positive well is now taken into account, as well as the error between replicates. The second difference is that the more efficient reactions have much higher transformation efficiencies calculated by the new method. For example, transformation of the *0%-200bp* construct generated BLE-resistant clones at a rate of 22.203 +/- SE 7.475 transformants x 10^{-6} cells in the *HTUB wt* cell line using the new calculation method, as compared to 9.833 +/- SE 0.423 transformants x 10^{-6} cells using the old method. This was to be expected, as again, here the probability of the wells from these very efficient transformations containing multiple transformants is high. This effect is less striking for the *HTUB MSH2-/-* cell line, where the estimate of transformation efficiency for the *0%-200bp* construct is only increased to 10.696 +/- SE 4.029 transformants x 10^{-6} cells, as compared to 9.13 +/- SE 0.754 using the old calculation method. In contrast, the less efficient reactions (where most of the wells can be expected to contain a single transformed cell) are altered less drastically by the revised calculations: for example, transformation of the *0%-50bp* construct generated BLE-resistant clones at a rate of 1.138 +/- SE 0.485 transformants x 10^{-6} cells in the *HTUB wt* cell line using the new calculation method, as compared to 0.912 +/- SE 0.175 transformants x 10^{-6} cells using the old method.

Despite these differences in transformation efficiency, when the values are plotted on a log-linear graph, the shape looks similar to the graph plotted from the original calculations. There is a steep drop-off in transformation efficiency below a substrate length of 50 bp for all cell-lines, as seen before. In the revised graph, however, the values plateau at 150 bp, instead of 200 bp as with the old calculations. At intermediate substrate lengths the transformation efficiencies derived from the *HTUB MSH2-/-* cell line are perhaps slightly lower than those from the *HTUB wt* and *HTUB MSH2+/-* cell lines, but the overlapping standard error bars show that this is not statistically significant.

It is worth noting that the revised calculations, by increasing the estimates of transformation efficiencies of the more efficient reactions, can lead to more dramatic reductions in recombination efficiency when comparing shorter substrates to *0%-200bp*. For example (see figures 3.6 and 3.9), recombination using the *0%-25bp* substrate is 295-fold less efficient than using *0%-200bp* by the old calculations, and 637.9-fold less efficient by the new method.

3.2.2.2 5% divergence

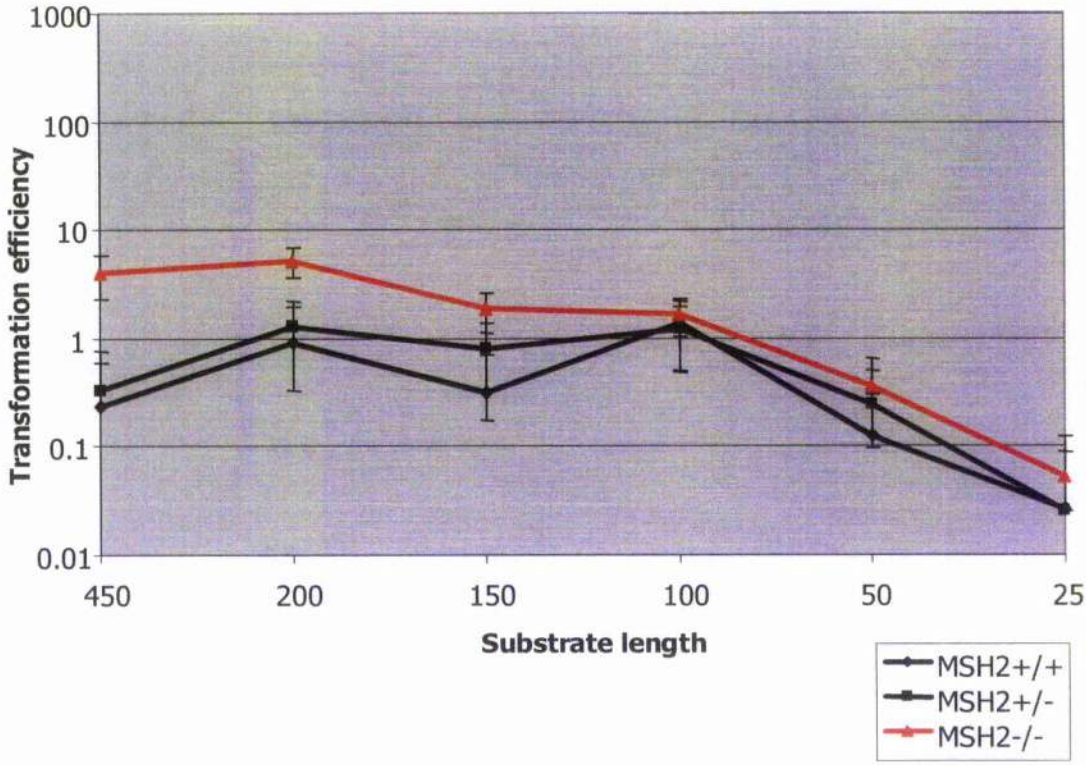
Transformation efficiencies were calculated using data obtained by transformation of constructs derived from the *pHYG05* construct, using the binomial distribution as described above (shown in graphical form in figure 3.10). The same conclusions can be drawn as from the original calculations: firstly, that there is a considerable reduction in recombination efficiency compared to substrates perfectly matched to the genomic target sequence; and secondly, that *MSH2* mutation influences recombination using longer substrates.

The reduction in recombination efficiency compared to the perfectly-matched substrates is more pronounced when the new calculation method is used. There is now a 23-fold reduction in transformation efficiency (a reduction from 22.205 to 0.914 x transformants 10^{-6} cells), as compared to 13-fold when the old method is used. This is because, as discussed above, the less efficient reactions using the 5% substrates are less affected by the statistical adjustments taking into account the possibility of multiple transformants in each well.

The *HTUB MSH2-/-* cell line still displays a different pattern of transformation efficiency to the *HTUB wt* cell line. The *HTUB MSH2-/-* cell line undergoes transformation at a higher rate at all substrate lengths, but this is more pronounced when longer substrates are used. A 5.726 and 5.911-fold increase in transformation efficiency when *HTUB MSH2-/-* is compared to *HTUB wt* when *5%-200bp* and *5%-150bp* are used, is reduced to 1.167-fold, 2.886-fold and 1.986-fold reductions for *5%-100bp*, *5%-50bp* and *5%-25bp*, respectively. Therefore, we can still conclude that *MSH2* mutation allows the recombination machinery to act with greater efficiency for substrates longer than 100 bp in length. A reduction in transformation efficiency when substrate lengths are reduced below 100 bp is seen for all cell lines, as with the original calculation method.

3.2.2.3 11% divergence

Transformation efficiencies were calculated using data obtained by transformation of constructs derived from the *pHYG11* construct, using the binomial distribution as described above (shown in graphical form in figure 3.11). The revised calculations lead to higher estimations of transformation efficiency than the original method for the *11%-50bp* and *11%-25bp* constructs in the *HTUB-MSH2-/-* cell line (0.28 compared to 0.013 transformants x 10^{-6} cells, and 0.269 compared to 0.031 transformants x 10^{-6} cells,

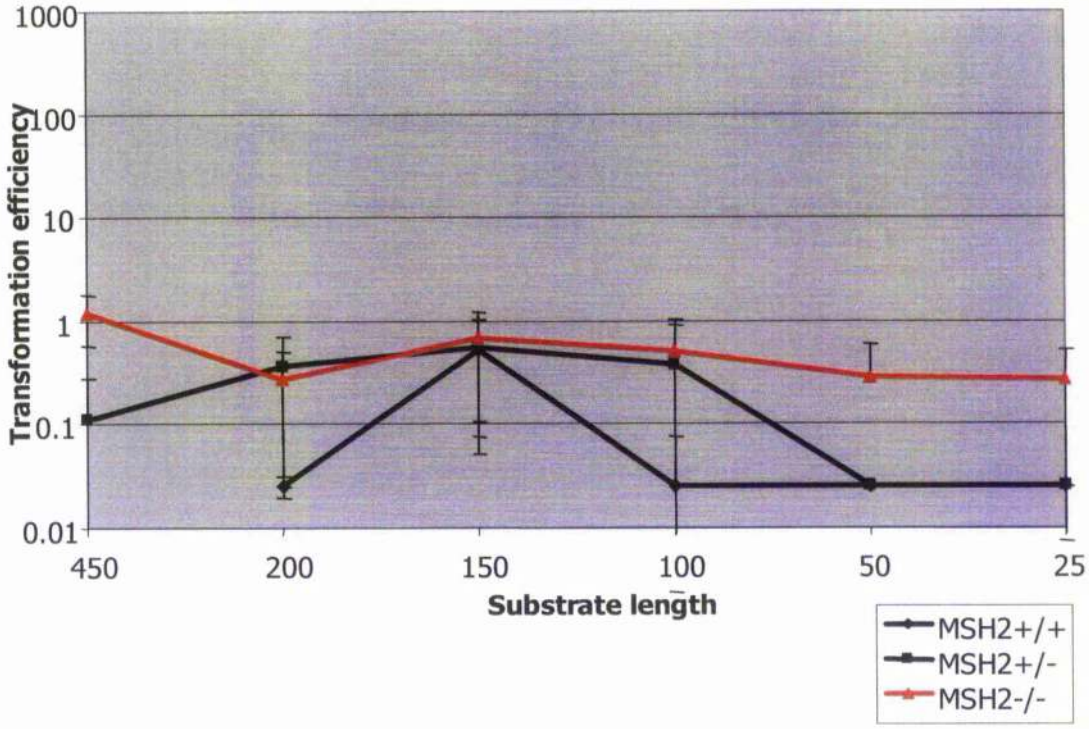


b)

	MSH2+/+		MSH2-/-		
Length (bp)	Transformants/10^6 cells	Reduction	Transformants/10^6 cells	Reduction	MMR-/MMR+
200	0.914		5.232		5.726
150	0.313	2.916	1.852	2.825	5.911
100	1.398	0.653	1.632	3.207	1.167
50	0.127	7.167	0.368	14.220	2.886
25	0.026	35.003	0.052	100.948	1.986

	MSH2+/-		MSH2-/-		
Length (bp)	Transformants/10^6 cells	Reduction	Transformants/10^6 cells	Reduction	MMR-/MMR+
200	1.290		5.232		4.055
150	0.787	1.638	1.852	2.825	2.352
100	1.239	1.042	1.632	3.207	1.317
50	0.245	5.268	0.368	14.220	1.502
25	0.017	75.889	0.052	100.948	3.049

Figure 3.10. Recombination mediated by substrates 5% diverged from the target sequence; transformation efficiencies calculated using the binomial distribution. a) a log-linear plot of transformation efficiency (in transformations x 10^{-6} cells) vs substrate length (in bp). b) mean transformation efficiencies of wild-type (MSH2+/+; top), MSH2+/- (bottom) and MSH2-/- (both) for each substrate length. The column marked Reduction shows the fold reduction relative to substrates with 200 bp



b)

	MSH2+/+		MSH2-/-		
Length (bp)	Transformants/10^6 cells	Reduction	Transformants/10^6 cells	Reduction	MMR-/MMR+
200	0.025		0.266		10.649
150	0.517	0.048	0.676	0.394	1.307
100	0.026	0.979	0.511	0.521	20.023
50	0.017	1.471	0.280	0.950	16.478
25	0.017	1.471	0.269	0.991	15.805

	MSH2+/-		MSH2-/-		
Length (bp)	Transformants/10^6 cells	Reduction	Transformants/10^6 cells	Reduction	MMR-/MMR+
200	0.365		0.266		0.730
150	0.545	0.669	0.676	0.394	1.240
100	0.382	0.955	0.511	0.521	1.340
50	0.017	21.444	0.280	0.950	16.478
25	0.017	21.444	0.269	0.991	15.805

Figure 3.11. Recombination mediated by substrates 11% diverged from the target sequence; transformation efficiencies calculated using the binomial distribution. a) a log-linear plot of transformation efficiency (in transformations x 10^{-6} cells) vs substrate length (in bp). b) mean transformation efficiencies of wild-type (MSH2+/+; top), MSH2+/- (bottom) and MSH2-/- (both) for each substrate length. The column marked Reduction shows the fold reduction relative to substrates with 200 bp

respectively. This means that although a reduction in transformation efficiency with substrate length can still be observed in the MMR-proficient cell lines, this effect is no longer really seen in the *HTUB MSH2-/-* cell line: in this case, recombination does not seem to be affected by substrate length. However, analysis of this is very difficult due to the low numbers of transformants recovered.

3.2.3 Characterisation of integration by hygromycin resistance

A difficulty with the above analysis is that it is assumed that all integrations occur by HR into the *HTUB* target. It is possible, however, that other forms of recombination (see section 1.4) could invalidate these results. For example, it is known in *T. brucei* that microhomology-mediated recombination can occur into unexpected locations (Conway *et al.*, 2002b). To examine this, the antibiotic resistance of the transformants was examined. Recombination by HR of the constructs from this assay into the *HTUB* locus could lead to loss of hygromycin resistance due to disruption of the *HYG* ORF, whereas aberrant recombination into other loci should allow the cells to survive in the presence of this antibiotic. Therefore, cloned transformants were tested for hygromycin resistance by passage into 1.5 ml of HMI-9 medium containing 5 $\mu g.ml^{-1}$ of hygromycin, and assessing growth of the trypanosomes after 3 days. The proportions of hygromycin resistant (Hyg^R) and hygromycin sensitive (Hyg^S) transformants from each class of transformation are recorded in figures 3.9, 3.10 and 3.11. In many cases, a proportion of transformants retain hygromycin resistance. In the most efficient transformation reactions (with >100 bp of perfect homology), this was between 4.8-21.8% of transformants. No clear alteration in the proportions of Hyg^R transformants was apparent in transformations from the whole range of substrate lengths and sequence divergence, and in all three cell lines. Although some classes of transformant were dominated by Hyg^R cells, the small sample sizes in many cases make it difficult to analyse this in more detail.

To attempt to examine this further, the frequency of recombination leading to Hyg^R transformants was calculated from the total recombination frequency (calculated using the original calculation method) and the proportion of Hyg^R clones, and is shown in graphical form in figures 3.12, 3.13 and 3.14. In all cell lines assayed, the different sized PCR products derived from *pHYGwt* and *pHYG05* (with the exception of the longer 5% substrates as mentioned below) displayed a log-linear reduction in recombination

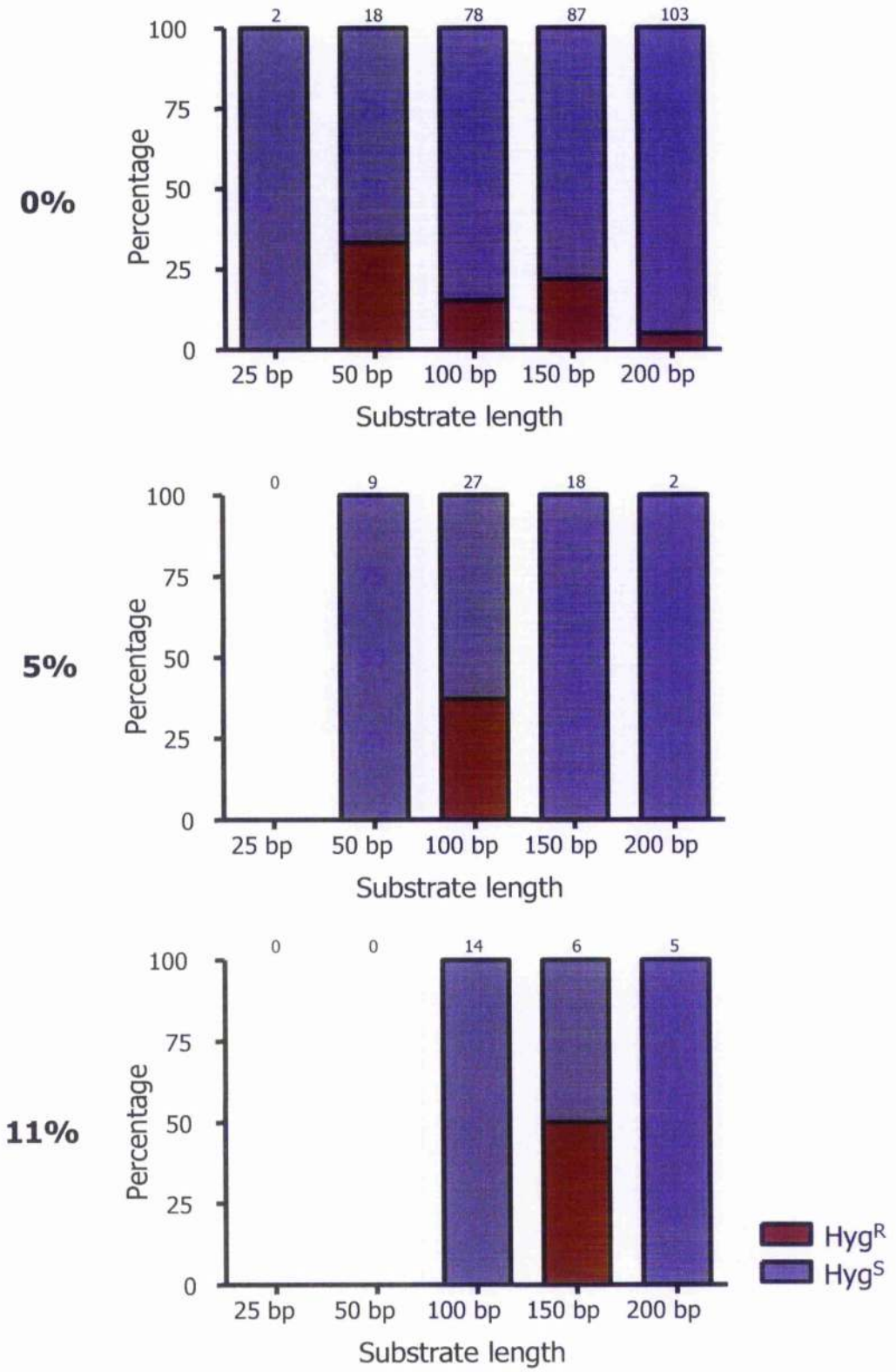


Figure 3.12. Bar charts showing hygromycin resistance. The proportion of hygromycin resistant (HygR; maroon) and sensitive (HygS; light blue) cloned transformants recovered from recombination of constructs of varying lengths (shown along the x-axis) with 0%, 5% and 11% divergence from the genomic *HYG* sequence into *HTUB wt* cells.

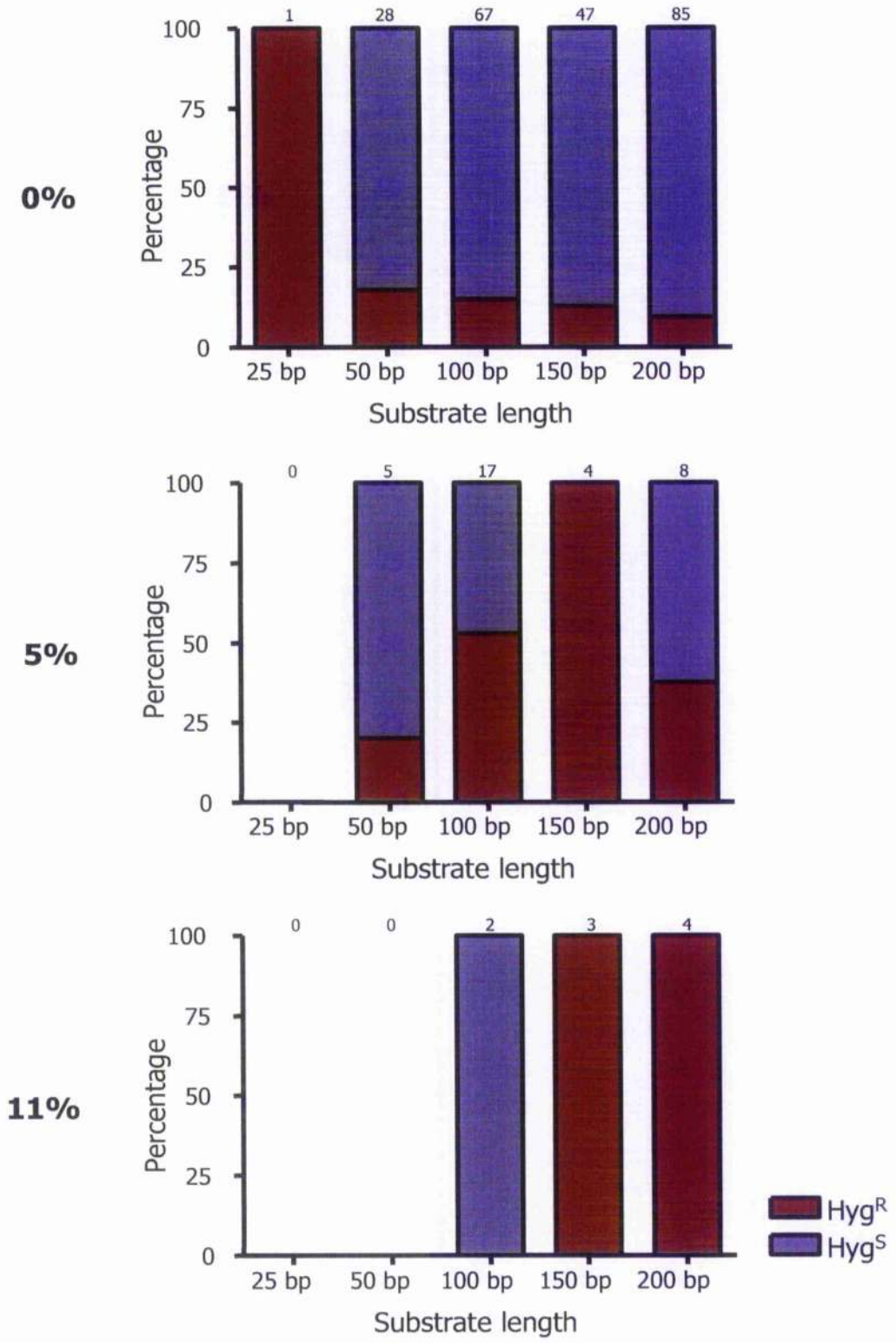


Figure 3.13. Bar charts showing hygromycin resistance. The proportion of hygromycin resistant (HygR; maroon) and sensitive (HygS; light blue) cloned transformants recovered from recombination of constructs of varying lengths (shown along the x-axis) with 0%, 5% and 11% divergence from the genomic *HYG* sequence into *HTUB MSH2+/-* cells.

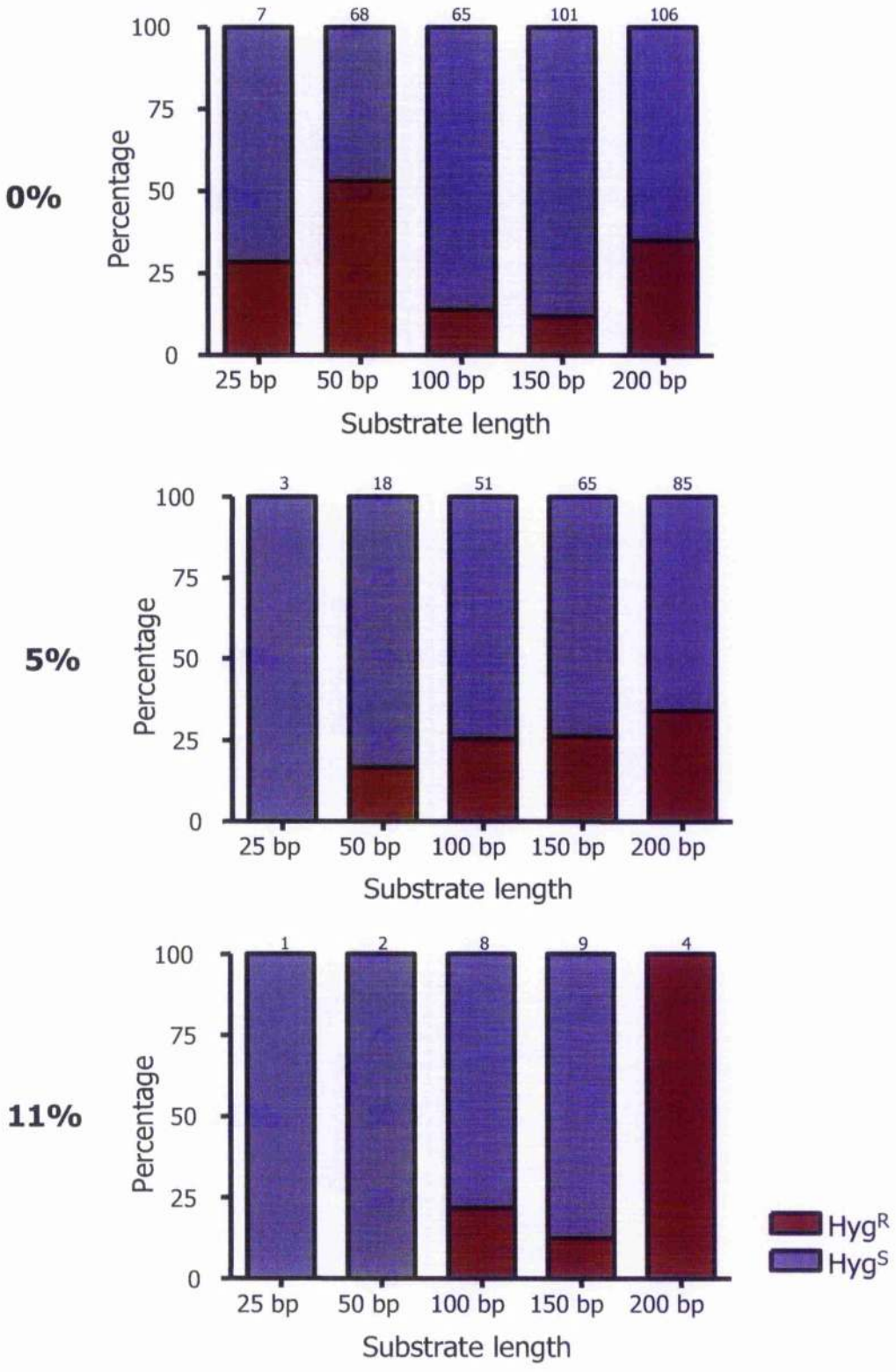


Figure 3.14. Bar charts showing hygromycin resistance. The proportion of hygromycin resistant (HygR; maroon) and sensitive (HygS; light blue) cloned transformants recovered from recombination of constructs of varying lengths (shown along the x-axis) with 0%, 5% and 11% divergence from the genomic *HYG* sequence into *HTUB MSH2-/-* cells.

efficiency, ranging from ~1 transformant x 10^{-6} cells for 200 bp substrates, down to undetectable transformation for the shortest substrate. These reactions do not seem to be dependent on MMR, as there did not appear to be any significant difference between the cell lines assayed, and the 0% and 5% constructs gave very similar graphs (see figures 3.12 and 3.13). An exception to this are the transformations of constructs containing 5% divergence into the *HTUB wt* cell line; here, Hyg^R transformants were only identified in the 100 bp reactions. However, as the *HTUB MSH2+/-* cell line gave very similar results to the *HTUB MSH2-/-* cell line, this can probably be dismissed as an experimental discrepancy due to the low numbers of cloned transformants available for analysis in many cases. Transformation of constructs derived from *pHYG11* (see figure 3.14) gave extremely few transformants at substrate lengths shorter than 150 bp, but again the small sample sizes means this could reflect random fluctuation.

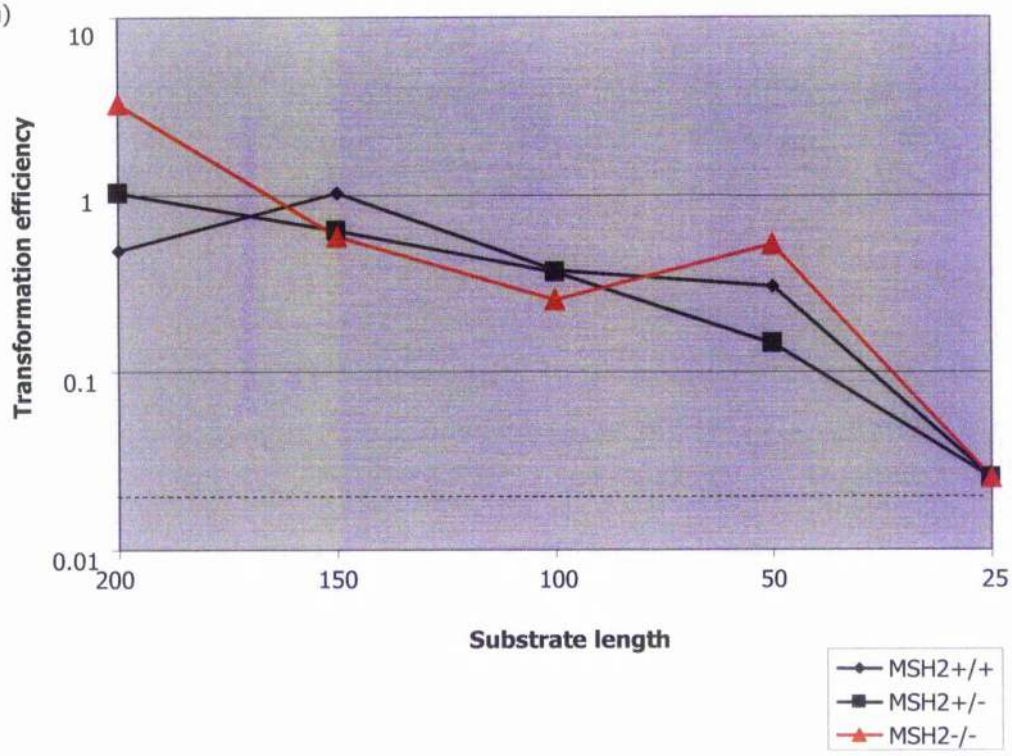
The frequency of recombination leading to Hyg^R transformants was calculated from the total recombination frequency (calculated using the new, binomial distribution calculation method) and the proportion of Hyg^R clones, and is shown in graphical form in figures 3.18 - 3.20. As with the calculations for total transformation efficiency, this gives very similar results to the original calculations.

In general, it can be concluded that a proportion of Hyg^R transformants are generated by recombination of all constructs, regardless of length or sequence divergence.

3.2.4 Southern analysis of the integrations

A number of Hyg^S and Hyg^R transformants were analysed by Southern blot, using the strategy shown in figure 3.15, to look at the integration events in more detail and to attempt to define how Hyg^R transformants arose. Genomic DNA from 24 Hyg^S and 24 Hyg^R

a)



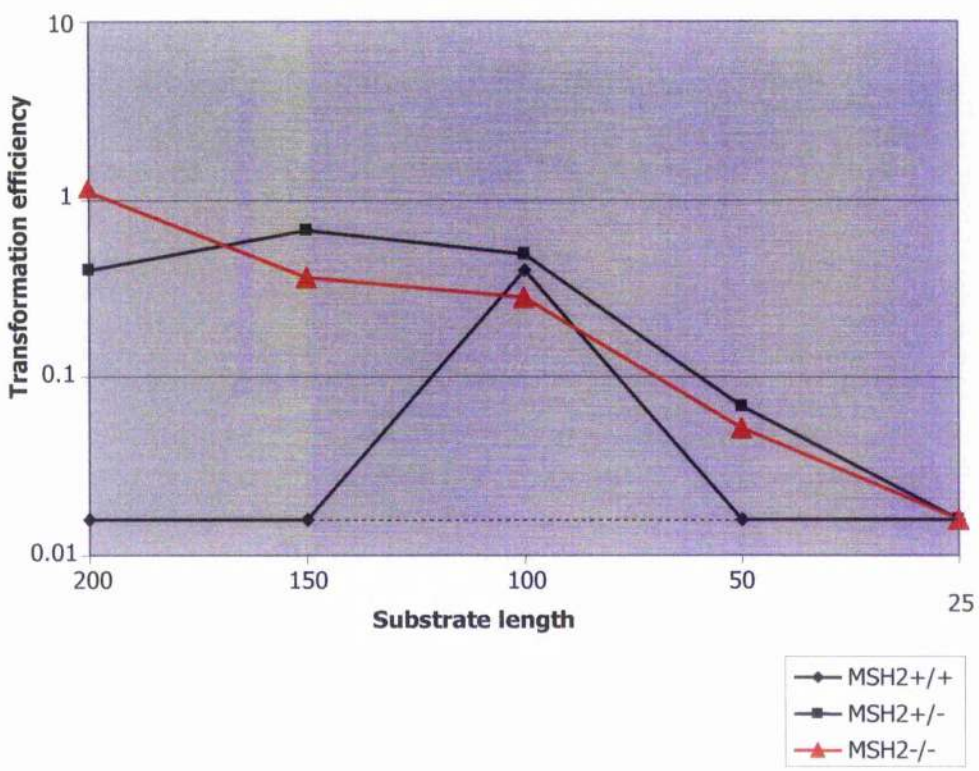
b)

	MSH2+/+		MSH2-/-		
Length (bp)	Transformants/10^6 cells	Reduction	Transformants/10^6 cells	Reduction	MMR-/MMR+
200	0.477		3.19		6.688
150	1.019	0.468	0.572	5.577	0.561
100	0.372	1.282	0.257	12.412	0.691
50	0.306	1.559	0.516	6.182	1.686
25	0.025	19.080	0.025	127.600	1.000

	MSH2+/-		MSH2-/-		
Length (bp)	Transformants/10^6 cells	Reduction	Transformants/10^6 cells	Reduction	MMR-/MMR+
200	1.019		3.19		3.131
150	0.622	1.638	0.572	5.577	0.920
100	0.365	2.792	0.257	12.412	0.704
50	0.145	7.028	0.516	6.182	3.559
25	0.025	40.760	0.025	127.600	1.000

Figure 3.15. Frequency of hygromycin-resistant transformants resulting from recombination mediated by substrates perfectly matched to the target sequence (original calculations). a) a log-linear plot of transformation efficiency (in transformations x 10-6 cells) versus integration flank (in bp). b) mean transformation efficiencies of wild-type (MSH2+/+; top), *MSH2+/-* (bottom) and *MSH2-/-* (both) for each substrate length. The column marked Reduction shows the fold reduction relative to substrates with 200 bp targeting flanks, whereas the MMR-/MMR+ column shows the fold increase in transformation efficiency of *MSH2-/-* cells relative to *MSH2+/+* or *MSH2+/-* cells.

a)



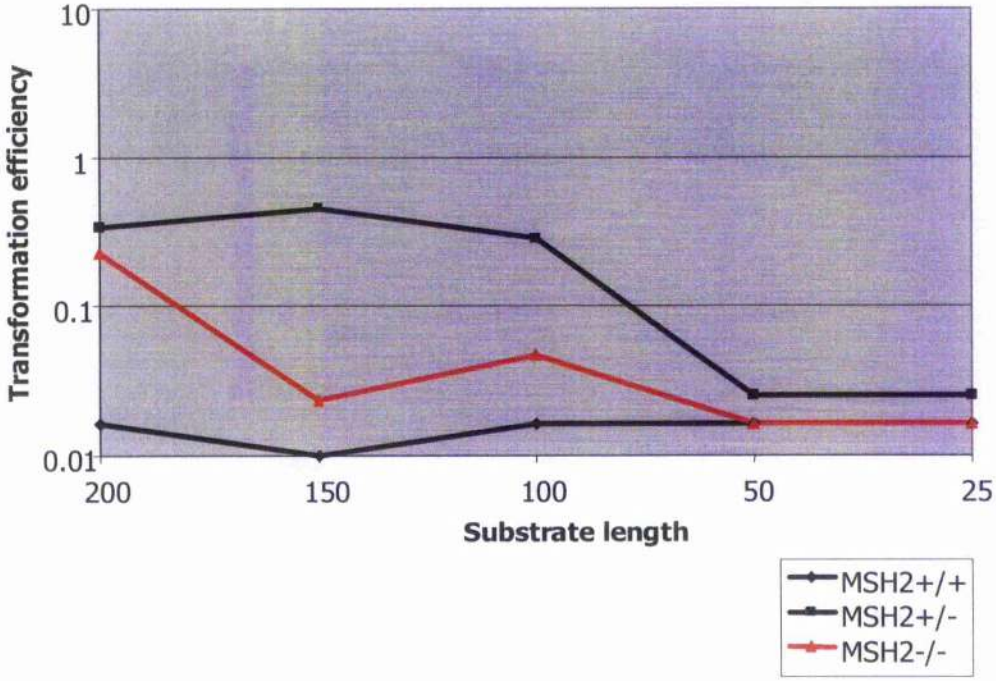
b)

	MSH2+/+		MSH2-/-		
Length (bp)	**Transformants/10^6 cells**	**Reduction**	**Transformants/10^6 cells**	**Reduction**	**MMR-/MMR+**
200	0.016		1.098		68.625
150	0.016	1.000	0.366	3.000	22.875
100	0.402	0.040	0.28	3.921	0.697
50	0.016	1.000	0.052	21.115	3.250
25	0.016	1.000	0.016	68.625	1.000

	MSH2+/-		MSH2-/-		
Length (bp)	**Transformants/10^6 cells**	**Reduction**	**Transformants/10^6 cells**	**Reduction**	**MMR-/MMR+**
200	0.406		1.098		2.704
150	0.666	1.640	0.366	3.000	0.550
100	0.496	0.819	0.28	3.921	0.565
50	0.069	0.170	0.052	21.115	0.754
25	0.016	0.039	0.016	68.625	1.000

Figure 3.16. Frequency of hygromycin-resistant transformants resulting from recombination mediated by substrates 5% divergent from the target sequence (original calculations). a) a log-linear plot of transformation efficiency (in transformations x 10-6 cells) versus integration flank (in bp). b) mean transformation efficiencies of wild-type (MSH2+/+; top), *MSH2+/-* (bottom) and *MSH2-/-* (both) for each substrate length. The column marked Reduction shows the fold reduction relative to substrates with 200 bp targeting flanks, whereas the MMR-/MMR+ column shows the fold increase in transformation efficiency of *MSH2-/-* cells relative to *MSH2+/+* or *MSH2+/-* cells.

a)



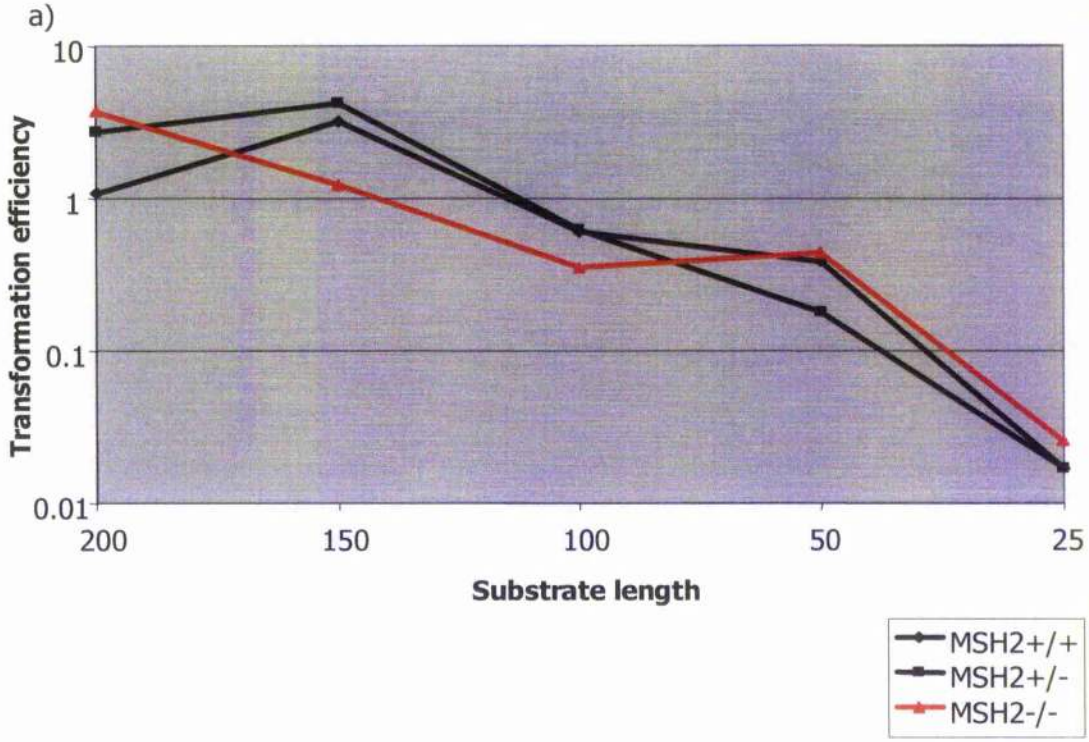
b)

	MSH2+/+		MSH2-/-		
Length (bp)	Transformants/10^6 cells	Reduction	Transformants/10^6 cells	Reduction	MMR-/MMR+
200	0.016		0.222		13.875
150	0.010	1.600	0.023	9.652	2.300
100	0.016	1.000	0.047	4.723	2.938
50	0.016	1.000	0.016	13.875	1.000
25	0.016	1.000	0.016	13.875	1.000

	MSH2+/-		MSH2-/-		
Length (bp)	Transformants/10^6 cells	Reduction	Transformants/10^6 cells	Reduction	MMR-/MMR+
200	0.333		0.222		0.667
150	0.458	1.600	0.023	9.652	0.050
100	0.283	0.057	0.047	4.723	0.166
50	0.025	0.640	0.016	13.875	0.640
25	0.025	0.640	0.016	13.875	0.640

Figure 3.17. Frequency of hygromycin-resistant transformants resulting from recombination mediated by substrates 11% divergent from the target sequence (original calculations). a) a log-linear plot of transformation efficiency (in transformations x 10-6 cells) versus integration flank (in bp). b) mean transformation efficiencies of wild-type (MSH2+/+; top), *MSH2+/-* (bottom) and *MSH2-/-* (both) for each substrate length. The column marked Reduction shows the fold reduction relative to substrates with 200 bp targeting flanks, whereas the MMR-/MMR+ column shows the fold increase in transformation efficiency of *MSH2-/-* cells relative to *MSH2+/+* or *MSH2+/-* cells.

a)



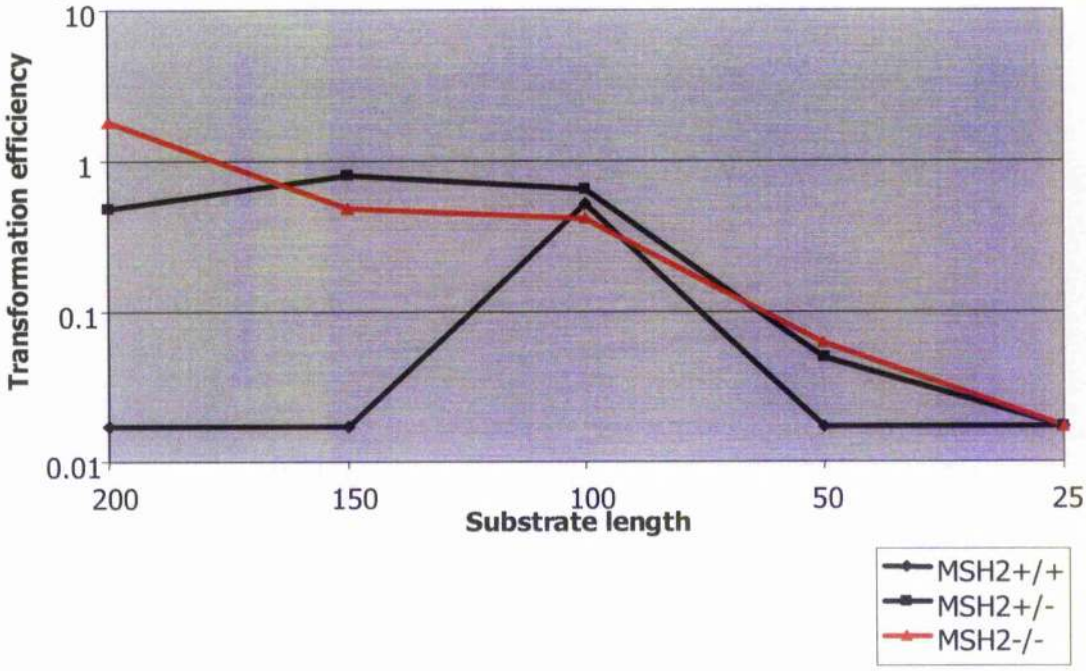
b)

	MSH2+/+		MSH2-/-		
Length (bp)	Transformants/10^6 cells	Reduction	Transformants/10^6 cells	Reduction	MMR-/MMR+
200	1.077		3.734		3.467
150	3.218	0.335	1.221	3.059	0.379
100	0.603	1.785	0.348	10.734	0.577
50	0.379	2.841	0.434	8.598	1.146
25	0.017	63.350	0.026	145.314	1.512

	MSH2+/-		MSH2-/-		
Length (bp)	Transformants/10^6 cells	Reduction	Transformants/10^6 cells	Reduction	MMR-/MMR+
200	2.765		3.734		1.351
150	4.249	0.651	1.221	3.059	0.287
100	0.625	4.426	0.348	10.734	0.557
50	0.178	15.504	0.434	8.598	2.435
25	0.017	162.391	0.026	145.314	1.509

Figure 3.18. Frequency of hygromycin-resistant transformants resulting from recombination mediated by substrates perfectly matched to the target sequence (original calculations). a) a log-linear plot of transformation efficiency (in transformations x 10-6 cells) versus integration flank (in bp). b) mean transformation efficiencies of wild-type (MSH2+/+; top), *MSH2+/-* (bottom) and *MSH2-/-* (both) for each substrate length. The column marked Reduction shows the fold reduction relative to substrates with 200 bp targeting flanks, whereas the MMR-/MMR+ column shows the fold increase in transformation efficiency of *MSH2-/-* cells relative to *MSH2+/+* or *MSH2+/-* cells.

a)



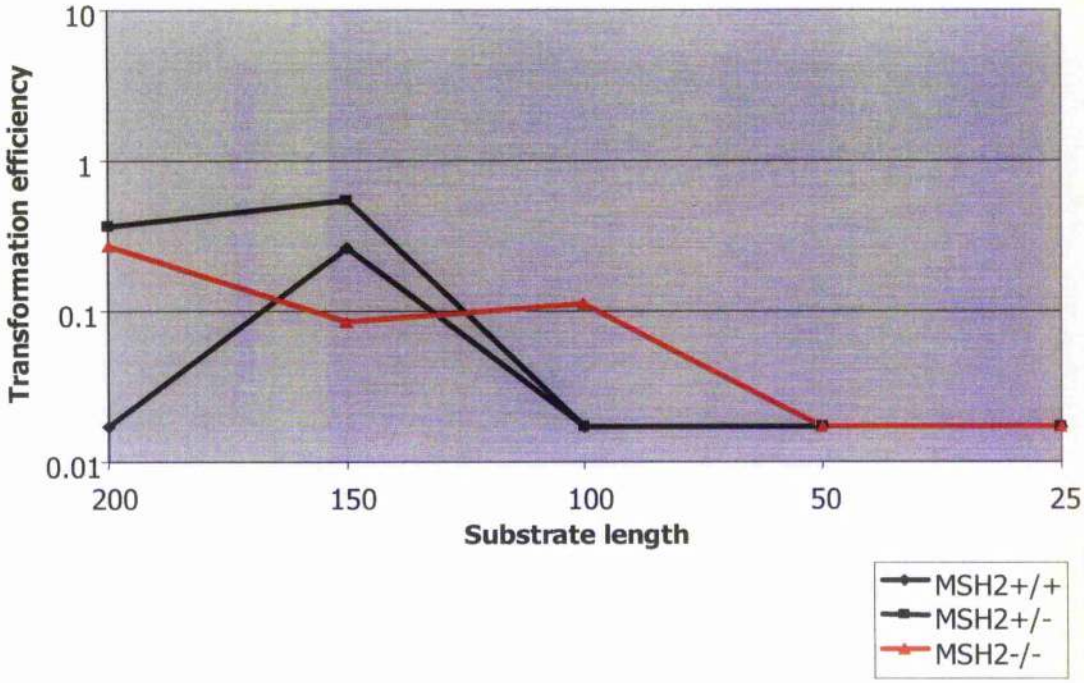
b)

	MSH2+/+		MSH2-/-		
Length (bp)	Transformants/10^6 cells	Reduction	Transformants/10^6 cells	Reduction	MMR-/MMR+
200	0.017		1.785		105.009
150	0.017	1.000	0.484	3.686	28.492
100	0.518	0.033	0.416	4.292	0.803
50	0.017	1.000	0.061	29.112	3.607
25	0.017	1.000	0.017	105.009	1.000

	MSH2+/-		MSH2-/-		
Length (bp)	Transformants/10^6 cells	Reduction	Transformants/10^6 cells	Reduction	MMR-/MMR+
200	0.484		1.785		3.690
150	0.787	0.614	0.484	3.686	0.615
100	0.656	0.738	0.416	4.292	0.634
50	0.049	9.877	0.061	29.112	1.252
25	0.017	28.459	0.017	105.009	1.000

Figure 3.19. Frequency of hygromycin-resistant transformants resulting from recombination mediated by substrates 5% divergent from the target sequence (revised calculations). a) a log-linear plot of transformation efficiency (in transformations x 10-6 cells) versus integration flank (in bp). b) mean transformation efficiencies of wild-type (MSH2+/+; top), *MSH2+/-* (bottom) and *MSH2-/-* (both) for each substrate length. The column marked Reduction shows the fold reduction relative to substrates with 200 bp targeting flanks, whereas the MMR-/MMR+ column shows the fold increase in transformation efficiency of *MSH2-/-* cells relative to *MSH2+/+* or *MSH2+/-* cells.

a)



b)

	MSH2+/+		MSH2-/-		
Length (bp)	Transformants/10^6 cells	Reduction	Transformants/10^6 cells	Reduction	MMR-/MMR+
200	0.017		0.266		15.660
150	0.259	0.066	0.085	3.150	0.327
100	0.017	1.000	0.112	2.380	6.581
50	0.017	1.000	0.017	15.660	1.000
25	0.017	1.000	0.017	15.660	1.000

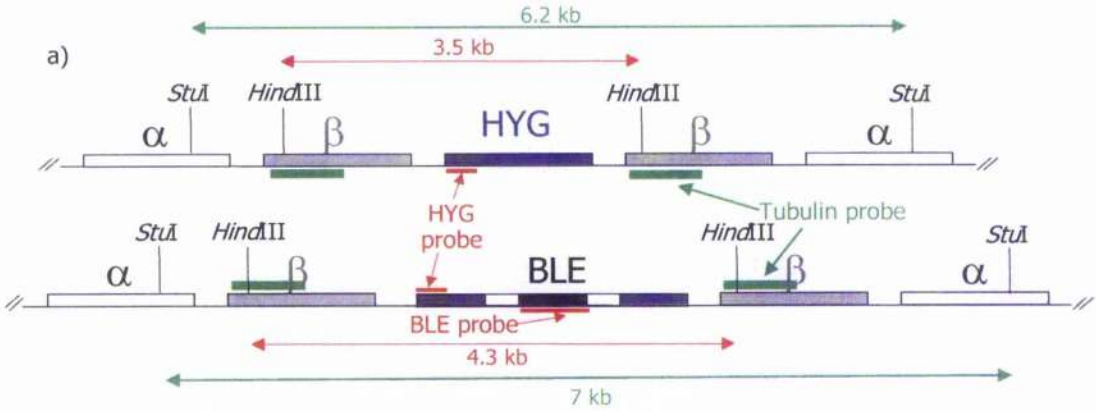
	MSH2+/-		MSH2-/-		
Length (bp)	Transformants/10^6 cells	Reduction	Transformants/10^6 cells	Reduction	MMR-/MMR+
200	0.365		0.266		0.730
150	0.545	0.066	0.085	3.150	0.155
100	0.017	1.000	0.112	2.380	6.581
50	0.017	1.000	0.017	15.660	1.000
25	0.017	1.000	0.017	15.660	1.000

Figure 3.20. Frequency of hygromycin-resistant transformants resulting from recombination mediated by substrates 11% divergent from the target sequence (revised calculations). a) a log-linear plot of transformation efficiency (in transformations x 10-6 cells) versus integration flank (in bp). b) mean transformation efficiencies of wild-type (MSH2+/+; top), *MSH2+/-* (bottom) and *MSH2-/-* (both) for each substrate length. The column marked Reduction shows the fold reduction relative to substrates with 200 bp targeting flanks, whereas the MMR-/MMR+ column shows the fold increase in transformation efficiency of *MSH2-/-* cells relative to *MSH2+/+* or *MSH2+/-* cells.

clones from a representative cross-section of transformations into either *HTUB wt* or *HTUB MSH2-/-* cells, as well as the untransformed *HTUB wt* or *HTUB MSH2-/-* strains, was digested with *Hin*dIII, separated by agarose gel electrophoresis and Southern blotted. The entire, 384 bp *BLE* ORF was amplified from *pRM481* construct DNA using primers Ble-PstI and Ble-SphI (see appendix 1), and the PCR product used as a probe for the Southern blots (figure 3.16). A *Hin*dIII site can be found in every copy of the *β tubulin* gene in the repetitive tubulin array, so the expected restriction fragment generated by correct integration into the *HTUB* locus measures 4.3 kb, containing one copy of *β tubulin* with its processing flanks, the disrupted *HYG*, and *BLE* with its processing flanks. The *BLE* probe annealed to a single fragment of this size in all the Hyg^S clones studied, from both *HTUB wt* and *HTUB MSH2-/-* cell lines. This fragment was also seen in almost all Hyg^R clones, implying correct integration of the PCR products into the *HYG* locus in these clones, despite retention of hygromycin resistance. In four of the Hyg^R clones, all of which were derived from transformation into *HTUB MSH2-/-* cells (using PCR products *0%-50bp*, *11%-100bp* (twice) and *11%-200bp*), the 4.3 kb restriction fragment was not present, but instead single fragments of different sizes were seen. Most likely, these correspond to aberrant integration of the transformation constructs into other genomic locations, presumably employing short regions of homology to the *HYG* integration flanks (Conway *et al.*, 2002b).

The Southern blots were then stripped of radioactive material to allow further analysis. A 228 bp section of the *HYG* ORF (upstream of the region of the gene that was the target used for integration in these experiments) was amplified from *HTUB wt* genomic DNA using primers Hyg5' and Hygprobe-3' (see appendix 1) and the PCR product was used to re-probe the Southern blots. This probe is expected to anneal to the same 4.3 kb fragment described in the previous section if correct integration into the *HYG* locus has taken place. As for *BLE*, this single band was again seen in all of the HYG^S clones, apart from in two cases. These cases resulted from transformation of PCR products *0%-50bp* and *5%-150bp* (lanes 10 and 11, *HTUB wt* Hyg^S, figure 3.17) where an additional restriction fragment of ~3 kb was seen, perhaps corresponding to the undisturbed *HYG* locus (as described below). These isolated discrepancies between the *HYG* and *BLE* probes was not analysed in more detail. The expected restriction fragment generated by *Hin*dIII digestion of the wild-type, undisturbed *HYG* locus is 3.5 kb, corresponding to one copy of *β tubulin* with its processing flanks, and *HYG* (see figure 3.15). In the four Hyg^R clones with aberrant integrations (lanes 3, 12, 13 and 23, *HTUB MSH2-/-* Hyg^R, figure 3.17), a single 3.5 kb restriction fragment corresponding to the intact *HTUB* locus was seen. However, in the

a)



b)

	HTUB MSH2 wt		HTUB MSH2-/-	
	HygS	**HygR**	**HygS**	**HygR**
1	*0%-50bp*	*0%-100bp*	*0%-50bp*	*0%-50bp*
2	*0%-50bp*	*0%-100bp*	*0%-50bp*	*0%-50bp*
3	*0%-100bp*	*5%-100bp*	*0%-100bp*	*0%-50bp*
4	*0%-100bp*	*5%-100bp*	*0%-100bp*	*0%-100bp*
5	*0%-200bp*	*0%-100bp*	*0%-150bp*	*0%-100bp*
6	*0%-200bp*	*0%-100bp*	*0%-150bp*	*0%-150bp*
7	*0%-200bp*	*0%-150bp*	*0%-200bp*	*0%-150bp*
8	*5%-50bp*	*0%-150bp*	*0%-200bp*	*0%-200bp*
9	*5%-100bp*	*0%-150bp*	*5%-50bp*	*0%-200bp*
10	*5%-150bp*	*0%-150bp*	*5%-50bp*	*5%150bp*
11	*0%-50bp*	*0%-150bp*	*5%-100bp*	*5%-200bp*
12	*0%-50bp*	*0%-150bp*	*5%-100bp*	*11%-100bp*
13	*0%-100bp*	*0%-200bp*	*5%-150bp*	*11%-100bp*
14	*0%-100bp*	*0%-200bp*	*5%-200bp*	*11%-200bp*
15	*0%-150bp*	*5%-100bp*	*5%-200bp*	*0%-100bp*
16	*0%-150bp*	*5%-100bp*	*0%-100bp*	*0%-150bp*
17	*0%-150bp*	*5%-100bp*	*0%-100bp*	*0%-150bp*
18	*0%-150bp*	*5%-100bp*	*0%-150bp*	*5%-100bp*
19	*0%-200bp*	*5%-100bp*	*0%-150bp*	*5%-100bp*
20	*0%-200bp*	*5%-100bp*	*5%-100bp*	*5%-200bp*
21	*0%-200bp*	*5%-200bp*	*5%-100bp*	*5%-200bp*
22	*0%-200bp*	*11%-150bp*	*5%-200bp*	*5%-200bp*
23	*5%-100bp*	*11%-150bp*	*5%-200bp*	*11%-100bp*
24	*11%-150bp*	*11%-150bp*	*5%-200bp*	*11%-150bp*
HygR 1		*0%-200bp*		
HygR 2		*5%-100bp*		
HygR 3		*0%-100bp*		
HygR 4		*0%-150bp*		
HygR 5		*0%-100bp*		
HygS 1	*0%-200bp*			
HygS 2	*11%-100bp*			
HygS 3	*5%-150bp*			

Figure 3.21. Southern analysis of cloned transformants. a) a shematic representation of the strategy used to analyse cloned transformants. The α and β tubulin, *HYG* and *BLE* ORFs are represented by white, grey, blue and black boxes, respectively, *Hind*III and *Stu*I restriction sites by vertical lines, probe fragments by red lines (*HYG* or *BLE*) or green lines (tubulin) and expected restriction fragment sizes by horizontal arrows of the appropriate colour. b) a summary of the PCR products used to generate the transformants analysed in this section.

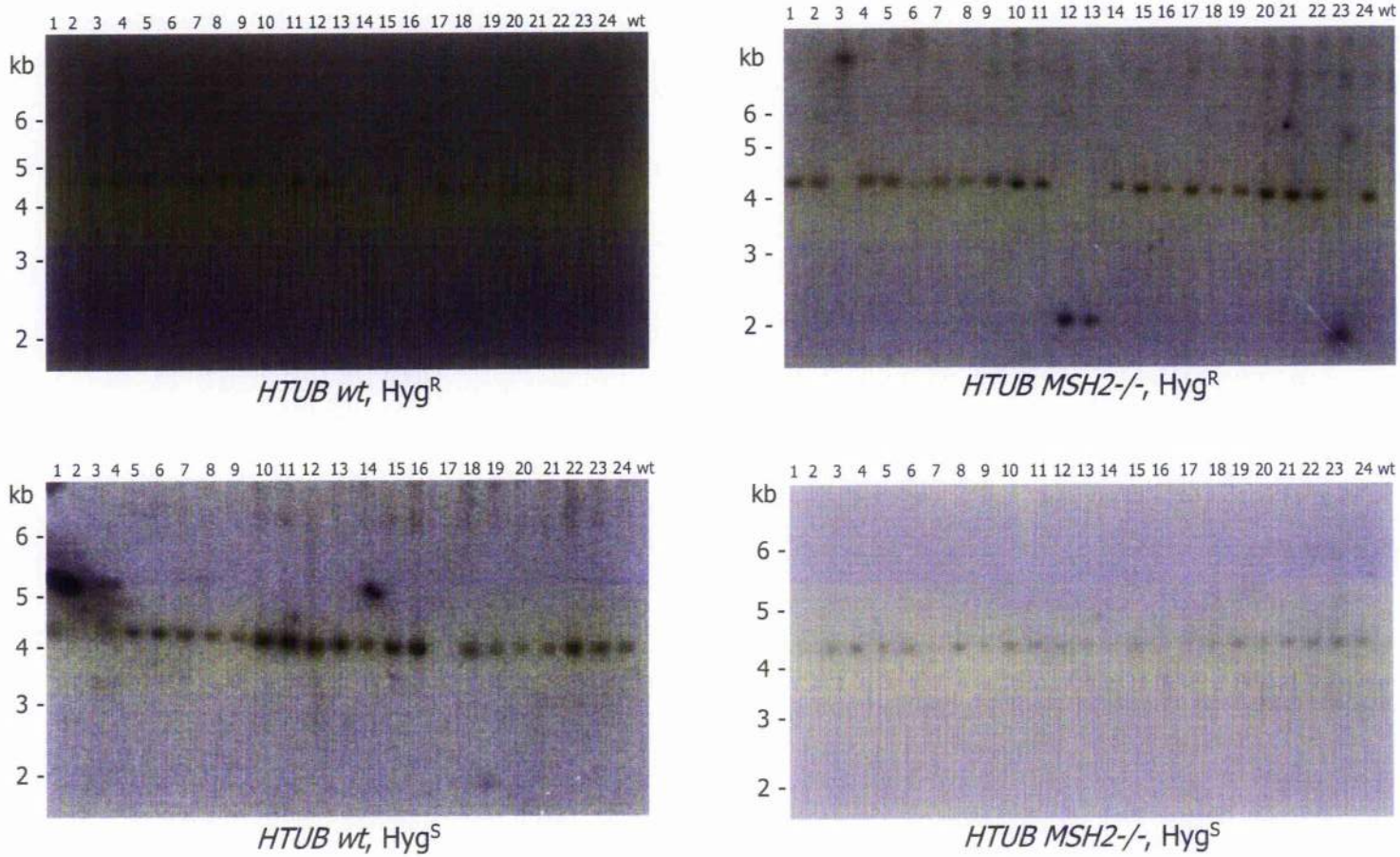


Figure 3.22. Southern analysis of cloned transformants from the recombination efficiency assay. Genomic DNA was prepared from hygromycin resistant (HygR) or sensitive (HygS) clonal cell lines derived from transformation of different constructs into *HTUB wt* or *HTUB MSH2-/-* cell lines (1-24; see part a) and from untransformed *HTUB wt* and *HTUB MSH2-/-* cells. 5 µg of genomic DNA was digested with *Hin*dIII and run on a 0.8% agarose gel. The DNA was Southern blotted and probed with a 386 bp product PCR-amplified from the *BLE* ORF on the *pRM481* plasmid, and washed to 0.2x SSC, 0.1% SDS at 65 °C.

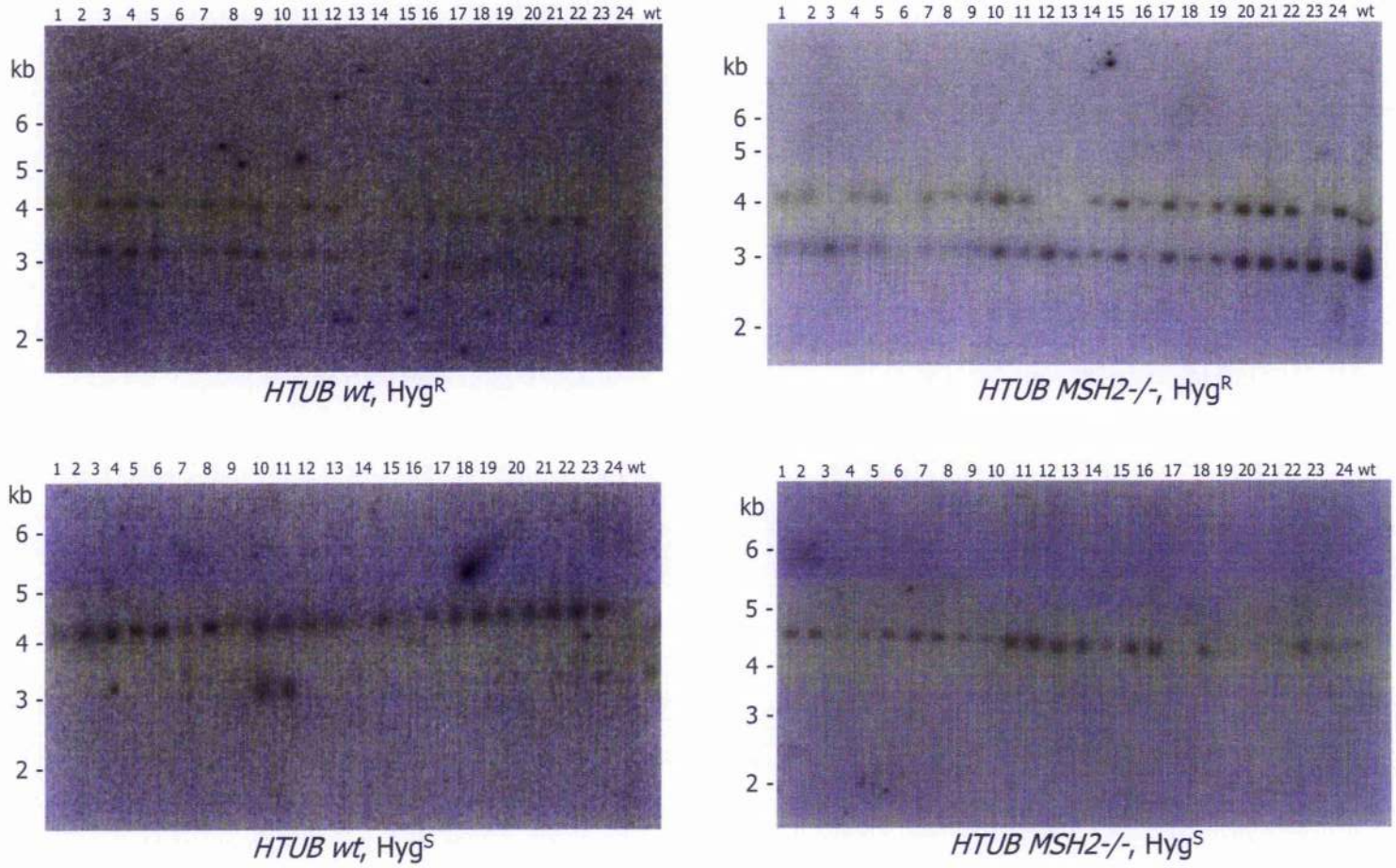


Figure 3.23. Southern analysis of cloned transformants from the recombination efficiency assay. Blots from figure 16 were stripped of radioactivity and reprobed using a 228 bp probe PCR-amplified from the *HYG* ORF of *HTUB wt* genomic DNA, and washed to 0.2x SSC, 0.1% SDS at 65 °C.

rest of the Hyg^R clones, two restriction fragments were seen, one of 3.5 kb and one of 4.3 kb, suggesting duplication of the *HTUB* genomic region in these transformants.

In order to confirm the pattern of integration in the *HTUB* locus observed by Southern blot, a sample of the transformants were analysed by diagnostic PCR, as summarised in figure 3.18 a. For genomic DNA samples from the different classes of transformants (lane 2 *HTUB wt* Hyg^S: Hyg^S; lane 1 *HTUB MSH2-/-* Hyg^R: Hyg^R duplication; lane 12 *HTUB MSH2-/-* Hyg^R: Hyg^R aberrant: see figure 3.18 b), as well as no-DNA and *HTUB wt* controls (figure 3.18 c), three PCR reactions were carried out. Firstly, the *HYG* locus was amplified using primers HYG5' and HYG3' (see appendix 1), which amplified the complete ORF (861 bp). Secondly, linkage of the *BLE* cassette to the tubulin array was assessed by PCR using primers βαtub and Midbleo. Thirdly integrity of the genomic DNA template was assayed by amplification of a 471 bp central section of the RNA polymerase I ORF using primers PolI5' and PolI3' (see appendix 1). In the Hyg^S clone, the *HYG* locus is expected to be disrupted by integration of the *BLE* construct and therefore to yield a PCR product of ~2 kb; the *BLE* ORF is also expected to be linked to the tubulin array yielding a PCR product of 1.3 kb. Hyg^R transformants can be divided into two classes. In the Hyg^R cells where the *HYG* locus is thought to be duplicated (*i.e.*, the majority of clones, yielding two restriction fragments annealed by the *BLE* probe), the above products are expected, as well as an 861 bp PCR product corresponding to an undisturbed *HYG* gene. On the other hand, in the Hyg^R clone where bleomycin resistance has been obtained by putatively aberrant integration into other genomic locations: therefore, amplification of the *HYG* gene should yield an 861 bp product, whereas linkage of *BLE* and tubulin should not be possible. Representative PCR products for each class of transformant are shown in figure 3.18, and in all cases products of the predicted sizes were amplified, confirming the interpretations made from the Southern analysis.

The observed duplication of at least a section of chromosome I could be caused by three things. Firstly, in a subset of the parasite population, trisomy of the marked chromosome I, and potentially other chromosomes, could be present, allowing integration of *BLE* into one copy of the *HYG* locus, while the other copy remains intact. Secondly, duplication of the *HTUB* locus could have occurred within the repetitive tubulin array either previous to, or during, recombination. Thirdly, a mechanism similar to break-induced replication (BIR; see section 1.2.2) could have occurred during *BLE* integration, causing duplication of a large section of the chromosome or even the complete chromosome. To differentiate between these mechanisms, a representative set of transformants were subjected to further analysis.

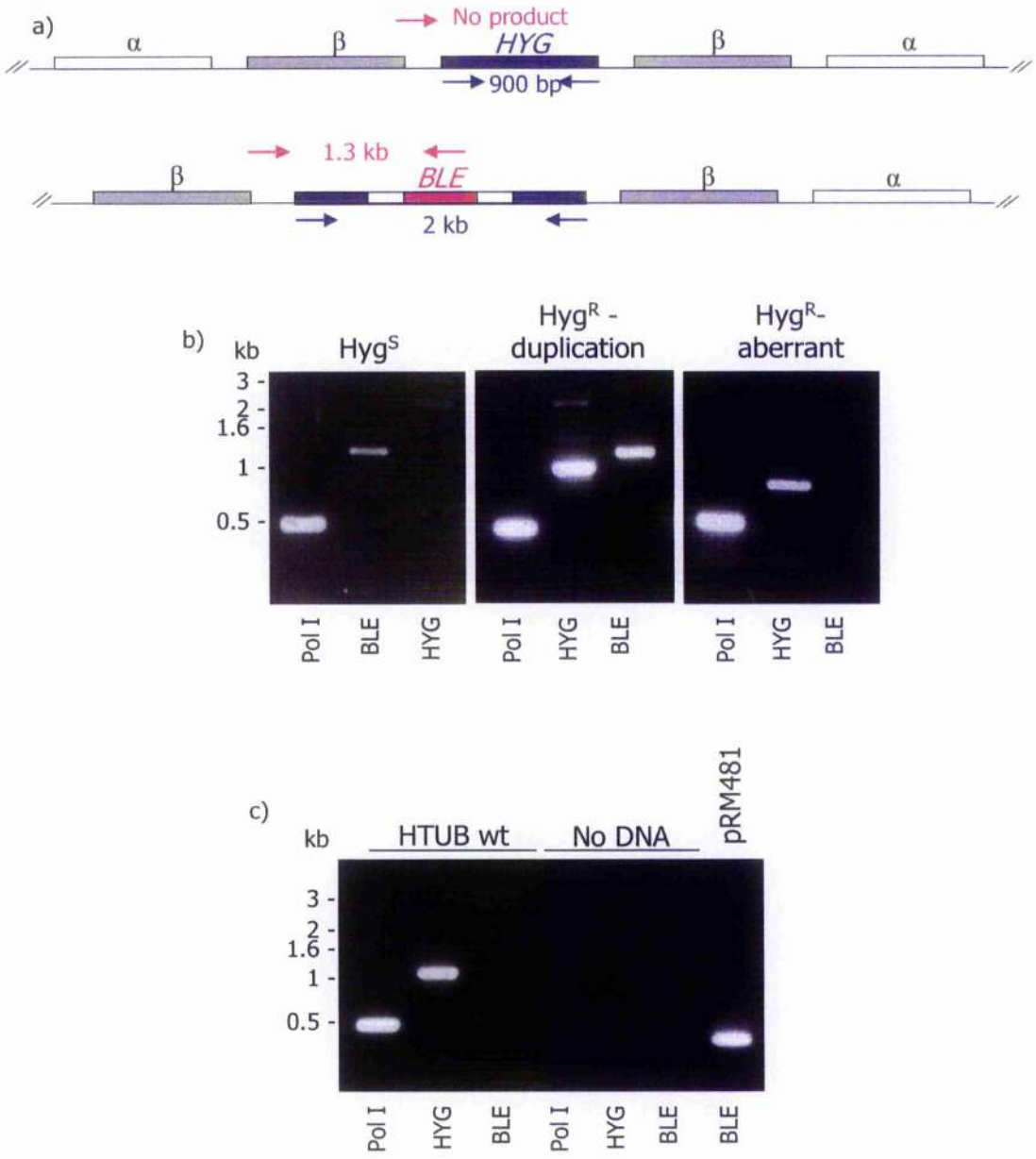


Figure 3.24. Analysis of cloned transformations from the transformation efficiency assay by diagnostic PCR. a) A schematic representation of the PCR reactions carried out. α and β tubulin, *HYG* and *BLE* ORFs are represented by white, grey, blue and purple boxes respectively; primers used for the "HYG" PCR reactions are indicated by blue arrows and the expected sizes of the products generated in blue text; primers used for the "BLE" PCR reactions are shown by purple arrows and the expected sizes of the products generated in purple text. b) Representative PCR reactions from each class of transformant: hygromycin sensitive (HygS), hygromycin resistant with duplication of the HTUB locus (HygR-duplication), and hygromycin resistant due to aberrant integration of the transformation construct (HygR-aberrant). 10 µl of each diagnostic PCR reaction was separated on an agarose gel and stained using ethidium bromide. c) Control PCR reactions: the same reactions as above were performed using *HTUB wt* genomic DNA and no DNA. In addtion, the *BLE* ORF was amplified from plasmid *pRM481*.

3.2.5 Further analysis of Hyg^R transformants

Further analysis was carried out on a new representative selection of *HTUB wt* transformants: four Hyg^R clones with apparent duplication of a genomic region (*Hyg^R 1-4*); one Hyg^R clone with aberrant integration of the *BLE* construct (*Hyg^R 5*); and three Hyg^S clones (*Hyg^S 1-3*). In addition, *HTUB wt* cells were examined as a control (see figure 3.15). The Southern blot analysis carried out for the previous set of transformants was repeated on these transformants in order to confirm that they were representative of the larger sample of clones analysed previously (see figure 3.19). Genomic DNA from the clones was digested with *Hin*dIII, separated by agarose gel electrophoresis, Southern blotted, and probed with *BLE* or *HYG*, giving the results expected based on previous analysis (see section 3.2.3). Clones *Hyg^R 1-4* generated a single restriction fragment of 4.3 kb when probed with the *BLE* ORF, and two fragments of 3.5 kb and 4.3 kb when probed with *HYG* sequence, indicating duplication of the *HYG* locus; *Hyg^R 5* showed a fragment of approximately 13 kb when probed with the *BLE* ORF, indicating a putative aberrant integration, and a fragment of 3.5 kb when probed with *HYG* sequence corresponding to the undisturbed *HTUB* locus; *Hyg^S 1-3* showed the same 4.3 kb restriction fragment with both probes, indicating that the *HYG* loci had been targeted without duplication. These results corroborated that these integrations were indeed representative.

The karyotype of the transformants studied in the previous section was next examined by pulse field gel electrophoresis (see figure 3.20), in order to identify any duplications of large genomic regions which may be indicative of BIR. If locus duplication causing hygromycin resistance is generated by BIR, linear DNA molecules of several kb would be expected to appear in clones *Hyg^R 1-4*, but not in *Hyg^R 5* (thought to result from an aberrant integration) or the four Hyg^S clones. Genomic plugs were made from each of the transformants, and whole chromosomes were separated by pulse field gel electrophoresis and stained with ethidium bromide. Two different sets of conditions were used for the pulse field gel electrophoresis, allowing clear separation of megabase chromosomes in one case, and of the intermediate chromosomes and minichromosomes in the other. In the ethidium bromide-stained gel, all the megabase chromosomes appeared unchanged in all transformants, either Hyg^R or Hyg^S, apart from *Hyg^R 5*. In this clone, where hygromycin resistance is conferred by aberrant integration of the construct, an additional band of approximately 370 kb could be seen. Slight differences in migration of the intermediate chromosomes are probably due to tandem repeat variation within the population of *HTUB wt* cells.

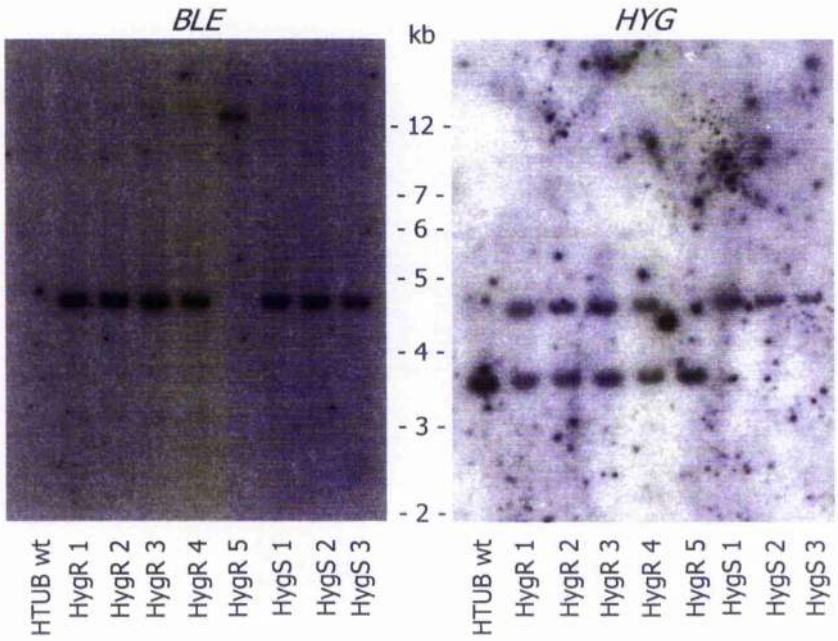


Figure 3.25. Southern analysis of cloned transformants from the recombination efficiency assay. Genomic DNA was prepared from five hygromycin resistant (HygR 1-5) and three hygromycin sensitive (HygS 1-3) clonal cell lines derived from transformation of different constructs into *HTUB wt* cells, and from untransformed *HTUB wt* cells. 5 µg of genomic DNA was digested with *Hin*dIII and run on a 0.8% agarose gel. The DNA was Southern blotted and probed with a 386 bp product PCR-amplified from the *BLE* ORF on the *pRM481* plasmid (left), and washed to 0.2x SSC, 0.1% SDS at 65 °C. Blots were stripped of radioactivity, and reprobed using a 228 bp section of the *HYG* ORF PCR-amplified from *HTUB wt* unstransformed genomic DNA (right), and washed to 0.2x SSC, 0.1% SDS at 65 °C.

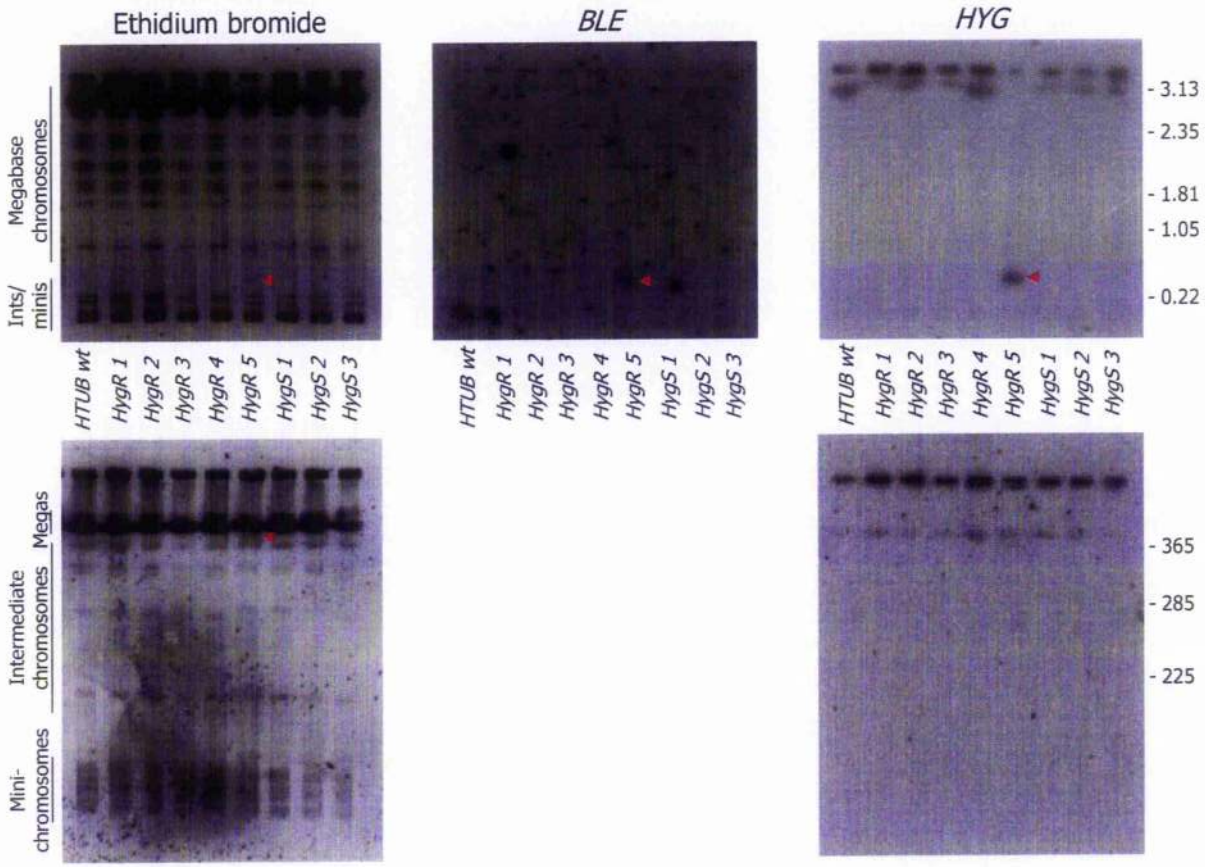
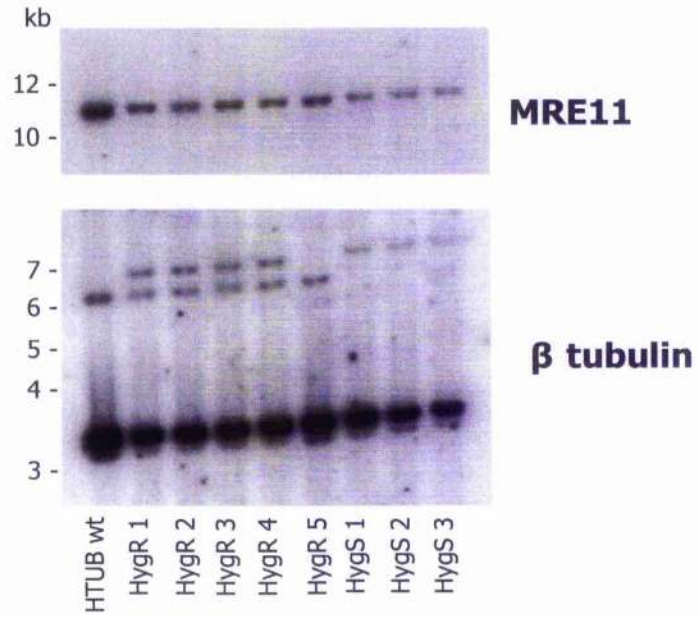


Figure 3.26. Pulse field gel electrophoresis analysis of a subset of transformants. Genomic plugs were prepared from 9 different cell lines (see figure 15), and separated under conditions allowing separation primarily of either the megabase chromosomes (top) or the minichromsomes and intermediate chromosomes (bottom). Pulse field gels were stained with ethidium bromide (left), Southern blotted and probed firstly with a 384 bp product PCR-amplified from the *BLE* ORF (middle) and secondly with a 228 bp product PCR-amplified from the 5' region of the *HYG* ORF (left), and in both cases washed to 0.2x SSC, 0.1% SDS at 65 °C. Pink triangles indicate the extra DNA molecule referred to in the text.

The pulse field gels were next Southern blotted. The 384 bp *BLE* ORF was amplified from *pRM481* plasmid DNA using primers Ble-PstI and Ble-SphI (see appendix 1), and the PCR product used as a probe for the Southern blots (figure 3.20). When probed with *BLE*, large DNA species, corresponding to the unseparated larger megabase chromosomes (Melville *et al.*, 2000), can be seen; this is seen in all cases apart from *HygR 5*, where a DNA molecule corresponding to the extra ~370 kb band is bound by the *BLE* probe. The larger, megabase chromosome band from *HygR 5* does not seem to anneal to this probe, but this is possibly due to the presence of less material in this lane (see ethidium bromide-stained picture). The Southern blots were then stripped of labelled probe fragments, and the whole 861 bp *HYG* ORF was amplified from Lister 427 *HTUB wt* genomic DNA and the PCR product was used to re-probe the stripped Southern blots. Here, annealing to the megabase chromosome DNA is repeated for most clones and the ~370 kb band is hybridised in clone *HygR 5*. The entire *HYG* ORF was used as a probe in this experiment, and therefore hybridisation to the integrated *BLE* constructs, as well as the genomic *HTUB* locus, occurred.

Finally, the transformants were subjected to further Southern analysis to assess whether duplication of the entire tubulin array had taken place. To do this, the intensity of β tubulin (present on chromosome I) signal was compared to the *MRE11* locus (which is located on chromosome II). If an extra copy of the array or the whole chromosome, generated by a BIR-like mechanism, is present in the Hyg^R cells, the amount of tubulin DNA would be expected to increase by 50% compared with sequence on another chromosome (in this case, *MRE11*). Genomic DNA from each clone was digested with *Stu*I, separated by agarose gel electrophoresis and Southern blotted (see figure 3.21). A 432 bp section of the *MRE11* ORF was amplified from Lister 427 genomic DNA using primers MRE11probe-5' and MRE11probe-3' (see appendix), and the PCR product used as a probe for the Southern blot. In each case, the probe hybridised to a single fragment of approximately 11 kb. The intensities of these bands were quantified using ImageQuant analysis software (Amersham Biosciences) and normalised to the band from *HTUB wt* DNA (see figure 3.21 b), to account for differences in DNA loading. The Southern blot was stripped of radioactivity; a 631 bp section of the *β tubulin* gene was amplified from Lister 427 genomic DNA using primers βtubulin5'-3' and βtubulin3'-5' (see appendix), and the PCR product was used as a probe for the stripped Southern blot (see figure 3.21). A *Stu*I restriction site is found in every copy of *α tubulin* within the repetitive tubulin array (see figure 3.15), so the expected sizes of the restriction fragments generated that will be bound by a probe homologous to *β*

a)



b)

	Intensities		
	MRE11	**βtub**	**MRE11/βtub**
wt	100	100	1
HygR 1	57.17	66.04	0.866
HygR 2	39.92	55.53	0.719
HygR 3	51.58	59.90	0.861
HygR 4	47.61	57.85	0.823
HygR 5	58.45	89.66	0.652
HygS 1	29.80	40.71	0.732
HygS 2	22.03	49.12	0.449
HygS 3	20.32	34.24	0.594

Figure 3.27. Quantitative Southern analysis of a selection of cloned transformants from the recombination efficiency assay. a) Genomic DNA was prepared from five hygromycin resistant (HygR 1-5) and three hygromycin sensitive (HygS 1-3) clonal cell lines derived from transformation of different constructs into *HTUB wt* cell lines which were used for pulse field gel analysis, and from untransformed *HTUB wt* cells. 5 µg of genomic DNA was digested with *Stu*I and run on a 0.8% agarose gel. The DNA was Southern blotted and probed with the 432 bp product PCR-amplified from the *MRE11* ORF of Lister 427 genomic DNA (top) and washed to 0.2x SSC, 0.1% SDS at 65 °C. Blots were stripped of radioactivity, and reprobed using a 631 bp section of the β tubulin ORF PCR-amplified from Lister 427 genomic DNA (bottom), and washed to 0.2x SSC, 0.1% SDS at 65 °C. b) Intensities of the bands from the Southern blots in a) were normalised to the band generated by *HTUB wt* DNA (Intensities) and the ratio of MRE11 to β tubulin was calculated (MRE11/βtub).

tubulin are 3.5 kb, corresponding to one copy of a complete tubulin repeat, 6.2 kb, corresponding to the undisturbed *HTUB* locus plus a complete tubulin repeat, and/or 7 kb, corresponding to the *HTUB* locus with integrated *BLE* plus a complete tubulin repeat. The main 3.5 kb band was quantified using ImageQuant (Amersham Biosciences), and again all bands were normalised to the band generated from *HTUB wt* DNA. The normalised intensities from both probings of each lane were then compared, and the ratios are shown in figure 3.21 b. If duplication of the entire tubulin array had taken place, the MRE11/βtub ratio would be expected to be 0.666. This is not what is seen, with the Hyg^S transformants generating lower numbers than the Hyg^R ones. Quantification in this case may, however, be unreliable due to saturation of the β tubulin signal, as this is present in multiple copies, which may reflect the large degree of divergence seen between clones *Hyg^R 1-4*. However, it would appear that if anything, there has been a reduction in the amount of β tubulin DNA present in the Hyg^R clones relative to the Hyg^S and *HygR 5* clones. No firm conclusions can therefore be drawn from this experiment. However, considering the observations from the pulse field gels, no clear evidence has emerged that that BIR accompanies *BLE* integration in the Hyg^R cells that have not undergone aberrant integration.

3.3 Discussion

An assay system based on electroporation of linear DNA homologous to a unique site in the *T. brucei* genome (first described in (Bell and McCulloch, 2003) was used to study the requirements for substrate length and homology in the presence and absence (caused by genetic interruption of the *MSH2* locus) of a functional MMR system. This transformation assay has already been validated as a system for studying recombination efficiency, as it was useful in the detection of differences in efficiency between linear constructs of a constant length (Bell and McCulloch, 2003).To the best of our knowledge, this is the first study in any organism where both substrate length and homology have been assayed simultaneously in a systematic manner. Substrate length was studied in the range 25–200 bp, and sequence divergence in the range 0–11%.

Two different methods were used to calculate transformation efficiency from the number of positive wells counted following selection of transformed trypanosomes. Firstly, a simple calculation was carried out where the number of positive wells was divided by the number of cells put on selection. However, this does not take into account the fact that a positive well could contain cells derived from one, or from several, successful

transformations. Therefore, the binomical distribution was used to make a more accurate estimate of the real transformation efficiency. The overall conclusions to be drawn using both calculation methods are very similar, but there are small differences between the transformations obtained: when the new calculation method is used, the more efficient reactions have higher transformation efficiency compared to the original, simple method. This is to be expected because in these reactions it is more likely that there will be multiple transformants in each well. However, the general conclusions that can be drawn are very similar, and the discussion that follows draws on the original calculations.

Transformations of fifteen constructs into MMR-proficient and MMR-deficient *T. brucei* were carried out. In *T. brucei*, recombination of perfectly-matched (0% divergent) substrates decreases with substrate length in a log-linear relationship. Transformation efficiencies decrease from approximately 10 transformants x 10^{-6} cells to approximately 0.9 transformants x 10^{-6} cells in the range 200–50 bp, with a plateau above this and a steep drop-off below 50 bp. These requirements for substrate length are strikingly similar to results from *E. coli* by Lovett *et al* (Lovett *et al.*, 2002), and similar to the situation seen in *S. cerevisiae*, where there is a log-linear reduction in recombination frequency when substrates between 960 bp and 80 bp are used (Jinks-Robertson *et al.*, 1993). The presence or absence of MMR in this substrate range had absolutely no effect on recombination efficiency, in contrast to previous observations (based on longer substrates) both in *T. brucei* (Bell and McCulloch, 2003) and in *S. cerevisiae* (Datta *et al.*, 1996; Negritto *et al.*, 1997; Chen and Jinks-Robertson, 1998) where removal of MMR lead to a slight increase in the recombination even of perfectly-matched substrates. This could mean that the previous observation in *T. brucei* was an experimental artefact, or that MMR only acts to suppress recombination on sequence matched substrates when they are of a significant length, perhaps because they are more prone to secondary structure during strand exchange.

Recombination of substrates with 5% divergence from the genomic target locus yielded much lower transformation efficiencies than substrates with no divergence, with 200 bp substrates recombining approximately ten times less efficiently than their perfectly matched counterparts. Substrate length also comes into play at this level of divergence. When substrates above ~100bp are used, inactivation of MMR appears to alleviate this reduction somewhat, and the transformation efficiencies for constructs *5%-200bp* and *5%-150bp* are 4.292 and 5.04-fold higher in *HTUB MSH2-/-* than in *HTUB wt* cells. In contrast to what has been observed previously (Bell and McCulloch, 2003), a certain degree of haploinsufficiency in these reactions is suggested by the fact that for these same constructs, the transformation efficiencies are only 2.971 and 2.1-fold higher in *HTUB*

MSH2-/- than in *HTUB MSH2+/-* cells. When recombination substrates below ~100 bp in length are used, there does not appear to be any significant difference between the three cell lines. There is sharp decline in transformation efficiency below this length, similar to that seen below 50 bp in the sequence matched substrates. Recombination of substrates with 11% divergence from the genomic target locus again yielded lower transformation efficiencies. Here, the inactivation of MMR appears to have no effect on transformation efficiencies (when the overall shape of the graph rather than individual data points is taken into account), with a gradual reduction in efficiency with decreasing substrate length observed in all cell lines. No or very few transformants were recovered from transformations with substrates of 50 or 25 bp.

The following overall picture of HR in *T. brucei* emerges from the data described above. It is clear that recombination is dependent on substrate length, at least over the range 25-200 bp. Recombination dependent on substrates longer than 200 bp does not show increased frequency, although it is impossible to tell whether this is due to a true plateau in recombination efficiency, or to the limits of transformation in this assay. Previous observations (Bell and McCulloch, 2003) that *T. brucei* recombination is sensitive to base mismatches are confirmed by these results, with large reductions in transformation efficiency at all substrate lengths when the amount of divergence is raised to 5 and 11%. MMR contributes to the prevention of homeologous recombination, at least for substrates with 5% mismatches that are longer than ~100 bp, although the reduction in transformation efficiency is not completely alleviated. However, when the amount of divergence rises to 11%, disruption of MMR no longer has an effect on transformation efficiency. In contrast to previous data (Bell and McCulloch, 2003), disruption of MMR does not increase the rate of transformation of perfectly matched substrates, as discussed above. Neither does this disruption lead to a decrease in recombination, as is seen in transformation of linear DNA in *S. cerevisiae*, which is dependent on Msh2-Msh3 as well as Rad1-Rad10 (Langston and Symington, 2005). When perfectly matched substrates are used, recombination becomes very inefficient at substrate lengths below ~50 bp. However, the presence of 5% mismatches increases this minimal length for efficient recombination to ~100 bp; presumably, this reflects the MEPS being lengthened by the presence of mismatches. The *T. brucei* MEPS value calculated previously was 142 bp (44–210 at 95% confidence interval) in *MSH2+/+* cells, and 103 bp (62–135 bp at 95% confidence interval) in *MSH2-/-* cells (Bell and McCulloch, 2003). The results shown here demonstrate that recombination can in fact take place using shorter substrates than this.

The analysis of recombination in 5% mismatched substrates below approximately 100 bp, and on 11% mismatched substrates below approximately 200 bp, reveals MMR-independent recombination. This may be reminiscent of the situation in *S. cerevisiae* (Spell and Jinks-Robertson, 2003), where Rad51-dependent and Rad51-independent HR are affected in different ways by MMR, and may reveal a RAD51-independent reaction in *T. brucei*. Rad51-independent recombination in *S. cerevisiae* is dependent on Rad59, and has less stringent substrate requirements, and is not regulated by MMR. Although the *T. brucei* genome does not contain a homologue of Rad59, or indeed Rad52 (El Sayed *et al.*, 2005), it can be hypothesised that an unidentified factor, or factors, is catalysing these reactions. The fact that most integrations by this putative pathway occur by homology into the *HTUB* locus, rather than recombining into aberrant locations, argues that microhomology-mediated end-joining is not responsible (discussed further below). The maximum transformation efficiency rate for this pathway appears to be approximately ~1 transformant x 10^{-6} cells, with a substrate length-dependent reduction in efficiency below ~50 bp (0% substrates) or ~100 bp (5% substrates). The presumption that this pathway is RAD51-independent is in accordance with observations that substrates containing 11% divergence also appear to be unaffected by MMR and that purified *E. coli* RecA protein is incapable of catalysing *in vitro* strand exchange of substrates with this level of divergence (DasGupta and Radding, 1982) and that *S. cerevisiae* recombination of substrates with more than 9% divergence is independent of MMR control (Datta *et al.*, 1996).

Cloned transformants were analysed for hygromycin resistance. While correct integration of the *BLE* constructs into the *HTUB* locus should lead to loss of resistance to this antibiotic, aberrant integration into other loci could leave the *HYG* ORF undisturbed and allow survival of the transformants in the presence of this antibiotic. Retention of a linear episome in the *T. brucei* nucleus causing hygromycin resistance is very unlikely (Wickstead *et al.*, 2003). Naturally occurring episomes are rare in *T. brucei*, and no examples of stably maintained bacterial plasmids have been described in *T. brucei*. Surprisingly, in almost all classes of transformant, some clones retained hygromycin resistance. These reactions also appear to retain some length-dependence, suggesting that they constitute a roughly constant proportion of the transformants in each class of transformants. Southern analysis showed that in almost all cases, hygromycin resistance was not due to aberrant integration of the *BLE* construct, but instead was apparently due to the presence of two copies of the *HTUB* locus.

The possibility that the observed duplication of at least a section of chromosome I had been caused by BIR was investigated. This is especially relevant in *T. brucei* as this is one

of the mechanisms proposed to catalyse *VSG* switching (Barry and McCulloch, 2003). Pulse field gel electrophoresis was carried out using genomic DNA from a selection of transformants. If BIR had taken place, then an extra DNA molecule arising from duplication of a large chromosomal region would be expected to be observed by ethidium bromide staining. This does not appear to be the case, and the only karyotypic differences seen were in fact in the aberrant transformant (as discussed below). Analysis of the tubulin locus by quantitative Southern blot was carried out, and although inconclusive may confirm that the retention of hygromycin resistance observed in some transformants is not due to BIR. It is not possible to state at this point whether hygromycin resistance in these transformants is caused by duplication of the *HTUB* locus during or prior to integration of the *BLE* constructs, or to trisomy of chromosome I (and presumably of other megabase chromosomes) within the cultured Lister 427 trypanosome population.

In a small number of cases, 4/5 of which were detected in an *MSH2-/-* genetic background, retention of hygromycin resistance was not due to the presence of two copies of the *HYG* locus, but instead was caused by aberrant integrations of the *BLE* construct. Although the integration sites were not analysed, it is probable that these reactions occurred by the same RAD51-independent microhomology-mediated end-joining as has previously been observed in *RAD51-/-* cells (Conway *et al.*, 2002b). When the karyotype of one transformant that had undergone aberrant integration was analysed by pulse field gel electrophoresis, a new DNA molecule of approximately 370 kb was observed, seemingly containing the *BLE* and possibly the *HYG* ORFs, although Southern analysis showed that it was not as part of the *HTUB* locus. The most likely cause for the appearance of this DNA molecule is a gross chromosomal rearrangement associated with integration of the construct at an aberrant location. This *de novo* appearance of DNA following transformation is similar to observations made in *L. major* (Beverley and Coburn, 1990), where linear DNA molecules of a similar size, derived from one of the large chromosomes but leaving that chromosome unaltered, were detected in a population following selection with the drug methotrexate. Linear amplified DNA molecules have also been identified in *L. tarentolae* cells following transformations to mutate the *JBP1* ORF (Genest *et al.*, 2005). These new DNA molecules of ~100 kb in length contain a copy of *JBP1*, a copy of the neomycin resistance cassette, and other repetitive sequences in a palindromic arrangement, and cannot be maintained in the absence of drug. Formation of these molecules is thought to be associated with the transformation event itself, rather than the selective pressure, and is proposed to be mediated by template switch within a replication fork. Telomeric sequences had been added to the ends of the *L. major* linear DNA

molecules in both cases; whether this is also true of the DNA molecule observed here is unknown and requires further study.

Despite the importance of *T. brucei* HR in antigenic variation (McCulloch and Barry, 1999), it does not appear to be regulated differently than in other eukaryotes that have been studied. The results described in this chapter add to the growing body of evidence (including (Robinson *et al.*, 2002; Conway *et al.*, 2002b; Bell and McCulloch, 2003)) showing overwhelmingly that HR is carried out in a remarkably similar way in *T. brucei* and in other organisms. This conservation of the recombinational mechanism does not come as a surprise given the very strong conservation of the structure of the RecA and Rad51 proteins, filaments, and strand exchange mechanism (Shinohara *et al.*, 1992; Sung and Robberson, 1995; Brendel *et al.*, 1997; Seitz *et al.*, 1998) throughout the kingdoms of life. Previous authors (Shen *et al.*, 2001; Wickstead *et al.*, 2003) have claimed that *T. brucei* recombination efficiency is not affected by substrate length; the results presented here are in direct opposition to these findings. This could be due to the fact that the transformations described here were carried out in a systematic manner, with electroporation of the different constructs into a single cell population at a defined population density, a factor that anecdotally has a very strong effect on *T. brucei* transformation efficiency.

To illustrate the conservation of the HR mechanism, a detailed comparison of data from *T. brucei* and from other organisms (see figure 3.22) was conducted using published data. A comparison of HR efficiency, in MMR-proficient and MMR-deficient cells, of substrates of fixed size (350 bp, 589 bp and 450 bp, respectively) but increasing sequence divergence in *S. cerevisiae* (Datta *et al.*, 1997), *A. thaliana* (Li *et al.*, 2006) and *T. brucei* (Bell and McCulloch, 2003) was carried out, as an extension of previous analysis by Li *et al* (Li *et al.*, 2006). To allow comparison of different organisms and experimental systems with quite different absolute rates of recombination, all recombination efficiencies were normalised to that of the perfectly matched substrate; the natural logarithm of this recombination ratio was plotted against sequence divergence. In all three organisms, the results were remarkably similar, with the natural log of recombination ratio reducing with sequence divergence in a linear fashion, at least above ~2% sequence divergence (the reduction in efficiency loses its linearity at the lowest levels of divergence). Disruption of the MMR system significantly elevates the recombination rate in all three species at al levels of divergence. However, subtle differences may allow us to draw conclusions about the differences between these distinct recombinational mechanisms. Firstly, HR in wild-type cells from these three organisms shows different sensitivity to mismatches, with yeast

being the most sensitive, followed by *T. brucei*, then *Arabidopsis*. Secondly, the fact that genetic interruption of MMR has a bigger effect in increasing the rate of homeologous recombination in *S. cerevisiae* than in *A. thaliana* or *T. brucei* suggests that the recombinational machinery, rather than MMR, has a bigger effect on plant and trypanosome recombination compared to yeast.

The experiments described here all studied HR at an interstitial genomic location, mimicking the HR required for the repair of most spontaneous DSBs. However, recombination mediating *VSG* switching events occurs at subtelomeric Bloodstream Expression Sites, within specialised Expression Site Bodies (see section 1.1.4), and therefore this recombination may well be regulated differently. For instance, recombination between divergent sequences such as the 70 bp repeats is an important feature of *VSG* switching. Differential use of recombination pathways and/or suppression of MMR (Blundell *et al*., 1996) in *VSG* switching reactions could allow the use of such

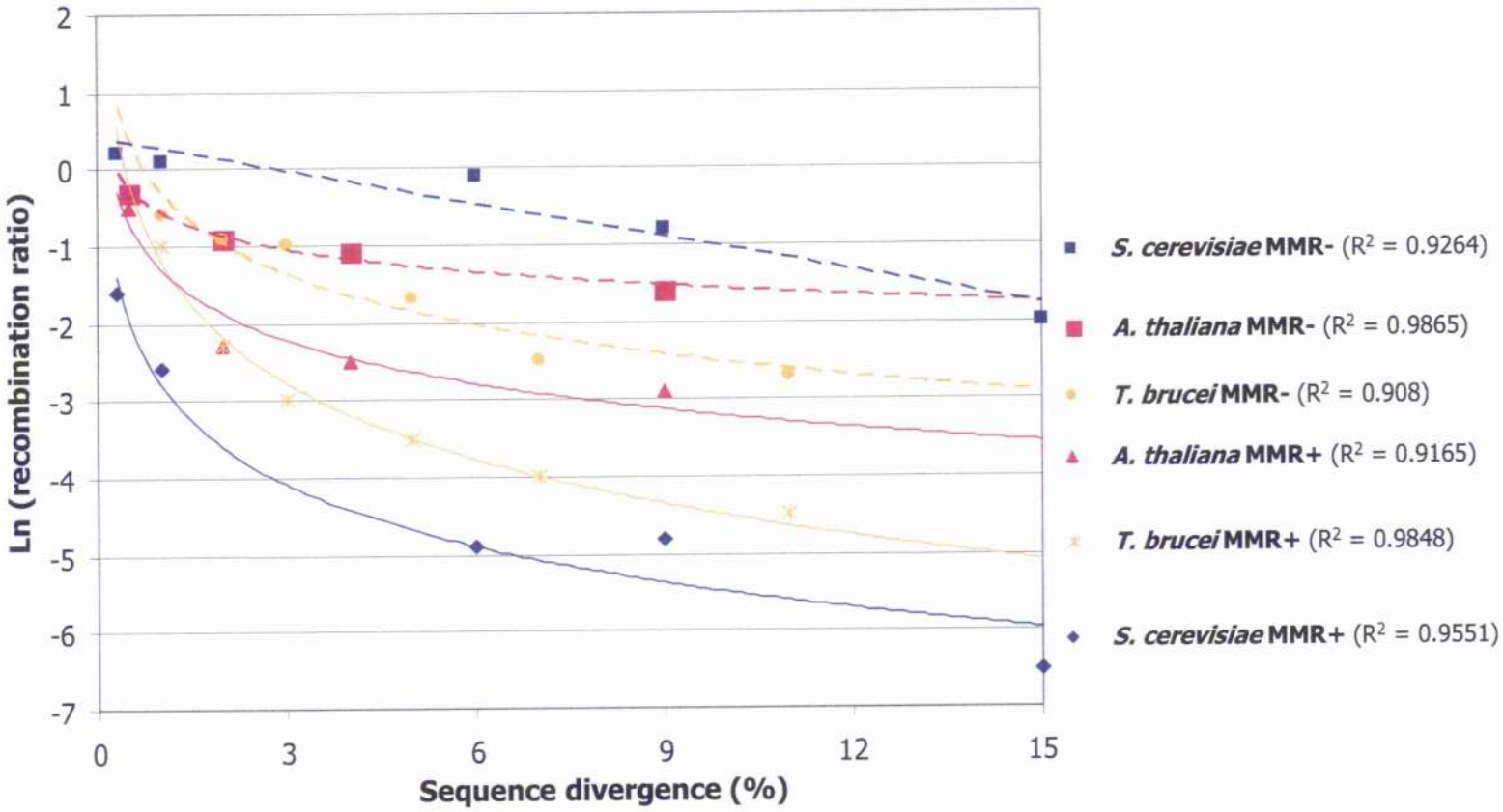


Figure 3.28. A graph showing natural log of the recombination ratio against the percentage sequence divergence from recombination in MMR+ and MMR- cells in *S. cerevisiae*, *A. thaliana* and *T. brucei*. Based on Li *et al.*, 2006.

substrates. This could be caused by the subtelomeric location of the BESs; preferential use of different DNA repair pathways, perhaps with different substrate requirements, depending on the distance from the telomere has been documented in other systems (Ricchetti *et al.*, 2003). Alternatively, the ESB could somehow control the access of DNA repair factors to the active BES. Finally, the use of RNA polymerase I for *VSG* transcription (Gunzl *et al.*, 2003), as opposed to RNA polymerase II for the rest of the genome could somehow mark out the BES for differential repair.

A general model for homologous recombination both at interstitial genomic sites and at bloodstream expression sites is shown in figure 3.23. At interstitial sites, it is known that RAD51 catalyses recombination using longer substrates (Conway *et al.*, 2002b); the MEPS for this reaction has been estimated at 142 bp. However, recombination using shorter substrates is still possible. This is hypothesised to occur *via* two RAD51-independent DNA repair pathways. Firstly, an unknown factor is thought to carry out a role analogous to Rad59 in *S. cerevisiae* in catalysing reactions using shorter DNA substrates, and perhaps with less stringent homology requirements. In addition, a microhomology-mediated end-joining pathway (Burton McBride, Wilkes, Barry and McCulloch, in preparation) is able to carry out DSB repair reactions using only a few base pairs of homology (Conway *et al.*, 2002b). Work in this thesis suggests that at interstitial sites, MMR only monitors RAD51-mediated HR. In the specialised situation of the BES, however, the situation is much less well understood. The fact that mutation of *T. brucei RAD51* (McCulloch and Barry, 1999) or *RAD51-3* (Proudfoot and McCulloch, 2005) dramatically decreases both recombinational and transcriptional pathways of *VSG* switching implies that both RAD51-dependent and RAD51-independent repair pathways have roles in all switching mechanisms. This could be explained by RAD51 having a direct role in the catalysis of *in situ* switching, or by a co-operative role for RAD51-dependent and RAD51-independent recombination pathways in *VSG* switching, with RAD51 regulating the catalysis being carried out by the second mechanism. Microhomology-mediated end-joining may have specialised roles in the formation of mosaic *VSG*s from *VSG* pseudogenes late in chronic infection. The importance of a potential BIR pathway is not known in either interstitial or subtelomeric recombination in trypanosomes.

A number of follow-up experiments could be carried out to test some of the predictions made by this model. Firstly, RAD51-independent recombination at interstitial sites could be studied by creating *RAD51* mutants within the *HTUB wt* cell line. Recombination in these cells would be hypothesised to be less dependent on both the length and homology levels of the recombination substrates. Similarly, the importance of BIR in *T. brucei* could

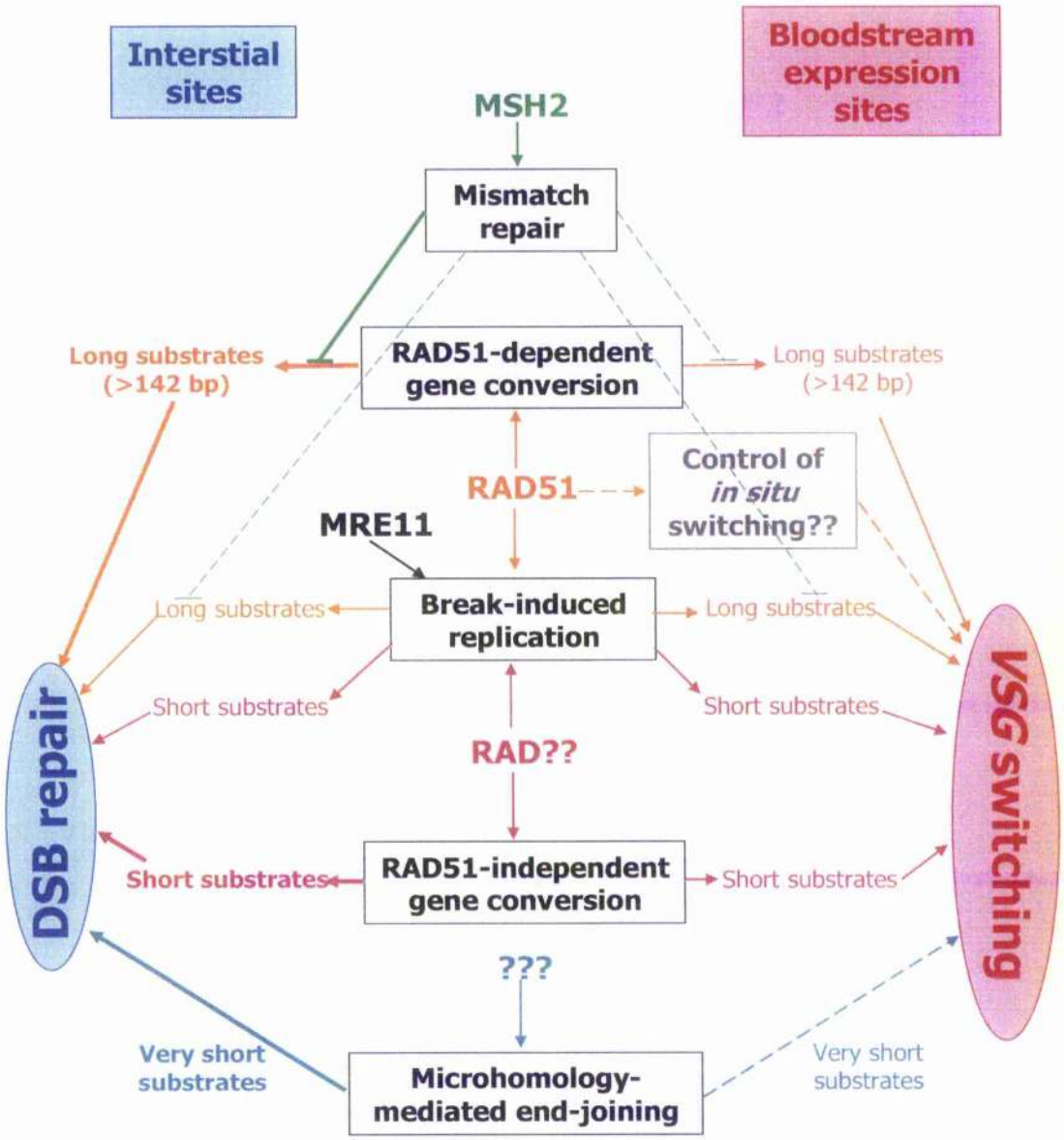


Figure 3.29. A model describing the roles of different DNA repair pathways in *T. brucei*. Repair pathways (shown in the central boxes) thought to be involved in interstitial DSB repair (left; blue) and in *VSG* switching (right; pink) are shown. Proteins involved in these processes are shown in green (MSH2), orange (RAD51), black (MRE11), purple (RAD??, an as-yet-unidentified factor needed for RAD51-independent HR) and turquoise (???, an as-yet-unidentified factor needed for MMEJ). Relationships between these factors are represented by heavy arrows (interactions that are known), light arrows and dashed arrows (interactions that will be assessed by subsequent experiments).

be studied by looking at *MRE11* mutants, as this protein is essential for all forms of yeast BIR (McEachern and Haber, 2006). Secondly, a system analogous to the HTUB assay could be established to study HR at the active BES of Lister 427 bloodstream stage cells, by integration of *HYG* into this locus. Recombination using substrates of varying lengths and homology could be assayed in wild-type cells, and this could be compared to *RAD51* and *MSH2* mutants to establish whether these DNA repair proteins act differently at interstitial and BES locations. Thirdly, an RNAi screen for genes causing sensitivity to DNA damaging agents could allow the identification of factors involved in RAD51-independent recombination and/or MMEJ; a considerable proportion of the genes in the *T. brucei* genome are annotated as hypothetical ORFs with no obvious homologues in other organisms, making database mining only of limited use in this regard. Work currently being carried out in our group on BRCA2 and on the RAD51 paralogues should also add a great deal to our understanding of both general HR and of *VSG* switching in *T. brucei*. In conclusion, although our understanding of HR in *T. brucei* is reasonably well characterised, a number of additional experiments are needed before its specialised role in *VSG* switching can be fully understood.

CHAPTER 4

T. BRUCEI HOMOLOGUES OF MSH4 AND MSH5 AND THEIR POTENTIAL ROLE IN MEIOTIC RECOMBINATION

4 *T. brucei* homologues of MSH4 and MSH5 and their potential role in meiotic recombination

4.1 Introduction

4.1.1 Genetic exchange in the Kinetoplastida

Several groups have found evidence of genetic exchange in wild populations of *T. brucei* (Tait, 1980), *T. cruzi* (Bogliolo *et al.*, 1996; Carrasco *et al.*, 1996; Machado and Ayala, 2001) and *L. major* (Kelly *et al.*, 1991; Dujardin *et al.*, 1995): using multilocus enzyme electrophoresis, molecular karyotyping and random amplification of polymorphic DNA, hybrid phenotypes were seen after coexistence of two parasite strains in the same geographic location. Genetic exchange serves to provide sequence reassortment, and in *T. brucei* it has been proposed that this has the useful function of creating novel of variable antigen types present in neither parent, thus increasing the repertoire (Turner *et al.*, 1991).

Although genetic exchange has been shown to occur in the trypanosomatids, their population structure has also been described as largely clonal (Tibayrenc *et al.*, 1986; Machado and Ayala, 2001; Tibayrenc and Ayala, 2002, Njiokou *et al.*, 2004). Facultative sexual reproduction, where organisms are able to reproduce asexually for long periods in-between rounds of sexual reproduction, is common amongst unicellular eukaryotes (Dacks and Roger, 1999). However, others have argued for a panmictic (freely mating) model of *T. brucei* population structure (*e.g.*, Cibulskis 1988). *T. brucei* population structure seems to depend on the subspecies under investigation. *T. b. rhodesiense* has a clonal population structure, whereas *T. b. brucei* undergoes mating in the wild giving an epidemic population structure (MacLeod *et al.*, 2000). The geographical location of the population can also affect its structure, and even in a single location the population structure can change over time (A. MacLeod, pers. comm.). The existence of hybrid parasites in the wild suggests that genetic exchange has at least occurred in the past, if not in an ongoing fashion in the population. A species' ability to carry out genetic exchange in the laboratory demonstrates an extant capacity for this process.

4.1.1.1 *T. brucei* hybrids can be formed in the laboratory

It has been possible to obtain crosses in the laboratory for *T. brucei* and *T. cruzi* (as described below) but not *L. major* (Panton *et al.*, 1991; Victoir and Dujardin, 2002).

Cotransmission of two uncloned *T. brucei* field isolates in a laboratory setting, the nearest possible approximation to the situation in the wild, has yielded some hybrid cells (Degen *et al.*, 1995). *T. brucei* crosses have been made in the laboratory (Jenni *et al.*, 1986; Sternberg *et al.*, 1989; Gibson, 1989; Turner *et al.*, 1990; Gibson and Whittington, 1993; MacLeod *et al.*, 2005a) by feeding mixtures of parasite stocks to tsetse flies, letting the cells develop in the fly and then sampling metacyclic cells. Cloned progeny can then be analysed for heterozygosity in isoenzyme and restriction fragment length polymorphism analyses. In these experiments, genetic exchange appears to take place at a specific stage of the life cycle: transmission of strains marked with drug resistance markers (Gibson and Whittington, 1993) or fluorescence (Bingle *et al.*, 2001) through the tsetse fly indicates that genetic exchange probably occurs at the epimastigote stage of the life cycle. Mating is non-obligatory, and only a minority of the clones obtained from laboratory crosses are in fact hybrids: metacyclic populations are mixtures of hybrid and parental trypanosomes. The number of hybrids produced could be limited by the relatively short lifespan of the tsetse host (Jenni *et al.*, 1986; Turner *et al.*, 1990).

Meiotic genetic exchange following laboratory crosses has allowed to construction of a genetic map in *T. brucei*. This genetic map (MacLeod *et al.*, 2005b) comprised of 182 markers in 11 linkage groups corresponding to the 11 megabase chromosomes. Comparing this genetic map to the available physical map showed that the average size of a recombination unit is 15.6 kb/cM.

4.1.1.2 Models for the mechanism of *T. brucei* genetic exchange

A number of models have been put forward to explain the mechanism of genetic exchange in *T. brucei*. The presence of a number of hybrid clones containing an increased amount of genetic material, consistent with triploidy (Jenni *et al.*, 1986; Gibson *et al.*, 1992; Tait *et al.*, 1996; Hope *et al.*, 1999), has suggested a model of fusion of diploid parental nuclei followed by progressive chromosome loss during vegetative growth, as proposed initially by Paindavoine et al (Paindavoine *et al.*, 1986). Allelic segregation and reassortment in *T. brucei* have been shown to take place in a Mendelian fashion (Turner *et al.*, 1990; MacLeod *et al.*, 2005a), with the implication that genetic exchange occurs by a meiotic mechanism. The metacyclic cells sampled are thought to be the F1 progeny following meiotic division. However, no haploid (gamete) stages have ever been found (Gibson, 2001; MacLeod *et al.*, 2005a); intial fluorimetric studies (Zampetti-Bosseler *et al.*, 1986) suggested that the metacyclic life cycle stage was haploid, but this was later shown by genetic (Tait *et al.*, 1989) and further fluorimetric (Kooy *et al.*, 1989) methods not to be the

case. The demonstration of a Mendelian mechanism of genetic exchange suggests that the triploid cells isolated following genetic exchange could have arisen by chromosome non-disjunction during meiosis (Tait *et al.*, 1993). Such a process is rather frequent in meiosis, and could have increased precedence in *T. brucei* as strains diverge through asexual growth. Two other models remain possible, however. Firstly, a "ciliate/flagellate" model has been proposed (Gibson, 1995), whereby diploid parents merge then undergo meiosis giving eight haploid nuclei within a single cell, two of which fuse to give a diploid progeny and the rest of which are lost. The second model is a true meiosis that produces (as yet undetected) haploid gametes, which fuse in the salivary gland of the tsetse. This model seems to be preferable for two reasons: firstly, it is the simpler of the two to fit the observed data, and secondly, the ciliate/flagellate model could not explain the relative absence of self-fertilised progeny isolated from crosses (MacLeod *et al.*, 2005a). Kinetoplast DNA is inherited from both parents by a poorly understood mechanism (Gibson and Garside, 1990; Gibson *et al.*, 1997).

4.1.1.3 Genetic exchange in other trypanosmatids

Laboratory crosses by one group (Gaunt *et al.*, 2003) yielded hybrid cells after co-passage through the life cycle of *T. cruzi* cells from different lineages which had been transfected with genes conferring resistance to different antibiotics. As in *T. brucei*, mating is not obligatory. Observations from this experiment suggest a model of fusion of parental genomes followed by allele loss during the intracellular stage within the mammalian host. This is in contrast with the *T. brucei* model, which is based on data from a number of experimental crosses; further work is needed to establish whether the different observations in these closely related species truly reflect a difference in mechanism rather than experimental artefact. Although *L. major* hybrids have been isolated from wild populations, successful laboratory crosses are yet to be reported (Panton *et al.*, 1991). As the species *T. equiperdum* and *T. evansi* do not undergo cyclical transmission in the tsetse fly, it is assumed that they do not undergo genetic exchange (Gibson and Stevens, 1999; Gibson, 2001). On the other hand, *T. vivax* does show some degree of genetic complexity in its population structure (Gibson and Stevens, 1999), and has a metacyclic stage in its life cycle (Tetley *et al.*, 1981), suggesting that genetic exchange could take place in this species.

4.1.2 Meiotic recombination

The mechanism of meiosis has been extensively characterised in model organisms such as *S. cerevisiae* and *M. musculus*, and in human cells. To consider whether this be applicable to the trypanosomatids, this section provides an overview of the process.

4.1.2.1 The stages of meiosis

Meiosis is a specialised cell division, consisting of a single round of replication followed by two reductive rounds of cell division, leading to the formation of haploid gametes. Meiosis appears to be eukaryote-specific, and initiates in diploid cells. During meiosis, large-scale changes in chromosomal structure occur, and are tightly coordinated with a specialised recombination reaction. Whereas in mitotic cells homologous recombination is used to repair some forms of damage to DNA, recombination in meiosis is crucial for the pairing and correct segregation of homologous chromosomes.

The two rounds of division in meiosis are referred to as Meiosis I and Meiosis II, and both follow the same basic stages as mitotic cellular division. Meiosis I (see figure 4.1) begins with genome replication at S phase. The cells then undergo a specialised prophase, consisting of four discrete stages, known as leptotene, zygotene, pachytene and diplotene. These are differentiated by chromosome morphology as observed under a light microscope (Page and Hawley, 2003; Richardson *et al.*, 2004; Svetlanov and Cohen, 2004). At leptotene, sister chromatids (*i.e.* a chromosome and its newly replicated copy) are paired. At zygotene, pairs of sister chromatids are brought together, leading to pachytene, which is characterised by the tight association (synapsis) of homologous chromosomes in a structure called the bivalent, involving the Synaptonemal Complex (SC, see section 4.1.2.2.1). At this stage, the chromosomes are generally compacted and aligned along their lengths. Recombination occurs in meiosis to yield visible exchanges of sequence between homologous chromosomes in structures termed chiasma. At diplotene, the SC is partially degraded, though the chromosomes remain joined by the chiasma. Prophase is followed by diakinesis, when chromosomes condense ready for metaphase. Here, chiasma are resolved and homologous chromosomes are separated to opposite poles by microtubules attached to kinetochores at the centromeres. Finally, at anaphase I, sister chromatid cohesion is broken. During the second round of meiotic division, Meiosis II, sister chromatids are separated to make haploid gametes. It is a significant feature of meiotic division that there is no replication between Meiosis I and Meiosis II. Beyond this, sister chromatid segregation in Meiosis II appears to be like that found in mitosis.

4.1.2.2 Differences between meiotic and mitotic recombination

The recombinational processes in meiosis are very similar to mitotic recombination as described in section 1.2, but with some important differences, detailed below. As in mitotic recombination, repair proteins are often found grouped in foci. The factors involved in stages of meiotic recombination are summarised in figure 4.2.

4.1.2.2.1 Meiotic recombination takes place within the framework of the synaptonemal complex

The Synaptonemal Complex (SC) is a "zip-like" proteinaceous structure whose role is to hold together the bivalent structure (Page and Hawley, 2003; Page and Hawley, 2004). The first stage of SC formation is the appearance of the axial elements (AEs) in early prophase, which hold sister chromatids together. After homologous chromosomes have been brought into proximity at leptotene, some AEs are converted to axial associations, which will later become the sites of recombination (chiasma). Other AEs are converted to lateral elements (LEs) before pachytene, in a process requiring cohesin and condensin proteins. The meiotic cohesin complex contains Smc1 and Smc3 like its mitotic counterpart, but other proteins differ between the complexes. LEs are joined together by transverse filaments, bridges composed of coiled-coil proteins such as Zip1 (in *S. cerevisiae*) and SCP1 (in human cells). Surprisingly, although the SC is required in most organisms for successful meiotic division, in a small number of organisms including *A. nidulans* and *S. pombe*, the SC is not detectable and meiosis appears to be able to take place without it, although a discontinuous linear element can be observed along the aligned chromosomes.

4.1.2.2.2 Meiotic recombination is initiated by DSBs created by SPO11

Initiation of meiotic recombination, at leptotene (Padmore *et al.*, 1991), is caused by the deliberate creation of DSBs catalysed by the transesterase Spo11 (Keeney *et al.*, 1997). Spo11 protein is then removed in a Rad50 and Mre11-dependent, asymmetrical manner, creating 3'single-stranded ends at either side of the DSB (Neale *et al.*, 2005). It is apparent that the function of Spo11 is to create DSBs, as the absence of Spo11 in some mutants can be partially rescued by creation of breaks by other means (Thorne and Byers, 1993; Romanienko and Camerini-Otero, 2000; Celerin *et al.*, 2000). In some organisms, including fungi and mammals (Romanienko and Camerini-Otero, 2000; Celerin *et al.*, 2000; Mahadevaiah *et al.*, 2001), this Spo11-mediated creation of DSBs is necessary for

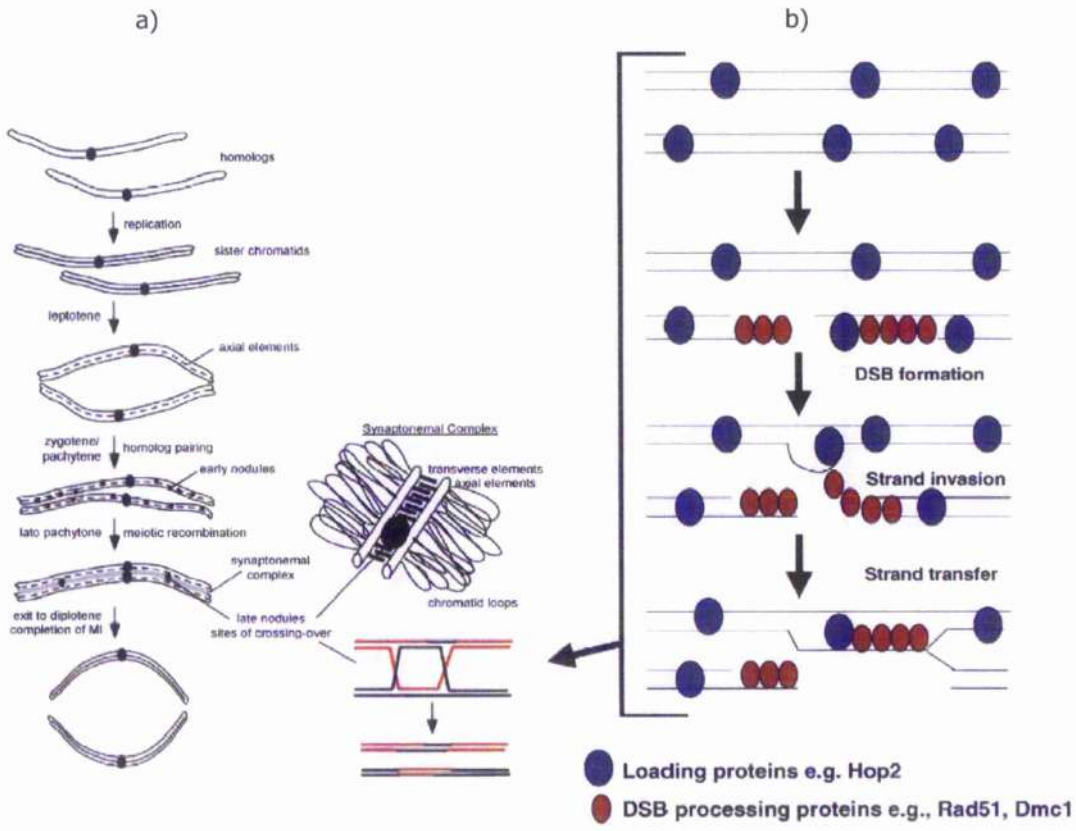
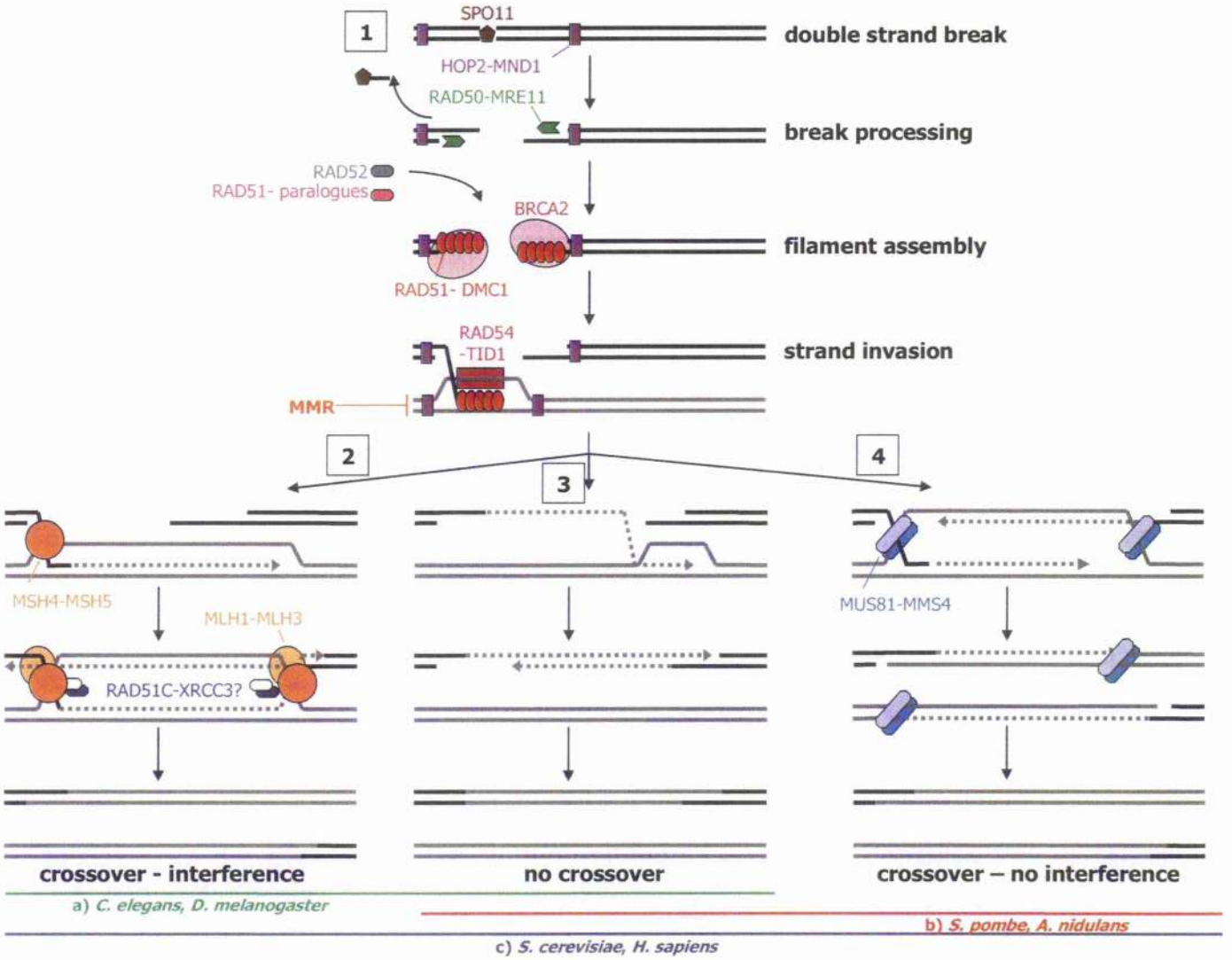


Figure 4.1. Chromosome segregation and recombination in meiosis. a) different stages of chromosomal morphology during meiotic prophase I is indicated. Homologous chromosomes (homologs) are replicated, and sister chromatids remain associated, as in mitosis. Axial elements are formed between the sister chromatids at the leptotene stage, and homologous chromosomes are then paired in a bivalent structure at the zygotene and pachytene stages. Early recombination nodules (small black dots) are visible at early pachytene, whereas late recombination nodules are visible at late pachytene. At pachytene, the chromosomes are bound together by the synaptonemal complex, composed of a variety of proteinaceous substructures. At the end of prophase, the synaptonemal complex is degraded, although homologous chromosomes remain bound together. Chromosome segregation is mediated by the kinetochores (large black dots).b) loading of proteins that facilitate meiotic recombination at a molecular level. Proteins, such as Hop2 (blue circle), are loaded onto the chromosomes before the creation of a DSB, whereas proteins that promote strand exchange, such as Rad51 and Dmc1 (red circle), are loaded onto the ssDNA tails created subsequent to DSB formation. Adapted from Richardson *et al.*, 2004.


1
SPO11
HOP2-MND1
RAD50-MRE11
double strand break
break processing
RAD52
RAD51- paralogues
BRCA2
RAD51- DMC1
filament assembly
RAD54
-TID1
MMR
strand invasion
2
3
4
MSH4-MSH5
MLH1-MLH3
RAD51C-XRCC3?
MUS81-MMS4
crossover - interference
no crossover
crossover – no interference
a) C. elegans, D. melanogaster
b) S. pombe, A. nidulans
c) S. cerevisiae, H. sapiens

(previous page) **Figure 4.2. A generalised scheme of the different pathways of meiotic recombination.** 1) All pathways are initiated by Spo11-mediated DSB formation. The Hop2-Mnd1 dimer binds to chromosomes to promote pairing specifically between homologous chromosomes. Spo11 is removed by Rad50 and Mre11, in an asymmetrical fashion, revealing 3' single-strand DNA overhangs. These are bound by Rad51 and Dmc1, forming nucleoprotein filaments that promote strand invasion. Rad51 and Dmc1 are regulated by Brca2, Rad52 and Rad51 paralogues (as in mitotic recombination). Strand invasion is controlled by Rad54 and its meiosis-specific homologue Tid1, and monitored by the mismatch repair system (MMR), which prevents recombination between homeologous substrates. 2) In the crossover pathway that undergoes interference, Holliday Junctions are bound and stabilised by the Msh4-Msh5 heterodimer. Double Holliday Junctions are resolved to yield recombination products, possibly directly by Rad51C-Xrcc3. 3) Noncrossover recombination events are repaired by a synthesis-dependent strand annealing (SDSA) reaction. 4) In the crossover pathway without interference, strand exchange intermediates are cleaved by Mus81-Mms4 before they reach the Double Holliday Junction stage. The pathways are used to different extents in different organisms, as indicated. Adapted from McCulloch *et al.*, 2006.

the formation of the SC. In other organisms, including *D. melanogaster* (McKim and Hayashi-Hagihara, 1998) and *C. elegans* (Dernburg *et al.*, 1998), the presence of DSBs is not necessary for chromosome synapsis or for the early stages of meiosis (Colaiacovo *et al.*, 2003).

4.1.2.2.3 Strand invasion during meiotic recombination is catalysed by DMC1 as well as RAD51

DSBs formed by Spo11 can be repaired in wild-type cells by one of three recombinational pathways, discussed in more detail below (section 4.1.2.2.4), some of which involve crossing over and exchange of DNA sequence between homologous chromosomes. All are initiated by invasion of the 3' single-stranded ends of the DSB into the corresponding region of the intact homologous chromosome (Hunter and Kleckner, 2001). This is catalysed by Rad51, which is central to mitotic recombination, (see section 1.2.1). Meiotic strand exchange in many organisms is distinguished by the necessity also for Dmc1, a meiosis-specific paralogue of Rad51. Dmc1 is capable of performing strand exchange in the same way as Rad51 (Masson and West, 2001), so why the two related proteins act together remains unclear (see figure 4.2).

C. elegans and *D. melanogaster* appear not to encode DMC1 homologues, indicating that its function in meiosis is not universal. Indeed, impaired meiotic recombination in *S. cerevisiae* Dmc1 mutants can be rescued by overexpression of Rad51 or Rad54 (Stahl *et al.*, 2004), implying at least partial overlap in the function of Rad51 and Dmc1.

4.1.2.2.4 Multiple pathways exist for the repair of SPO11-induced DSBs

Three routes for the repair of DSBs created by Spo11 have been described (see figure 4.2). Crossing over is defined as reciprocal recombination between nonsister chromatids, leading to an exchange of genetic information. Although typical of meiotic recombination, crossover (CO) does not occur at each DSB to avoid the disruption of favourable linkage groups. A synthesis-dependent strand annealing (SDSA) pathway very similar to mitotic repair can repair DSBs without creating COs (noncrossovers; NCOs) (Paques and Haber, 1999; Allers and Lichten, 2001; Hollingsworth and Brill, 2004); no meiosis-specific factors for this have been discovered. In addition, two separate pathways that generate COs have been identified, distinguished by the presence or absence of a phenomenon known as CO interference (COI). In this, COs discourage the formation of other COs nearby, leading to less double COs within a given genetic interval than would be predicted from the frequency of simple COs (Novak *et al.*, 2001). Stahl's counting model of COI suggests that adjacent COs are separated by a specific number of NCO events (Novak *et al.*, 2001).

Distinct genetic factors are required for the two CO pathways. The COI pathway requires a number of factors including the meiosis-specific MutS homologues Msh4 and Msh5 (see section 4.1.2.3), which bind to Holliday Junction (HJ) strand exchange intermediates, preventing them from being unwound to NCO intermediates (Moens *et al.*, 2002; Snowden *et al.*, 2004). Proteins specifically needed for the CO pathway that does not undergo interference include the endonuclease Mus81 (Haber and Heyer, 2001) and its partner, known as Mms4 in *S. cerevisiae* and Eme1 in *S. pombe* (Interthal and Heyer, 2000; de los Santos *et al.*, 2003). Mus81 is related to the nucleotide excision repair endonuclease Rad1, and is expressed at, but is not exclusive to, meiosis (Boddy *et al.*, 2001). This dimer promotes CO by cleaving recombination intermediates in such a way that a CO is formed without a HJ intermediate (Heyer *et al.*, 2003; Osman *et al.*, 2003; Hollingsworth and Brill, 2004). In the absence of both Msh5 and Mms4 (*i.e.*, both CO pathways described above) in *S. cerevisiae*, a small number of COs is still detected. However, viability of these cells is severely reduced, implying that inappropriate pairing must be occurring (de los Santos *et al.*, 2003; Argueso *et al.*, 2004).

Although NCO repair of meiotic DSBs appears to be ubiquitously conserved, the different CO pathways are used to differing extents in different organisms. In some, such as *S. pombe* and *A. nidulans*, the Msh4 and Msh5 genes are absent from the genome, and none of the COs are subject to interference. In some other organisms, such as *C. elegans* and *D.*

melanogaster, only the COI pathway is used. In still others, including *S. cerevisiae* and *H. sapiens*, both pathways are used (Hollingsworth and Brill, 2004). Different classes of meiotic recombination events are temporally distinct; organisms that use both Msh4-dependent and Mus81-dependent CO pathways, such as *S. cerevisiae*, undergo two rounds of DSBR during meiosis, first a pre-synaptic round that is not subject to interference, then a second round that is subject to interference at pachytene (Stahl *et al.*, 2004).

4.1.2.2.5 Meiotic recombination takes place between nonsister chromatids

It is important that recombination takes place between homologous chromosomes within the bivalent structure. The basis for this must lie, at least in part, with the meiosis-specific proteins involved. Key to this may be the Hop2-Mnd1 dimer, which interacts with chromosomes before and during synapsis (Leu *et al.*, 1998; Petukhova *et al.*, 2003) and is necessary for pairing of homologous chromosomes at the early stages of meiosis (Petukhova *et al.*, 2003). This heterodimer interacts directly with Rad51 and Dmc1 to stimulate D-loop formation (Petukhova *et al.*, 2005) specifically with the homologous chromosome (Zierhut *et al.*, 2004). Nevertheless, Hop2 and Mnd1 appear to be absent from the *D. melanogaster* and *C. elegans* genomes, as is Dmc1, implying that these factors cannot be absolutely required for meiotic recombination.

Beyond the recombination process, the SC is an important determinant of substrate choice in recombination. As well as having roles in the mediation of axial length compaction, the SC protein condensin controls homologue pairing (Yu and Koshland, 2003). It mediates the correct localisation of Red1 and Hop1, whose physical interaction (de los Santos and Hollingsworth, 1999; Woltering *et al.*, 2000) is important for correct chromosome pairing (Schwacha and Kleckner, 1997). Red1-Hop1 also acts as a structural restraint on Dmc1-independent (but still Rad51-dependent) strand invasion, and recruits the checkpoint kinase Mek1 (Wan *et al.*, 2004).

4.1.2.3 Meiosis-specific MutS homologues

The meiosis-specific MutS homologues Msh4 and Msh5 were first discovered in screens for meiotic genes in *S. cerevisiae* (Ross-Macdonald and Roeder, 1994; Hollingsworth *et al.*, 1995), and orthologues have since been identified in a number of eukaryotes, including *H. sapiens* (Paquis-Flucklinger *et al.*, 1997; Her and Doggett, 1998; Winand *et al.*, 1998; Santucci-Darmanin *et al.*, 1999; Bocker *et al.*, 1999), *M. musculus* (Her *et al.*, 1999; Santucci-Darmanin *et al.*, 2001), *A. thaliana* (Higgins *et al.*, 2004), and *C. elegans* (where Msh4 has been named him-14, high incidence of males) (Zalevsky *et al.*, 1999; Kelly *et*

al., 2000). They are not conserved universally, however, as they are not detected in *D. melanogaster* or *S. pombe* (Hoffmann and Borts, 2004). Although identifiably MutS homologues, both proteins lack both the N-terminal mismatch interaction domain found in other members of the family (Culligan *et al.*, 2000; Obmolova *et al.*, 2000). The absence of this domain suggests they do not recognise base mismatches, and have adapted to bind other structures, as confirmed by Snowden et al (Snowden *et al.*, 2004). Mutation of Msh4 or Msh5 in *S. cerevisiae* has no effect on the repair of heteroduplex DNA in mitotic or meiotic cells (Ross-Macdonald and Roeder, 1994; Hollingsworth *et al.*, 1995; Zalevsky *et al.*, 1999). In addition, mice lacking these proteins show no increase in tumorigenesis (Edelmann *et al.*, 1999; de Vries *et al.*, 1999a; Kneitz *et al.*, 2000), a phenotype associated with loss of mismatch repair (see section 1.3). A number of experiments indicate meiosis-specific functions: *msh5* mutant mice are infertile due to defects in spermatogenesis (Nakagawa *et al.*, 1999; de Vries *et al.*, 1999b), and both *MSH4* and *MSH5* are important for chromosome pairing in mice (Kneitz *et al.*, 2000). Biochemical studies show that the MSH4-MSH5 dimer binds to the centre of HJ structures, forming an ATP hydrolysis-dependent sliding clamp which stabilises the HJ (Snowden *et al.*, 2004). Finally, the MSH4-MSH5 heterodimer has an important role in the regulation of crossing over during meiotic recombination (see section 4.1.2.2.4).

The transcription of Msh4 and Msh5 is has been shown to be modulated between mammalian tissues, and stages of the yeast life cycle. In *S. cerevisiae*, *Msh4* and *Msh5* mRNA and protein are detected only in meiotic tissues (Ross-Macdonald and Roeder, 1994; Pochart *et al.*, 1997). In mammals, the situation is slightly more complex. Northern analysis of human tissues shows that both *Msh4* and *Msh5* mRNA can be detected in meiotic testis and ovary (Her and Doggett, 1998; Winand *et al.*, 1998; Santucci-Darmanin *et al.*, 1999; Bocker *et al.*, 1999; Her *et al.*, 1999; Santucci-Darmanin *et al.*, 2001) and Msh4 and Msh5 proteins can be observed colocalised to synapsed chromosomes (Ross-Macdonald and Roeder, 1994; Winand *et al.*, 1998; Santucci-Darmanin *et al.*, 1999; Bocker *et al.*, 1999). However, *Msh5* mRNA has also been detected in a wide range of other tissues such as lymph nodes and thymus (Winand *et al.*, 1998; Bocker *et al.*, 1999) and Msh5 protein has been detected in B- and T-cell tumour cell lines (Bocker *et al.*, 1999), whereas Msh4 mRNA cannot be detected in non-meiotic cells (Winand *et al.*, 1998; Santucci-Darmanin *et al.*, 2001). RT-PCR of the Msh4 gene product shows that small amounts of transcription occur in non-meiotic tissues, but these are an alternatively spliced form that is incapable of dimerisation with MSH5 (Santucci-Darmanin *et al.*, 1999). In general, Msh4 expression levels have been remarked to be lower than Msh5 levels (Santucci-Darmanin *et al.*, 1999).

Msh4 and Msh5 proteins interact with each other, as shown by yeast two-hybrid analysis and coimmunoprecipitation (Pochart *et al.*, 1997; Bocker *et al.*, 1999; Her *et al.*, 1999). The formation of this dimer appears to be human-specific, since both proteins are incapable of forming homodimers (Winand *et al.*, 1998) or of interacting with other MMR-related MutS homologues (Pochart *et al.*, 1997). Interactions with proteins from other families are possible, however: *in vivo* co-localisation of Msh4 with the MutL homologues Mlh1 (Santucci-Darmanin *et al.*, 2000) and Mlh3 (Santucci-Darmanin *et al.*, 2002), the SC proteins Zip1 (Novak *et al.*, 2001) and Cor1 (Kneitz *et al.*, 2000), has been observed and Msh4 has been observed in Rad51 repair foci (Lipkin *et al.*, 2002; Neyton *et al.*, 2004; Svetlanov and Cohen, 2004). Yeast two-hybrid assays have shown interactions of Msh4 with Rad51 and Dmc1 (Neyton *et al.*, 2004), and VBP1, which acts in microtubule assembly during meiosis and may provide competition with Msh5 (Her *et al.*, 2003), and interaction between Msh5 and a central regulator of DNA repair Mre11 (Uetz *et al.*, 2000).

The interaction of MutL homologues with the Msh4/Msh5 heterodimer reveals functions in meiotic recombination in addition to their roles in post-replicative MMR (Hoffmann and Borts, 2004); separation-of-function mutant alleles of the MutS homologues have been isolated showing a difference between the structural role they play in meiotic recombination and the catalytic one they play in post-replicative MMR (Argueso *et al.*, 2002; Argueso *et al.*, 2003). Sterility is seen in mice lacking the MLH1-MLH3 dimer (Lipkin *et al.*, 2000). MLH1 and MLH3 are found as a heterodimer in meiotic cells (Wang and Kung, 2002), are required for a subset of COs (Wang *et al.*, 1999) and appear to be epistatic to MSH4 in the regulation of CO formation (Abdullah *et al.*, 2004). These observations suggest that this heterodimer could provide an intermediate between the MSH4-MSH5 heterodimer and other factors, a role analogous to that of the MutL orthologues in mitotic MMR. Studies looking for interactions between the MLH1-MLH3 and MSH4-MSH5 heterodimers have not always shown co-localisation (Hoffmann and Borts, 2004); this could suggest a transient reaction releasing the MSH4-MSH5 sliding clamp from DNA structures as proposed by Snowden *et al.* (Snowden *et al.*, 2004). This model is consistent with the observations of Argueso *et al.* (Argueso *et al.*, 2003), that some interference is retained in the absence of Mlh1 and Mlh3.

4.1.3 Identification of meiotic genes in the trypanosome genome

Sequencing of the nuclear genomes of *T. brucei*, *T. cruzi* and *L. major* (Berriman *et al.*, 2005; El Sayed *et al.*, 2005; Ivens *et al.*, 2005) has revealed that a number of meiotic genes are present. BLAST searches and phylogenetic analysis identified putative orthologues

encoding Spo11, Dmc1, Mnd1, Hop1, Msh4 and Msh5, and possibly Mus81 (see table 4.1). These proteins constitute a "core meiotic machinery" of proteins common to all or most meiotic systems (Villeneuve and Hillers, 2001), a concept explored further below (see section 4.3). However, at the outset of this work the expression and functionality of none of these genes had been explored in any of the trypanosomatids, nor in any other protists. Subsequently, DMC1 has been examined in *T. brucei* (Proudfoot and McCulloch, 2006). In this chapter, the putative *T. brucei* orthologues of Msh4 and Msh5 have been chosen for more detailed examination.

		SC	DSB initiation			strand exchange				resolution		
		Hop1*	Spo11*	Rad50	Mre11	Rad51	Dmc1*	Hop2*	Mnd1*	Msh4*	Msh5*	Mus81
protozoa	T. brucei	+	+(2)	+	+	+	+	-	+	+	+	+
	L.major	+	+(2)	+	+	+	+	-	+	+	+	+
	T.cruzi	+	+(2)	+	+	+	+	-	+	+	+	+
	G.intestinalis	+	+	+	+	-	+(2)	+	+	-	-	+
	P.falciparum	+	+	+	+	+	+	+	+	-	-	+
metazoa	H.sapiens	+	+	+	+	+	+	+	+	+	+	+
	D.melanogaster	+	+	+	+	+	-	-	-	-	-	+
	C.elegans	+	+	+	+	+	-	-	-	+	+	+
fungi	S.cerevisiae	+	+	+	+	+	+	+	+	+	+	+
	S.pombe	+	+	+	+	+	+	+	+	-	-	+
	E.cuniculi	-	+	+	+	+	-	+	+	-	-	+
plant	A.thaliana	+	+(3)	+	+	+	+	+	+	+	+	+

Table 4.1. A table summarising the presence of members of the core meiotic machinery in a range of eukaryotic organisms. The genomes of *Trypanosoma brucei*, *Leishmania major*, *Trypanosoma cruzi*, *Giardia intestinalis*, *Plasmodium falciparum*, *Homo sapiens*, *Drosophila melanogaster*, *Caenorhabiditis elegans*, *Saccharomyces cerevisiae*, *Schizzosaccharomyces pombe*, *Encephalitozoon cuniculi* and *Arabidopsis thaliana* were searched for the presence of genes required for synaptonemal complex formation (Hop1), DSB initiation (Spo11, Rad50 and Mre11), strand exchange (Rad51, Dmc1, Hop2 and Mnd1) and the resolution of strand exchange intermediates (Msh4, Msh5 and Mus81). Presence of a gene in the genome is indicated by +, and where more than one copy appears to be present, the number of copies is indicated in brackets, whereas absence of a gene is indicated by -. An asterisk next to the name of the gene shows that its product is only active in meiotic processes.

4.2 Results

4.2.1 Identification of T. brucei *homologues of MSH4 and MSH5*

Similarity searching of the TIGR and Sanger *T. brucei* sequencing databases were preformed using BLAST algorithms and the MSH2 polypeptide sequences from *H. sapiens* and *S. cerevisiae* revealed the *T. brucei* genes encoding homologues of MSH2, MSH3 and MSH6 identified previously by our group (Bell et al 2004; see section 1.4), as well as two other protein sequences. One predicted a 3495 bp ORF (XM_817292) located on chromosome 10 encoding a 1165 aa protein (XP_822385), and the second a 2370 bp ORF (AAX79333) located on chromosome 3 encoding a 732 aa protein (XP_825491).

The presence of these genes in the *T. brucei* genome database strain was confirmed by independent sequencing. Primers MSH4-1 to MSH4-11 and MSH5-1 to MSH5-7, homologous to the TREU 927 sequence, shown in appendix 3 and 4, were used to PCR-amplify the *MSH4* and *MSH5* ORFs from Lister 427 genomic DNA. PCR products were sequenced by the Molecular Biology Sequencing Unit (MBSU), University of Glasgow, and assembled using ContigExpress (VNTI). The sections of sequence available were in all cases identical in the Lister 427 and TREU 927 genomes, apart from a single synonomous difference in *MSH5*.

in silico translations of the predicted polypeptides for XM_817292 and AAX79333 were next used in BLAST searches of the NCBI database, and the results are summarised in tables 4.2 and 4.3. For XM_817292, the sequences in the database found to have the highest similarity were predicted MutS homologues from *T. cruzi* and *L. major*. For AAX79333, only a *T. cruzi* predicted gene was identified from the trypanosomatids. Beyond these close homologues (see below), XM_817292 revealed similarity to a predicted MSH4 from the sea urchin *Strongylocentrotus purpuratus* and MutS homologues from a range of organisms, whereas AAX79333 revealed MSH5 homologues from a variety of eukaryotes as well as human G7 protein. These results indicate that each gene encodes a MutS homologue distinct from the previously described MSH2, MSH3 and MSH6(8) *T. brucei* genes.

***T. brucei* MSH4**	XP_822385	
XP_822385	mismatch repair protein [Trypanosoma brucei]	0
XP_815283	mismatch repair protein MSH4, putative [Trypanosoma cruzi]	0
CAJ04396	mis-match repair protein, putative [Leishmania major]	8.00E-74
XP_780878	similar to MutS protein homolog 4 [Strongylocentrotus purpuratus]	6.00E-42
EAL92596	DNA mismatch repair protein Msh2, putative [Aspergillus fumigatus Af293].	1.00E-36
BAB06088	DNA mismatch repair protein (mismatch recognition step) [Bacillus halodurans]	7.00E-35
ZP_00564383	MutS 1 protein [Methylobacillus flagellatus KT]	3.00E-31
NP_816771	DNA mismatch repair protein HexA [Enterococcus faecalis V583]	5.00E-31
ZP_00315170	Mismatch repair ATPase (MutS family) [Microbulbifer degradans 2-40]	2.00E-30
YP_014020	DNA mismatch repair protein MutS [Listeria monocytogenes str. 4b F2365]	2.00E-30

***T. cruzi* MSH4**	XP_815283	
XP_815283	mismatch repair protein MSH4, putative [Trypanosoma cruzi]	0
XP_822385	mismatch repair protein [Trypanosoma brucei]	0
CAJ04396	mis-match repair protein, putative [Leishmania major]	4.00E-75
XP_780878	similar to MutS protein homolog 4 [Strongylocentrotus purpuratus]	2.00E-35
XP_762508	hypothetical protein UM06361.1 [Ustilago maydis 521]	1.00E-31
BAB06088	DNA mismatch repair protein (mismatch recognition step) [Bacillus halodurans]	1.00E-31
XP_688406	PREDICTED: similar to MutS protein homolog 4 [Danio rerio]	1.00E-31
AAT70180	MSH4 [Arabidopsis thaliana]	2.00E-31
AAK79801	Mismatch repair protein MutS, ATPase [Clostridium acetobutylicum ATCC 824]	4.00E-31
YP_147159	DNA mismatch repair protein [Geobacillus kaustophilus HTA426]	7.00E-31

***L. major* MSH4**	CAJ04396	
CAJ04396	mis-match repair protein, putative [Leishmania major]	0
XP_822385	mismatch repair protein [Trypanosoma brucei]	1.00E-58
XP_815283	mismatch repair protein MSH4, putative [Trypanosoma cruzi]	1.00E-44
O13396	DNA mismatch repair protein msh-2 Neurospora crassa	1.00E-22
XP_662610	hypothetical protein AN5006_2 [Aspergillus nidulans FGSC A4]	5.00E-22
EAA74947	MSH2_NEUCR DNA mismatch repair protein MSH2 [Gibberella zeae PH-1]	9.00E-21
CAJ05982	DNA mismatch repair protein, putative; MSH2 [Leishmania major]	5.00E-20
AAK08648	putative mismatch repair protein MSH2 [Trypanosoma brucei]	6.00E-20
XP_368365	hypothetical protein MG00879.4 [Magnaporthe grisea 70-15]	1.00E-19
XP_819877	DNA mismatch repair protein MSH2, putative [Trypanosoma cruzi]	2.00E-19

Table 4.2. Results of BLAST searches of the NCBI database using MSH4 protein sequences from *T. brucei*, *T. cruzi* and *L. major* as query sequences. Accession numbers of the result sequences are shown in the left-hand column, the brief description of each gene product as shown in the NCBI database is shown in the middle column, and the E-value of each result (indicating the probability that sequences could be that similar by chance) in the right-hand column.

***T. brucei* MSH5** XP_825491

XP_825491	mismatch repair protein MSH5 [Trypanosoma brucei]	0
XP_809831	mismatch repair protein MSH5, putative [Trypanosoma cruzi]	0
NP_038628	mutS homolog 5 [Mus musculus]	3.00E-52
EAA67360	hypothetical protein FG01367.1 [Gibberella zeae PH-1]	5.00E-52
CAB52406	G7 protein [Homo sapiens]	9.00E-52
AAH41031	MSH5 protein [Homo sapiens]	1.00E-51
NP_751897	mutS homolog 5 isoform b [Homo sapiens]	1.00E-51
XP_532080	PREDICTED: similar to mutS homolog 5 isoform c isoform 1 [Canis familiaris]	1.00E-51
CAE83984	mutS homolog 5 [E. coli] [Rattus norvegicus]	1.00E-51
AAP36674	Homo sapiens mutS homolog 5 (E. coli) [synthetic construct]	1.00E-51

***T. cruzi* MSH5** XP_809831

XP_809831	mismatch repair protein MSH5, putative [Trypanosoma cruzi]	0
XP_825491	mismatch repair protein MSH5 [Trypanosoma brucei]	0
XP_847837	MutS-like protein [Leishmania major]	1.00E-68
CAE83984	mutS homolog 5 [E. coli] [Rattus norvegicus]	2.00E-56
NP_038628	mutS homolog 5 [Mus musculus]	3.00E-56
XP_532080	PREDICTED: similar to mutS homolog 5 isoform c isoform 1 [Canis familiaris]	5.00E-56
AAH41031	MSH5 protein [Homo sapiens]	2.00E-55
NP_751897	mutS homolog 5 isoform b [Homo sapiens]	2.00E-55
CAB52406	G7 protein [Homo sapiens]	2.00E-55
NP_079535	mutS homolog 5 isoform a [Homo sapiens]	5.00E-55

***L. major* MSH5** XP_847837

XP_847837	MutS-like protein [Leishmania major]	0
XP_809831	mismatch repair protein MSH5, putative [Trypanosoma cruzi]	6.00E-36
XP_825491	mismatch repair protein MSH5 [Trypanosoma brucei]	3.00E-32
BAB02831	unnamed protein product [Arabidopsis thaliana]	3.00E-16
EAA67360	hypothetical protein FG01367.1 [Gibberella zeae PH-1]	2.00E-15
AAS76767	At3g20475 [Arabidopsis thaliana]	5.00E-15
BAD95388	DNA mismatch repair protein [Arabidopsis thaliana]	5.00E-15
AAP97415	msh5 [Coprinopsis cinerea]	2.00E-14
XP_475492	unknown protein [Oryza sativa (japonica cultivar-group)]	4.00E-13
XP_447404	unnamed protein product [Candida glabrata]	1.00E-12

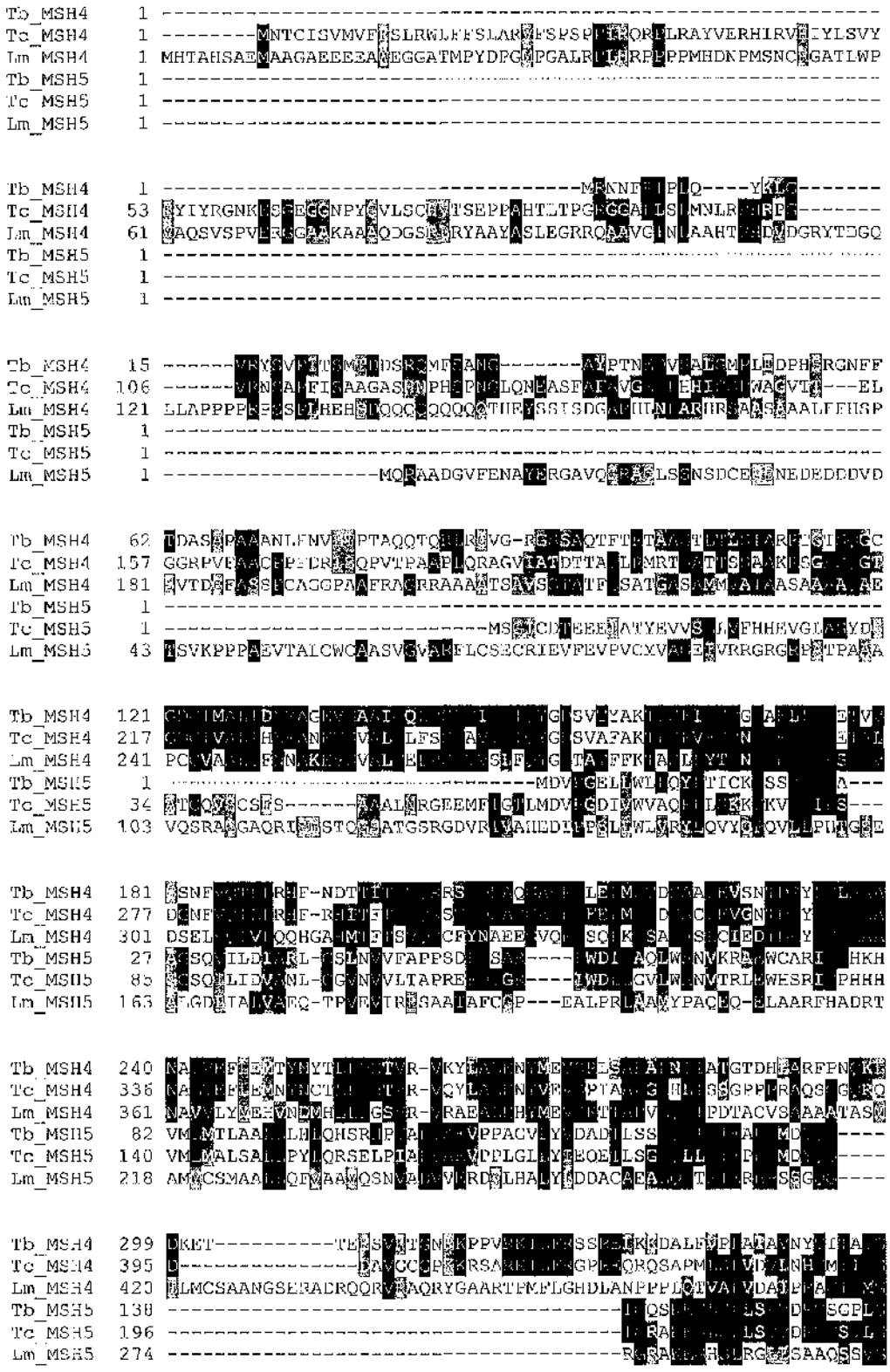
Table 4.3. Results of BLAST searches of the NCBI database using MSH5 protein sequences from *T. brucei*, *T. cruzi* and *L. major* as query sequences. Accession numbers of the result sequences are shown in the left-hand column, the brief description of each gene product as shown in the NCBI database is shown in the middle column, and the E-value (indicating the probability that sequences could be that similar by chance) of each result in the right-hand column.

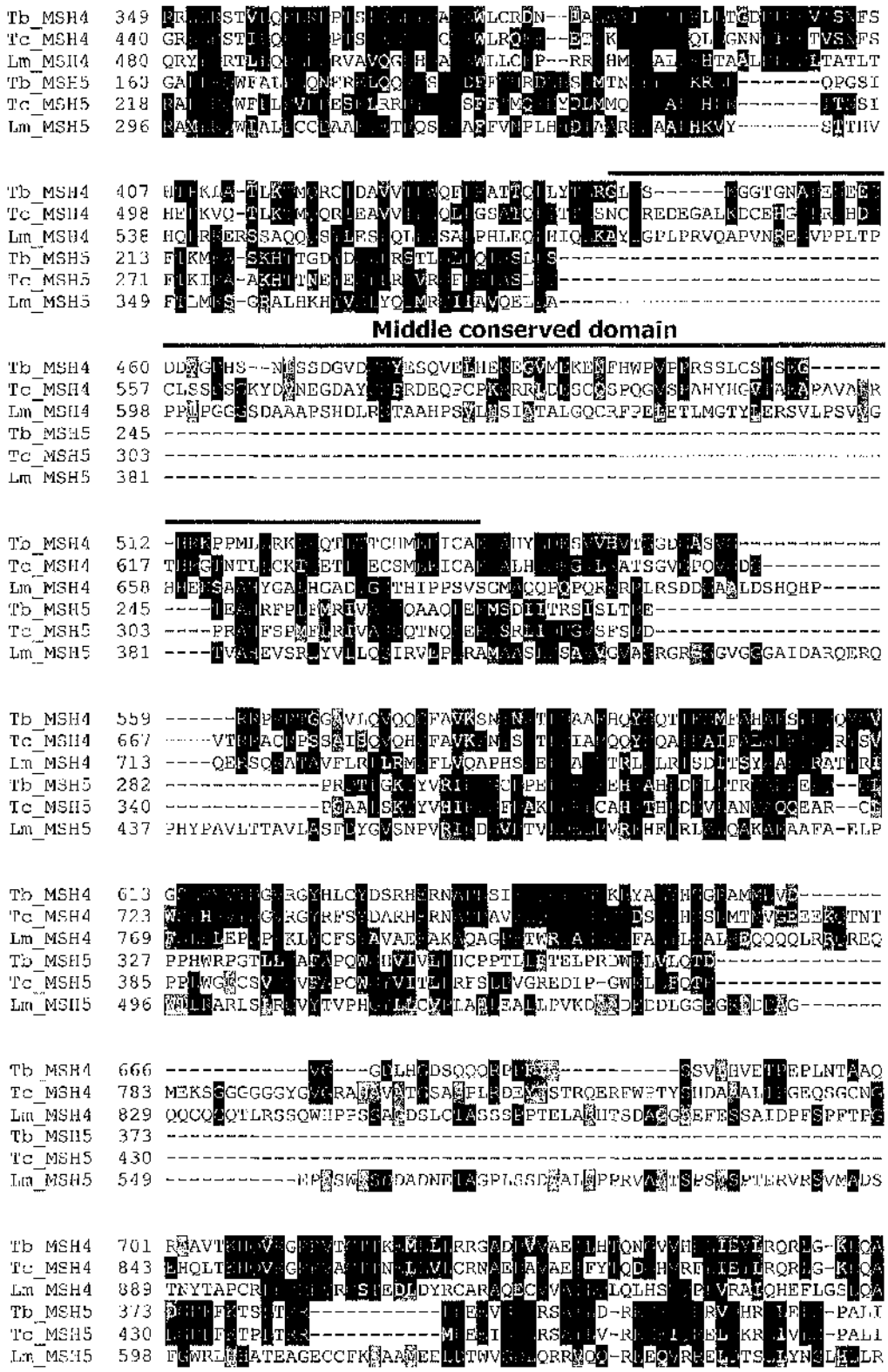
4.2.2 Identification of putative T. cruzi and L. major homologues of MSH4 and MSH5

As discussed above, BLAST searching of NCBI revealed putative orthologues of *T. brucei* gene XP_822385 in *Trypanosoma cruzi* and *Leishmania major*, and a putative orthologues of *T. brucei* gene AAX79333 in *T. cruzi*. To examine this further, the translated polypeptides of the *T. brucei* genes were used to search the *T. cruzi* and *L. major* genome databases, using BLAST searches (see tables 4.2 and 4.3). Two genes were thus identified in *L. major*, a 4712 bp ORF located on chromosome 21 predicted to encode a 1377 aa protein (CAJ04396) and orthologous to XP_822385, and a 3287 bp ORF located on chromosome 29 predicted to encode a 1096 aa protein (XP_847837) and orthologous to XP_825491. Similiarly, two orthologous genes were identified in *T. cruzi*, though their chromosomal locations are unknown. A 3773 bp ORF predicted to encode a 1264 aa protein (XP_815283) is orthologous to XP_822385, and a 2750 bp ORF predicted to encode a 784 aa protein (XP_809831) is orthologous to XP_825491. Accession numbers for all the above sequences are shown in appendix 2.

For the *T. cruzi* and *L. major* genes, no PCR confirmation of their presence in the genome was performed. Nevertheless, *in silico* translations of these genes were used to search the NCBI database, and the results are summarised in tables 4.2 and 4.3. As for the *T. brucei* proteins, the most related sequences in the database were the orthologues from the other kinetoplastids. Beyond this, *T. cruzi* XP_815283 identified predicted MSH4 from the sea urchin *Strongylocentrotus purpuratus* and other MutS homologues from a range of organisms, whereas XP_809831 identified MSH5 homologues from a variety of eukaryotes as well as human G7 protein. For *L. major* protein CAJ04396 identified MSH2 sequences from a range of eukaryotic organisms, including the kinetoplastids, whereas XP_847837 identified predicted MSH5 from the fungus *Coprinopsis cinerea* and a number of unannotated hypothetical ORFs from a range of organisms. This analysis again suggests that we have identified two further MutS-like genes in each of *T. cruzi* and *L. major*.

The six MutS-related protein sequences identified above were compared by performing a sequence alignment using ClustalX (Thompson *et al.*, 1997), which was visualised using Boxshade (http://www.ch.embnet.org/software/BOX_form.html), as shown in figure 4.3. This shows that the proteins show significant sequence conservation with each other, and the middle conserved domain and ATPase/helix-turn-helix domains common to all MutS homologues, as defined in Obmolova *et al.* 2000 (Obmolova *et al.*, 2000) can be identified




Middle conserved domain

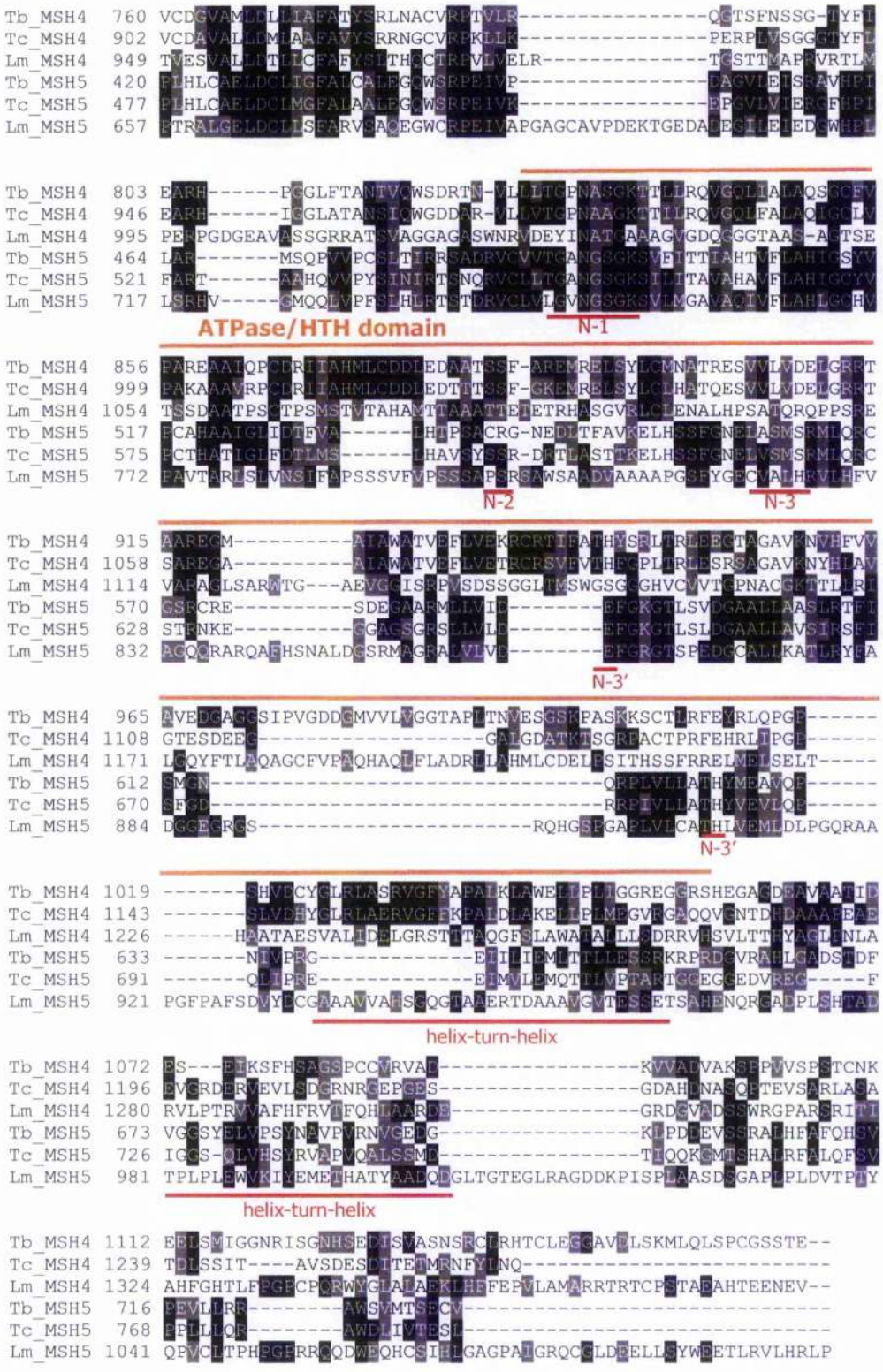


Figure 4.3. Global multiple alignments of putative kinetoplastid homologues of MSH4 and MSH5. Multiple sequence alignment of putative *T. brucei* (Tb), *T. cruzi* (Tc) and *L. major* (Lm) MSH4 and MSH5 polypeptide sequences. Sequences were aligned using ClustalX (Thompson *et al.*, 1997) and shaded using the Boxshade server (http://www.ch.embnet.org/software/BOX_form.html): identical residues in 30% of the sequences are shaded in black, and conserved residues in grey. The middle conserved domain and ATPase/helix-turn-helix domains, as defined by Obmolova *et al.* (2000), are shown by underlining in blue and orange respectively, and the important motifs of the ATPase an helix-turn-helix domain, as described in the text, are shown underlined in red.

(see figure 1.10). It also shows that gene products XP_822385, XP_815283 and CAJ04396 from *T. brucei*, *T. cruzi* and *L. major* are more related to each other, whilst gene products XP_825491, XP_809831 and XP_847837 from *T. brucei*, *T. cruzi* and *L. major* are orthologues. For instance, each of the latter three is significantly larger than the former three. Within each class, the *L. major* protein is more diverged than *T. brucei* and *T. cruzi*. The alignments also reveal inserted sequences without homology in the other kinetoplastid proteins. In *L. major* CAJ04396, there is an insertion of approximately 100 amino acids between the middle conserved domain and the ATPase domain. In *L. major* XP_847837, there are a number of insertions: N-terminal and C-terminal extensions of 73 and 59 amino acids respectively (compared to the *T. brucei* sequence); three small insertions between the middle conserved domain and the ATPase domain; and four small insertions within the ATPase domain, but between the conserved motifs of this region. In *T. cruzi* XP_809831, there is an N-terminal extension of 55 amino acids (in relation to *T. brucei*), homologous to a section of the *L. major* N-terminal extension. The functional significance, if any, of these divergences is unknown.

4.2.3 Bioinformatic assignment of the kinetoplastid MutS-related proteins as MSH4 and MSH5

4.2.3.1 Sequence similarity based on whole-protein pairwise alignments

To attempt to assign functional identities to the MutS homologues from *T. brucei*, *T. cruzi* and *L. major* described above, all MutS-related proteins from the kinetoplastids were compared by pairwise alignment using ClustalW (VNTI), and the percentage sequence identities calculated (see table 4.4). The level of sequence identity between members of this protein family, both within and between species, was generally low (7.2% - 20.6% sequence identity, excluding identities between likely orthologues). The orthologous proteins are therefore clearly revealed, as their sequence identities are significantly higher (21.9-64.9% identity). The exception to this is the putative *L. major* MSH4 and MSH5 proteins, which showed only 21.9–23.4% and 21.9–23% identity to the *T. brucei* and *T. cruzi* putative orthologues, respectively. In general, the *L. major* MutS-related proteins are the most diverged, for reasons that are unclear.

In addition, the MutS homologues from *T. brucei* were compared to the MutS homologues from *H. sapiens* by pairwise alignment using ClustalW (VNTI), and the percentage sequence identities calculated (see table 4.5). The level of sequence identity is equivalent to that found for the kinetoplastid homologues (8 – 19.9% sequence identity, excluding

	Tb MSH2	Tb MSH3	Tb MSH4	Tb MSH5	Tb MSH8	Tc MSH2	Tc MSH3	Tc MSH4	Tc MSH5	Tc MSH6	Lm MSH2	Lm MSH3	Lm MSH4	Lm MSH5	Lm MSH8
Tb MSH2	100	18.8	15.6	13.5	18.9	64.9	18.2	16.8	13.9	18	55.6	18.9	13.8	13.8	16.1
Tb MSH3	18.6	100	14.1	14.8	19.9	19.5	58.3	14.6	16.6	20.5	18.9	44.7	13.1	13.7	20.6
Tb MSH4	15.6	14.1	100	12.4	13.9	17.3	14.7	44	14.2	10.9	16.9	13.6	21.9	13.7	13
Tb MSH5	13.5	14.8	12.4	100	13.8	13.3	14.2	10.2	53.7	11.7	13.7	12.9	7.2	21.9	12.9
Tb MSH8	18.9	19.9	13.9	13.8	100	18.4	19.3	15.4	14.4	59.8	19	19.6	13.2	13.5	44.6
Tc MSH2	64.9	19.5	17.3	13.3	18.4	100	18.7	16.9	14.6	17.8	57.3	18.2	13.8	13.1	17.2
Tc MSH3	18.2	58.3	14.7	14.2	19.3	18.7	100	15.7	16	18.9	18.7	46.6	12.7	12.5	19.9
Tc MSH4	16.8	14.6	44	10.2	15.4	16.9	15.7	100	13.6	13.3	16.4	13.7	23.4	10.1	13
Tc MSH5	13.9	16.6	14.2	53.7	14.4	14.6	16	13.6	100	14.9	13.2	16	9.2	23	13.6
Tc MSH6	18	20.5	10.9	11.7	59.8	17.8	18.9	13.3	14.9	100	17.6	18.4	12.2	12.3	45.4
Lm MSH2	55.6	18.9	16.9	13.7	19	57.3	18.7	16.4	13.2	17.6	100	19.4	13.3	14.8	17.5
Lm MSH3	18.9	44.7	13.6	12.9	19.6	18.2	46.6	13.7	16	18.4	19.4	100	12.4	15	16.3
Lm MSH4	13.8	13.1	21.9	7.2	13.2	13.8	12.7	23.4	9.2	12.2	13.3	12.4	100	11.5	11.5
Lm MSH5	13.8	13.7	13.7	21.9	13.5	13.1	12.5	10.1	23	12.3	14.8	15	11.5	100	15.1
Lm MSH8	16.1	20.6	13	12.9	44.6	17.2	19.9	13	13.6	45.4	17.5	16.3	11.5	15.1	100

Table 4.4. Percent sequence identities from pairwise comparisons of MutS homologues from *T. brucei* (Tb), *T. cruzi* (Tc) and *L. major* (Lm). Each pair of protein sequences was aligned using AlignX (Vector NTI) and the percent sequence identities obtained are shown. Identities between two MSH2 proteins are highlighted in pink, between two MSH3 proteins in orange, between two MSH4 proteins in green, between two MSH5 proteins in blue and between two MSH6/8 proteins in lilac.

	Tb MSH2	Tb MSH3	Tb MSH4	Tb MSH5	Tb MSH8	Hs MSH2	Hs MSH3	Hs MSH4	Hs MSH5	Hs MSH8
Tb MSH2	100	18.8	15.6	13.5	18.9	30	17	18.4	12	14.2
Tb MSH3	18.8	100	14.1	14.8	19.9	17.3	21.5	18.9	16.6	17.1
Tb MSH4	15.6	14.1	100	12.4	13.9	14	12.8	18.3	14.2	12
Tb MSH5	13.5	14.8	12.4	100	13.8	12	11.5	12.9	19	8
Tb MSH8	18.9	19.9	13.9	13.8	100	19	18.6	18.1	15.4	18.4
Hs MSH2	30	17.3	14	12	19	100	18.8	15.6	13.5	18.9
Hs MSH3	17	21.5	12.8	11.5	18.6	18.8	100	14.1	14.8	19.9
Hs MSH4	18.4	18.9	18.3	12.9	18.1	13.5	14.8	100	12.4	13.8
Hs MSH5	12	16.6	14.2	19	15.4	13.5	14.8	12.4	100	13.8
Hs MSH8	14.2	17.1	12	8	18.4	18.9	19.9	13.9	13.8	100

Table 4.5. Percent sequence identities from pairwise comparisons of MutS homologues from *T. brucei* (Tb) and *H. sapiens* (Hs). Each pair of protein sequences was aligned using AlignX (Vector NTI) and the percent sequence identities obtained are shown. Identities between two MSH2 proteins are highlighted in pink, between two MSH3 proteins in orange, between two MSH4 proteins in green, between two MSH5 proteins in blue and between two MSH6/8 proteins in lilac.

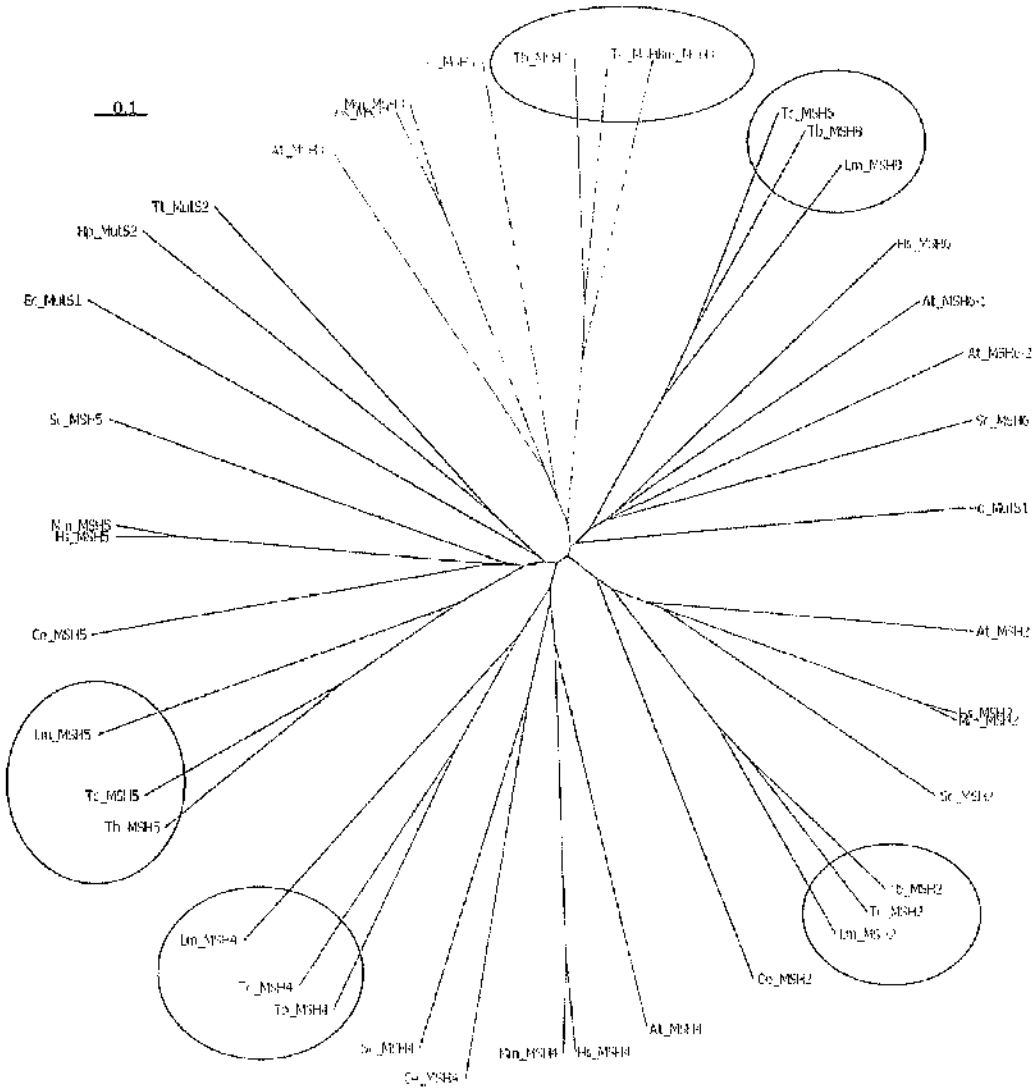


Figure 4.4. A phylogenetic tree of prokaryotic and eukaryotic MutS homologues. The whole polypeptide sequences of 2 MutS1 , 2 MutS2, 8 MSH2, 7 MSH3, 8 MSH4, 7 MSH5 and 7 MSH6/MSH8 homologues from *Arabidopsis thaliana* (At), *Caenorhabiditis elegans* (Ce), *Esherischia coli* Ec), *Haemophilus ducreyi* (Hd), *Homo sapiens* (Hs), *Helicobacter pylori* (Hp), *Leishmania major* (Lm), *Mus musculus* (Mm), *Saccharomyces cerevisiae* (Sc), *Thermus thermophilus* (Tt), *Trypanosoma brucei* (Tb), and *Trypanosoma cruzi* (Tc) were compared by ClustalX (Thompson *et al.*, 1997). The sequence comparison was then used to generate a phylogenetic tree which was visualised using Hypertree (http://hypergraph.sourceforge.net/hypertree.html). The MutS sequences are shown in black, MSH2 in red, MSH3 in orange, MSH4 in green, MSH5 in blue and MSH6/MSH8 in green. The subgrouping of the kinetoplastid orthologues within each branching is circled. The distance corresponding to 10 aa changes per 100 residues is indicated.

identities between likely orthologues). The orthologous proteins are again clearly revealed, with significantly higher sequence identities (18.3–30% sequence identity).

4.2.3.2 Construction of a phylogenetic tree

To attempt to assign identities more clearly to the trypanosomatids MutS-related proteins relative to a wide range of such factors, available protein sequences of all putative MutS homologues from the trypanosomatids, from other eukaryotes (*Arabidopsis thaliana, Caenorhabiditis elegans, Homo sapiens, Mus musculus* and *Saccharomyces cerevisiae*), MutS1 sequences from *Escherischia coli* and *Thermus thermophilus* and MutS2 sequences from *Haemophilus ducreyi* and *Hylobacter pylori*, were aligned using ClustalX (Thompson *et al.*, 1997), and an unrooted phylogenetic tree derived using Hypertree (http://hypergraph.sourceforge.net/hypertree.html) (see figure 4.4). In this analysis, predicted orthologues of each MutS family member grouped together, with a subgroup made up of kinetoplastid proteins found within each group of orthologues, indicating strongly that the new MutS-related proteins identified here encode MSH4 and MSH5 homologues.

4.2.3.3 Protein alignments of conserved domains

Protein sequences of all MutS homologues from *T. brucei*, *H. sapiens* and *S. cerevisiae* were aligned using ClustalX (Thompson *et al.*, 1997) and visualised using Boxshade (http://www.ch.embnet.org/software/BOX_form.html). Functional domains as defined by Oblomova *et al.* (Obmolova *et al.*, 2000) were isolated and are shown in figures 4.5 – 4.10. As has been reported previously (Bell *et al.*, 2004), *T. brucei* MSH2, MSH3 and MSH8 display considerable conservation with their yeast and human homologues in the mismatch interaction domain. Within this domain, The conserved Phe and Glu residues that bind to the mismatched base in *Thermus aquaticus* MutS (Obmolova *et al.*, 2000) are conserved in *T. brucei* MSH8 (Phe69 and Glu71of this protein), *H. sapiens* and *S. cerevisiae*, though not in MSH2 or MSH3. Mutations in these residues have been linked to MMR deficiency in yeast and HNPCC in humans (Obmolova *et al.*, 2000), indicating an important role in protein function. Two other residues identified for *T. aquaticus* MutS are required for DNA recognition and are conserved in *T. brucei* MSH3 (Gln118 and Arg136) and MSH8 (Gln 129 and Arg145), as well as in the human and yeast homologues. As has been observed for their homologues in other organisms (Eisen, 1998; Culligan *et al.*, 2000; Obmolova *et al.*, 2000), the *T. brucei* MSH4 and MSH5 proteins display only limited sequence homology in this conserved mismatch interaction domain. As for MSH2, none

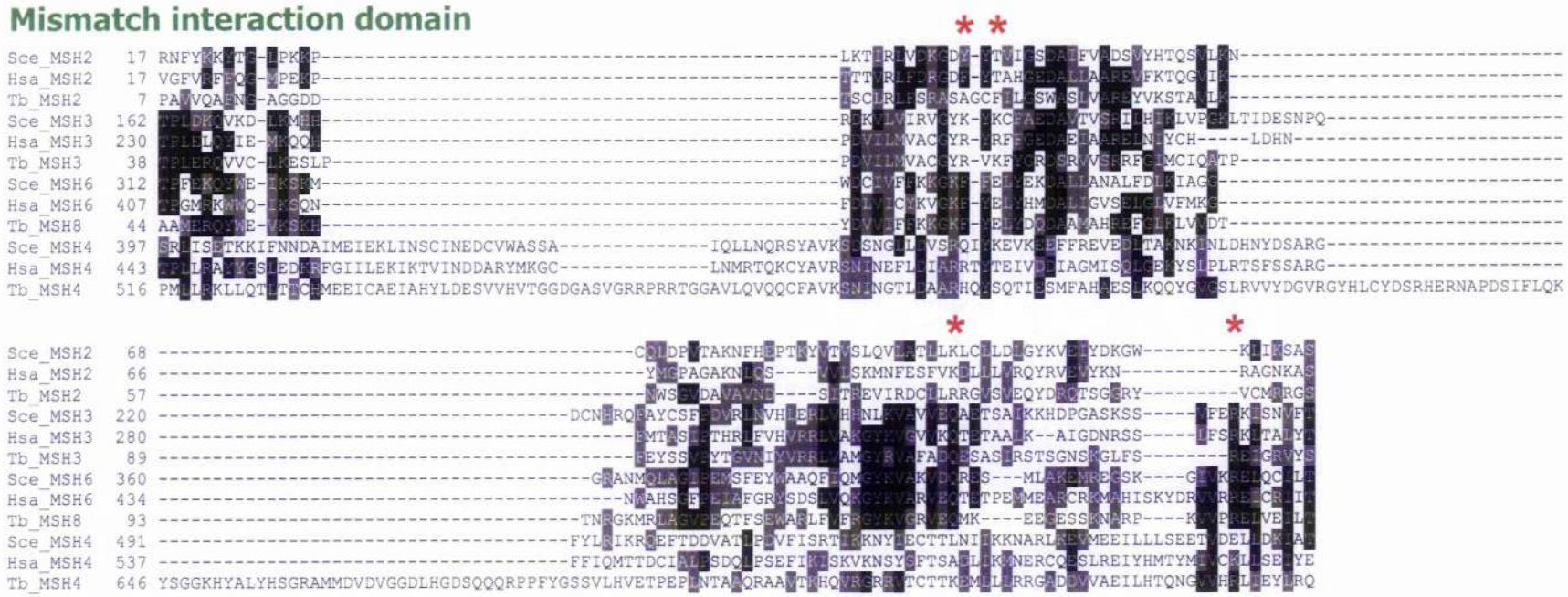


Figure 4.5. Multiple alignment of *T. brucei* MSH4 polypeptides with a range of MutS homologues. Multiple sequence alignment of the *T. brucei* MSH4 polypeptide with MSH2, MSH3 and MSH4 homologues from *S. cerevisiae* (Sc), *H. sapiens* (Hsa) and *T. brucei* (Tb). Sequences were aligned using ClustalX (Thompson et al 1997) and shaded using the Boxshade server (http://www.ch.embnet.org/software/BOX_form.html): identical residues found in at least 30% of residues are shaded in black and conserved residues in grey. Important residues and motifs (as described in the text; Obmolova *et al.*, 2000) are highlighted by red asterisks and underlining. To align the Mismatch Interaction Domain, this region was identified in the MSH2, MSH3 and MSH8 polypeptides and aligned in isolation with the entire protein sequences from the MSH4 homologues.

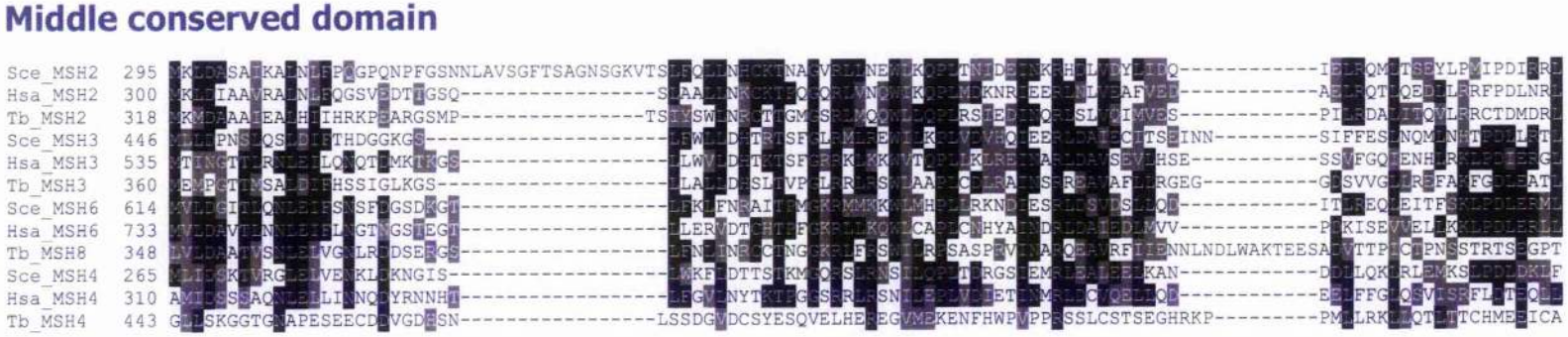


Figure 4.6. Multiple alignment of *T. brucei* MSH4 polypeptides with a range of MutS homologues. Multiple sequence alignment of *T. brucei* MSH4 polypeptide with MSH2, MSH3 and MSH4 homologues from *S. cerevisiae* (Sc), *H. sapiens* (Hsa) and *T. brucei* (Tb). Sequences were aligned using ClustalX (Thompson et al 1997) and shaded using the Boxshade server (http://www.ch.embnet.org/software/BOX_form.html): identical residues are shaded in black and conserved residues in grey. To obtain the middle conserved domain, global multiple alignments were carried out.

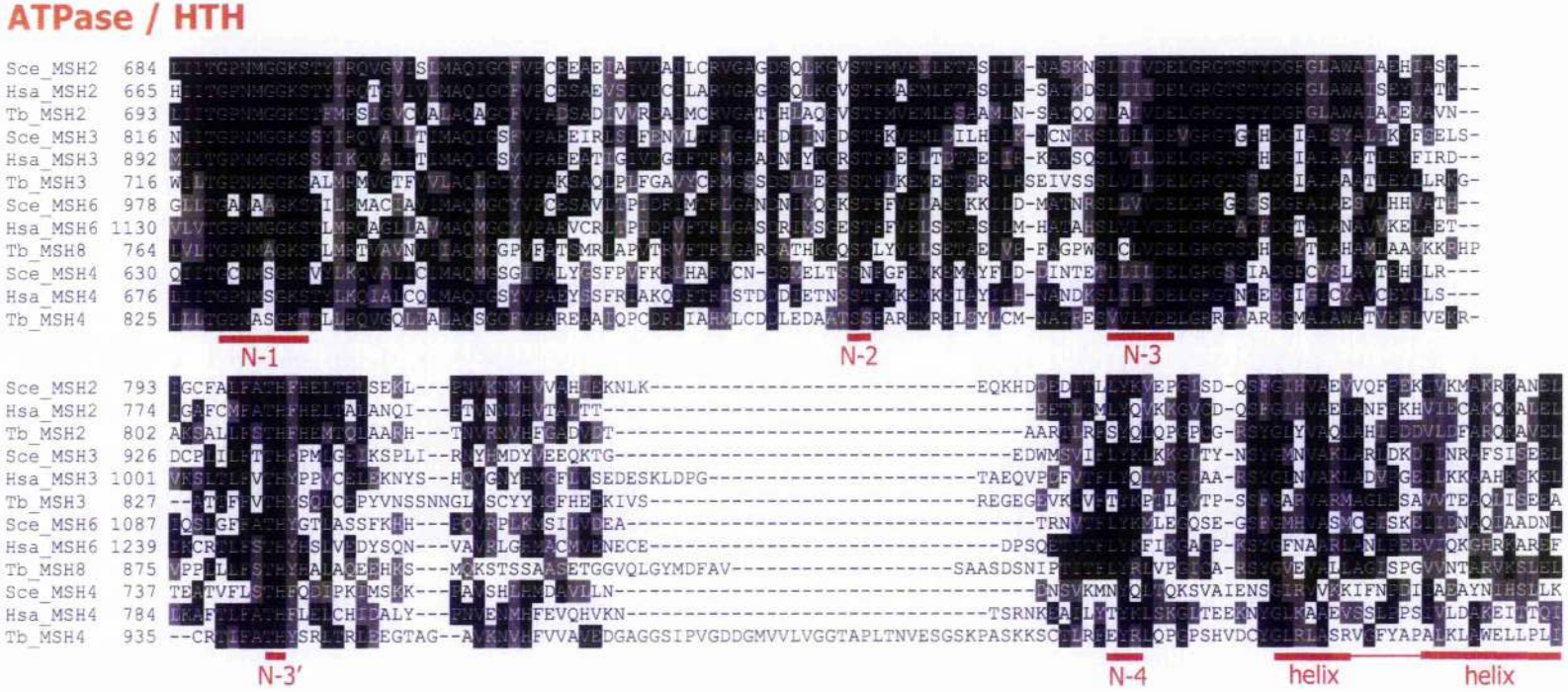


Figure 4.7. Multiple alignment of *T. brucei* MSH4 polypeptides with a range of MutS homologues. Multiple sequence alignment of *T. brucei* MSH4 polypeptide with MSH2, MSH3 and MSH4 homologues from *S. cerevisiae* (Sc), *H. sapiens* (Hsa) and *T. brucei* (Tb). Sequences were aligned using ClustalX (Thompson et al 1997) and shaded using the Boxshade server (http://www.ch.embnet.org/software/BOX_form.html): identical residues are shaded in black and conserved residues in grey. Important residues and motifs as described in the text are highlighted by red asterisks and underlining. To obtain the ATPase/helix-turn-helix domain, global multiple alignments were carried out and the relevant section cut out.

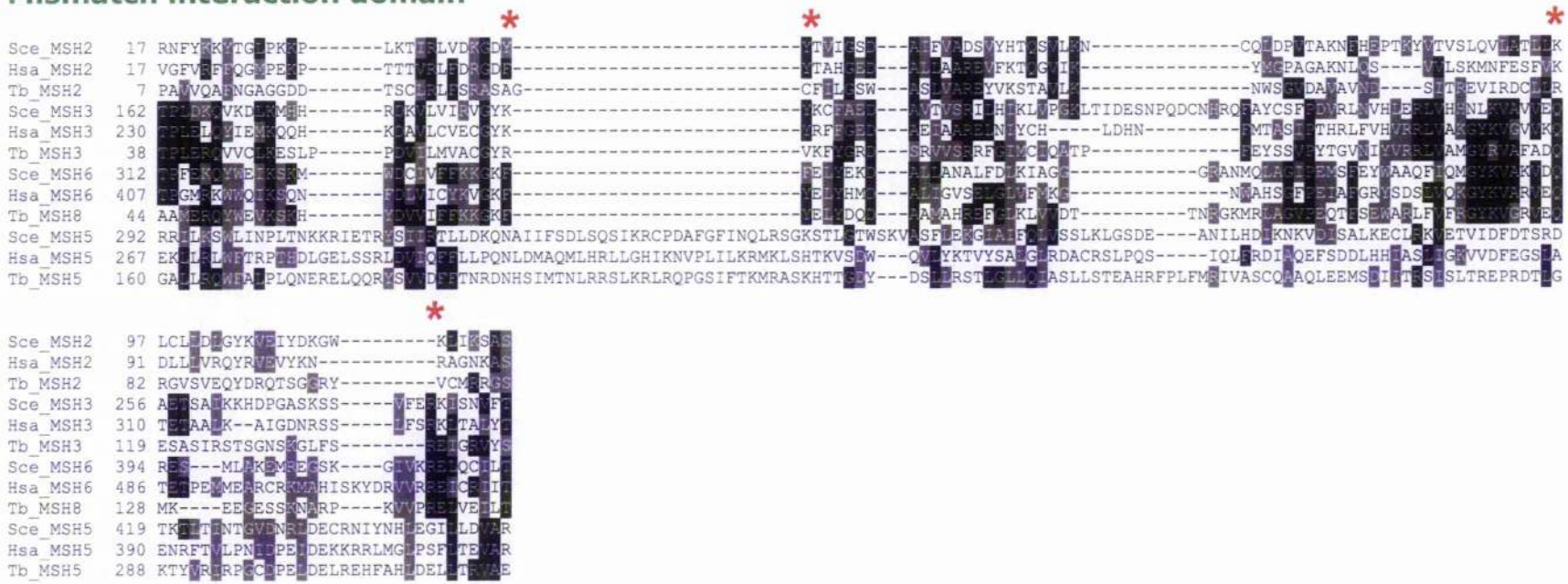


Figure 4.8. Multiple alignment of *T. brucei* MSH5 polypeptides with a range of MutS homologues. Multiple sequence alignment of the *T. brucei* MSH5 polypeptide with MSH2, MSH3 and MSH5 homologues from *S. cerevisiae* (Sc), *H. sapiens* (Hsa) and *T. brucei* (Tb). Sequences were aligned using ClustalX (Thompson et al 1997) and shaded using the Boxshade server (http://www.ch.embnet.org/software/BOX_form.html): identical residues found in at least 30% of sequences are shaded in black and conserved residues in grey. Only the conserved domains defined by Obmolova et al (2000) are shown in this figure; important residues and motifs (as described in the text) are highlighted by red asterisks and underlining. To align the Mismatch Interaction Domain, this region was identified in the MSH2, MSH3 and MSH8 polypeptides and aligned in isolation with the entire protein sequences from the MSH5 homologues.

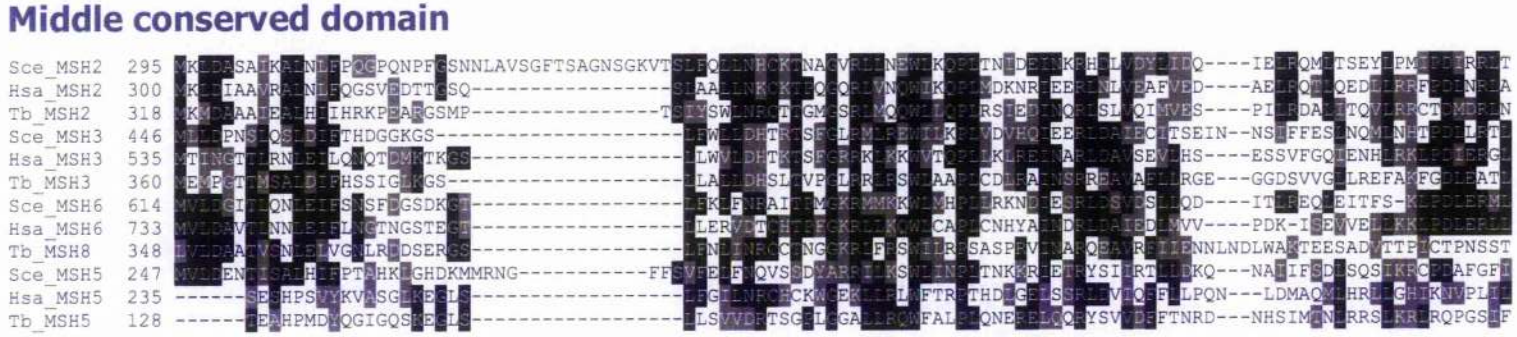


Figure 4.9. Multiple alignment of *T. brucei* MSH5 polypeptides with a range of MutS homologues. Multiple sequence alignment of *T. brucei* MSH5 polypeptides with MSH2, MSH3 and MSH5 homologues from *S. cerevisiae* (Sc), *H. sapiens* (Hsa) and *T. brucei* (Tb). Sequences were aligned using ClustalX (Thompson et al 1997) and shaded using the Boxshade server (http://www.ch.embnet.org/software/BOX_form.html): identical residues found in at least 30% of sequences are shaded in black and conserved residues in grey. To obtain the middle conserved domain, global multiple alignments were carried out.

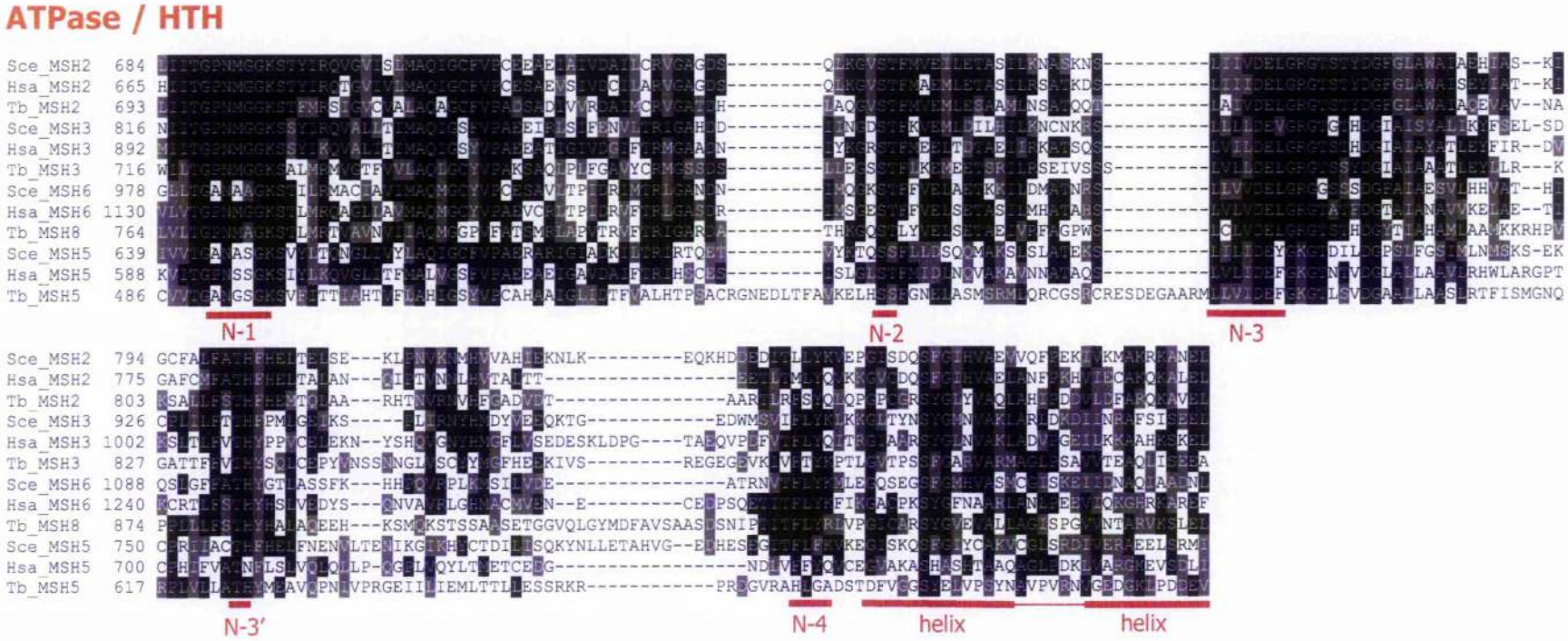


Figure 4.10. Multiple alignment of *T. brucei* MSH5 polypeptides with a range of MutS homologues. Multiple sequence alignment of *T. brucei* MSH5 polypeptides with MSH2, MSH3 and MSH5 homologues from *S. cerevisiae* (Sc), *H. sapiens* (Hsa) and *T. brucei* (Tb). Sequences were aligned using ClustalX (Thompson et al 1997) and shaded using the Boxshade server (http://www.ch.embnet.org/software/BOX_form.html): identical residues conserved in at least 30% of sequences are shaded in black and conserved residues in grey. Important residues and motifs (as described in the text; Obmolova *et al.*, 2000) are highlighted by red asterisks and underlining. To obtain the ATPase/helix-turn-helix domains, global multiple alignments were carried out.

of the important conserved residues can be identified. Beyond this, the overall level of homology is lower and all the MSH4 and MSH5 proteins contain significant sequence insertions relative to the MMR proteins.

In contrast to the lack of conservation of the mismatch interaction domains, the middle conserved domain and the C-terminal ATPase/helix-turn-helix domain are well conserved in all the *T. brucei* MutS family members, including MSH4 and MSH5. The nucleotide-binding motifs within the ATPase domain (N-1, N-2, N-3 and N-3', located at the dimer interface of the folded protein) appear to be intact in each protein: the N-1 and N-3 regions correspond to the widespread Walker A and Walker B motifs, whereas the N-2 and N-3' motifs are unique to the ABC ATPase superfamily to which these proteins belong. The function of the middle conserved domain is not yet known (see section 1.3.2), meaning that crucial residues are not identifiable. Nevertheless, this region of the MSH5 proteins, including *T. brucei*, displays substantial conservation. The *T. brucei* MSH4 appears, however, to be somewhat more diverged than that in humans and yeast. Taken together, the above data appear to confirm that the *T. brucei* genes are functional, but typically lack the N-terminal domains for mismatch interaction.

4.2.3.4 A note on annotation

When putative meiotic genes were being annotated in the trypanosomatids, it was stated that *L. major* lacked an *MSH4* homologue (Ivens *et al.*, 2005). However, during this analysis a gene has been identified that likely encodes MSH4: it groups with the MSH4 sequences in phylogenetic analysis, aligns well with MSH4 proteins from *T. brucei* and *T. cruzi*, and has higher sequence identity scores when aligned with MSH4 than with other MutS homologues. In addition, the gene is orthologous based on genetic synteny (see below) with the three trypanosomatids. Thus, all three sequenced kinetoplastid genomes appear to contain homologues of MSH2, 3 and 6, suggesting conservation of the MMR system, and of MSH4 and 5, and therefore conservation of putative meiotic functions.

4.2.4 Analysis of the genomic environment of MSH4 *and* MSH5

4.2.4.1 *In silico* sequence analysis

The areas of the *T. brucei*, *T. cruzi* and *L. major* genomes surrounding the *MSH4* and *MSH5* ORFs were analysed for the presence of ORFs (although unfortunately, the region upstream of *T. cruzi MSH4* is not yet available; see figures 4.11 and 12). Large regions of

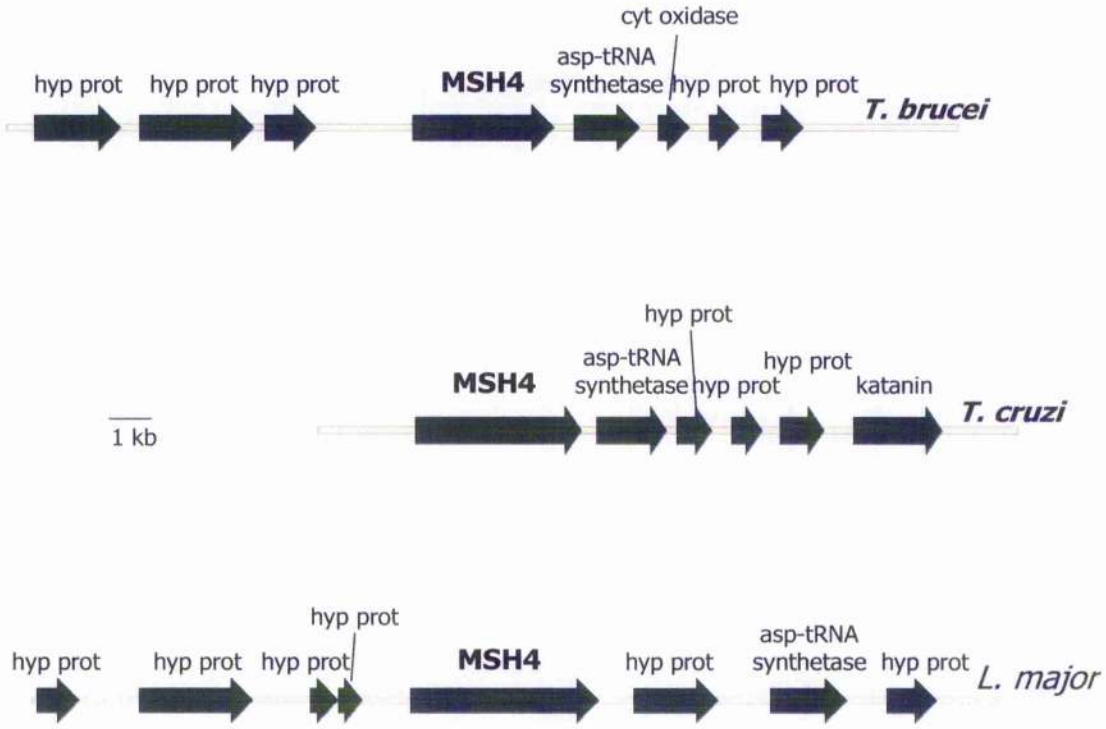


Figure 4.11. The genomic environment of the *MSH4* ORFs in *T. brucei*, *T. cruzi* and *L. major*. The genomic environment of the *MSH4* ORFs in each organism were analysed for the presence of ORFs using Vector NTI, and were then annotated by BLAST searching of the NCBI database. The chromosomes are indicated by yellow horizontal lines, and predicted ORFs by blue arrows. In addition to MSH4 from each organism, the following ORFs are indicated: asparty;-tRNA synthetase (asp-tRNA synthetase), cytochrome oxidase (cyt oxidase), katanin, and several hypothetical proteins (hyp prot).

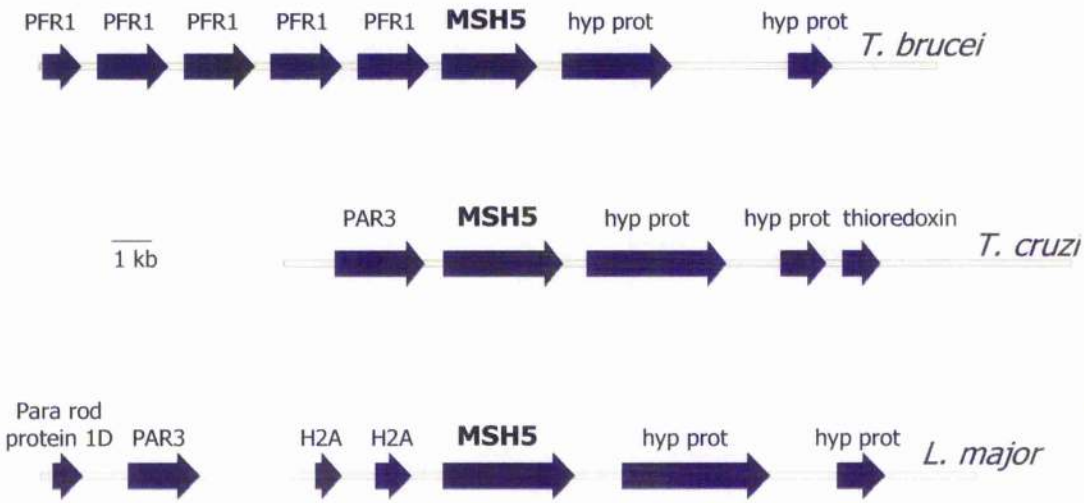


Figure 4.12. The genomic environment of the *MSH5* ORFs in *T. brucei*, *T. cruzi* and *L. major*. The genomic environment of the *MSH5* ORFs in each organism were analysed for the presence of ORFs using Vector NTI, and were then annotated by BLAST searching of the NCBI database. The chromosomes are indicated by yellow horizontal lines, and predicted ORFs by blue arrows. In addition to MSH5 from each organism, the following ORFs are indicated: Paraflegellar rod protein 1 (PFR1), paraflagellar rod protein 3 (PAR3), thioredoxin, paraflagellar rod protein 1D (Para rod protein 1D), histone H2A (H2A) and several hypothetical proteins (hyp prot).

genomic sequence were extracted from geneDB and converted to Vector NTI files for annotation. Putative ORFs surrounding *MSH4* and *MSH5* were identified by using their polypeptide sequences as queries for BLAST searches of the NCBI database. As has been observed for a significant amount of the three trypanosomatid genomes, synteny between the regions surrounding *MSH4* and *MSH5* is apparent. The region upstream of the *MSH4* ORFs, at least in *T. brucei* and *L. major*, contains a number of hypothetical ORFs that might encode proteins of unknown function. An aspartyl-tRNA synthetase gene is found directly downstream of *MSH4* in *T. brucei* and *T. cruzi*, although a hypothetical ORF may be inserted between these genes in *L. major*. Downstream of the aspartyl-tRNA synthetase ORF is an ORF annotated as a putative cytochrome oxidase in *T. brucei*, but as hypothetical ORFs in *T. cruzi* and *L. major*. However, sequence identities calculated from pairwise alignments of the predicted translations of these ORFs (performed using ClustalW, VNTI) were 61.5% between *T. brucei* and *T. cruzi*, 30.5% between *T. brucei* and *L. major*, and 31.5% between *T. cruzi* and *L. major*, indicative of a conserved gene. The genomic region surrounding the *MSH5* ORFs contains more characterised genes. Upstream of the *T. brucei* gene is an array of four identical copies of the paraflagellar rod component *PFR1*, presumably the result of a *T. brucei*-specific gene amplification. Paraflagellar rod proteins are also found upstream of *MSH5* in *T. cruzi* and *L. major*: both these genomes contain a copy of *PAR3*, and in addition, upstream of this in *L. major* is a gene called paraflagellar rod protein 1D (although it should be noted that this gene has no homology to *PFR1*). Downstream of MSH5 are two hypothetical proteins. Sequence identities calculated from pairwise alignments of the predicted translations of the ORF immediately downstream of *MSH5* (performed using ClustalW, VNTI) were 36.2% between *T. brucei* and *T. cruzi*, 24.1% between *T. brucei* and *L. major*, and 26.1% between *T. cruzi* and *L. major*, suggesting a conserved gene. Sequence identities calculated from pairwise alignments of the second ORF (performed using ClustalW, VNTI) were 11% between *T. brucei* and *T. cruzi*, 48.6% between *T. brucei* and *L. major*, and 9% between *T. cruzi* and *L. major*, suggesting a conserved gene only between *T. brucei* and *L. major*.

4.2.4.2 Southern analysis of the *T. brucei MSH4* and *MSH5* loci

Genomic DNA samples from *T. brucei* strains Lister 427 and IITat 1.2 were digested with a panel of restriction enzymes, separated by agarose gel electrophoresis and Southern blotted. A 467 bp region of the *MSH4* ORF and a 427 bp region of the *MSH5* ORF were amplified from Lister 427 genomic DNA using primers MSH4-4 and MSH4-5 or MSH5-4 and MSH5-5, and each PCR product used as a probe in separate Southern blots. In each case, the probes hybridised to a single restriction fragment, confirming that both genes are

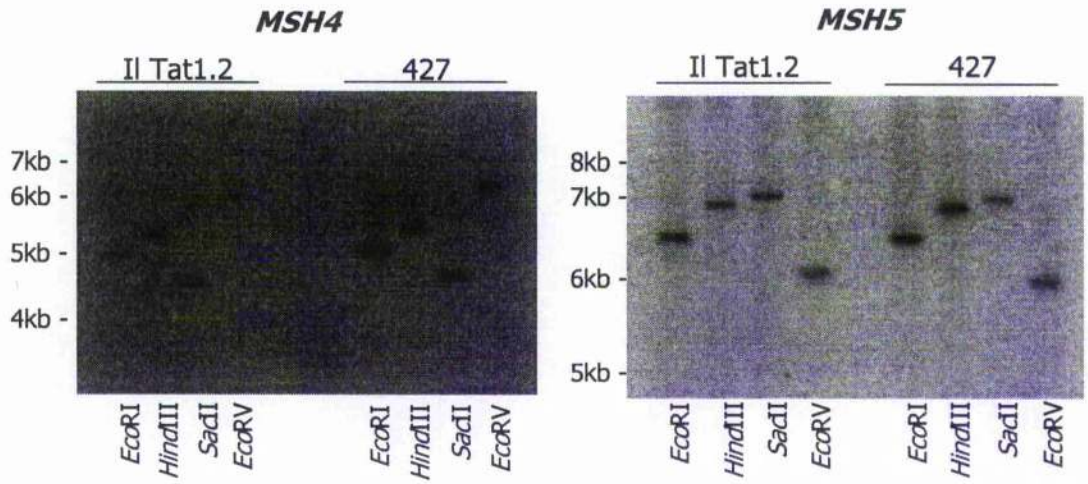


Figure 4.13. Genomic Southern blots probed for *T. brucei MSH4* and *MSH5* genes. Southern blots of Lister 427 and IlTat1.2 genomic DNA was restriction-digested with the enzymes shown, separated on a 0.8% agarose gel, probed with a 467 bp PCR product amplified from the *T. brucei MSH4* gene or a 427 bp product amplified from the *T. brucei MSH5* gene and washed to 0.2x SSC, 0.1% SDS at 65°C.

present in a single copy in the strains tested, as predicted by the genome project (see figure 4.13). The identical restriction patterns in Lister 427 and IlTat 1.2 strains indicates that the genomic environment of *MSH4* and *MSH5* in these two strains, are highly related.

4.2.5 *Analysis of* MSH4 *and* MSH5 *expression in procyclic form and bloodstream stage* T. brucei

In organisms where Msh4 and Msh5 are known to have a role in the control of meiotic recombination, the expression of Msh4 and to a lesser degree Msh5 is limited to meiotic tissues or life cycle stages (see section 4.1.2.3). This is in contrast to other MutS homologues, which, although they can have meiotic functions, have roles in MMR and are therefore expressed ubiquitously. It is already known that this generalised expression of MMR genes appears to be conserved in *T. brucei*, as *MSH2* and *MSH8* are expressed in both procyclic form and bloodstream stage cells (J. Bell, PhD thesis, University of Glasgow). The experiments in this section aim to ascertain whether expression of *T. brucei MSH4* and *MSH5* mirrors this generalised expression, or if it is limited to particular life cycle stages. Genetic exchange in *T. brucei* most likely takes place in the epimastigote life cycle stage (see section 4.1.1), implying that this is where meiosis-specific genes might be predicted to be expressed. Unfortunately, it is difficult to acquire sufficient numbers of such life cycle stage cells for analysis. Expression of *MSH4* and *MSH5* was examined in the culturable procyclic and bloodstream stages: expression in these, presumably non-meiotic, life-cycle stages might suggest that the proteins are involved in general MMR and not in the same specific, meiosis-specific way as their homologues in other organisms.

4.2.5.1 Analysis of *MSH4* and *MSH5* expression by RT-PCR

The expression of MSH4 and MSH5 was first analysed, in a non-quantitative manner, by reverse transcriptase PCR (RT-PCR). Total RNA was isolated from Lister 427 procyclic form, EATRO 795 procyclic form and Lister 427 bloodstream stage cells that had been grown to concentrations of approximately 5 x 10^6 (procyclic form) and 1.5 x 10^6 (bloodstream stage) cells.ml^{-1}. The total RNA was converted to cDNA, which was then used as the template for RT-PCR using primers specific to the *MSH4* and *MSH5* ORFs (MSH4-4 and MSH4-5, and MSH5-4 and MSH5-5, see appendix). A single product of the expected size (467 bp for *MSH4* and 427 bp for *MSH5*) was seen in each reaction, showing that transcript from both genes is present in all three RNA samples (see figure 4.14).

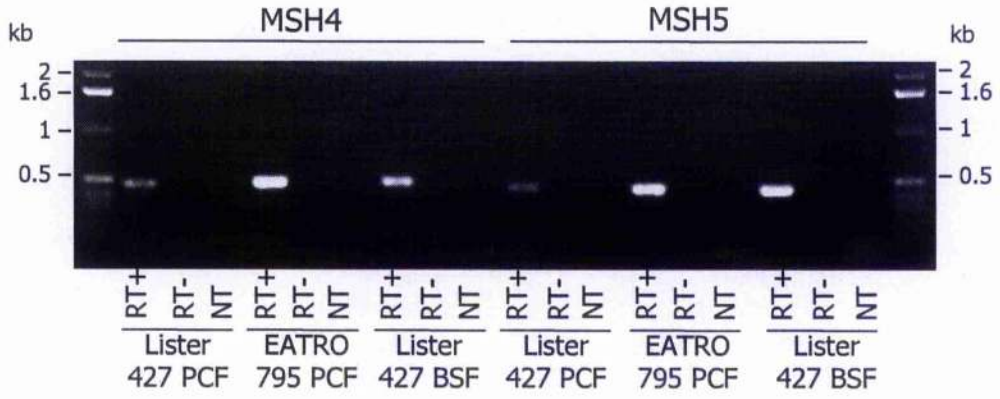


Figure 4.14. Analysis of the expression of *T. brucei MSH4* and *MSH5* by RT-PCR. RT-PCR was performed on RNA isolated from Lister 427 procyclic (PCF), EATRO 795 procyclic and Lister 427 bloodstream form (BSF) cells, using primers complementary to the *MSH4* and *MSH5* ORFs. For each RNA and primer pair, RT positive (RT+), RT negative (RT-) and no template (NT) reactions were carried out as described in the text.

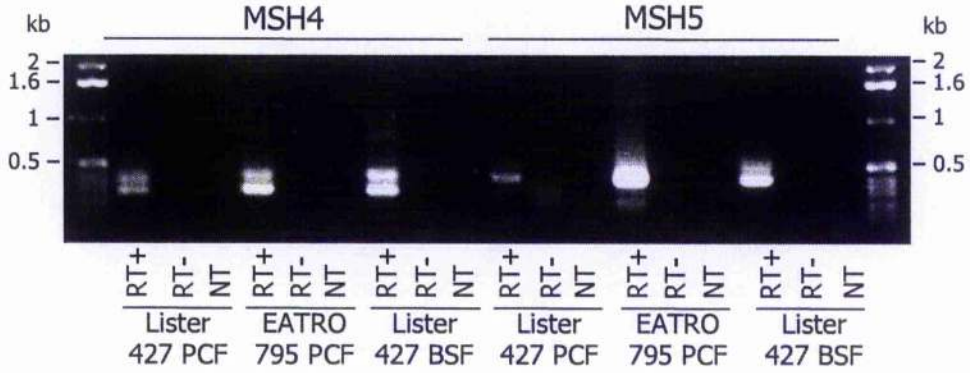


Figure 4.15. Analysis of the expression of *T. brucei MSH4* and *MSH5* by semi-nested RT-PCR. Two rounds of RT-PCR were performed on RNA isolated from Lister 427 procyclic (PCF), EATRO 795 procyclic and Lister 427 bloodstream form (BSF) cells. All RT-PCR reactions used the same 5' primer complementary to the spliced leader sequence, but nested 3' primers complementary to the ORF. For each reaction, RT positive (RT+), RT negative (RT-) and no template (NT) reactions were carried out as described in the text.

Negative control reactions were also performed. Firstly, reactions were carried out by generating cDNA using the same total RNA samples as above, but without addition of reverse transcriptase at the appropriate stage. RT-PCR reactions using this substrate did not yield any product detectable on an ethidium bromide-stained agarose gel, showing that the PCR products do not arise from contamination by genomic DNA in the cDNA samples. Secondly, RT-PCR reactions were carried using water instead of cDNA as the PCR substrate. Again, no product was detectable from these reactions, showing a lack of contamination in the oligonucleotide primers and other reagents used in the reactions.

The above analysis cannot determine if mature, capped and polyadenylated mRNA is generated for *MSH4* and *MSH5*. Transcription in the kinetoplastids is unusual among eukaryotes in that genes are arranged in large directional gene clusters, and primary RNA transcripts contain coding regions for more than one mRNA (Imboden *et al.*, 1987; Tschudi and Ullu, 1988). The polycistronic transcript is divided into individual transcripts by *trans*-splicing, leading to the addition of a conserved 39-nucleotide splice leader RNA sequence to the 5' end of each processed transcript. This splice leader sequence has the function of adding a cap to the mRNA; *trans*-splicing is necessary also for polyadenylation (Ullu *et al.*, 1993). Therefore, RT-PCR reactions that link the splice leader sequence to a sequence specific to a particular transcript should show the presence of mature, processed mRNA. *Trans*-splicing in *T. brucei* is thought to occur most frequently at the first AG dinucleotide following a polypyrimidine tract of 8 to 25 nucleotides upstream of the start codon, giving 5' UTRs of a median length of 68 nucleotides (Benz *et al.*, 2005). Based on sequence analysis, appropriate signals for the splice site could be identified 110 bp upstream of the *MSH4* start codon, and either 9 bp or 121 bp upstream of the predicted *MSH5* start codon.

To assess if these predictions are correct, the presence of mature MSH4 and MSH5 mRNA was measured by semi-nested RT-PCR, performed on RNA isolated from Lister 427 procyclic, EATRO 795 procyclic and Lister 427 bloodstream stage cells. A first RT-PCR reaction was performed using one primer (SL, see appendix 1) complementary to the splice leader sequence and one complementary to a sequence in the 5' region of the *MSH4* and *MSH5* ORFs (MSH4-3 and MSH5-3, see appendix 1). The product from the first reaction was then used, undiluted, as the template for a second round of RT-PCR, using the same 3' primer, but 5' primers closer to the start of each ORF (MSH4-2 and MSH5-2, see appendix); PCR products were separated on a 0.8% agarose gel, stained by ethidium bromide, and visualised using UV light (see figure 4.15). The expected sizes of PCR products for these reactions were therefore 391 bp for *MSH4*, and either 392 bp or 504 bp

for *MSH5*. The *MSH4* reactions all yielded two products visible on the ethidium bromide-stained agarose gel, of approximately 400 bp and 200 bp. PCR products were incubated with Taq polymerase and cloned into the TOPO TA vector, and cloned PCR products were sequenced by the MBSU, University of Glasgow. No full-length sequences definitively linking *MSH4* sequence to the splice leader sequence were obtained; however, *MSH4* sequence was detected from some of the cloned PCR products from each *T. brucei* strain. From other PCR products, sequence was obtained was shown by BLAST searching of the NCBI database to be homologous to part of *T. brucei* chromosome 7, and unannotated sequence from other organisms. These results suggest that the larger product visible on the ethidium bromide-stained agarose gel is derived from *MSH4* cDNA (the size of this product is consistent with the predicted size of the PCR product) and the smaller one corresponds to the unknown sequence. The *MSH5* reaction from Lister 427 procyclic form RNA yielded a single product of approximately 450 bp, from EATRO 795 procyclic form RNA yielded two products of approximately 200 bp and 450 bp, and Lister 427 bloodstream stage RNA yielded two products of approximately 450 bp and 500 bp. These PCR products were also incubated with Taq polymerase, but cloning into the TOPO TA vector was unsuccessful; however, it is probable that the 450 bp PCR product was amplified from *MSH5* cDNA. The same negative controls were performed as for RT-PCR using gene-internal primers, and again showed a lack of contamination from genomic DNA.

4.2.5.2 Analysis of *MSH4* and *MSH5* expression by northern blot

The putative expression of MSH4 and MSH5 in procyclic form of *T. brucei* strains STIB 247, EATRO 795 and TREU 927, and in bloodstream stage of *T. brucei* from the Lister 427 strain, was analysed by northern blot to see if RNA was detectable (see figure 4.17). The STIB 247 and TREU 927 strains were studied because they are known to be competent for genetic exchange, and therefore capable of meiotic recombination (Tait et al., 1993). EATRO 795 and Lister 427 are the laboratory-adapted strains used by our group and their maintenance in culture is likely to have precluded their capacity for tsetse transmission, which may have consequences for meiotic recombination and, indeed, for *MSH4* and *MSH5* expression. Total RNA was extracted from the *T. brucei* cells at approximately 5 x 10^6 (procyclic form) and 1.5 x 10^6 (bloodstream stage) cells.ml^{-1}, and set amounts were run on a denaturing formaldehyde gel and blotted. A 467 bp region of the *MSH4* ORF and a 427 bp region of the *MSH5* ORF were amplified from Lister 427 genomic DNA using primers MSH4-4 and MSH4-5 or MSH5-4 and MSH5-5, and each PCR product used as a probe in separate northern blots. An RNA species of a

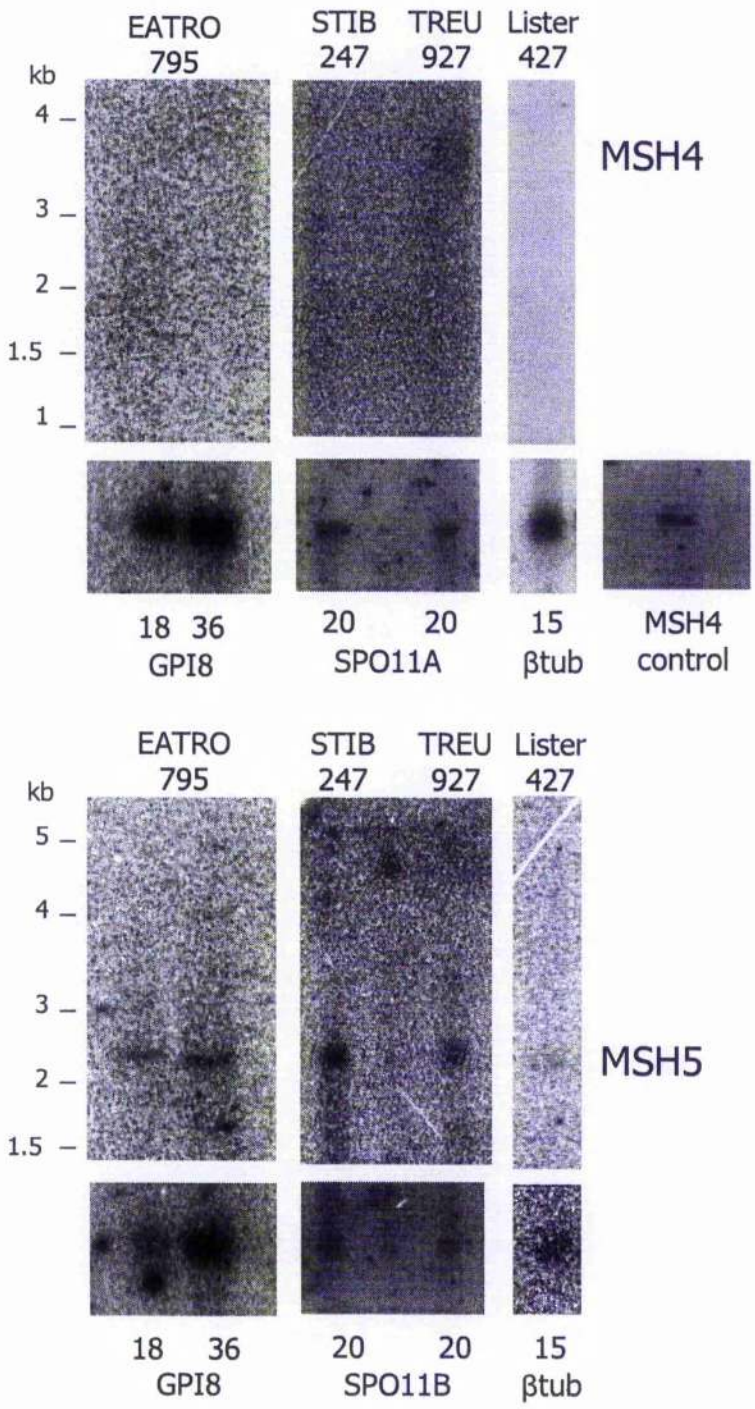


Figure 4.16. Northern analysis of *T. brucei MSH4* and *MSH5* expression in procyclic form and bloodstream stage cells. Total RNA was extracted from EATRO 795, STIB 247 and TREU 927 procyclic form and Lister 427 bloodstream stage cells, separated by agarose gel electrophoresis and northern blotted and hybridised with MSH4 or MSH5 ORF-specific probe (upper panels). Blots were stripped and re-hybridised with probes specific to different ORFs: GPI8 for the EATRO 795 blots, β tubulin (βtub) for Lister 427 blots, SPO11A for 247/927 MSH4, SPO11B for 247/927 MSH5. The amount of RNA run in each lane (μg) is shown at the bottom of each lane. Control Southern blots, with digested genomic DNA probed with the same MSH4 ORF-specific probe, are shown to the left.

size potentially consistent with mature mRNA (~2.5 kb) was detected for *MSH5* in all cases, whereas no clear signal was detected for *MSH4*, with the possible exception of the TREU 927 sample. The blots were then stripped of labelled probe fragment, and re-probed with PCR products derived from other ORFs to control for RNA loading (probes derived from *T. brucei GPI8*, *SPO11A* and *SPO11B* were gifts from Christopher Proudfoot, whereas a probe was derived from the *T. brucei β tubulin* gene was made by PCR amplification of a central region of the ORF using the primers βtubulin5'-3' and βtubulin3'-5' (see appendix 1), giving a PCR product of 631 bp). In all cases, signals were detected, indicating the integrity of the RNA. As a further positive control, small amounts of the PCR product used to make the *MSH4* probe was separated on an agarose gel, blotted and probed in the same tube as the northern blots. Here again, a specific signal was detected, showing that the *MSH4* probe fragment was not defective. These experiments indicate that *MSH5* mRNA is expressed to levels detectable by northern blot in both bloodstream and procyclic life cycle stages. In contrast, *MSH4* appears not to be detectably expressed in at least 3 of the 4 *T. brucei* strains examined.

4.2.6 *An attempt to generate* MSH4 *and* MSH5 *knockout bloodstream stage* T. brucei

MSH4 and *MSH5* genes have been mutated in a number of eukaryotic organisms (see section 4.1.2.3). In these cases, all phenotypes observed were linked to the regulation of meiotic recombination; deficiencies in MMR were never seen (Ross-Macdonald and Roeder, 1994; Hollingsworth *et al.*, 1995; Zalevsky *et al.*, 1999; Edelmann *et al.*, 1999; de Vries *et al.*, 1999a; Kneitz *et al.*, 2000). Given the precedent for the generation of such mutants, this section describes attempts to generate gene "knockout" MSH4 and MSH5 cell lines in *T. brucei* bloodstream stage cells. Although this will not allow a study to be made of meiotic recombination (since this is not thought to take place in these cells and this strain is monomorphic and incapable of transmission to epimastigotes), such an approach would allow the elimination of the possibility that these proteins have functions in general MMR.

4.2.6.1 Design and generation of *MSH4* and *MSH5* knockout constructs

Constructs were designed to completely eliminate the *MSH4* and *MSH5* ORFs from the genome (see figure 17) following homologous integration. To do this, targeting flanks were cloned, corresponding to sequence immediately upstream and downstream of the

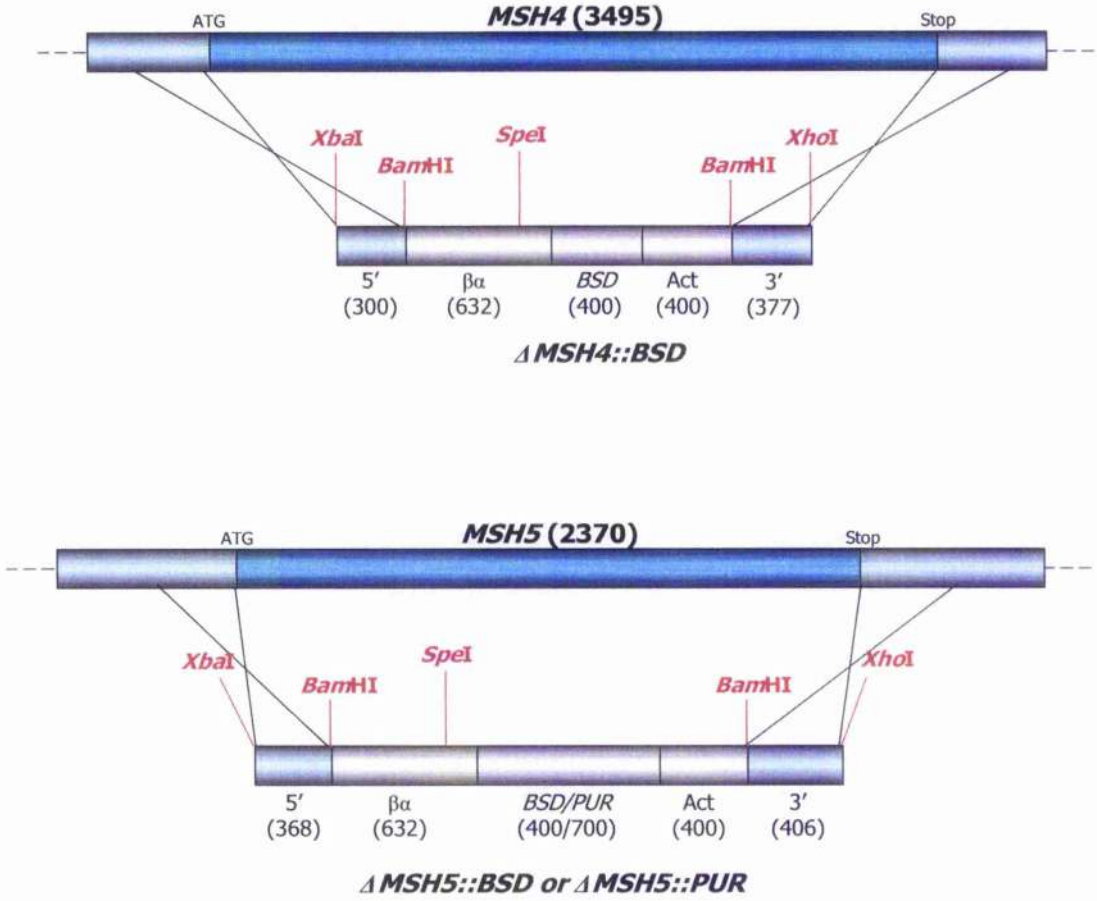


Figure 4.17. A strategy for the generation of *T.brucei MSH4* and *MSH5* knockout mutants. Two constructs were made to attempt to delete the *MSH4* or *MSH5* ORFs (dark blue; ORF length in bp indicated) and replace them with either the blasticidin S deaminase (*BSD*) or puromycin N acetyltransferase (*PUR*) gene. To allow processing into mature mRNAs, both antibiotic resistance genes are flanked, upstream, by a sequence derived the β and α tubulin intergenic region (βα) and, downstream, by a sequence derived from the actin intergenic region (Act). Integration of the constructs was intended to occur by homologous recombination (represented by black crosses) using targeting sequences (light blue) derived from genomic sequence just overlapping the start and stop codons of the *MSH4* or *MSH5* ORF, which are located at each end of the constructs. The gene knockout constructs were linearised by *Xba*I and *Xho*I restriction digests prior to *T. brucei* transformation. Sizes of DNA sequences (in bp) are shown in brackets. Locations of restriction sites used for diagnosic restriction digests are indicated.

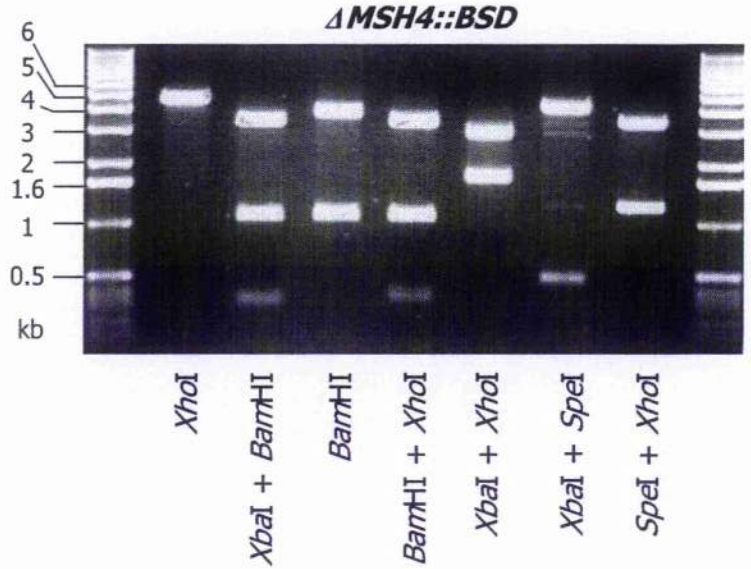
ORFs, were designed to integrate the construct into the genome. Between these flanks were inserted either the blasticidin S deaminase (*BSD*) or puromycin N acetyltransferase (*PUR*) ORF, flanked by processing signals from the tubulin and actin loci to allow production of a stable, mature mRNA.

The targeting flanks were PCR-amplified from Lister 427 genomic DNA. Primers MSH4 5'1 and MSH4 5'2 generated a 300 bp 5' flank for *MSH4*, while MSH4 3'1 and MSH4 3'2 generated a 377 bp 3' flank. For *MSH5*, primers MSH5 5'1 and MSH5 5'2 generated a 368 bp 5' flank, and MSH5 3'1 and MSH5 3'2 generated a 406 bp 3' flank. For each gene, the two targeting flanks were cloned sequentially into the pBluescript II KS cloning vector. The *BSD* and *PUR* ORFs, plus processing signals, were PCR-amplified from the constructs pCP101 and pCP121, respectively, using primers βα 5' *Bam*HI and Act 3' *Bam*HI (see appendix 1) and giving products of 966 bp or 1279 bp. They were then cloned into *Bam*HI restriction sites separating the *MSH4* or *MSH5* flanks to generate the *ΔMSH4::BSD, ΔMSH5::BSD and ΔMSH5::PUR* constructs. Diagnostic restriction digests were carried out, and in all cases gave restriction fragments of the expected sizes (see figures 4.18 and 4.19), showing that organisation of the final constructs were as expected. In addition, all three constructs were sequenced by the MBSU, University of Glasgow and shown to be correct. For reasons discussed below, the *ΔMSH4::PUR* construct was never completed.

4.2.6.2 An attempt to generate *MSH4* and *MSH5* mutant cell lines in *T. brucei*

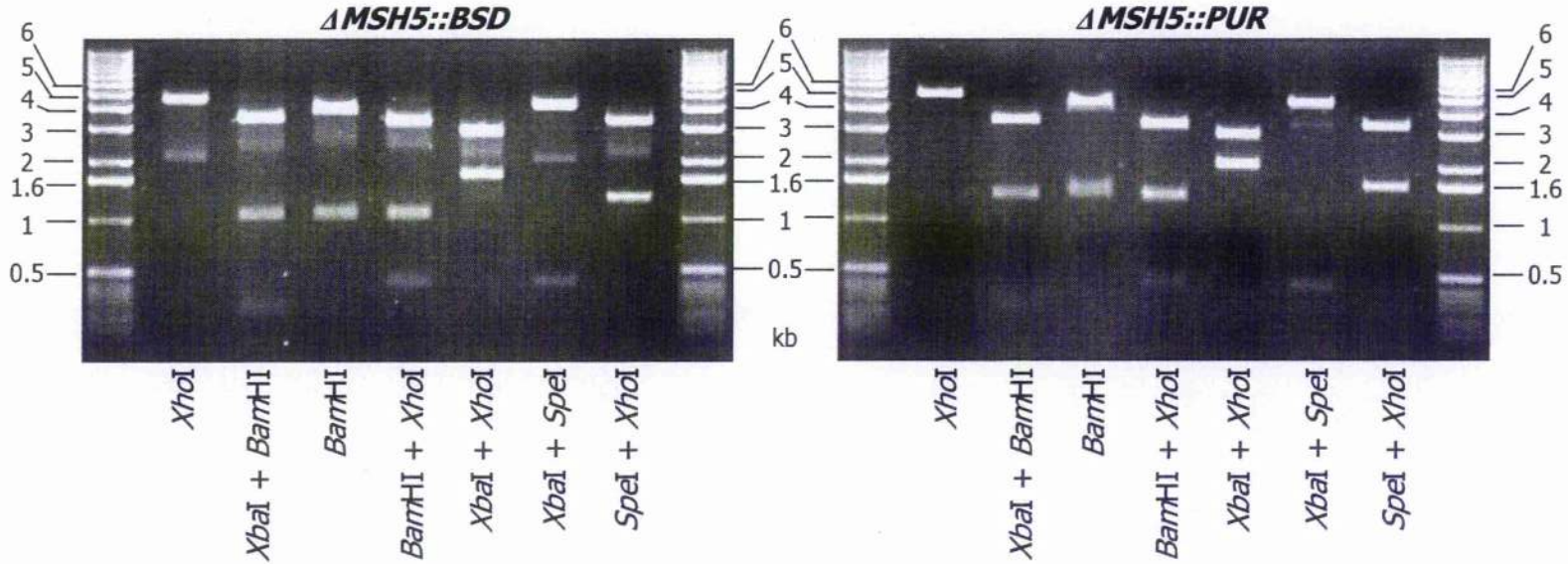
Numerous attempts were made to generate *MSH4* and *MSH5* mutants in Lister 427 bloodstream form cells by transformation by the constructs described in the previous section. The knockout constructs were separated from the pBluescript II KS plasmid by double digestion with *Xba*I and *Xho*I (see figure 4.17), purified (see section 2.1.3.2) and electroporated using standard conditions (see section 2.1.3.1). A number of transformations were carried out for each construct, and selection for transformants, using puromycin dihydrochloride or blasticidin S hydrochloride, was performed at a range of drug concentrations (see table 4.6).

In only one transformation, using *ΔMSH5::BSD* and 5 $\mu g.ml^{-1}$ BSD selection were antibiotic resistant clones recovered. These few transformant clones were analysed for correct integration of the *ΔMSH5::BSD* knockout construct by Southern blotting (see figure 4.20). Untransformed Lister 427 genomic DNA and genomic DNA samples from the putative transformants were digested with *Stu*I, separated by agarose gel



Δ MSH4::BSD	
Enzyme(s)	**Expected sizes (bp)**
*Xho*I	5100
*Xba*I + *Bam*HI	300 + 1400 + 3400
*Bam*HI	1400 + 3700
*Bam*HI + *Xho*I	400 + 4700
*Xba*I + *Xho*I	2100 + 3000
*Spe*I + *Xba*I	700 + 4400
*Spe*I + *Xho*I	1300 + 3800

Figure 4.18. Analysis of *MSH4* knockout construct by restriction digest. Approx 50ng *ΔMSH4::BSD* plasmid DNA was singly and doubly digested with a range of restriction enzymes as indicated, and fragments separated on a 0.8% agarose gel, stained with ethidium bromide and visualised using UV light (left). Restriction enzymes used and expected sizes of fragments are shown in the table (right).



ΔMSH5::BSD	
Enzyme(s)	**Expected sizes (bp)**
*Xho*I	5200
*Xba*I + *Bam*HI	400 + 1400 + 3400
*Bam*HI	1400 + 3800
*Bam*HI + *Xho*I	400 + 1400 + 3400
*Xba*I + *Xho*I	2200 + 3000
*Spe*I + *Xba*I	800 + 4400
*Spe*I + *Xho*I	1400 + 3800

ΔMSH5::PUR	
Enzyme(s)	**Expected sizes (bp)**
*Xho*I	5500
*Xba*I + *Bam*HI	400 + 1700 + 3400
*Bam*HI	1700 + 3800
*Bam*HI + *Xho*I	400 + 1700 + 3400
*Xba*I + *Xho*I	2500 + 3000
*Spe*I + *Xba*I	800 + 4700
*Spe*I + *Xho*I	1700 + 3800

Figure 4.19. Analysis of *MSH5* knockout constructs by restriction digest. Approx 50ng *ΔMSH5::BSD* or *ΔMSH5::PUR* plasmid DNA was singly and doubly digested with a range of restriction enzymes, as indicated, and fragments separated on a 0.8% agarose gel, stained with ethidium bromide and visualised using UV light (above). Restriction enzymes used and expected sizes of fragments obtained are shown in the table (below).

Construct	Transformations	Drug concentrations	Total transformants	Correct transformants
ΔMSH4::BSD	2	2.5 μg.ml^{-1}	0	0
ΔMSH4::BSD	1	2.5 μg.ml^{-1}	0	0
ΔMSH4::BSD	2	5 μg.ml^{-1}	0	0
ΔMSH4::BSD	6	0.5/1/2/3/4/5 μg.ml^{-1}	0	0
ΔMSH5::BSD	2	2.5 μg.ml^{-1}	16	0
ΔMSH5::BSD	1	5 μg.ml^{-1}	0	0
ΔMSH5::PUR	2	1 μg.ml^{-1}	0	0
ΔMSH5::PUR	2	2 μg.ml^{-1}	0	0

Table 4.4.6. A summary of attempts to delete the *T. brucei MSH4* and *MSH5* ORFs by integration of the construct *ΔMSH4::BSD*, *ΔMSH5::BSD* and *ΔMSH5::PUR* following electroporation. The number of transformations, drug concentrations used, and number of correct transformant clones are shown.

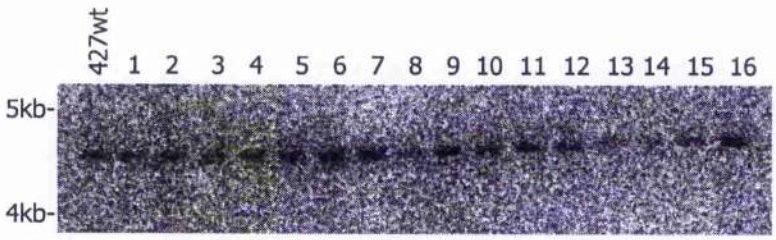


Figure 4.20. Southern analysis of potential *MSH5* heterozygote cell lines. Genomic DNA isolated from blasticidin-resistant clones transformed with *ΔMSH5::BSD* was digested with *Stu*I, separated on a 0.8% agarose gel, Southern blotted and probed with a 368 bp product corresponding to the 5' integration flank from the *ΔMSH5::PUR* construct and washed to 0.2x SSC, 0.1% SDS at 65°C. The blot was visualised after overnight exposure to a phosphoimager screen.

electrophoresis, and Southern blotted. The 300 bp 5' integration flank of the *ΔMSH5::BSD* construct was PCR-amplified from Lister 427 genomic DNA using primers MSH5 5'1 and MSH5 5'2 (see appendix 1), and the PCR product used as a probe for the Southern blot. In all cases, only the 4.5 kb fragment corresponding to the undisturbed wild-type genomic locus is detected, indicating either that these cells were able to grow due to incomplete antibiotic selection, or that the construct had integrated anomalously into a locus other than *MSH5*. At the time that these transformations were being performed, a large number of transformation efficiency experiments were being carried out successfully, showing that the electroporation apparatus is not faulty.

4.2.7 Generation and analysis of trypanosomes containing ectopic copies of MSH4 and MSH5

As it appeared not to be possible to delete either the *MSH4* or the *MSH5* ORF by reverse genetics, an alternative strategy for the disruption of *MSH4* and *MSH5* expression was developed whereby an attempt was made to increase the expression of each gene by integration of each ORF into the tubulin array of wild-type Lister 427 bloodstream form cells, potentially allowing stable expression of each RNA and protein. It is possible that the ectopic expression of either component of the MSH4-MSH5 heterodimer in bloodstream from cells could have a disruptive effect, for instance by inappropriately interacting with the other MutS family members. Alternatively, MSH4 or MSH5 may be able to function in MMR, and overexpression might actually improve the efficiency of the reaction.

4.2.7.1 Design and generation of constructs for ectopic expression of *MSH4* and *MSH5*

Constructs were designed to integrate the *MSH4* and *MSH5* ORFs into the tubulin array (see figure 4.21). These constructs were designed to allow integration by homologous recombination using terminal βα and αβ intergenic sequences, thereby replacing an α tubulin ORF. As the α tubulin gene is multicopy, it is not expected that this should cause an adverse effect on cell growth, a prediction demonstrated experimentally (Bell and McCulloch, 2003). Selection for integration of the *MSH4* or *MSH5* gene utilised an upstream bleomycin resistance gene (*BLE*) ORF. The organisation of the constructs means that processing of *BLE* occurs using the βα tubulin intergenic region (*βα*) upstream and the actin intergenic region (*act*) downstream; processing of *MSH4* or *MSH5* occurs using *act*

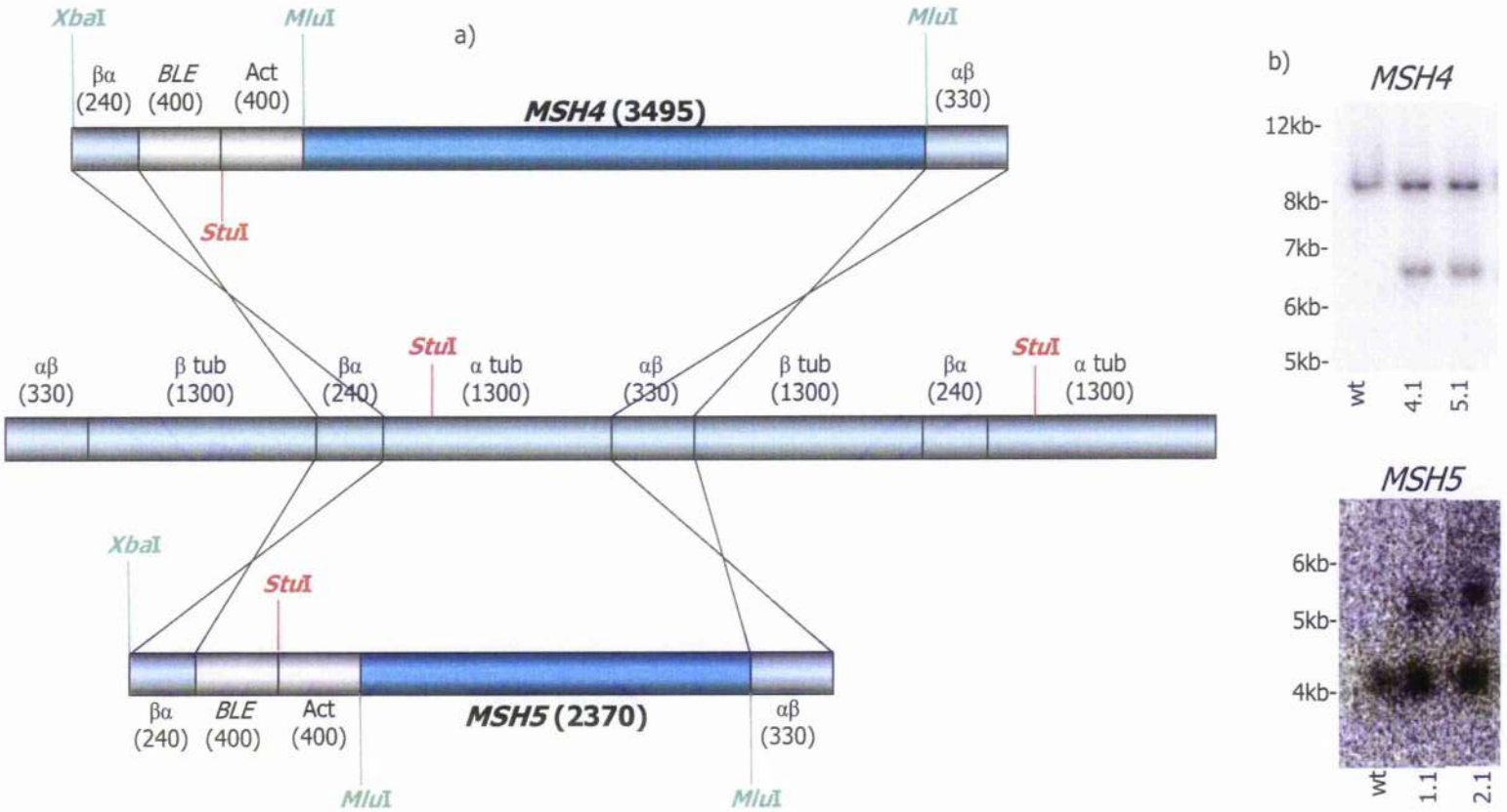


Figure 4.21. Integration of the *MSH4* and *MSH5* ORFs into the tubulin array. a) A strategy for the integration of the *MSH4* and *MSH5* ORFs into the tubulin array. Constructs were generated to integrate the *MSH4* (above) and *MSH5* (below) ORFs, associated with the bleomycin resistance protein gene (*BLE*), into the tubulin array. Integration occurs by recombination (represented by black crosses) using βα and αβ tubulin intergenic sequences. Processing of *BLE* occurs using the βα tubulin intergenic region (βα) upstream and the actin intergenic region (act) downstream; processing of *MSH4* or *MSH5* occurs using act upstream and the αβ tubulin intergenic region (αβ) downstream. Sizes of DNA sequences (in bp) are shown in brackets. Locations of restriction sites used for Southern blot analysis of transformants are shown in pink, and locations of *Mlu*I restriction sites used for cloning in green. b) Southern analysis of transformants containing *MSH4* and *MSH5* integrated into the tubulin array. Genomic DNA was digested with *Stu*I, separated on a 0.8% agarose gel, Southern blotted and probed with a 467 bp product PCR-amplified from the *T. brucei MSH4* gene or a 427 bp product PCR-amplified from the *T. brucei MSH5* gene and washed to 0.2x SSC, 0.1% SDS at 65oC. In each case, untransformed Lister 427 wild-type cells (wt) were analysed along with two independent transformants for each gene: *MSH4 4.1* and *5.1*, and *MSH5 1.1* and *2.1*.

upstream and the αβ tubulin intergenic region (*αβ*) downstream. In both cases, transcription relies on read-through by RNA Pol II rather than introduced promoters.

To make the constructs, the *MSH4* and *MSH5* ORFs were PCR-amplified, using high-fidelity Herculase polymerase, from Lister 427 genomic DNA, using the primers MSH4overexp-5'-Mlu and MSH4overexp-3'-Mlu and MSH5overexp-5'-Mlu and MSH5overexp-3'-Mlu, respectively (see appendix 1). Each product was then cloned into the vector pRM481 (R. McCulloch, unpublished) using the *Mlu*I site between *act* and *αβ* processing sequences. The constructs were confirmed by restriction digestion, but were not sequenced.

4.2.7.2 Generation of cell lines ectopically expressing *MSH4* and *MSH5* in *T. brucei*

The constructs described above were transformed into wild-type Lister 427 bloodstream form *T. brucei*. Both constructs were linearised using *Xba*I, purified (see section 2.1.3.2) and transformed as described in section 2.1.3.1. Clonal transformants were selected in 24-well plates, in HMI-9 containing 2.5 $\mu g.ml^{-1}$ phleomycin. To determine if the constructs had integrated as expected, genomic DNA from phleomycin-resistant transformants was digested with *Stu*I, separated by agarose gel electrophoresis and Southern blotted. A 467 bp region of the *MSH4* ORF and a 427 bp region of the *MSH5* ORF were amplified from Lister 427 genomic DNA using primers MSH4-4 and MSH4-5 or MSH5-4 and MSH5-5, and each PCR product used as a probe in separate Southern blots.

Figure 4.21 b shows the results of Southern analysis for untransformed Lister 427 trypanosomes, as well as two independent transformants of the pRM481-*MSH4* construct (cell lines *MSH4* 4.1 and 5.1) and of the pRM481-*MSH5* construct (cell lines *MSH5* 1.1 and 2.1). In untransformed cells, an 11 kb restriction fragment was detected, corresponding to the wild-type *MSH4* locus. This fragment remains undisrupted in the putative *MSH4* expressor cells, but an additional 7 kb fragment is visible in each clone, corresponding to a copy of *MSH4* integrated into the tubulin array. For the *MSH5* experiment, a 4.5 kb restriction fragment was detected in untransformed cells, corresponding to the wild-type *MSH5* locus. As for *MSH4*, this fragment remains undisrupted following integration of the pRM481-*MSH5* construct, and an additional 5.7 kb fragment, corresponding to a copy of *MSH5* integrated into the tubulin array, arises.

To determine whether or not the ectopically introduced gene copies result in changed *MSH4* and *MSH5* mRNA levels, total RNA was isolated from untransformed Lister 427 cells, and from each of the transformants for *MSH4* and *MSH5* described above. RNA in each case was prepared for cells grown to approximately 1.5 x 10^6 cells.ml^{-1}. 15 μg of RNA was separated on a denaturing formaldehyde gel and northern blotted (see figure 4.22). A 467 bp region of the *MSH4* ORF and a 427 bp region of the *MSH5* ORF were amplified from Lister 427 genomic DNA using primers MSH4-4 and MSH4-5 or MSH5-4 and MSH5-5, and each PCR product used as a probe in separate Southern blots. As a control for the presence of RNA, the blots were stripped of labelled probe fragment and re-probed with a 631 bp PCR product derived from β tubulin using primers βtub 5' and βtub 3' (see appendix 1), showing integrity of the RNA. As a further control, small amounts of the PCR product used to make the *MSH4* probe was run on an agarose gel and blotted. These Southern blots were probed in the same tube as the northern blots, and a signal was detected, showing that the probing process was not defective. This analysis showed an increase in the levels of *MSH5* mRNA in the transformed cells: from the quantity of total RNA as visualised by UV, and intensity of the β tubulin loading control, the amount of *MSH5* transcript appears to be present in greater abundance in the *MSH5* 1.1 and *MSH5* 2.1 cell lines compared with Lister 427 wild-type cells. In contrast, cell lines *MSH4* 4.1 and *MSH4* 5.1 did not produce detectable amounts of *MSH4* transcript, although loading controls and ethidium bromide stain controls showed the samples were intact. Why excision of the *MSH4* ORF from its endogenous 5' and 3' processing flanks, and expression from a known transcribed region of the *T. brucei* genome, failed to produce visible mRNA is unclear.

4.2.7.3 Growth of trypanosomes containing ectopic copies of *MSH4* and *MSH5 in vitro*

The *in vitro* growth rate of the *T. brucei* cells ectopically expressing *MSH5*, as well as the *MSH4* cells, which appeared to show no visibly increased expression, was calculated by diluting the *T. brucei* to a concentration of 2.5 x 10^5 cells.ml^{-1} in HMI-9 and measuring the cell density every 24 hours. This was carried out in duplicate on two separate occasions. None of the transformant cell lines ectopically expressing MSH4 and MSH5 showed any change in growth rate (see figure 4.23); the average population doubling times for untransformed wild-type cells were 8.52 +/- SE 0.057 h, compared with 8.40 +/- SE 0.127 h for cell line *MSH4* 4.1, 8.32 +/- SE 0.072 h for cell line *MSH4* 5.1, 8.36 +/- SE 8.52 h for *MSH5* 1.1, and 8.48 +/- SE 0.97 h for *MSH5* 2.1. It can be concluded that these genomic alterations to *MSH4* and *MSH5* have no noticeably disruptive effect on cell growth.

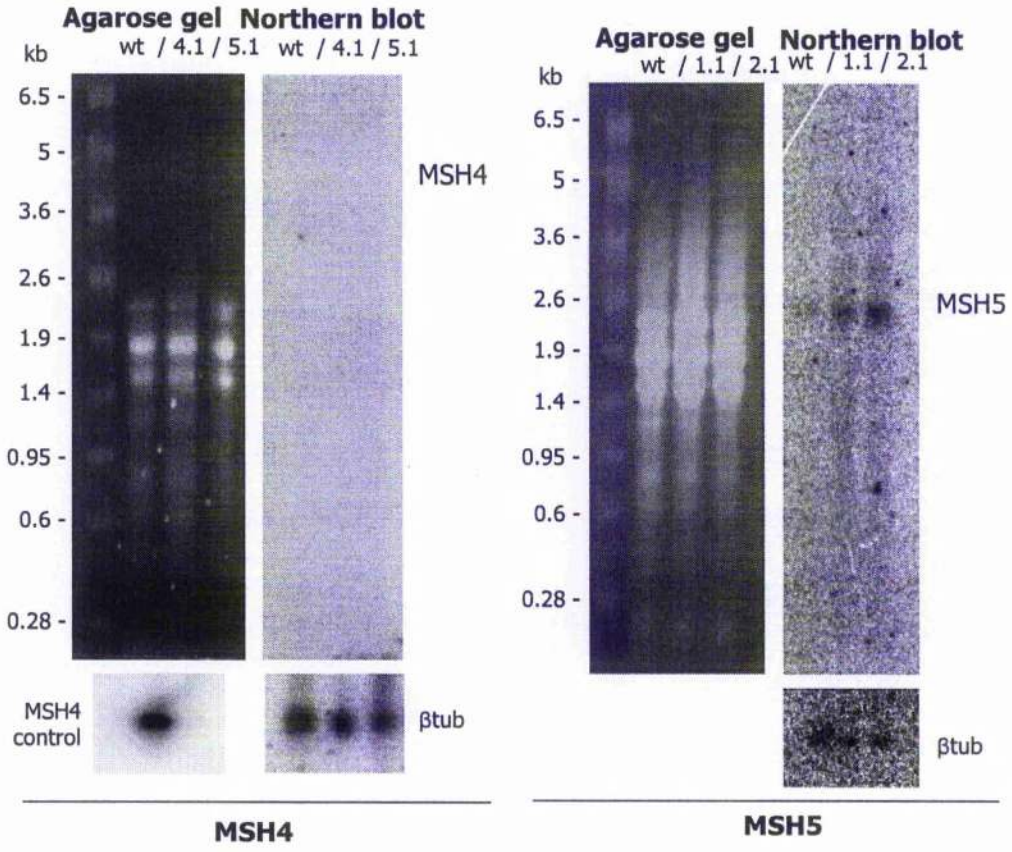
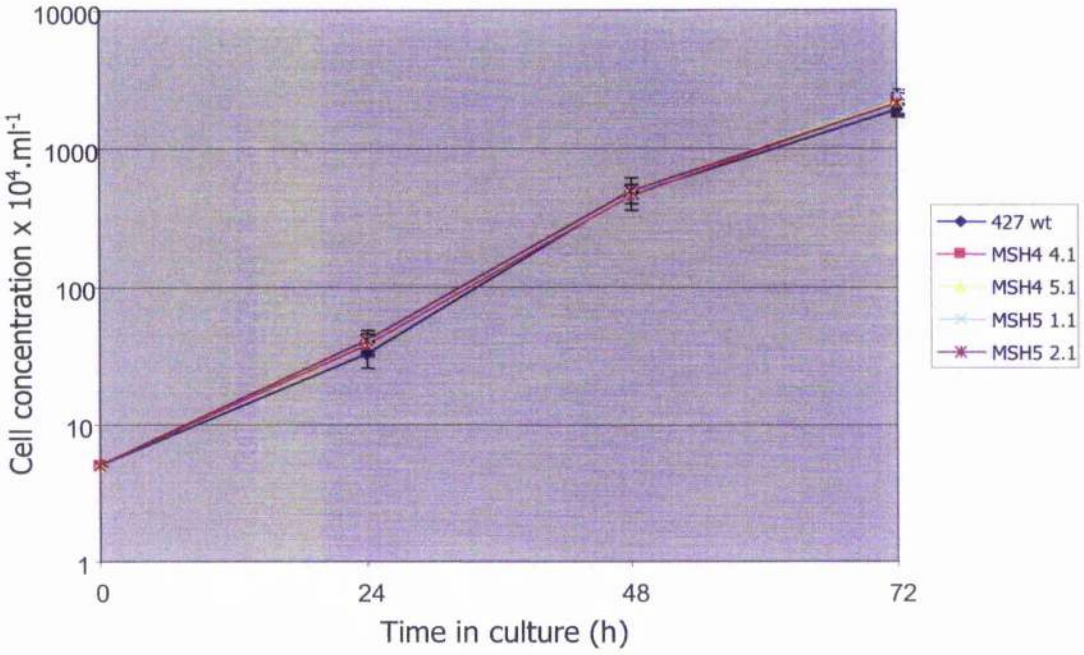


Figure 4.22. Northern analysis of *T. brucei MSH4* and *MSH5* expression in bloodstream form cells containing an ectopic copy of the genes. Total RNA was extracted from Lister 427 wild-type cells, as well as cell lines *MSH4 4.1* and *MSH4 5.1*, and cell lines MSH5 1.1 and MSH5 2.1 containing ectopic copies of *MSH4* and *MSH5*, respectively (see text). 15 µg of each sample was separated by agarose gel electrophoresis, visualised by ethidium bromide staining (shown to the left of each section) and northern blotted. Hybridisation with *MSH4* or *MSH5* ORF-specific probes are indicated, as is re-probing of the stripped blots. Blots were stripped and re-hybridised with a probe specific to the β tubulin ORF (βtub).



	Ave DT (h)	SE (h)
427 wt	8.52	0.057
MSH4 4.1	8.4	0.127
MSH4 4.2	8.32	0.072
MSH5 1.1	8.38	0.153
MSH5 2.1	8.48	0.097

Figure 4.23. Growth in culture of *T. brucei* cells containing an ectopic copy of *MSH4* and *MSH5*. Lister 427 wild-type cells, cell lines *MSH4 4.1* and *MSH4 5.1* and cell lines *MSH5 1.1* and *MSH5 2.1* containing ectopic copies of *MSH4* and *MSH5*, respectively, were grown *in vitro* from a starting density of 5 x 105 cells.ml-1 and cell concentration measured every 24 h. The log of cell concentration is shown against hours in culture.A table showing the average population doubling times (Ave DT) and standard errors (SE) calculated from this analysis is included.

4.2.7.4 Survival of trypanosomes containing ectopic copies of *MSH4* and *MSH5* in the presence of the alkylating agent, *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine (MNNG)

Cells containing ectopic copies of *MSH4* and *MSH5* were assayed for MMR efficiency by measuring their growth in the presence of the alkylating agent *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine (MNNG). This assay, described in more detail below, is based on the fact that cells show increased tolerance to MNNG if the MMR system is defective (described below). If the presence of an ectopic copy of *MSH4* or *MSH5* interferes with normal MMR function leading to an overall decrease in MMR efficiency, then MNNG tolerance should rise; on the other hand, if MMR efficiency improves, MNNG tolerance may be reduced.

N-methyl-*N'*-nitro-*N*-nitrosoguanidine (MNNG) is an alkylating agent that acts on guanine residues, adding O^6-methylguanine (O^6-meG) residues to genomic DNA. Cells with deficiencies in MMR genes are known to have increased resistance to MNNG in a variety of prokaryotic (Jones and Wagner, 1981; Karran and Marinus, 1982; Shi *et al.*, 2004) and eukaryotic (Goldmacher *et al.*, 1986; De Wind *et al.*, 1995; Ciotta *et al.*, 1998; Humbert *et al.*, 1999) organisms, although the exact reasons for this phenomenon (referred to as methylation tolerance) are as yet unclear.

The most widely accepted hypothesis to explain methylation tolerance is called the futile cycle model (see figure 4.24). In eukaryotes, damage caused by MNNG is repaired optimally by O^6-methylguanine DNA methyltransferase (MGMT) (Mitra *et al.*, 1982), which catalyses *in situ* methylation, transferring the methyl group to a cytosine residue within its own active site (Karran and Bignami, 1992). However, if an O^6-meG is present at replication, it is paired aberrantly with thymine residues. This mispairing is recognised by MSH2-MSH6 heterodimers (Levati *et al.*, 1998; Berardini *et al.*, 2000), setting off a futile cycle of repair: the T is excised and the region undergoes repair synthesis (Karran and Bignami, 1992; Griffin *et al.*, 1994; Humbert *et al.*, 1999), with incorporation of another T. If DSBs caused by this repair cycle are present at S phase, the cell will not be able to divide and will die (Griffin *et al.*, 1994). In addition, it has been shown in bacteria (Calmann and Marinus, 2005) and yeast (Durant *et al.*, 1999) that methylated DNA is perceived as homeologous and prevented from recombining by the MMR system; this reduction in HR is also detrimental to the cell's survival (Cejka *et al.*, 2005). Other models have been proposed to explain MNNG tolerance of MMR-deficient cells, however. These include the MMR proteins recognising and binding to the O^6-meG – T mismatches and

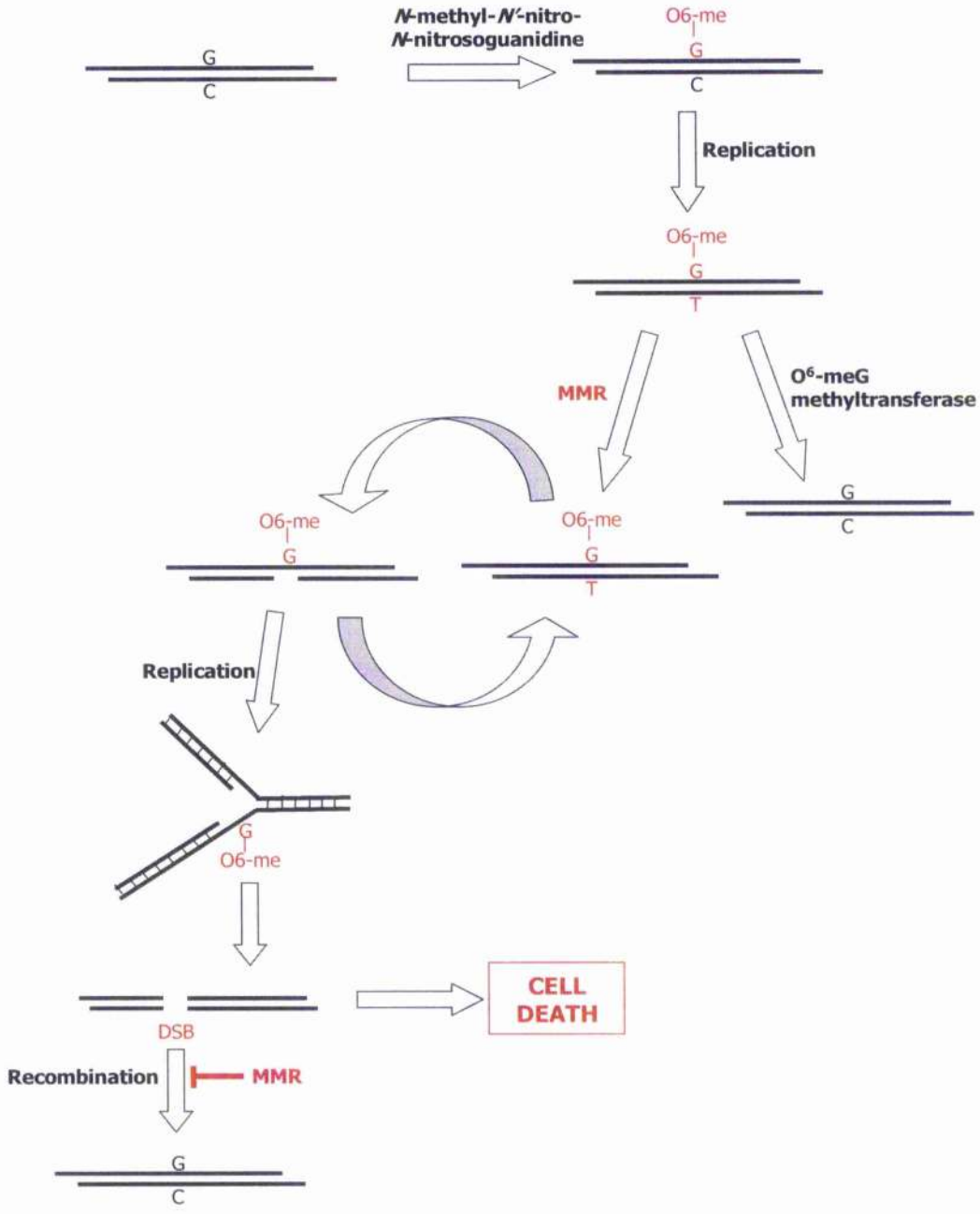
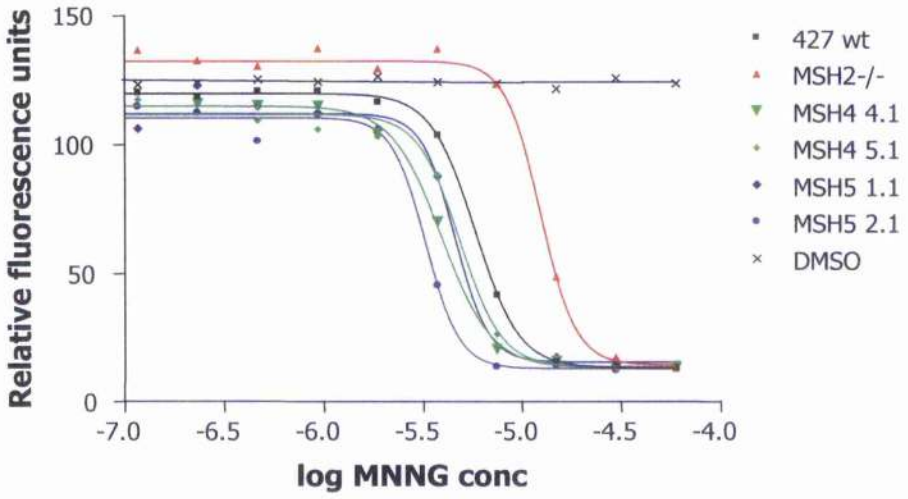


Figure 4.24. The futile cycle model to explain the toxicity of methylating agents. Recognition and excision of thymine (T) mispaired to O^6-methyl guanine (O^6-meG) by the mismatch repair system leads to a futile cycle of repair. Presence of a gap opposite the O^6-meG at replication will lead to a double strand break (DSB) and probable death of the cell. MMR also prevents homologous recombination using substrate containing O6-meG residues, which it perceives as homeologous. Adapted from Aquilina and Bignami, 2001 and Cejka *et al.*, 2005.

thus preventing access of MGMT (Mello *et al.*, 1996), or an MMR-dependent stalling of replication at these perceived mismatches triggering cell cycle arrest (Moreland *et al.*, 1999).

The effects of MNNG on MMR mutants in *T. brucei* have already been studied (Bell *et al.*, 2004), by adding the compound to bloodstream stage Lister 427 wild-type, *MSH2* or *MLH1* mutant cells and measuring their clonal growth. Mutating either *MSH2* or *MLH1* leads to significantly increased tolerance to MNNG, as predicted by observations in other organisms (see above). In this study, an alternative assay, rather than the clonal survival assay, was used to measure population growth in the presence of MNNG. This is based on the metabolism of the compound Resazurin, also known as Alamar Blue, which is a dye that is used as an indicator of cell growth. In the presence of living cells, blue, non-fluorescent resazurin is reduced, probably due to oxygen consumption during metabolism, to the pink, fluorescent compound resofurin, with a direct correlation to the amount of proliferating cells in culture (O'Brien *et al.*, 2000). This system has been widely used for measuring proliferation of bacterial (Collins and Franzblau, 1997; Martin *et al.*, 2005; Coban *et al.*, 2005) and mammalian (Ahmed *et al.*, 1994; de Fries and Mitsuhashi, 1995; Schreer *et al.*, 2005) cells following drug treatment, and has already been employed successfully to measure drug resistance in *T. brucei* (Raz *et al.*, 1997; Onyango *et al.*, 2000; Wallace *et al.*, 2002).

T. brucei bloodstream stage cells containing ectopic copies of *MSH4* and *MSH5* as well as Lister 427 wild-type and *MSH2-/-* cells, at a starting density of 10^5 cells.ml^{-1}, were grown in HMI-9 medium containing doubling dilutions of MNNG from 200 μM to 0.098 μM, in 96-well tissue culture plates (200 μl of cell culture per well). After 48 hours, 20 μl of Alamar Blue was added to each well and fluorescence measured after a further 24 hours of growth. Higher fluorescence is an indication of more metabolism of Alamar Blue and therefore more proliferating cells. Fluorescence was measured by a luminescence spectrometer (LS 55, Perkin Elmer) at an emission wavelength of 590 nm, and plotted on a log-linear graph, forming a sigmoidal curve as seen in the example in figure 4.25. IC50 values (*i.e.*, drug concentrations causing death of 50% of the cells) were then calculated from the fluorescence curve (calculations performed using Prism (GraphPad)). Each experiment was carried out in duplicate three times. MNNG is used as a solution in dimethyl sulfoxide (DMSO), and therefore control, duplicate experiments, identical to the ones described above except that DMSO at equivalent conentrations to those used in the MNNG dilutions, was added to the cell cultures. No significant reduction in cell



	Ave IC50 (μM)	St Error (μM)
427 wt	3.277	1.258
MSH2-/-	9.683	3.066
MSH4 4.1	1.197	0.125
MSH4 5.1	1.097	0.0163
MSH5 1.1	2.65	1.051
MSH5 2.1	2.162	0.712

Figure 4.25. An alamar blue assay measuring the tolerance of cell lines to MNNG when carrying an ectopic copy of *MSH4* or *MSH5*. Lister 427 wild-type cells, Lister 427 *MSH2-/-* cells, cell lines *MSH4 4.1* and *MSH4 5.1* and cell lines *MSH5 1.1* and *MSH5 2.1* carrying ectopic copies of *MSH4* and *MSH5*, respectively were grown in 96-well plates in the presence of doubling dilutions of MNNG for 48 h. Alamar blue was added to each well and the cells were grown for a further 24 h, and fluorescence was measured. A typical graph for this analysis is shown in which fluorescence is plotted against log of MNNG concentration. Fluorescence of a control sample in which DMSO was added to equivalent concentrations to MNNG is indicated. A table showing the average IC50 values (Ave IC50), and standard errors (St Error) calculated is shown for each cell line; the values were calculated for 6 repetitions of the data graphed.

proliferation was observed in this experiment (see figure 4.25), showing that the amounts of DMSO used in this assay had no significant effect on *T. brucei* growth.

Ectopic copies of *MSH4* and *MSH5* did not result in increased MNNG tolerance, as seen for *MSH2-/-* mutants. Indeed, no effect was seen apart perhaps from a very small reduction in sensitivity. Average IC50 values (see figure 4.25) observed were 1.197 +/- SE 0.125 x 10^{-6} μM for cell line *MSH4* 4.1, 1.097 +/- SE 0.163 x 10^{-6} μM for cell line *MSH4* 5.1, 2.650 +/- SE 1.051 x 10^{-6} μM for cell line *MSH5* 1.1, and 2.162 +/- SE 0.712 x 10^{-6} μM for cell line *MSH5* 2.1, compared with an average IC50 of 3.277 +/- SE 1.258 x 10^{-6} μM for wild-type Lister 427 cells, and 9.683 +/- SE 3.066 x 10^{-6} μM for the MMR-deficient *MSH2-/-* cell line. These results indicate that none of the cell lines display impaired MMR, and the possibility that MSH4 or MSH5 marginally improves the efficiency seems remote, as there was no evidence that either *MSH4* ectopic line has generated *MSH4* RNA.

4.3 Discussion

Homology searching of the *T. brucei, T. cruzi* and *L. major* genome databases, followed by confirmatory BLAST searches, has shown that these kinetoplastid parasites appear to have conserved a number of factors specific to meiosis. This chapter demonstrates that this meiotic repertoire includes the MutS homologues, *MSH4* and *MSH5*. These proteins, first discovered in *S. cerevisiae* (Ross-Macdonald and Roeder, 1994; Hollingsworth *et al.*, 1995) but subsequently found to be present in many eukaryotes, are not required for MMR but instead have an important role in the regulation of meiotic recombination. Biochemical and genetic evidence show that they bind to, and stabilise, recombination intermediates (Snowden *et al.*, 2004), promoting the crossover interference pathway for the repair of Spo11-induced (Moens *et al.*, 2002) DSBs. This difference in function compared to other MutS homologues is reflected in a difference in protein structure: MSH4 and MSH5 contain the conserved Middle Conserved and ATPase/Helix-Turn-Helix Domains found in all MutS homologues, but lack the N-terminal Mismatch Interaction Domain (Obmolova *et al.*, 2000). Protein sequence alignments show that this is also the case for MSH4 and MSH5 in *T. brucei*: Middle Conserved and ATPase/Helix-Turn-Helix Domains could be readily identified relative to MSH2, MSH3 and MSH6 (8) proteins for *T. brucei*, human and yeast. In contrast, little clear evidence for a conserved Mismatch Interaction Domain was observed for *T. brucei* MSH4 and MSH5. The apparent conservation of the rest of the two predicted proteins indicates that, if expressed, the *T. brucei* proteins are likely to be

functional. Further evidence for the identification of the proteins as MSH4 and MSH5 was provided by a phylogenetic comparison of a range of MutS-related sequences from a number of eukaryotes, where the parasite genes followed the grouping described in other organisms. This also indicated that MSH4 and MSH5 are present in *T. cruzi* and *L. major*, although detailed sequence comparisons have not been done to verify the functionality of the predicted proteins. Genome sequence comparisons showed that MSH4 and MSH5 are in regions of synteny in the three trypanosomatids, verifying their orthology, and Southern analysis confirmed that both genes are present in the predicted single copy in the *T. brucei* Lister 427 and IlTat 1.2 genomes.

To understand if *T. brucei MSH4* and *MSH5* expression was compatible with a putative role in meiosis, mRNA levels in bloodstream stage and procyclic form *T. brucei* was analysed by RT-PCR and northern analysis. Amplification of sequences from central sections of both *MSH4* and *MSH5* cDNA was possible. In addition, RT-PCR using the conserved splice leader sequence generated products of potentially correct sizes; sequencing of the PCR products revealed some spurious products as well as *MSH4* sequence, but was not possible for *MSH5*. A less sensitive analysis of *MSH4* and *MSH5* expression, allowing an approximate comparison of expression levels (although a quantitative analysis was not undertaken), was carried out by northern blot of total RNA from STIB 247, EATRO 795 and TREU 927 procyclic form cells and from Lister 427 bloodstream stage cells. This showed that whereas *MSH5* mRNA molecules were found to be present in all these samples, *MSH4* was not expressed sufficiently to be detected.

Meiosis is thought to take place in the epimastigote stage of the *T. brucei* life cycle (Bingle *et al.*, 2001), so it might be predicted that it is at this life cycle stage where *MSH4* and *MSH5* would be expressed. The observation that *MSH4* mRNA is not readily detectable by northern blot in (presumably) non-meiotic bloodstream stage and procyclic form *T. brucei*, whereas *MSH5* mRNA can be detected, is compatible with this conclusion. Indeed, this may be comparable to what has been observed in mammalian tissues. Expression of both *MSH4* and *MSH5* is strictly limited to meiotic cells in *S. cerevisiae*. On the other hand, in mammalian tissues, although both transcripts can be detected by RT-PCR in a variety of tissues, expression to a level detectable by northern blot can be seen for *MSH4* only in meiotic cells (*i.e.*, testes and ovary) but for *MSH5* in a number of other tissues as well (Her and Doggett, 1998; Winand *et al.*, 1998; Santucci-Darmanin *et al.*, 1999; Bocker *et al.*, 1999; Her *et al.*, 1999; Santucci-Darmanin *et al.*, 2001). Unfortunately, the epimastigote life cycle stage of *T. brucei* has not been grown in culture, meaning that northern blot analysis of *MSH4* and *MSH5* expression in this putative meiotic stage would

require dissection of a number of tsetse flies. Nevertheless, the fact that both genes' expression can be detected by RT-PCR means that a viable approach to look at putative stage-specific expression would be real-time RT-PCR on the small amounts of material dissected from the tsetse host. If it is correct that *T. brucei MSH4* mRNA is only present in epimastigote cells, the basis for this is unknown. Stage-specific degradation, or stabilisation, of mRNAs is well characterised in *T. brucei* (Clayton, 2002). Nevertheless, it is not possible to exclude that the *MSH4* mRNA expression is not stage-specific, but that the gene expression is constitutively below detection.

The high amounts of post-transcriptional regulation in *T. brucei* (Clayton, 2002) might mean that the quantity of mRNA in a cell does not necessarily correspond to the presence or absence of protein. For instance, it is not known if the *MSH5* mRNA is translated into protein in the life-cycle stages analysed, or if the gene's expression is genuinely non-stage specific. Moreover, it is possible that *MSH4* mRNA is efficiently translated to protein. Therefore, western blots are needed to analyse the expression of MSH4 and MSH5 satisfactorily. Anti-peptide antibodies for this purpose have been generated, but preliminary analysis has not revealed specific interactions with proteins of the correct sizes; further optimisation may be needed before they can be used (data not shown).

To attempt to exclude the possibility that MSH4 and MSH5 proteins are, in fact, present in bloodstream stage *T. brucei* and perform a function, constructs were made to delete the *MSH4* and *MSH5* ORFs through replacement with the blasticidin S deaminase and puromycin N acetyltransferase drug resistance cassettes in Lister 427 bloodstream stage cells. The organisation and sequence of these constructs was verified by diagnostic restriction digests and DNA sequencing, but it was not possible to obtain gene replacement transformants for either gene after a number of attempts, carrying out selection at a range of antibiotic concentrations. Some cells from one of the transformations survived selection, but when they were analysed by Southern blot it became apparent that growth of the cells was not due to integration of the construct into the correct genomic location, and may have been caused by problems with the antibiotic selection. This inability to mutate even one allele of either *MSH4* or *MSH5* is surprising. In other organisms, it has been possible to knock both genes out with no effect on non-meiotic cells (Ross-Macdonald and Roeder, 1994; Hollingsworth *et al.*, 1995; Zalevsky *et al.*, 1999; Edelmann *et al.*, 1999; de Vries *et al.*, 1999b), so it is unlikely that the genes are essential in *T. brucei*. One possibility is that both genes are located in polygenetic units with low endogenous levels of transcription, and so expression of the antibiotic resistance markers, dependent on read-through transcription, does not occur at high enough levels to allow selection at the

concentrations tested. The alternative is that each gene is in a region that is resistant to recombination, perhaps by virtue of condensed chromatin (Polach and Widom, 1995), or that the integration has lethal consequences for surrounding genes. One potential approach to examine these questions would be to induce the absence of *MSH4* and *MSH5* gene products by "knocking down" their expression by RNAi (Bosher and Labouesse, 2000), or by introducing an inducible copy of each ORF into the genome before attempting disruption of the two original copies.

Given the difficulties with mutating *MSH4* and *MSH5*, an alternative strategy for studying their function was adopted, where attempts were made to increase their expression levels by integration of an extra copy of each ORF into the tubulin array of Lister 427 bloodstream stage cells. Although *MSH5* mRNA abundance was shown by to be increased by the presence of this extra copy, *MSH4* mRNA remained at levels too low to be detected by northern blot. Again, the reasons for this are unclear, but may indicate very strong post-transcriptional control of its expression in bloodstream stage cells. Much of the post-transcriptional regulation in *T. brucei* is thought to reside in modulation of mRNA abundance *via* trans-splicing and polyadenylation efficiency, or variation in mRNA degradation rates, through signals in the 5' and 3' untranslated regions (Clayton, 2002). Since the ectopic expression strategy adopted here has replaced these presumptive signals for *MSH4* with intergenic sequences from *actin* and *tubulin*, it might be expected that any control strategy would be overridden. Why, then, *MSH4* mRNA is not generated is unclear. It is possible that the ectopic expression construct has a mutation that has not been noted. Alternatively, this may indicate that the *MSH4* gene contains ORF-specific features that prevent stable mRNA production, although what these might be is unclear. Finally, perhaps *MSH4* mRNA levels are controlled by an ectopic, sequence-specific process, such as RNAi, though there is no evidence at present to support this.

Despite these problems, two independent transformant cell lines for *MSH4* and *MSH5* were assayed for defects in growth and MMR. The growth rate of these cell lines did not differ significantly from that of wild-type bloodstream stage cells, and neither did their sensitivity to MNNG (corresponding to MMR efficiency). These findings appear consistent with united data from other organisms: overexpression of wild-type *MSH5* in *S. cerevisiae* had no effect on MNNG tolerance (Bawa and Xiao, 2003). A gain-of-function mutation in the helix-turn-helix of the yeast MSH5 protein has, however, been identified that causes an increase in MNNG tolerance (Bawa and Xiao, 2003), probably due to aberrant pairing with other MutS homologues interfering with their functions. Ectopic expression of both *MSH4* and *MSH5* together may be more likely to ensure an increase in

the amount of active heterodimer in the cell. Modification of the existing constructs could be used to achieve this, assuming the reason for lack of ectopic MSH4 expression can be understood.

The attempted knockouts and ectopic expression experiments described above were performed in Lister 427 bloodstream stage cells, which are not thought to be able to perform meiosis. Perhaps a more direct approach to understand MSH4 and MSH5 functions would be to alter the genes' expression, by mechanisms described above, in tsetse fly-transmissible, meiotically competent strains such as TREU 927 and STIB 386, and assess the effect of these changes on, for instance, fly transmission and the rate of genetic exchange. *MSH4* and *MSH5* knockout cells would be predicted to be able to be transmitted through the tsetse, as meiosis is non-obligatory in *T. brucei*, but might undergo a reduced rate of genetic exchange, whereas increased expression of these proteins would be unlikely to affect either process.

The starting point for these experiments was an inventory of core meiotic genes as outlined by Villeneuve and Hillers (Villeneuve and Hillers, 2001), which aimed to identify factors that are specifically needed for meiotic, rather than mitotic cell division and recombination. The presence of all or most of these factors has been interpreted as meaning that an organism is at least capable of carrying out meiosis, even if it occurs only rarely *in vivo*. A search for meiotic factors as evidence for a sexual cycle has already been carried out in other species, including the fungus *Candida albicans* (Tzung *et al.*, 2001). *C. albicans* is closely related to *S. cerevisiae*, but no sexual cycle had been identified in this organism at the time of genome sequencing. Analysis of its genome revealed that *C. albicans* possesses homologues of a number of proteins acting at various stages of *S. cerevisiae* meiosis, although some essential *S. cerevisiae* meiotic factors are missing. Molecular genotyping techniques such as multi-locus sequence typing subsequently showed that sexual reproduction makes a rare but significant contribution to the growth of this organism in the wild if not in the laboratory (Tavanti *et al.*, 2004), thus validating this approach as a way of looking for meiotically competent organisms. The genome of the anciently diverged protist parasite, *Giardia intestinalis*, has also been searched for meiotic genes (Ramesh *et al.*, 2005), again demonstrating that a large proportion of meiotic factors are present. Notably, Giardia appears not to encode MSH4 or MSH5, indicating an absence of the crossover interference pathway of meiotic recombination. Nevertheless, the presence of most other "core meiotic factors" in Giardia as well as in other protists including Entamoeba, has led to the suggestion that meiosis arose very early in eukaryotic

evoluation. Unlike Candida, however, no experimental evidence for genetic exchange has been presented for Giardia.

Meiosis-related genes, including MSH4 and MSH5, have been identified in the genomes of the kinetoplastids *T. cruzi*, *L. major* and *T. brucei*. The genomes of all three kinetoplastids contain homologues of a number of genes needed for synaptonemal complex formation, DSB creation, strand exchange and resolution stages of meiosis (see table 4.1). It should be noted that *T. brucei*, *T. cruzi* and *L. major* all lack the MutL homologue MLH3, which is known to also have roles in the regulation of crossover interference (Lipkin *et al.*, 2002); presumably, other MutL homologues are able to substitute for its function in the kinetoplastids. The *T. brucei* homologue of *DMC1* has been mutated in Lister 427 bloodstream stage cells (Proudfoot and McCulloch, 2006). The resulting *DMC1-/-* cells did not suffer from impaired growth or increased MMS sensitivity, indicating that the DNA repair pathways of these cells are intact, and had wild-type rates and profiles of *VSG* switching. These phenotypes are consistent with this *RAD51* homologue having a specific role in meiotic recombination at the epimastigote life cycle stage, and no general functions in recombination in bloodstream stage cells. However, it remains the case that no conclusive evidence links the functions of these putative meiotic genes to genetic exchange, and it theoretically remains possible that they perform other core cellular functions that have been co-opted into meiosis only in "higher" eukaryotes. In this regard, it is not clear that the form of genetic exchange observed in *T. cruzi* conforms to a meiotic process that would utilise these genes.

Understanding the molecular basis for genetic exchange in protists is of importance for two reasons. Firstly, the presence of meiotic genes is consistent with classical genetics data in *T. brucei* (MacLeod *et al.*, 2005a), which is the protist in which the best evidence for meiosis has been documented. Secondly, the conservation of meiotic factors in these anciently diverged eukaryotes (Hedges *et al.*, 2004) is informative with regard to the selective pressures for this process. It has been argued that *G. intestinalis* and the kinetoplastids have adopted the strategy of facultative sex (Dacks and Roger, 1999), whereby these organisms can survive for extended periods without undergoing sexual exchange, but retain the ability to do so when it becomes advantageous. This is an efficient strategy for a single-celled organism like *T. brucei*, as it allows some genetic exchange to take place without the potentially deleterious disruption of obligatory meiotic life cycle stages. Presumably, it implies also that punctuated bouts of meiosis select against those cells that have lost this capacity.

In summary, *T. brucei* has been shown to contain homologues of *MSH4* and *MSH5*, and no evidence indicates that they act in a way that is inconsistent with a role in meiotic recombination. Further work needs to be done, however, if their roles in crossover interference and *T. brucei* transmission through the tsetse fly vector are to be understood.

CHAPTER 5

SEARCHING FOR MUTATOR STRAINS IN *T. BRUCEI*

5 Searching for mutator strains in *T. brucei*

5.1 Introduction

5.1.1 Bacterial mutator strains

The majority of spontaneous mutations arising during replication are neutral, deleterious to the cell's survival or even lethal, disrupting crucial parts of the genome; very few mutations are likely to be beneficial to the cell (Taddei *et al.*, 1997). For this reason, it is thought that it is generally preferable for survival of a cell to keep mutation rates as low as the high metabolic cost of genomic repair allows (Drake *et al.*, 1998). However, it was observed as early as the 1950s (Giraud *et al.*, 2001b) that strains with increased mutation rates (known as mutators) are present in natural populations of a number of bacterial species (Shaver *et al.*, 2002; Denamur and Matic, 2006), including *Escherichia coli* (Denamur *et al.*, 2002), *Salmonella enterica* (LeClerc *et al.*, 1996), and *Pseudomonas aeruginosa* (Oliver *et al.*, 2000), as well as in cultured laboratory populations (Sniegowski *et al.*, 1997; LeClerc *et al.*, 1998; Boe *et al.*, 2000). Any population would be expected to contain some constitutive mutator cells due to spontaneous mutation of genes involved in DNA repair or the fidelity of replication; however, it has been shown both by computer modelling (Tenaillon *et al.*, 2001) and in the laboratory (Shaver *et al.*, 2002) that the observed proportion of mutators is higher than would be predicted by genetic drift. For instance, mutators are found among natural isolates of pathogenic *E. coli* and *S. enterica* at a frequency of approximately 1% (LeClerc *et al.*, 1996), while in *Pseudomonas aeruginosa* chronic infections of cystic fibrosis patients, the proportion can be as high as 53% (Macia *et al.*, 2005). It is common for mutator strains to be defective in an element of the MMR system, most often MutS or MutL (LeClerc *et al.*, 1996; Matic *et al.*, 1997; Oliver *et al.*, 2000; Giraud *et al.*, 2001a; Giraud *et al.*, 2001b; Richardson *et al.*, 2002; Prunier and Leclercq, 2005; Denamur and Matic, 2006). Depending on the MMR factor mutated, and the location of the mutation within the gene or regulatory regions, the mutation rate can be increased up to 1000-fold, by mechanisms described below (Radman *et al.*, 1995; Miller *et al.*, 2002).

Rather than being selected as advantageous in their own right (*e.g.* due to a decreased metabolic load), mutator alleles spread through populations thanks to the increased levels of sequence divergence, and therefore evolution, generated in mutator cells, allowing improved adaptation to hostile environments (Giraud *et al.*, 2001a). The mechanism by

which mutator alleles become fixed in bacterial populations is known as second-order selection. As bacteria are primarily asexual, an advantageous mutation remains linked to its mutator genetic background, which spreads through the population by "hitchhiking" with the beneficial mutations (Sniegowski *et al.*, 1997; Tenaillon *et al.*, 2001; Giraud *et al.*, 2001b; Chat *et al.*, 2006). A commonly found advantageous adaptation that allows mutator alleles to spread through a population is antibiotic resistance (Oliver *et al.*, 2000; Chopra *et al.*, 2003; Macia *et al.*, 2005; Baquero *et al.*, 2005; Denamur *et al.*, 2005; Ciofu *et al.*, 2005). Computer simulations show that if a mutator allele is separated from an associated advantageous mutation, its fixation in the population is prevented, showing that spread of mutator alleles is due to second-order selection (Tenaillon *et al.*, 2000). Fixation of a mutator allele in a population is dependent on the generation of advantageous mutations. If cells within the mutator subpopulation (rather than the wider nonmutator population) generate advantageous mutations that will increase the fitness of mutator cells, the mutator phenotype will become fixed in the population. Therefore, the larger the mutator subpopulation, the more likely it is to reach fixation as it becomes successively larger and accumulates more and more adaptive mutations (Tenaillon *et al.*, 1999; Tenaillon *et al.*, 2001; Giraud *et al.*, 2001a; Tanaka *et al.*, 2003). On the other hand, if nonmutator cells far outnumber the mutators, advantageous mutations will be more likely to arise in this population first despite their lower mutation rate (Giraud *et al.*, 2001a). Successive rounds of strong selection have been shown to lead to very rapid fixation of a mutator allele in an *in vitro* population, with 100% mutator cells after three successive rounds of selection (Mao *et al.*, 1997). While mutator strains with a strong mutator phenotype are more likely to generate advantageous mutations, they are also at higher risk of accumulating deleterious mutations before the allele reaches fixation, leading to a reduction in fitness and therefore a reduction in the number of mutators (Taddei *et al.*, 1997; Tenaillon *et al.*, 1999; Denamur *et al.*, 2002). Therefore, the most successful mutator strains tend to have intermediate mutation rates; in one study of clinical *E. coli* clinical isolates, 23% of strains were weakly hypermutable (rifampicin reversion rates of $4 \times 10^{-8} \leq f < 4 \times 10^{-7}$) whereas 0.7% were strong mutators (Baquero *et al.*, 2004).

As selection for mutator phenotypes is stronger in environments with more stringent selection, mutators are often found in bacterial populations inhabiting stressful environments. A well-documented example of this is the high levels of mutator cells in chronic infections of the bacterium *Pseudomonas aeruginosa* in the lungs of cystic fibrosis patients (Oliver *et al.*, 2000; Macia *et al.*, 2005; Hogardt *et al.*, 2006), which are subject to oxidative stress (Ciofu *et al.*, 2005) as well as repeated drug selection and a compartmentalised environment (Oliver *et al.*, 2000). Similarly, mutator bacteria are at a

selective advantage in the challenging environment of a late-stage urinary infection, and in growth in urine *in vitro*, but not in nutrient-rich LB broth (Labat *et al.*, 2005). The frequency of mutator strains is higher in pathogenic than commensal *E. coli* (Matic *et al.*, 1997). However, a mutator phenotype in itself does not lead to increased virulence: inactivation of MutS does not increase the virulence of *Salmonella typhimurium* in mouse infections (Zahrt *et al.*, 1999), and a mutator strain is in fact at a disadvantage compared to a nonmutator strain with the same virulence determinants (Picard *et al.*, 2001).

Although mutators are at an advantage when faced with novel or stressful environments, this short-term adaptation is associated in the long term with a loss of fitness in other environments (Giraud *et al.*, 2001a; Nilsson *et al.*, 2004). This may be due to a build-up of mutations that are neutral in one environment but detrimental in a second. Alternatively, it may be due to a phenomenon known as antagonistic pleiotropy (Cooper and Lenski, 2000), whereby the advantageous mutations associated with the fixation of the mutator allele are themselves detrimental to the organism's survival in other environments. The fixation of deleterious mutations by genetic drift is known as Muller's ratchet (Felsenstein, 1974), and this can be extremely powerful. For example, after 1000 generations of repeated colony isolations, 4% of *E. coli* mutator lineages had died out and 70% had defects in at least one sugar or catabolic pathway, whereas after the same period, only 3% of nonmutator strains showed alteration in any of the phenotypes tested (Funchain *et al.*, 2000).

Even in bacterial isolates without constitutively elevated mutation rates, in some circumstances the rate of genetic variation can be temporarily increased without genomic alteration of MMR genes (Rosenberg, 2001). This allows a short-term increase in the rate of evolution without the long-term detrimental effects of a constitutive mutator phenotype. The SOS response (Matic *et al.*, 1995; Bridges, 2001; ; He *et al.*, 2006Schlacher *et al.*, 2006) is a stress-triggered cellular response caused by DNA damage or interference in replication, for example in ageing *E. coli* colonies on an agar plate (Taddei *et al.*, 1997). The SOS response is controlled by RecA and by the LexA repressor, activated by binding of RecA to single-stranded DNA. The SOS response causes upregulation of at least 40 genes (Courcelle *et al.*, 2001). These include RecA itself, other factors promoting recombination (allowing an upregulation firstly of DNA repair and secondly of conjugation and transduction, permitting transfer of potentially advantageous genes), and several low-fidelity DNA polymerases, which are capable of bypassing lesions in the genomic DNA template while creating increased genetic variability. The SOS response has the twin advantages of allowing adaptation to increased levels of genomic damage allowing survival in these stressful situations, and of increasing the amount of ectopic

recombination and so potentially allowing adaptation to the stressful environment. Another response causing an increase in the mutation rate in stationary phase bacteria (Harris *et al.*, 1999) is a decrease in the amount of MutS, and to a lesser extent, MutH protein (Feng *et al.*, 1996). This phenomenon is seen in most natural *E. coli* isolates (Bjedov *et al.*, 2003), although it varies between strains: for example, the reduction in MutS levels is 26-fold greater in the pathogenic strain O157:H7 compared to the laboratory strain K-12 (Li *et al.*, 2003). Adaptive mutation has also been reported in a few yeast assay systems, thought to be caused by error-prone polymerases (Heidenreich and Wintersberger, 2001).

Back mutation to restore low mutation rates to constitutive mutators is very unlikely, and if a section of the gene has been deleted, impossible. However, a reduction in MMR efficiency also leads to an increase in recombination between divergent sequences (see section 3.1.1), allowing horizontal gene transfer of intact MMR genes from other species. Maximal levels of exchange are reached when the SOS response is activated in the absence of MMR (Matic *et al.*, 1995; Matic *et al.*, 2000). MMR genes from different *E. coli* lineages show a high level of mosaicism and phylogenetic discordance compared to other genes, suggesting frequent transfer of these genes over evolutionary time (Denamur *et al.*, 2000; Brown *et al.*, 2001). Rapid genomic divergence in mutator cells followed by restoration of MMR function can cause speciation in bacteria (Vulic *et al.*, 1999).

The loss of MMR function in HNPCC cancer cells leading to an increased rate of generation of cancer-causing mutations can be thought of as analogous to the generation of advantageous mutations in bacterial mutator strains (Venkatesan *et al.*, 2006). It has even been suggested that adaptive mutation in bacteria is a good model to study the multiple mutational origins of human cancer (Hall, 1995). Environmentally induced increase in the mutation rate has also been reported in multicellular eukaryotes, and could be mediated by error-prone translesion polymerases (Rosenberg, 2001).

5.1.2 Trypanosome mutator strains

The genome of the kinetoplastid *T. cruzi* contains the same repertoire of MutS homologues (see section 1.4) and indeed all MMR as seen in *T. brucei* (El Sayed *et al.*, 2005). In addition, the *MSH2* gene has been cloned and characterised (Augusto-Pinto *et al.*, 2001). An 829-bp section of the *TcMSH2* gene containing the essential ATPase domain (see section 4.2.3.3) was PCR-amplified from 13 different *T. cruzi* strains and sequenced, revealing 21 single-nucleotide polymorphisms (SNPs) in this region, five of which cause

amino acid substitutions (Augusto-Pinto *et al.*, 2003). These polymorphisms allow the division of the strains into three haplogroups corresponding to the three *T. cruzi* clades (Gaunt and Miles, 2000; Robello *et al.*, 2000; Machado and Ayala, 2001; Buscaglia and Di Noia, 2003; de Freitas *et al.*, 2006), with all strains from each haplogroup encoding the same isoform of the MSH2 protein. Representative strains from two clades (perhaps significantly, the ones which cause the most severe human infections (Di Noia *et al.*, 2002)) showed hydrogen peroxide-induced microsatellite instability and increased cisplatin resistance; both these phenotypes suggest that the observed differences in *MSH2* sequence may correspond to differences in MMR efficiency, at least under conditions of genotoxic stress. Preliminary analysis shows increased genetic variability in multi-copy gene families encoding surface antigens in the strains hypothesised to have less efficient MMR (Machado *et al.*, 2006). These phenotypes bear the hallmarks of mutator strains, and, if correct, would represent the first time that mutator strains have been observed in a eukaryotic pathogen. However, the mutations are yet to be linked directly to a difference in MMR efficiency.

The work in the following chapter follows on from these novel observations in *T. cruzi* to assess, firstly, whether polymorphisms are present also in the DNA and/or protein sequence of *T. brucei MSH2*. Secondly, the chapter assesses whether or not differences in MMR efficiency are found between *T. brucei* strains and subspecies.

5.2 Results

5.2.1 DNA sequence analysis

5.2.1.1 *Trypanosoma brucei* strains selected for analysis

The species *T. brucei* is divided into three morphologically indistinguishable subspecies: *T. b. brucei*, *T. b. rhodesiense*, and *T. b. gambiense* (Gibson, 2002). *T. b. rhodesiense* and *T. b. gambiense* are able to infect humans and some other primates (Pays *et al.*, 2006), while *T. b. brucei* is not. Trypanolytic activity is associated with the HDL fraction of human serum (Hager *et al.*, 1994), specifically the protein Apolipoprotein L-1 (APOL1) (Vanhamme *et al.*, 2003). APOL1 lyses trypanosomes by forming ionic pores in the cell membranes (Perez-Morga *et al.*, 2005), possibly in synergy with another high-density lipoprotein known as HPR (Shiflett *et al.*, 2005). *T. b. rhodesiense* neutralises APOL1 activity *via* a serum-resistance associated (*SRA*) protein, which is necessary and sufficient for growth in human serum and confers human serum resistance on *T. b. brucei*

Strain	Subspecies	Location	Host
B3200	*rhodesiense*	Uganda	Human
222	*rhodesiense*	Zambia	Human
208	*rhodesiense*	Zambia	Human
984	*brucei*	Kenya	Tsetse
S14	*brucei*	Uganda	Cow
427	*brucei*	monomorphic strain	
927	*brucei*	Kenya	Tsetse
Eliane	*gambiense*	Ivory Coast	Human
386	*gambiense*	Ivory Coast	Human

Table 5.1. *T. brucei* strains used in this sequencing study. The 9 *T. brucei* strains selected for use in this experiment are shown, along with their subspecies (*T. b. rhodesiense*, Rhodesiense; *T. b. brucei*, Brucei; or *T. b. gambiense*, Gambiense), the country from which the sample was isolated, and the host organism from which it was collected.

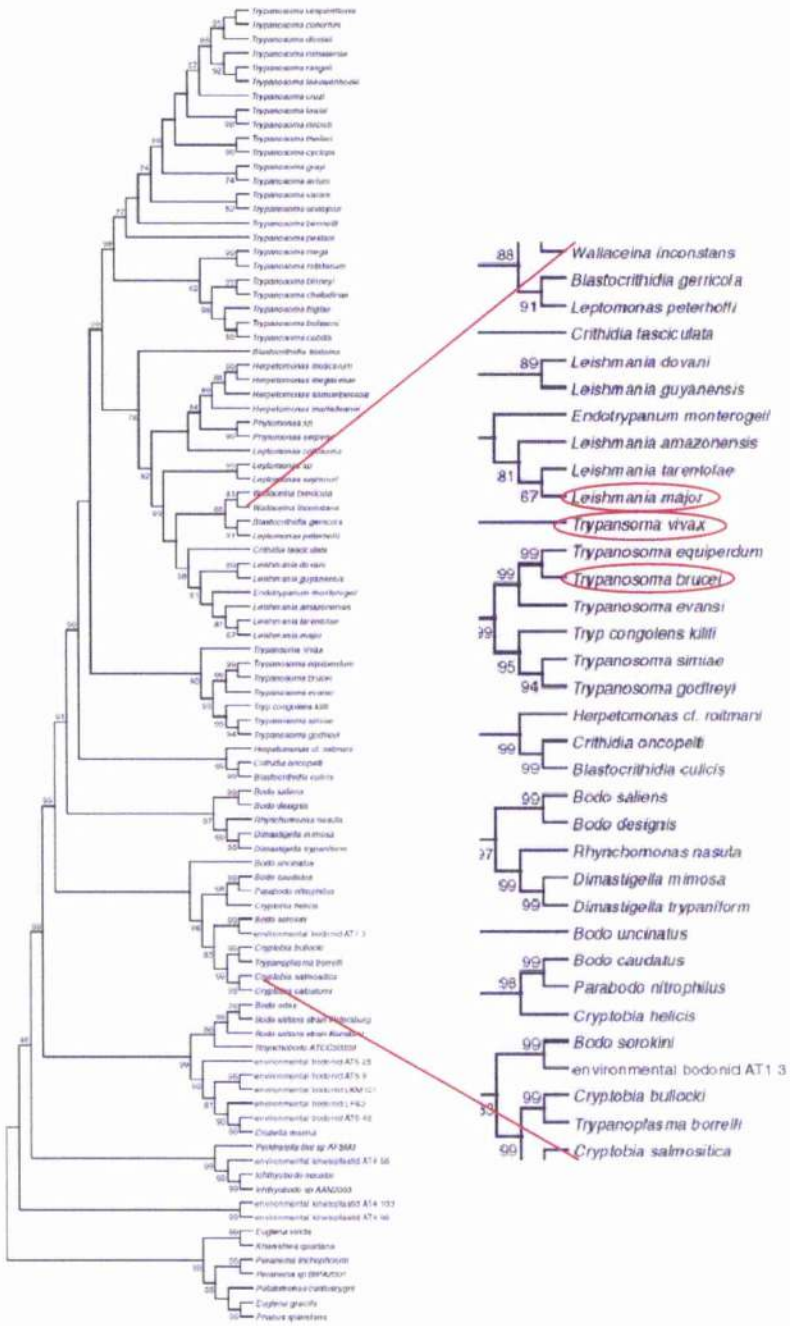


Figure 5.1. Phylogenetic tree of the Trypanosomatidae and Bodonidae families of the order Kinetoplastidae, based on 18S rRNA sequences. Species studied in the experiments in this chapter are circled in pink. Adapted from Piontkivska and Hughes 2005.

(Xong *et al.*, 1998). *SRA*, which is expressed as an *ESAG* (De Greef *et al.*, 1989), is related in sequence to *VSG* (De Greef and Hamers, 1994) and interacts with APOL1 in the lysosomes by contact between α-helical domains (Vanhamme *et al.*, 2003). On the other hand, despite its constitutive resistance to human serum (De Greef and Hamers, 1994), *T. b. gambiense* does not encode *SRA* (Gibson, 2002), implying its mode of resistance must be different; a gene similar to *SRA* has been identified in this subspecies, but on its own it is not sufficient to confer resistance on *T. b. brucei* (Berberof *et al.*, 2001). *T. b. gambiense* strains can be divided into two groups (Gibson, 1986): group 1 strains are avirulent, stably express serum resistance, and are homogeneous in marker analysis (Hide *et al.*, 1990; Truc and Tibayrenc, 1993); group 2 strains are more virulent, but lose serum resistance following serial passage in rodents (Pays *et al.*, 2006). *T. b. gambiense* can easily be differentiated from *T. b. brucei* and *T. b. rhodesiense* by molecular characterisation of polymorphic isozymes (Truc and Tibayrenc, 1993; Mathieu-Daude *et al.*, 1994). The latter two subspecies are more related to each other, and *T. b. rhodesiense* is thought to have evolved from a *T. b. brucei* clone that gained the ability to resist lysis by human serum (Mathieu-Daude *et al.*, 1994).

To examine the sequences of *MSH2*, nine *T. brucei* isolates were selected for analysis (genomic DNA provided by A. Macleod) as representative of the *T. brucei* species as a whole. These included four *T. b. brucei* isolates (two wild isolates from Kenya and one from Uganda, as well as the lab-adpated Lister 427 strain), three *T. b. rhodesiense* isolates (two from Uganda and one from Zambia), and two *T. b. gambiense* isolates from the Ivory Coast (one type 1 isolate (Eliane) and one type 2 isolate (STIB 386)). The names of these strains, the geographic location of their origins, and the host organism from which they were taken, are shown in table 5.1. Sequences from the genome projects of related kinetoplastids *L. major* and *T. vivax* were also included in the analysis, as well as *MSH2* sequence from the *T. b. brucei* strain ILTat 1.2, from which EATRO 795 is derived, obtained by J. Bell (PhD thesis, University of Glasgow). *T. vivax* was chosen for comparison rather than *T. cruzi* as it is more closely related to *T. brucei* (Piontkivska and Hughes, 2005; see figure 5.1). *L. major* is more diverged from *T. brucei* than either *T. vivax* or *T. cruzi*, and therefore provides a distant comparison.

5.2.1.2 Analysis of *MSH2* sequence from *T. brucei* isolates

To look for potential genetic changes indicative of mutators, the *MSH2* ATPase domain was selected for initial sequence analysis. MSH2 is central to the eukaryotic MMR system, and it is known that a large proportion of bacterial mutator strains contain

mutations in its homologue, MutS (see section 5.1.1). The ATPase domain (see figure 4.7; section 4.2.3.3) was chosen specifically for two reasons. Firstly, it is very well conserved, as ATPase function is essential for the function of the protein (see section 1.3.2). The majority of spontaneous mutations identified causing inactivation of the *S. cerevisiae* protein or HNPCC are found in the ATPase domain (Obmolova *et al.*, 2000). Secondly, studying the region of the gene that had previously been analysed in *T. cruzi* (Augusto-Pinto *et al.*, 2003) allows the most direct comparison of the levels of divergence in the two kinetoplastids.

The primers MSH2_ATPase_5 and MSH2_ATPase_3 (see appendix 1) were designed based on *MSH2* DNA sequence from the TREU 927 strain from the genome database. These primers amplify a 579-bp section of *MSH2*, containing most of the ATPase domain as well as a small section upstream of this conserved domain, which could be expected to be less well conserved (see figure 5.2). PCR was carried out using Herculase, a high-fidelity DNA polymerase, to amplify this fragment from genomic DNA of the nine strains described in the previous section. The same primers were also used to attempt to PCR-amplify the equivalent region of *MSH2* from *T. vivax* genomic DNA. In addition, a control PCR reaction was carried out using water instead of genomic DNA substrate. A small amount of each reaction was run on a 0.8% agarose gel, showing amplification of a single product of the correct size in all cases, apart from the reactions using *T. vivax* genomic DNA as substrate, which were unsuccessful, and the no-DNA control reaction. The lack of PCR-amplified product in the latter demonstrates that any product for the *T. brucei* strains did not result from contamination in the oligonucleotide primers, or other PCR reagents (figure 5.3; the PCR for TREU 927 is not shown in this figure). The lack of PCR product from *T. vivax* most likely reflects divergence in the *MSH2* sequence. No attempts to modify conditions to allow PCR amplification were performed.

The *MSH2* PCR products were next incubated with Taq DNA polymerase prior to cloning into the TOPO TA vector (Invitrogen). This step had the role of adding A residues to the ends of each PCR product; these single-base overhangs are present at the end of all Taq-generated but not Herculase-generated PCR products, and are essential for cloning of DNA fragments into the TOPO TA vector. At least two cloned PCR products from each strain were sequenced by the MBSU (University of Glasgow). These DNA sequences, as well as sequence from the *T. vivax* and *L. major* genome databases and from the *T. b. brucei* ILTat 1.2 strain obtained by J. Bell (PhD thesis, University of Glasgow), were compared by performing a sequence alignment using ClustalX (Thompson *et al.*, 1997), and visualised using Boxshade (http://ch.embnet.org/software/BOX_form.htm) as shown in figure 5.6.

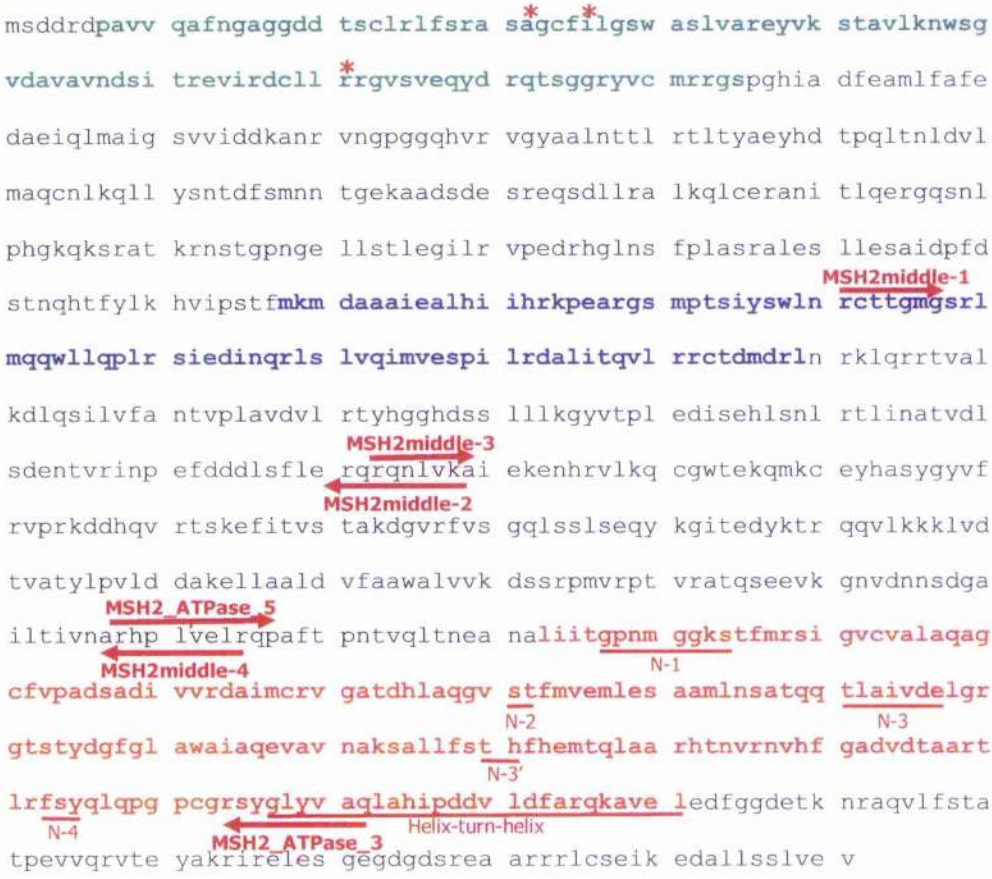


Figure 5.2. Primers used for sequencing of *T. brucei MSH2*. MSH2 protein sequence is shown in black text, with the mismatch interaction domain in green, middle conserved domain in blue, and ATPase/helix-turn-helix domain in orange. Conserved elements of the ATPase domain are underlined in red, important conserved residues indicated by red stars, and the locations of the primers are represented by pink arrows.

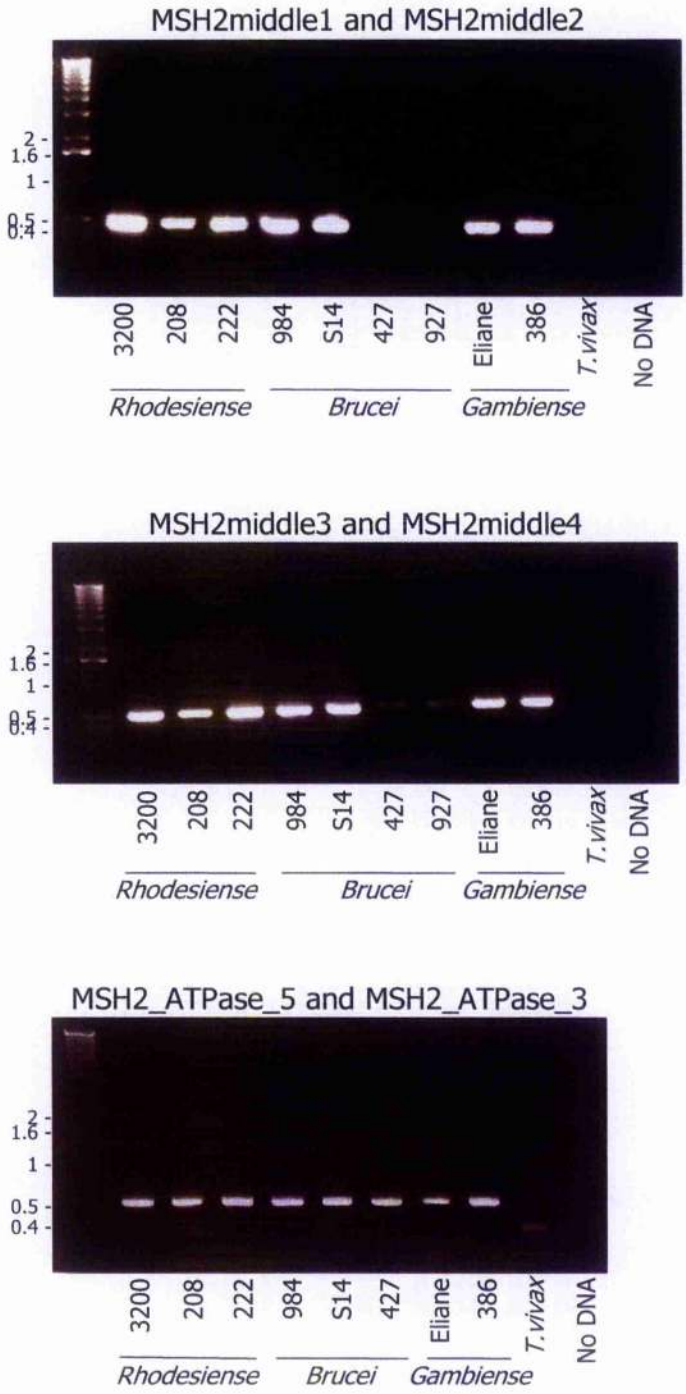
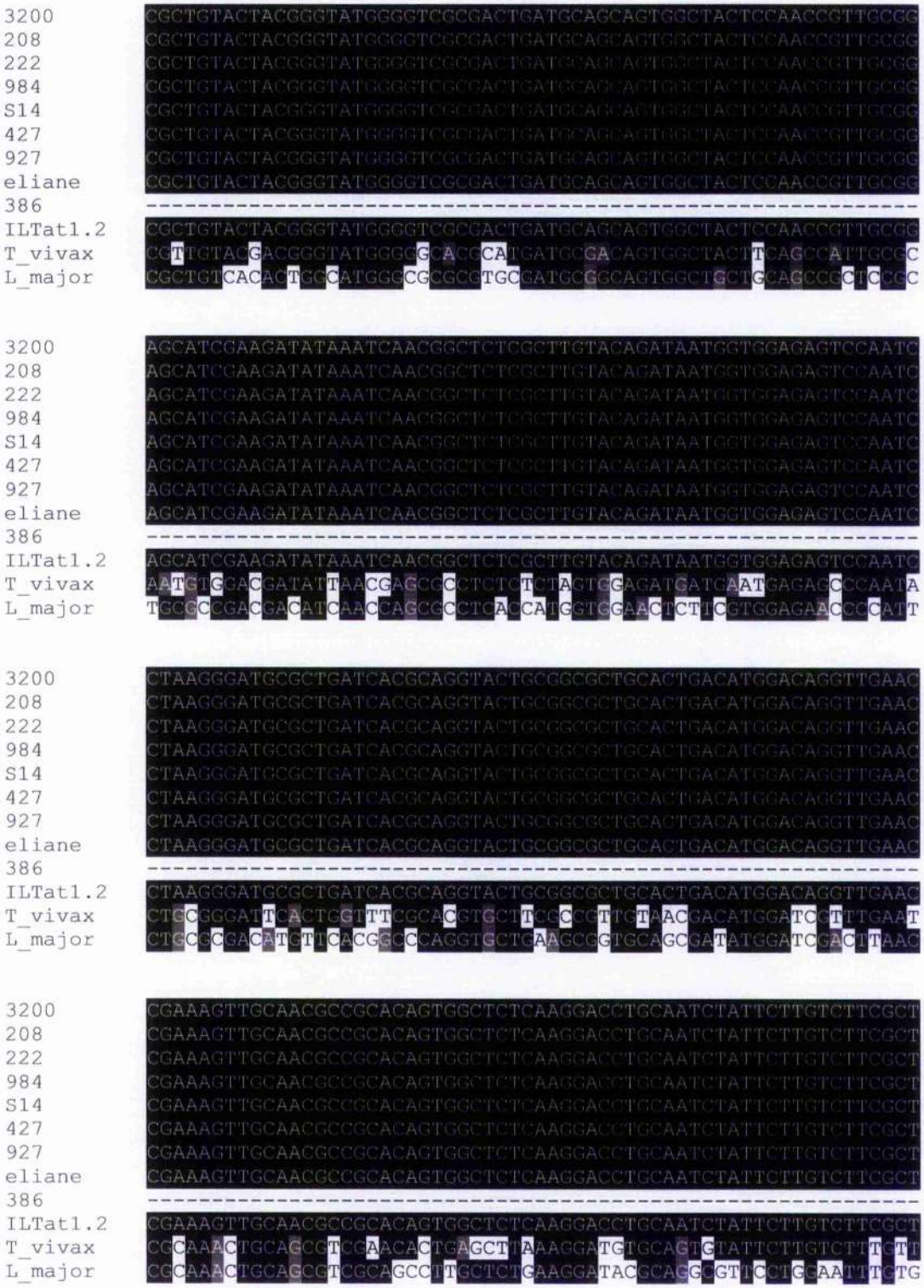


Figure 5.3. PCR amplification of three sections of *T. brucei MSH2*. PCR was performed on genomic DNA from 8 or 9 *T. brucei* strains from three different subspecies (indicated below the strain names) using three sets of primers. The reaction was also attempted on *T. vivax*. A no-DNA control was included. PCR products were separated on a 0.8% agarose gel.


3200
208
222
984
S14
427
927
eliane
386
ILTat1.2
T_vivax
L_major

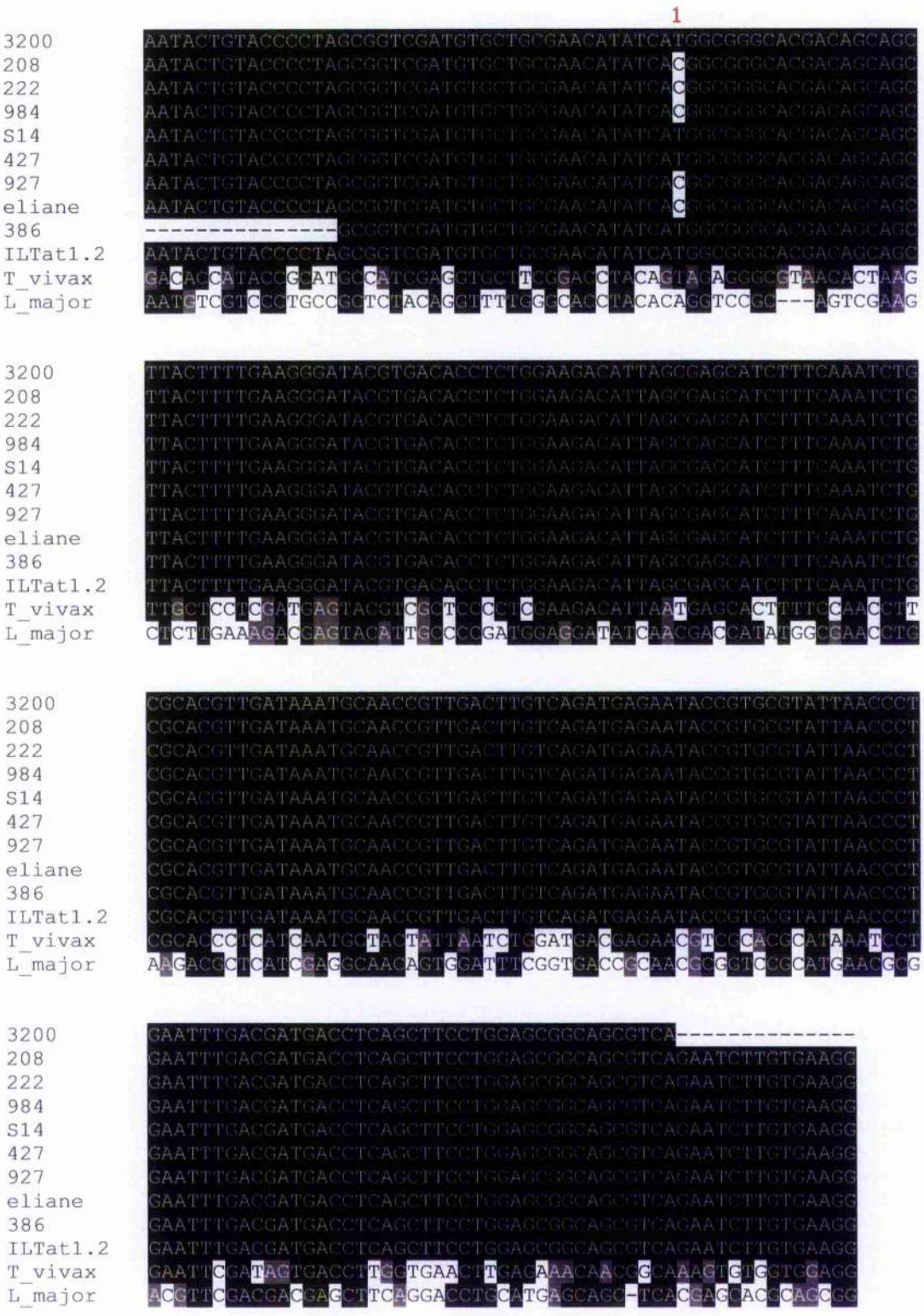
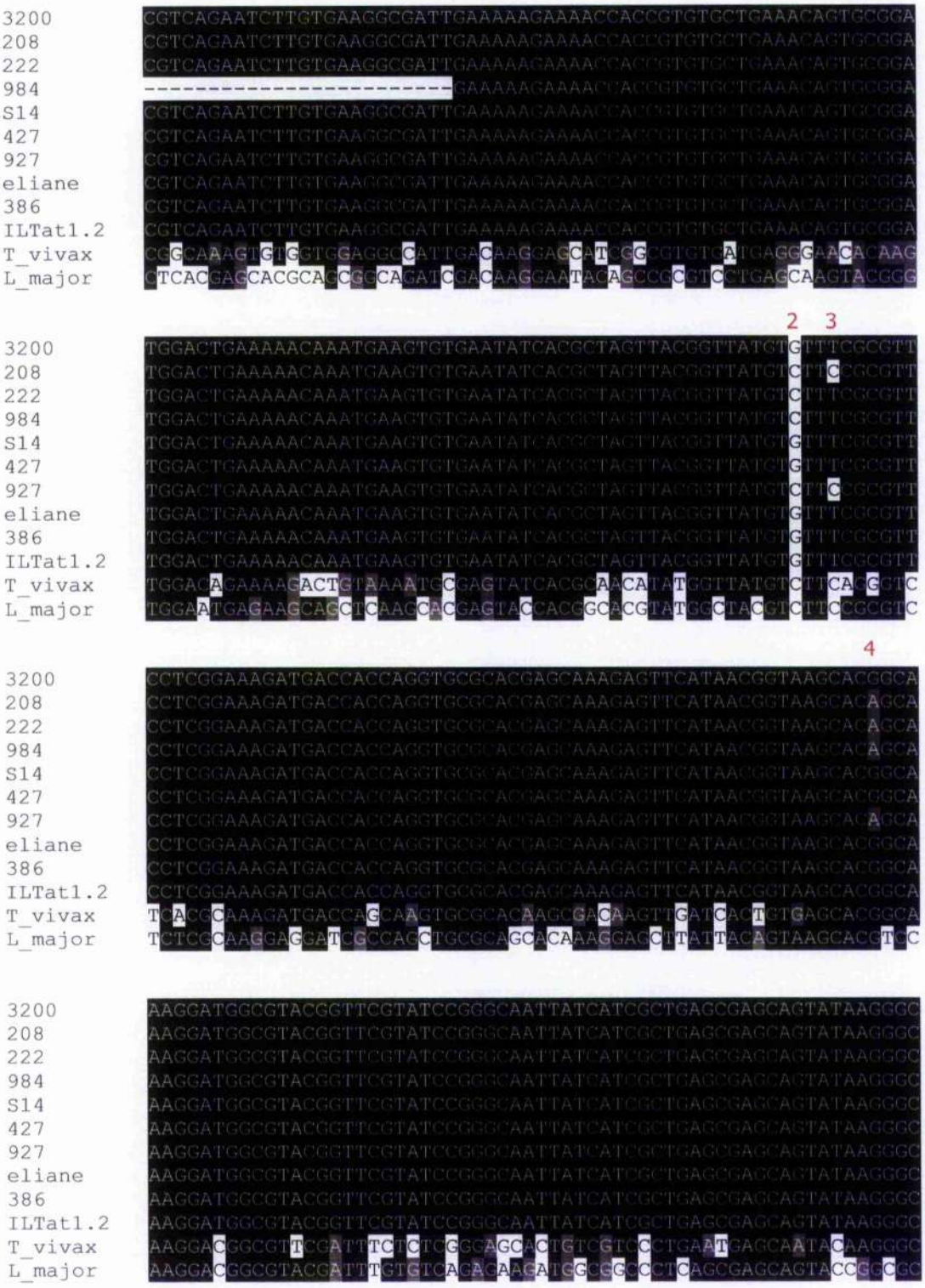
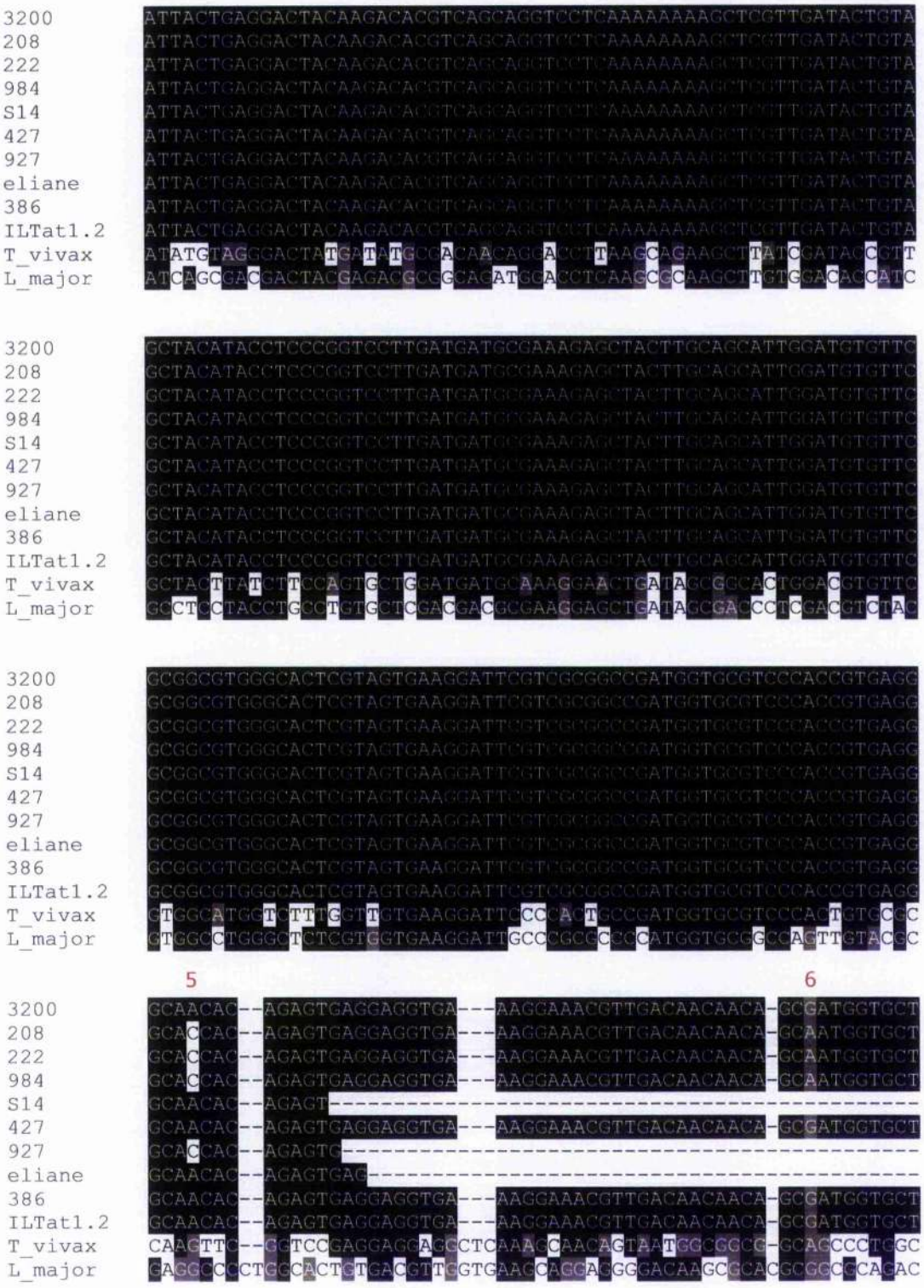


Figure 5.4. Multiple alignment of DNA sequence obtained by PCR amplification using primers MSH2middle-1 and MSH2middle-2 from 9 *T. brucei* strains, from the ILTat 1.2 *MSH2* gene sequenced by J. Bell, and from the *T. vivax* and *L. major* genome projects. Sequences were aligned using ClustalX (Thompson et al 1997) and shaded using the Boxshade server (http://ch.embnet.org/software/BOX_form.html): identical residues in 50% of the sequences are shaded in black, and conserved residues (purine or pyrimidine) in grey. Sites of SNPs are numbered in pink.


3200
208
222
984
S14
427
927
eliane
386
ILTat1.2
T_vivax
L_major
2
3
4


3200
208
222
984
S14
427
927
eliane
386
ILTat1.2
T_vivax
L_major
5
6

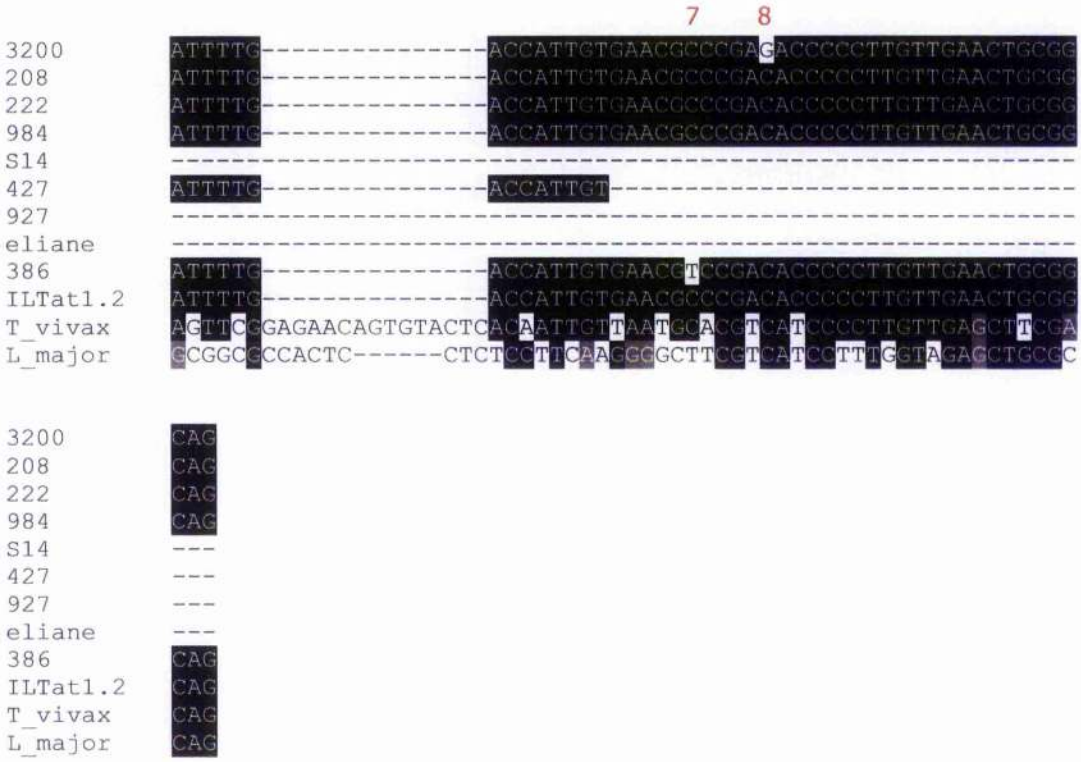
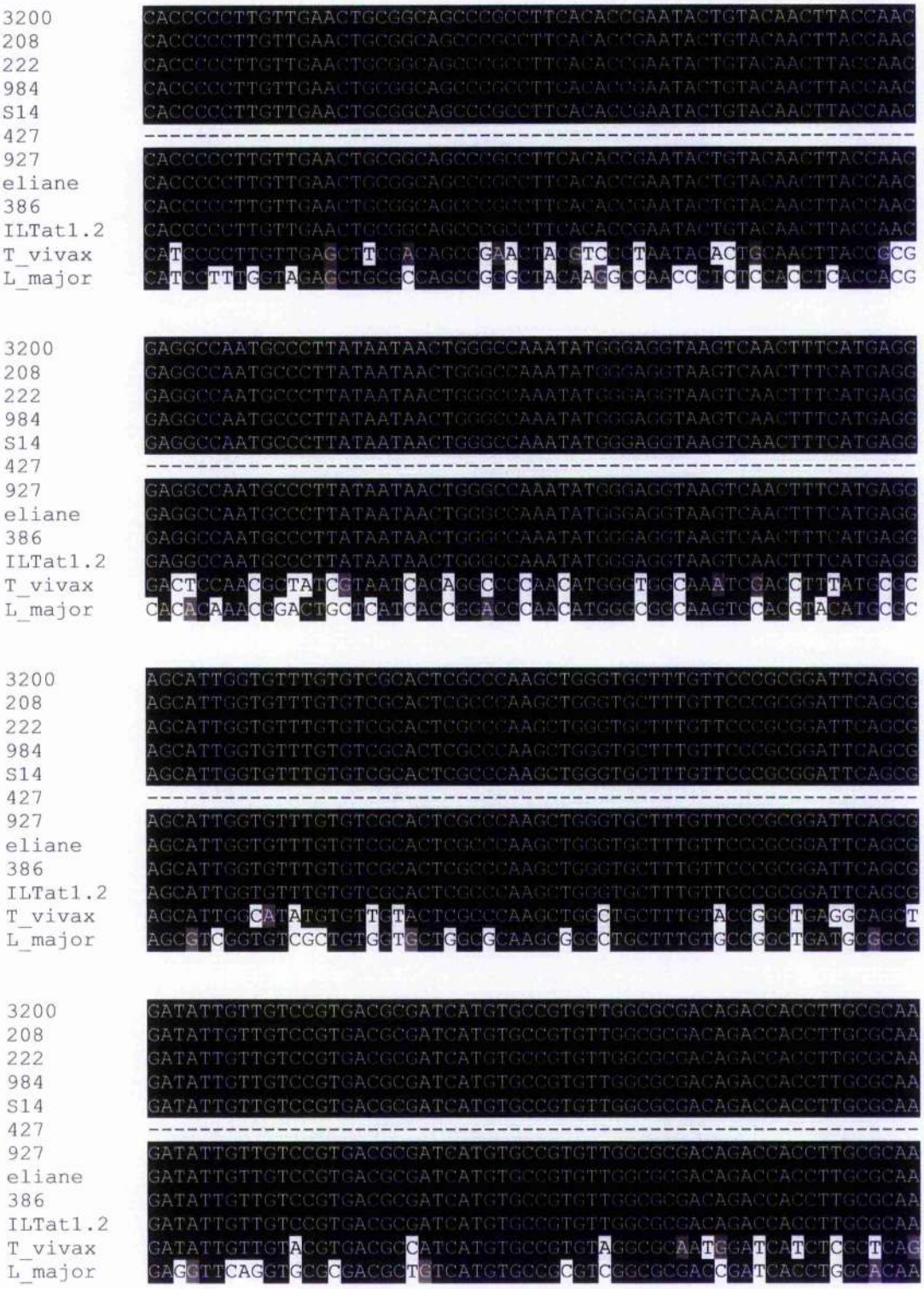
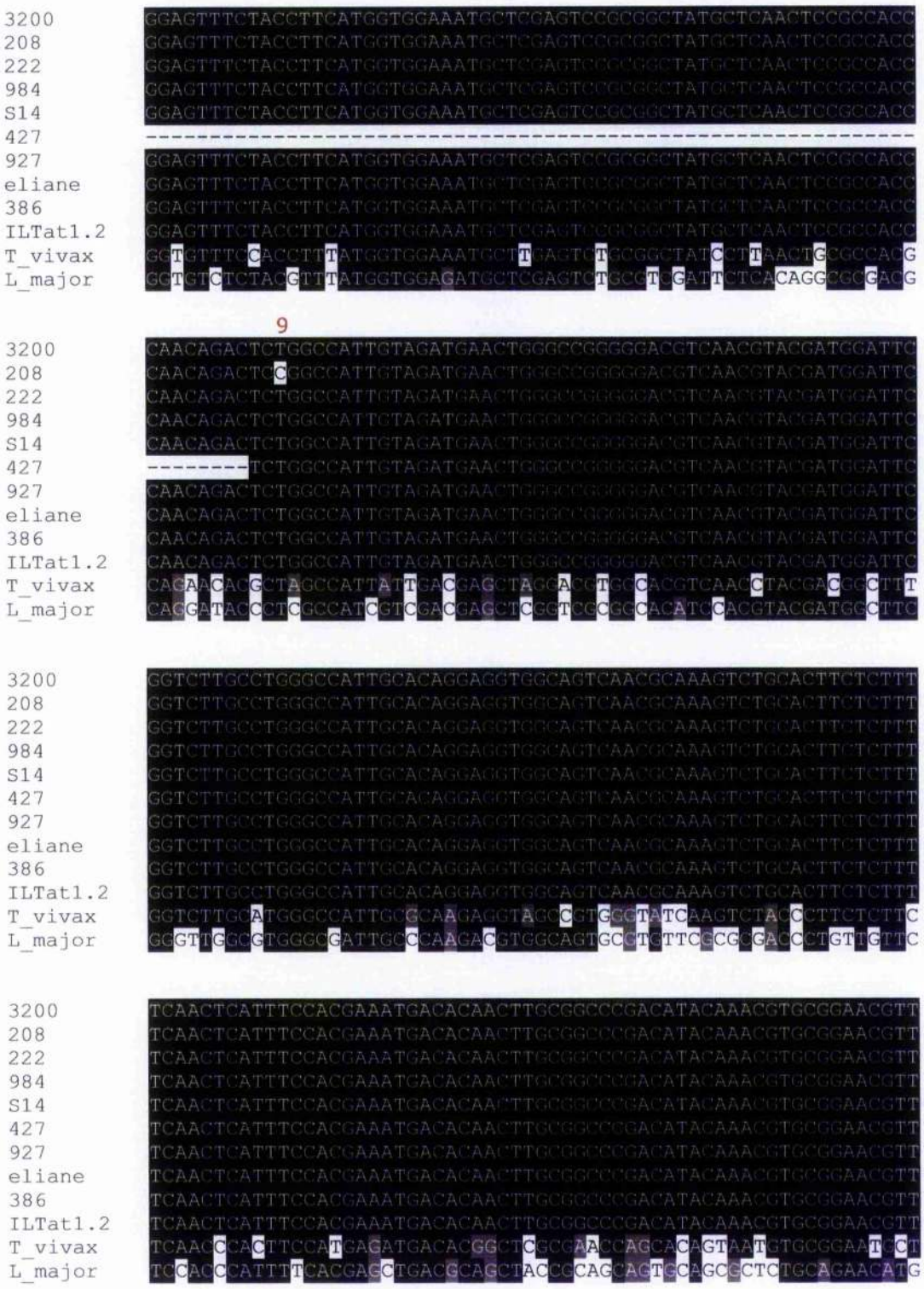


Figure 5.5. Multiple alignment of DNA sequence obtained by PCR amplification using primers MSH2middle-3 and MSH2middle-4 from 9 *T. brucei* strains, from the ILTat 1.2 *MSH2* gene sequenced by J. Bell, and from the *T. vivax* and *L. major* genome projects. Sequences were aligned using ClustalX (Thompson et al 1997) and shaded using the Boxshade server (http://ch.embnet.org/software/BOX_form.html): identical residues in 50% of the sequences are shaded in black, and conserved (purine or pyrimidine) residues in grey. Sites of SNPs are numbered in pink.


3200
208
222
984
S14
427
927
eliane
386
ILTat1.2
T_vivax
L_major


3200
208
222
984
S14
427
927
eliane
386
ILTat1.2
T_vivax
L_major
9

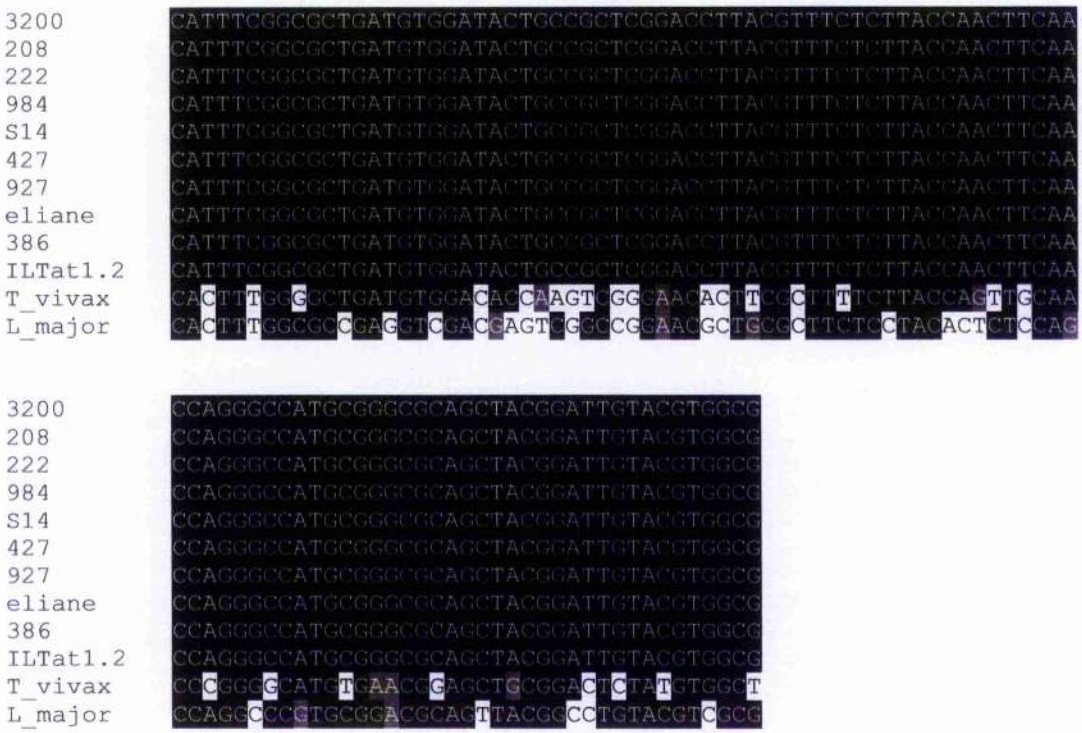


Figure 5.6. Multiple alignment of DNA sequence obtained by PCR amplification using primers MSH2_ATPase_5 and MSH2_ATPase_3 from 9 *T. brucei* strains, from the ILTat 1.2 *MSH2* gene sequenced by J. Bell, and from the *T. vivax* and *L. major* genome projects. Sequences were aligned using ClustalX (Thompson et al 1997) and shaded using the Boxshade server (http://ch.embnet.org/software/BOX_form.html): identical residues in 50% of the sequences are shaded in black, and conserved residues (purine or pyrimidine) in grey. Sites of SNPs are numbered in pink.

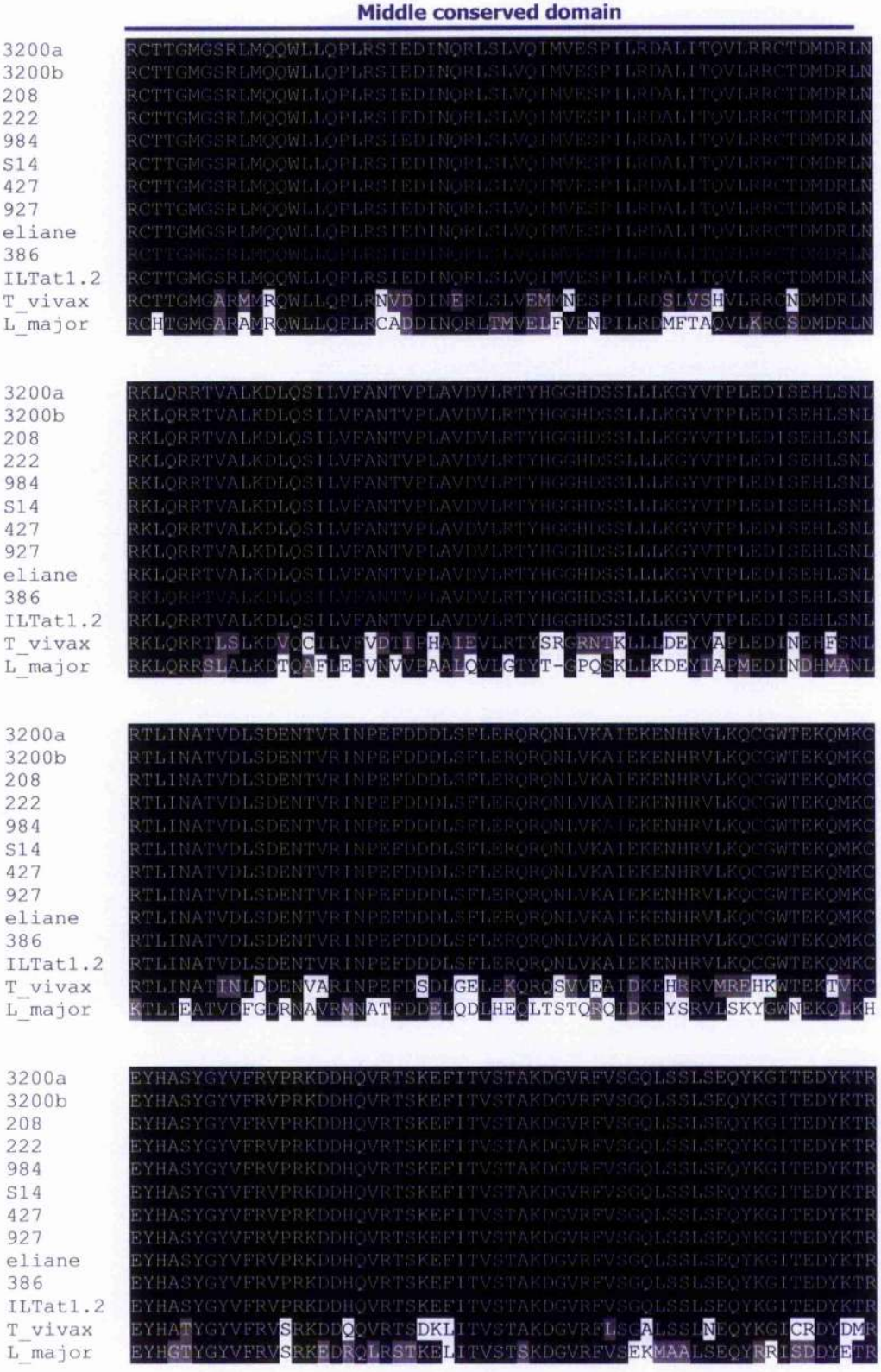
The consensus sequence from all the successful sequencing reactions is shown. Good quality sequence was not obtained from the Lister 427 DNA sample for 308 bp of upstream sequence of the PCR product, and this was omitted from the analysis. The alignment revealed that the sequences from all the *T. brucei* strains show remarkably high levels of sequence conservation: only a single polymorphism was seen at the DNA level, in the *T. rhodesiense* 208 strain (SNP 9). As this site is conserved in all the other *T. brucei* strains and in *T. vivax* and *L. major*, this suggests a potential PCR error in the 208 strain. In general, the *T. vivax* and *L. major* sequences, included for comparison, are recognisably homologous to the *T. brucei* sequence, although as would be expected the levels of homology between the distant species are much lower than between isolates of the *T. brucei* subspecies.

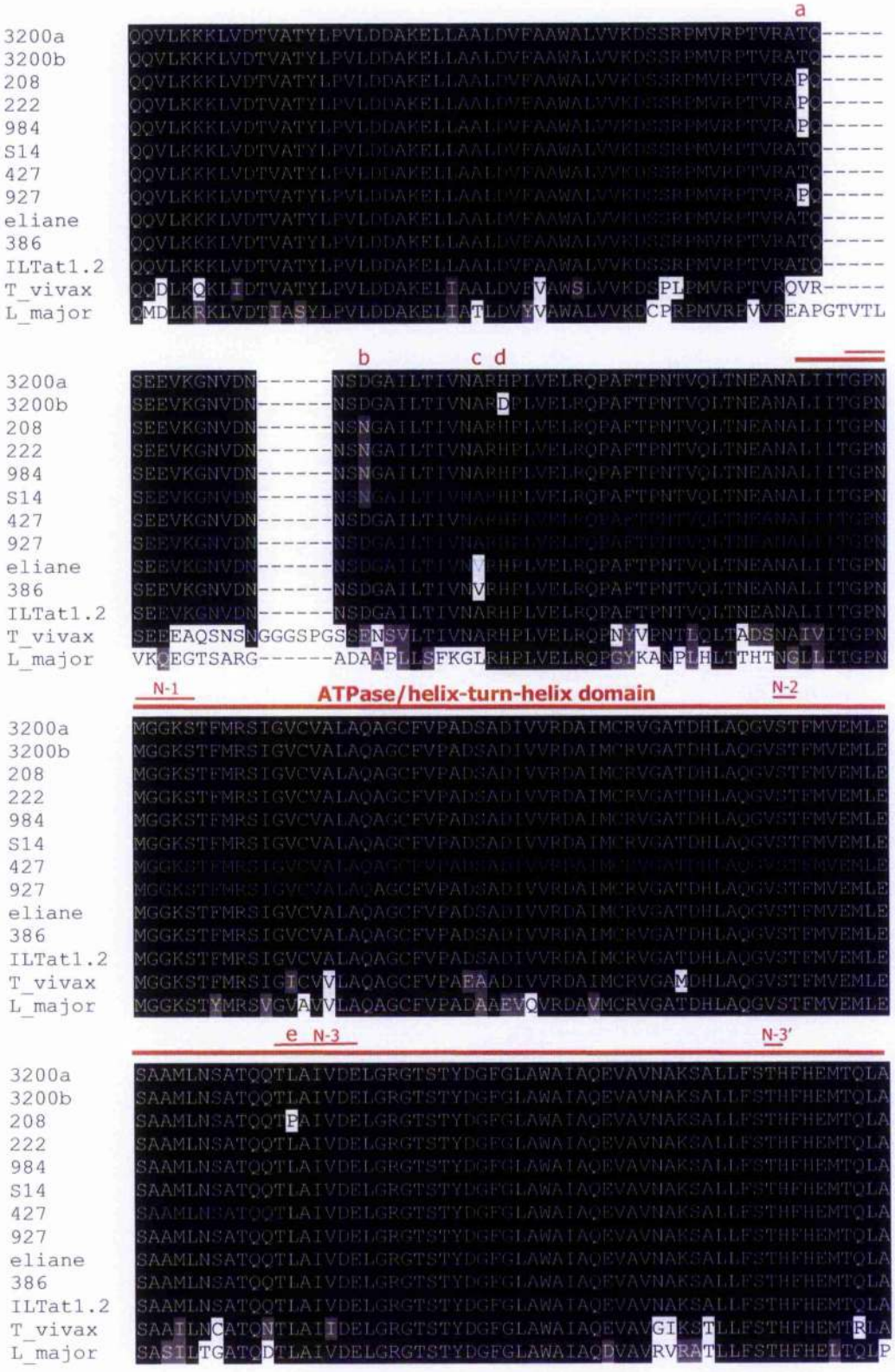
In order to assess whether this lack of sequence divergence is seen throughout the *MSH2* ORF, or whether it is specific to the ATPase domain, two pairs of primers (MSH2middle1 and MSH2middle2, and MSH2middle3 and MSH2middle4, see appendix 1) were designed based on *MSH2* DNA sequence from the TREU 927 strain from the genome database. These primers should PCR-amplify 476 bp and 522 bp sections, respectively, of the *MSH2* ORF upstream of the ATPase domain, covering half of the middle conserved domain as well as the poorly conserved region in-between the middle conserved and ATPase domains (see figure 5.2). The section of the gene PCR-amplified by primers MSH2middle3 and MSH2middle4 overlaps with those PCR-amplified by primers MSH2middle1 and MSH2middle2 and by primers MSH2_ATPase_5 and MSH2_ATPase_3, allowing assembly of a single, contiguous sequence. PCR was again carried out using Herculase on genomic DNA of the nine strains described in the previous section (section 5.2.1.1). The same primers were used to attempt to PCR-amplify the equivalent region of *T. vivax* genomic DNA. Finally, control PCR reactions were carried out using water instead of genomic DNA substrate. A small amount of each reaction was run on a 0.8% agarose gel, showing amplification of a single product of the expected size in all cases, apart from the reaction using *T. vivax* genomic DNA as substrate, which were unsuccessful, and the no-DNA control reactions (figure 5.3). As before, the PCR products were next incubated with Taq DNA polymerase prior to cloning into the TOPO TA vector (Invitrogen), and at least two cloned PCR products from each strain were sequenced by the MBSU (University of Glasgow). These DNA sequences, as well as sequence from the *T. vivax* and *L. major* genome databases and sequence from the ILTat 1.2 strain obtained by J. Bell (PhD thesis, University of Glasgow), were compared by performing a sequence alignment using ClustalX (Thompson *et al.*, 1997), visualised using Boxshade (http://ch.embnet.org/software/BOX_form.html) as shown in figure 5.4. The consensus

sequence from all successful sequencing reactions is shown. As for the sequences generated by primers MSH2_ATPase5' and MSH2_ATPase3', if sequencing did not yield readable sequence for specific clones, it was omitted.

The levels of polymorphism between the strains were also extremely low when these sequences were compared. Only one polymorphism (SNP 1) was seen in the section amplified by primers MSH2middle1 and MSH2middle2. 5 *T. brucei* isolates contained a C and 5 *T. brucei* strains (as well as *T. vivax*) contain a T residue, whereas *L. major* contained an A residue. In this case, data were not available for the beginning of the sequence derived from STIB 386 genomic DNA. The region amplified with primers MSH2middle3 and MSH2middle4, was more polymorphic, containing 7 SNPs. At SNP 2, 4 *T. brucei* isolates (as well as *T. vivax* and *L. major*), contain a C residue, and 6 *T. brucei* isolates contain a T residue. At SNP 3, 2 *T. brucei* isolates (and *T. vivax* and *L. major*) contain a C residue and 8 *T. brucei* isolates contain a T residue. At SNP 4, 4 *T. brucei* isolates contain an A residue and 6 *T. brucei* isolates (and *T. vivax* and *L. major*) contain a G residue. At SNP 5, 4 *T. brucei* isolates contain a C residue and 6 *T. brucei* isolates contain an A residue, whereas *T. vivax* and *L. major* contain a G residue. At SNP 6, of the sequences available, 4 *T. brucei* isolates (and *L. major*) contain a G residue and 3 *T. brucei* strains (and *T. vivax*) contain an A residue. At SNP 7 and 8, a single isolate differed from all the other *T. brucei* strains available, potentially suggesting PCR errors. The sequences for this PCR reaction from four strains (S14, 927, 427 and Eliane) are incomplete, however.

The three sections of *MSH2* sequence were next assembled into a single contiguous sequence and the predicted translated polypeptide sequences compared by performing a sequence alignment using ClustalX (Thompson *et al.*, 1997) and visualised using Boxshade (http://ch.embnet.org/software/BOX_form.htm) as shown in figure 5.7. Where DNA sequence was incomplete, data from another strain of the same subspecies was substituted in order to make up contiguous sequence in each case, and these instances have been highlighted. In total five polymorphic sites were identified at the protein level. At polymorphism a (corresponding to SNP 2), 4 *T. brucei* isolates encode proline, and 6 encode threonine, whereas *T. vivax* and *L. major* encode valine and alanine, respectively. At polymorphism b (corresponding to SNP 6), of the sequences available 3 *T. brucei* isolates encode asparagine, and 5 encode aspartate, whereas *T. vivax* and *L. major* encode glutamate and alanine, respectively. At polymorphism c (corresponding to SNP 7), a single *T. brucei* strain (STIB 386) encodes a valine and *L. major* encodes a leucine, whereas all the other *T. brucei* strains, as well as *T. vivax* encode an alanine.


Middle conserved domain
3200a
3200b
208
222
984
S14
427
927
eliane
386
ILTat1.2
T_vivax
L_major


a
3200a
3200b
208
222
984
S14
427
927
eliane
386
ILTat1.2
T_vivax
L_major
b
c d
N-1
ATPase/helix-turn-helix domain
N-2
e N-3
N-3'

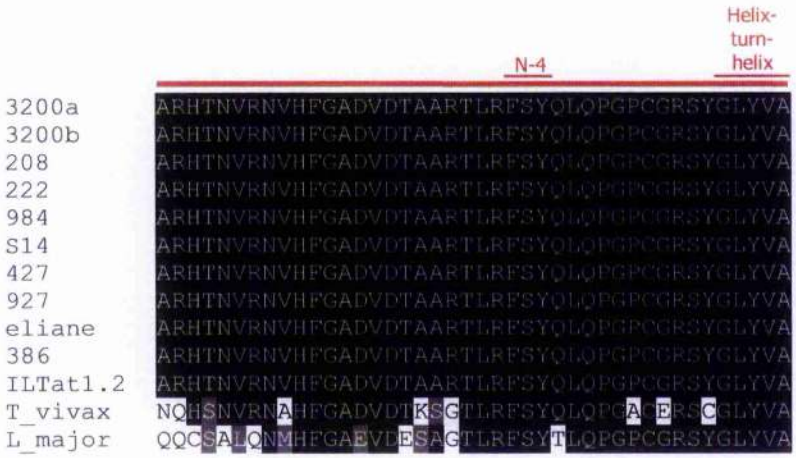


Figure 5.7. Multiple alignment of MSH2 polypeptide sequence obtained by translation of the DNA sequences shown in figures 4-6. Where it was necessary to infer sequence from other strains to give a full-length, contiguous sequence, as described in the main text, sequence is shown in light blue. Sequences were aligned using ClustalX (Thompson et al 1997) and shaded using the Boxshade server (http://ch.embnet.org/software/BOX_form.html): identical residues in 50% of the sequences are shaded in black, and conserved residues in grey. Sites of protein sequence polymorphism are labelled using pink letters.

Polymorphism d (corresponding to SNP 8) is found only in the overlap between the PCR products amplified with primer sets MSH2middle3 and MSH2middle4 and MSH2_ATPase_5 and MSH2_ATPase_3. As this polymorphism only appeared in one isolate, and was in the sequence amplified by the former pair of primers, whereas with the latter primer pair no change was seen, suggesting this polymorphism is due to a PCR or sequencing error, although it is also possible that it is due to the presence of another *MSH2* allele in the genome of this strain. At polymorphism e (corresponding to SNP 9), a single *T. brucei* isolate (208) encodes proline, whereas all the other *T. brucei* isolates, as well as *T. vivax* and *L. major*, encode leucine.

Assuming that all peptide polymorphisms are genuine (and for some, notably SNP d, and perhaps SNP e, this may be incorrect), it is apparent that most are likely not to affect function. SNPs a-d are found in the less conserved region between the middle conserved and ATPase domains, and SNPs a, b, and c are in residues that are variable in the *L. major* and *T. vivax* sequences. SNP e is a non-conservative change and does fall within an important conserved motif (the N-3 motif of the ATPase domain, a nucleotide-binding motif located at the dimer interface (Obmolova *et al.*, 2000; see section 4.2.3.3 and figure 4.7). In fact, the residue putatively altered, Leu772, is conserved in MSH2, MSH3 and MSH6(8) for *T. brucei*, *H. sapiens* and *S. cerevisiae*. This strong conservation of a residue in a conserved motif of the ATPase domain suggests that its mutation could have a detrimental effect on MMR efficiency. This change, if genuine, is specific to *T. b. rhodesiense* isolate 208, however, and is not common to the other examples of this subspecies. Overall, it is apparent that MSH2 from the different *T. brucei* subspecies and strains/isolates, in contrast to those from *T. cruzi*, are very well conserved. Furthermore, in only one case was a mutation found that might be predicted to cause a functional impairment.

5.2.1.3 Analysis of *RAD51* sequence from *T. brucei* isolates

In order to assess whether the low levels of polymorphism observed above in *MSH2* are due to a specific conservation of this protein, or to a generic high level of sequence conservation in all *T. brucei* coding regions, the sequence of the central recombination factor RAD51 was also studied. Primers RAD51-BamHI and RAD51-U3 (see appendix 1) were designed based on *RAD51* DNA sequence from the TREU 927 strain from the genome database (see figure 5.8). These primers should amplify a 715-bp section of *RAD51*, containing all four conserved domains of the protein: a helix-hairpin motif, a putative polymerisation motif, and Walker A and Walker B box nucleotide binding motifs

RAD51-BamHI

mntrtknkkr tkeviedevh diddtafdda avdavndntq emqqqvgdaa ggpsfrvlqi

menygvasad ikklmecgfl tvesvayapk ksilavkgis eakaekimae cckltpmgft

ratvfqeqrk etimvttgsr evdkllgggi evgsitelfg efrtgktqlc htlcvtcqlp

lsqgggegma lyidtegtfr perlvavaer ysldpeavle nvacaraynt dhqqqlllqa

satmaehrva iivvdsatal yrtdyngrge laarqmhlgk flrslrnlan eynvavvvtn

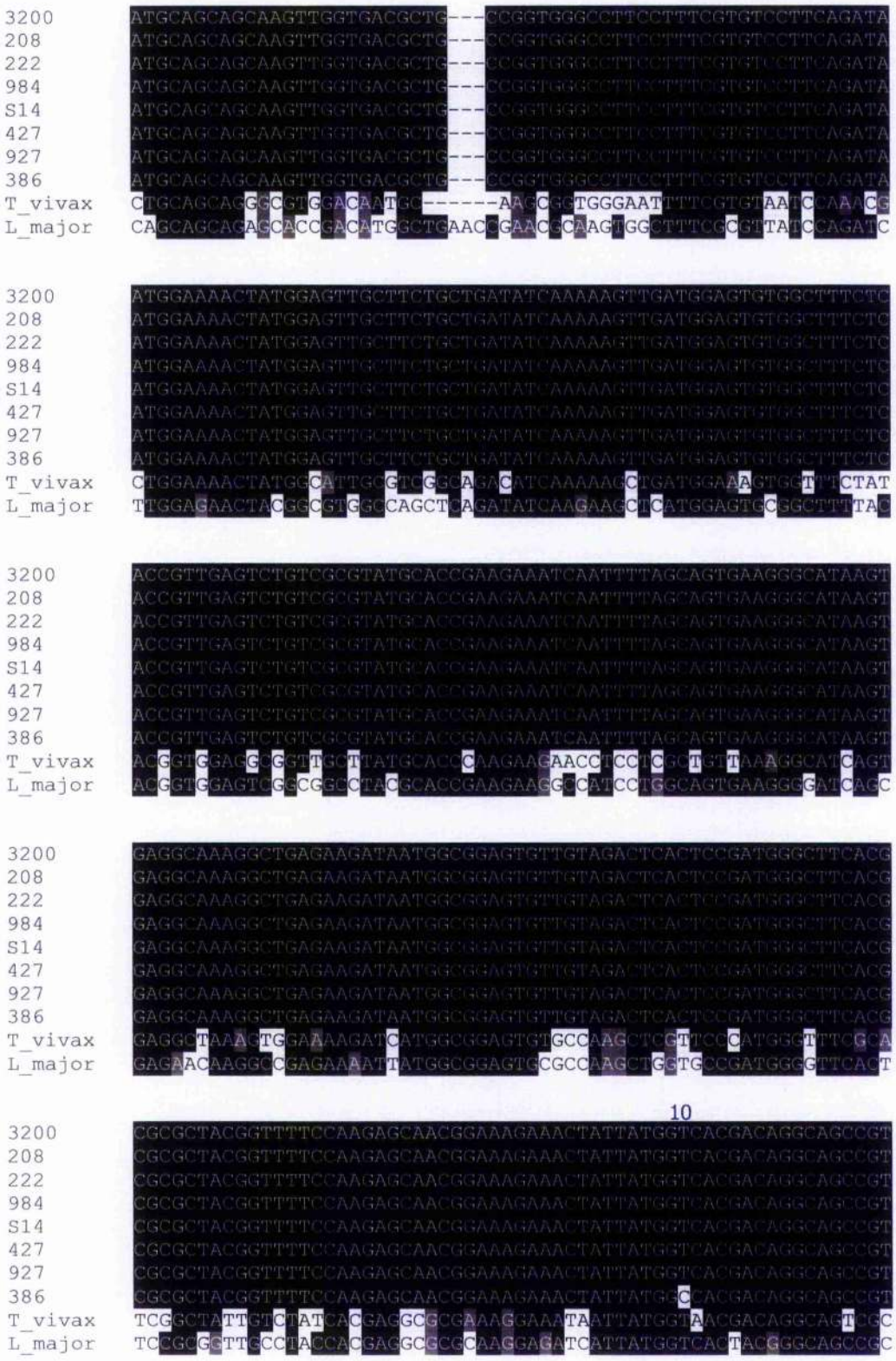
qvvanvdgaa ptfqadskkp igghimahas ttrlslrkgr geqriikvyd spclaeseai

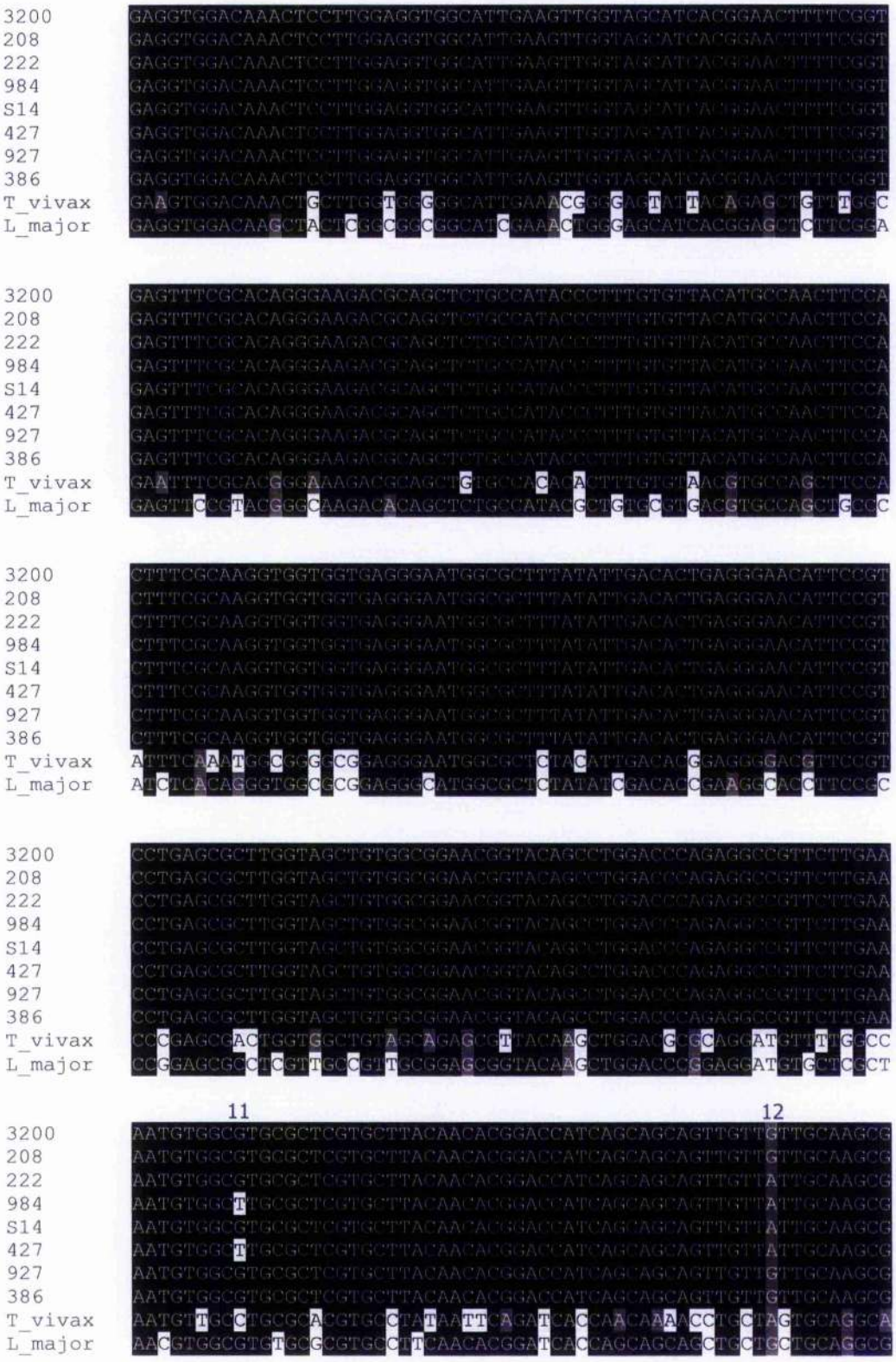
fgiyengvgd vrd

RAD51-U3

Figure 5.8. Primers used for sequencing of *T. brucei RAD51*. RAD51 protein sequence is shown in black text, with a helix-hairpin-helix motif in orange, a putative polymerisation domain in purple, a Walker A box in purple, and a Walker B box in green. Important conserved residues are indicated by red asterisks, and the locations of the primers are represented by pink arrows.

Figure 5.9. PCR amplification of a 719 bp fragment of *T. brucei RAD51*. PCR was performed on genomic DNA from 9 *T. brucei* strains from three different subspecies (indicated below the strain names). The reaction was also attempted on *T. vivax* genomic DNA. A no-DNA control was included. PCR products were separated on a 0.8% agarose gel.


3200
208
222
984
S14
427
927
386
T_vivax
L_major
10


3200
208
222
984
S14
427
927
386
T_vivax
L_major
11
12

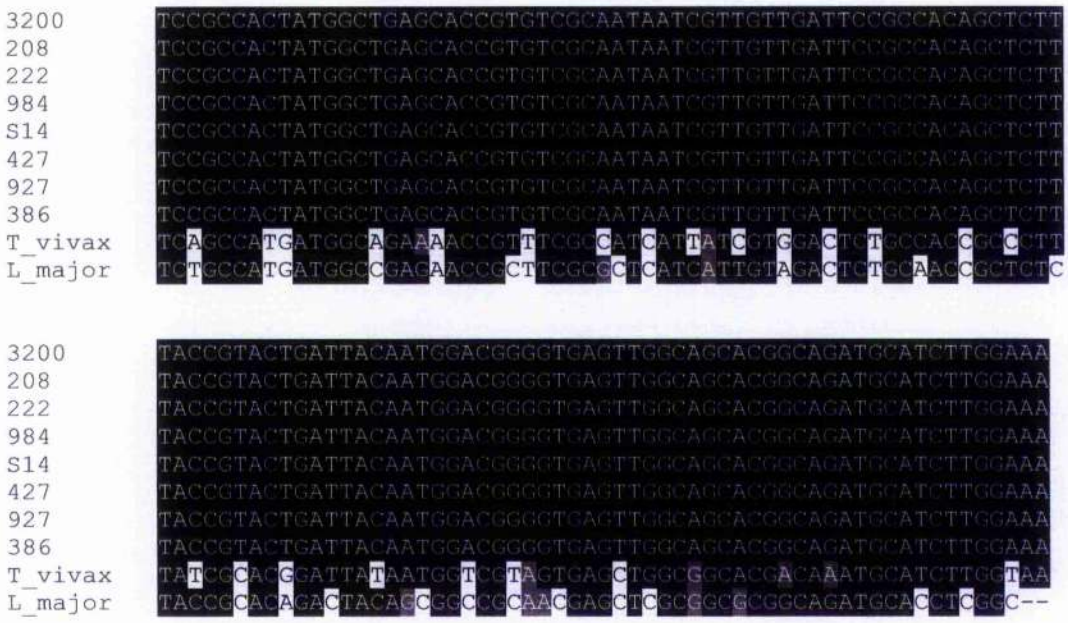


Figure 5.10. Multiple alignment of *RAD51* sequence obtained by PCR amplification using primers RAD51-BamHI and RAD51-U3 from 8 *T. brucei* strains and from the *T. cruzi* and *L. major* genome projects. Sequences were aligned using ClustalX (Thompson *et al.*, 1997) and shaded using the Boxshade server (http://ch.embnet.org/software/BOX_form.html): identical residues in 50% of the sequences are shaded in black, and conserved residues in grey. SNPs are numbered in blue.

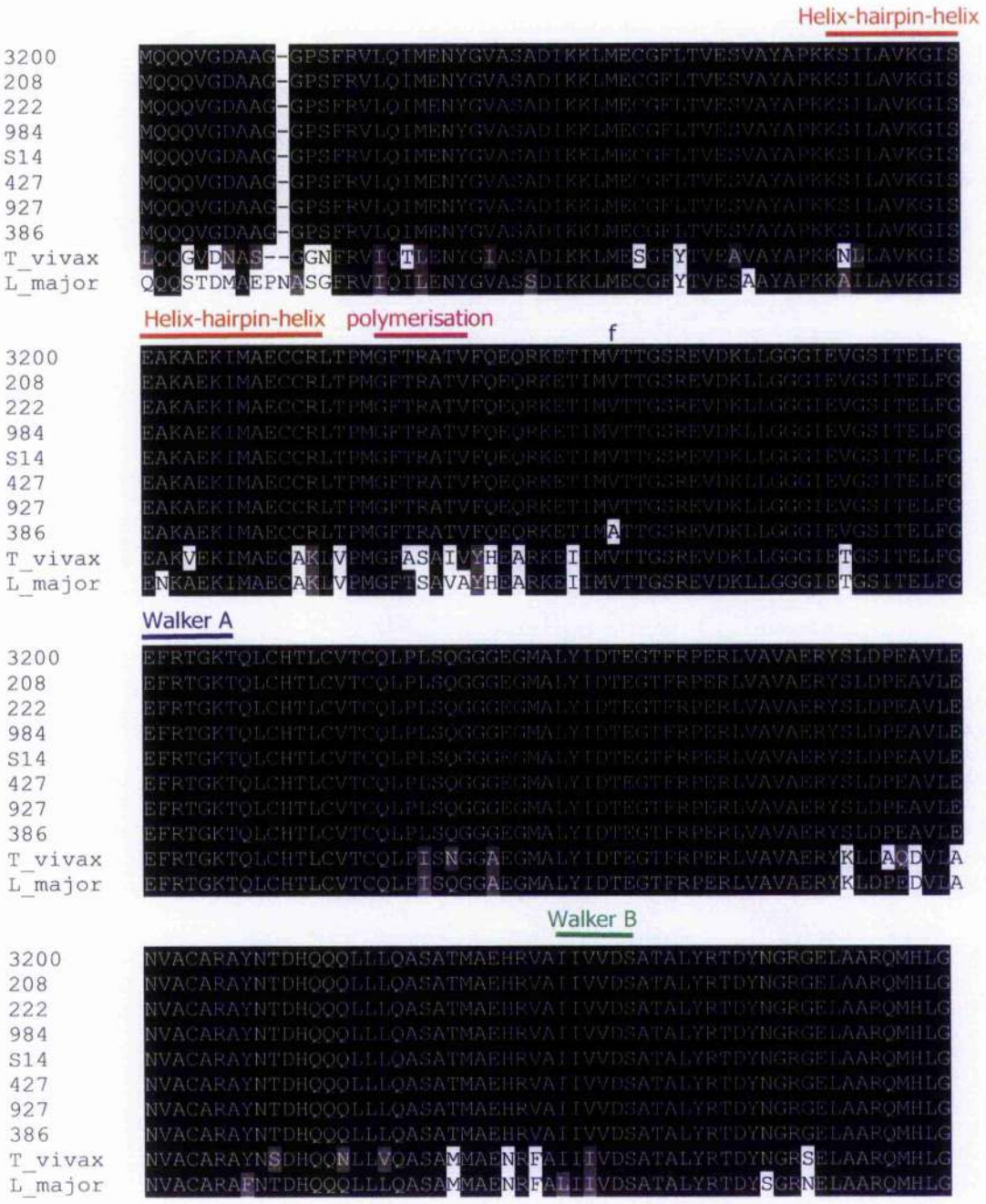


Figure 5.11. Multiple alignment of translations of the *RAD51* sequences shown in figure 5.10. Sequences were aligned using ClustalX (Thompson et al 1997) and shaded using the Boxshade server (http://ch.embnet.org/software/BOX_form.html): identical residues in 50% of the sequences are shaded in black, and conserved residues in grey. Protein sequence polymorphism is labelled with a blue letter.

(Proudfoot and McCulloch, 2005). As for *MSH2*, PCR was carried out using Herculase, a high-fidelity DNA polymerase, to amplify this fragment from genomic DNA of the nine isolates described in the previous section. The same primers were used to attempt to PCR-amplify the equivalent regions of *T. vivax* genomic DNA; in addition, a control PCR reaction was carried out using water instead of genomic DNA substrate. A small amount of each reaction was run on a 0.8% agarose gel, showing amplification of a single product of the expected size in all cases, apart from reaction using *T. vivax* genomic DNA as substrate, which was unsuccessful, and the no-DNA control reaction (figure 5.9). The PCR reaction using *T. b. gambiense* Eliane genomic DNA as a substrate was also unsuccessful, and therefore data from this strain is lacking in the following analysis. It is assumed that failure to amplify this section of the *T. vivax* gene reflects sequence divergence in the primer sequences, whereas failure to amplify this section of the *T. b. gambiense* gene is more likely simply due to mistakes setting up the reaction. These reactions were not repeated. All PCR products were incubated with Taq DNA polymerase prior to cloning into the TOPO TA vector (Invitrogen). At least two cloned PCR products from each strain were sequenced by the MBSU (University of Glasgow). These DNA sequences, as well as sequence from the *T. vivax* and *L. major* genome databases were compared by performing a sequence alignment using ClustalX (Thompson *et al.*, 1997), visualised using Boxshade (http://ch.embnet.org/software/BOX_form.htm) as shown in figure 5.10. The consensus sequence from all the successful sequencing reactions is shown.

Like the *MSH2* sequence, the amount of sequence divergence of *RAD51* is extremely low: polymorphism was only seen at three sites at the DNA level. At SNP 10, a single *T. brucei* isolate (STIB 386) contains a C residue, whereas all the other *T. brucei* isolates, as well as *T. vivax* and *L. major*, contain a T residue. At SNP 11, two *T. brucei* isolates contain a T residue, whereas all the other *T. brucei* isolates contain a G residue and *L. major* contains a C residue. At SNP 12, 4 *T. brucei* isolates (as well as *L. major*) contain a G residue, and five *T. brucei* isolates (as well as *T. vivax*) contain an A residue. Translations of each sequence were performed and the resultant polypeptides aligned as described previously for MSH2 (see figure 5.11). The only SNP that corresponded to a polymorphism at the protein level was SNP 10, with STIB 386 encoding an alanine whereas all the other *T. brucei* isolates (and *T. vivax* and *L. major*) encode valine. Although this falls in a region that is conserved with *T. vivax* and *L. major*, it does not affect any of the known functional domains of RAD51, and is not a residue conserved in RAD51 protein from other organisms (McCulloch and Barry, 1999). These results suggest that the lack of

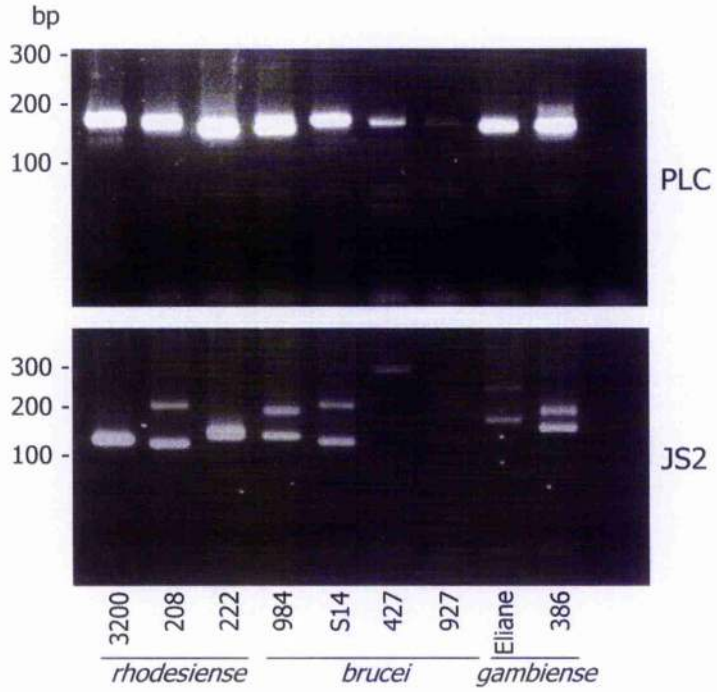


Figure 5.12. Amplification of the PLC and JS2 microsatellites. Microsatellites PLC and JS2 were amplified from 9 *T. brucei* strains as described in the text; the PCR products were then separated on an ethidium bromide-stained 3% Nusieve agarose gel.

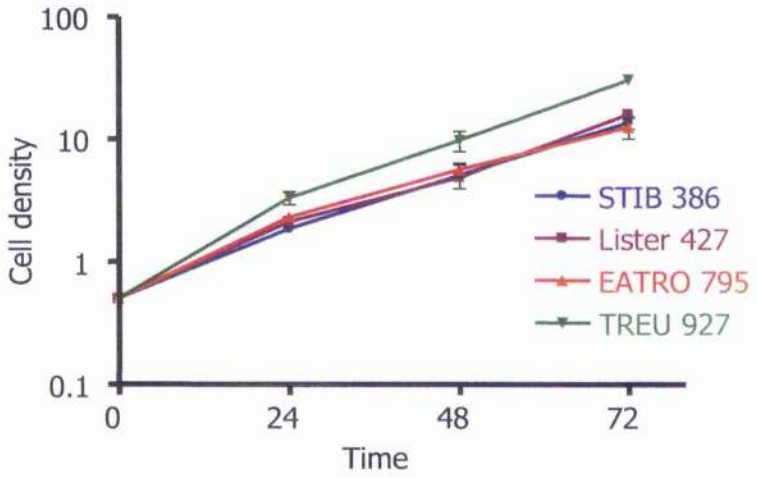
polymorphism is not specific to the *MSH2* gene, but extends at least to RAD51 also, and may be a general feature of coding sequences in the *T. brucei* genome.

5.2.1.4 Microsatellite analysis

To ensure that the lack of substantial sequence variation described above could not have arisen by cross-contamination of the genomic DNA samples, their purity was examined by the amplification of two microsatellites, named PLC and JS2 (see figure 5.12). These repetitive sequences composed of T-A and G-T dinucleotide repeats, respectively (Hope *et al.*, 1999) are located on chromosomes 2 and 5 respectively, and were PCR-amplified from genomic DNA using the primer pairs PLC-G and PLC-H3 and JS2-A and JS2-B, giving products between 100 bp and 300 bp. PCR was performed from all the *T. brucei* genomic DNA samples used in the previous analysis, using Taq DNA polymerase and a commercially made mixture of PCR reagents provided by A. MacLeod (see section 2.7.1). For the larger microsatellite, PLC, variation between the isolates is apparent, but difficult to resolve on this gel analysis. In contrast, for JS2, for most samples the different alleles were distinguishable, and size differences in one or both were discernible between the samples.

5.2.2 Study of a potential mutator phenotype in different strains and subspecies

The observed lack of differences in at least eight of the *MSH2* sequences analysed does not necessarily imply that MMR efficiency is equivalent in all strains, since only ~44% of the *MSH2* sequence has been examined, and potentially important mutations could have been missed in the N-terminal regions (Obmolova *et al.*, 2000). In addition, the other MMR genes have not been similarly analysed, and mutations in other factors are known to affect MMR (Bell *et al.*, 2004). For this reason, four strains for which procyclic form life cycle stage cells had been established in culture were assayed for MNNG, H_2O_2 and methyl methane sulfonate (MMS) sensitivity. The strains used were the *T. b. gambiense* strain STIB 386, and the *T. b. brucei* strains Lister 427, EATRO 795 and STIB 927. Although *MSH2* sequence in the EATRO 795 strain was not studied in the previous section, sequence from the ILTat 1.2 strain, derived from EATRO 795, was included in the analysis. Unfortunately, *T. b. rhodesiense* strain 208, for which an MSH2 mutation was putatively identified, was not available in the procyclic form. It was decided not to attempt this analysis using bloodstream stage cells, as the only strain available that can be grown readily in culture in the bloodstream form is Lister 427.



	Ave DT (h)	SE (h)
STIB 386	15.38	0.570
Lister 427	14.535	0.187
EATRO 795	16.29	0.997
TREU 927	12.23	0.056

Figure 5.13. Growth in culture of different *T. brucei* strains. Procyclic *T. brucei* of strains STIB 386, Lister 427, EATRO 795 and STIB 927 were grown in vitro from a starting density of 5 x 105 cells.ml-1 and cell concentration measured every 24 h. The log of cell concentration is shown against hours in culture.

5.2.2.1 Growth of different *T. brucei* strains and subspecies in culture

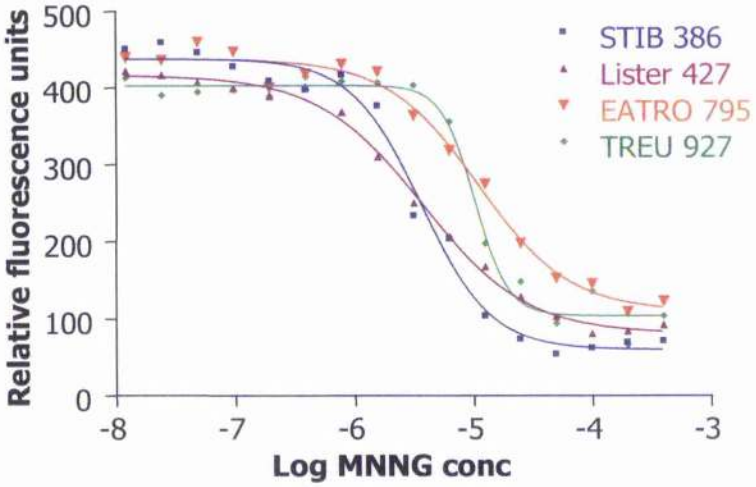
Before any drug sensitivity assays were carried out, the *in vitro* growth rate of procyclic form cells of strains STIB 386, Lister 427, EATRO 795 and TREU 927 was compared. Cells were diluted to a concentration of 5 x 10^5 cells.ml^{-1} in SDM-79 and the cell density was measured every 24 h. This was carried out in duplicate on two separate occasions. It is apparent from a log-linear plot of the cell concentrations (for averages and standard errors see figure 5.13) that although strains STIB 386, Lister 427 and EATRO 795 grew at approximately the same rate, population growth of TREU 927 was somewhat faster. The average population doubling times of the four strains in this analysis were 15.38 +/- SE 0.57 h for STIB 386, 14.535 +/- SE 0.187 h for Lister 427, 16.20 +/- SE 0.997 h for EATRO 795, and 12.23 +/- SE 0.056 h for TREU 927.

5.2.2.2 Tolerance of *T. brucei* strains and subspecies to MNNG

As described in section 4.2.7.4, tolerance to the alkylating agent MNNG is increased in cells with deficient MMR, including *T. brucei* and a range of prokaryotic and eukaryotic organisms. To assay the MNNG sensitivity of *T. brucei* strains STIB 386, Lister 427, EATRO 795 and TREU 927, procyclic form cells at a starting density of 5 x 10^5 cells.ml^{-1}, were grown in SDM-79 medium containing doubling dilutions of MNNG from 400 μM to 12.207 nM, in 96-well tissue culture plates (200 μl of cell culture per well). After 48 h, 20 μl of Alamar Blue (Resazurin; see section 4.2.7.4) was added to each well and fluorescence measured, using an LS55 luminescence spectrophotometer (Perkin Elmer) at an emission wavelength of 590 nm after a further 24 h of growth. Higher fluorescence is an indication of more metabolism of Alamar Blue and therefore more proliferating cells (see section 4.2.7.4). The fluorescence readings were plotted on a log-linear graph in relation to MNNG concentrations, and formed a sigmoidal curve as the drug impaired growth. An example of such a curve is shown in figure 5.14 a. This experiment was performed in duplicate four times for each strain, and IC50 values (*i.e.*, drug concentrations causing death of 50% of the cells) were then calculated from the fluorescence curves using Prism (GraphPad). These averages and standard errors are summarised in figure 5.14. In addition, data from section 4.2.7.4 on the tolerance of Lister 427 wild-type and *MSH2-/-* bloodstream stage cells were added to the analysis.

The four procyclic forms showed differences in their MNNG sensitivity. Average IC50 values (see figure 5.14 b) obtained were 3.959 +/- SE 1.4 μM for strain STIB 386, 4.998 +/- SE 1.767 μM for strain Lister 427, 10.703 +/- SE 3.784 μM for strain EATRO 795, and

a)



b)

	Ave IC50 (μM)	St Error (μM)
STIB 386 PCF	3.959	1.4
Lister 427 PCF	4.998	1.767
EATRO 795 PCF	10.7	3.784
TREU 927 PCF	14.921	5.276
Lister 427 wt BSS	3.277	1.258
Lister 427 MSH2-/- BSS	9.683	3.066

c)

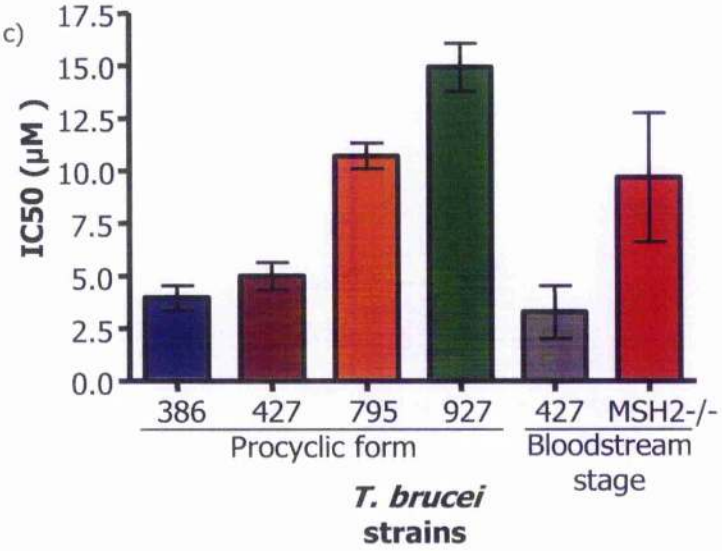


Figure 5.14. An alamar blue assay measuring the MNNG tolerance of different *T. brucei* strains. a) Procyclic *T. brucei* of strains STIB 386, Lister 427, EATRO 795 and STIB 927 were grown in 96-well plates in the presence of doubling dilutions of MNNG for 48 h. Alamar blue was added to each well and the cells were grown for a further 24 h before fluorescence was measured. A typical graph for this analysis is shown, in which fluorescence is plotted against the log of MNNG concentration. b) A table showing the average IC50 value for the four procyclic strains as well as wild-type and *MSH2-/-* bloodstream form cells, and standard errors calculated from 8 repetitions. c) a bar chart showing the average IC50 values from b).

14.921 +/- SE 5.276 μM for strain TREU 927, indicating that the latter two strains have 2.14-3.77-fold greater tolerance to MNNG in these conditions compared to the former two strains. This level of tolerance is, in fact, very similar to that seen when Lister 427 *MSH2-/-* bloodstream form cells (studied previously; see section 4.2.7.4), with an IC50 of 9.683 +/- SE 3.066 μM, are compared to wild-type Lister 427 bloodstream form cells (IC50 of 3.277 +/- SE 1.258 μM). Indeed, the level of sensitivity in STIB 386 and Lister 427 procyclic forms compare closely with the Lister 427 bloodstream stage cells, and the EATRO 795 and TREU 927 procyclic form cells are comparable to Lister 427 bloodstream stage *MSH2-/-* mutants. A one-way nonparametric analysis of variance test using Prism (GraphPad) of the mean IC50 values for procyclic form STIB 386 and Lister 427 and bloodstream stage Lister 427 wild-type cells gave a P value of 0.2944, and the same analysis for procyclic form EATRO 795 and TREU 927 and Lister 427 *MSH2-/-* bloodstream stage cells gave a P value of 0.0687. This means that the three strains in each group are not significantly different from each other in a 95% confidence interval. On the other hand, the same analysis comparing all six strains gave a P value of less than 0.0001, implying that all together, the strains are very significantly different from each other.

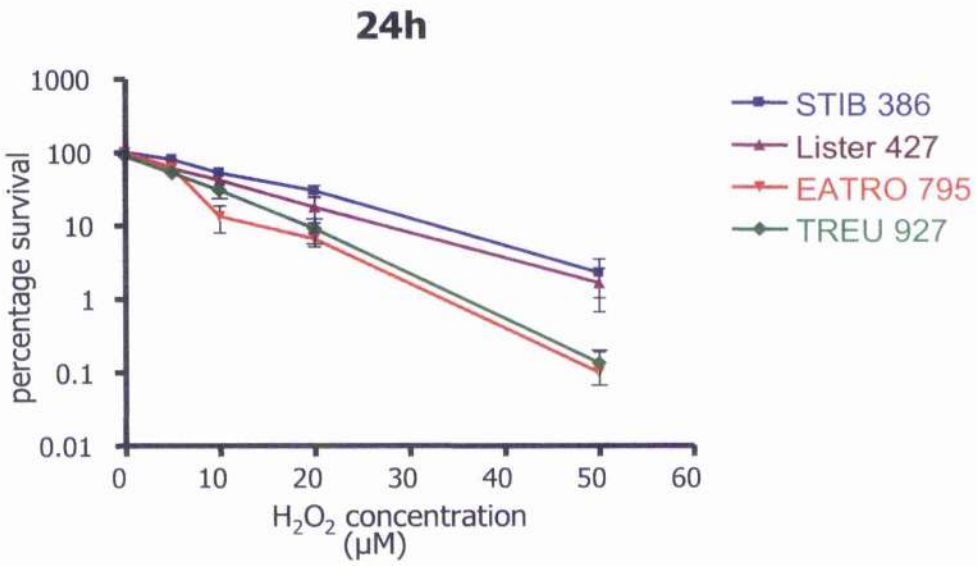
5.2.2.3 Tolerance of *T. brucei* strains and subspecies to H_2O_2

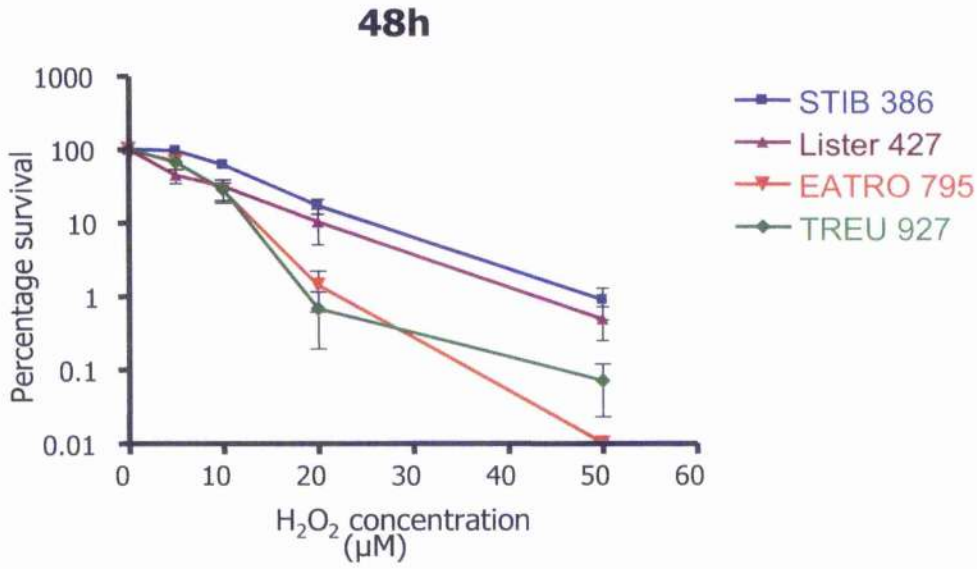
The most common lesion caused, both within DNA and in the cellular dNTP pool, by the alkylating agent H_2O_2 (and by naturally occurring oxidative damage in cells) is 7,8-dihydro-8-oxoguanine, commonly abbreviated to 8-oxoG (Wiseman and Halliwell, 1996; Slupphaug *et al.*, 2003). In DNA, this base is highly mutagenic as it pairs equally efficiently with A or C, causing GC -> TA transversions (Thomas *et al.*, 1997). Three specialised mechanisms exist in bacteria to deal with this damage (Slupphaug *et al.*, 2003; Fowler *et al.*, 2003; Barnes and Lindahl, 2004). In one strategy, the MutT enzyme prevents incorporation of 8-oxoGTP from the dNTP pool into DNA by converting it to GMP (Maki and Sekiguchi, 1992). The two other strategies involve two different glycosylases of the base excision repair (BER) machinery, triggering removal of a base from contaminated DNA: MutY initiates repair of A residues incorporated opposite 8-oxoG (Au *et al.*, 1989), whereas MutM initiates repair of 8-oxoG paired with C (Chung *et al.*, 1991). Mammalian genomes contain homologues of MutT, MutY and MutM, whereas *S. cerevisiae* has only a functional homologue of MutM, known as 8-oxoG glycosylase or OGG-1 (Fortini *et al.*, 2003). Kinetoplastid genomes also contain a homologue of MutM/OGG-1, the function of which has been verified in *T. cruzi* (Farez-Vidal *et al.*, 2001; Pena-Diaz *et al.*, 2004), and two MutY homologues are identifiable amongst the

factors involved in BER (El Sayed *et al.*, 2005), though they appear to lack a homologue of MutT. Together, this appears to indicate that all 8-oxoG in kinetoplastids is dealt with following its incorporation into DNA.

MMR has also been implicated in the repair of these lesions; 8-oxoG paired with A or C is recognised as a mismatch (Ni *et al.*, 1999; Mazurek *et al.*, 2002a), with the A or the 8-oxoG respectively being removed (Wyrzykowski and Volkert, 2003). Furthermore, incorporation of 8-oxoG from the dNTP pool causes genetic instability in MMR-deficient human cells (Russo *et al.*, 2004), and this mutator phenotype can be reduced when the pool is "sanitised" by overexpression of the MutT homologue MTH1 (Colussi *et al.*, 2002). Similarly, the mutator phenotype caused by mutation of MTH1 is partially alleviated by the MMR system (Egashira *et al.*, 2002). Both MSH2 and MLH1 are involved in MMR-mediated removal of 8-oxoG from mouse DNA, both before and after H_2O_2 treatment (Colussi *et al.*, 2002). Although some work (Colussi *et al.*, 2002) suggests that OGG1 and MMR functions are independent of each other, other observations point to a link between these pathways, with the MutY homologue MYH interacting with PCNA (which could act as a coordinator between the pathways; (Hayashi *et al.*, 2002)), and the MSH6 component of the MSH2-MSH6 heterodimer interacting with, and stimulating, the DNA binding and glycosylase activities of, MYH (Gu *et al.*, 2002; Mazurek *et al.*, 2002b). MSH3, on the other hand, does not seem to be involved in the repair of this oxidative damage (Gu *et al.*, 2002; Mazurek *et al.*, 2002a).

In *T. brucei*, it is known that mutation of *MSH2* causes increased sensitivity to H_2O_2 treatment, although preliminary results suggest that disruption of *MLH1* has no effect (A. da Silva Machado, pers. comm.). This appears to be compatible with the findings described in mammals, suggesting that MSH2 (possibly in partnership with MSH6) has a role in the repair of this damage, perhaps in signalling to the BER machinery. This could be especially important due to the putative lack of a MutT homologue in kinetoplastid genomes. To examine if the four *T. brucei* strains tested for MNNG sensitivity also displayed differential tolerance to H_2O_2, the percentage survival in the presence of increasing amounts of the alkylating agent was assayed (assay optimised by A. da Silva Machado, pers. comm.), in a comparable way to studies on mammalian cells (DeWeese *et al.*, 1998). Growth of procyclic form *T. brucei* from strains STIB 386, Lister 427, EATRO 795 and STIB 927 in the presence of a range of concentrations of H_2O_2 was measured by diluting the parasites from a culture in the log phase of growth (~5 x 10^6 cells.ml^{-1}) to 5 x 10^5 cells.ml^{-1} in SDM-79 medium containing 0, 5, 10, 20 or 50 μM of H_2O_2. The cell density was then measured every 24 h. This was carried out in duplicate on three separate


24h
1000
100
10
1
0.1
0.01
percentage survival
0
10
20
30
40
50
60
H_2O_2 concentration (μM)
STIB 386
Lister 427
EATRO 795
TREU 927


48h
1000
100
10
1
0.1
0.01
Percentage survival
0
10
20
30
40
50
60
H_2O_2 concentration (μM)
STIB 386
Lister 427
EATRO 795
TREU 927

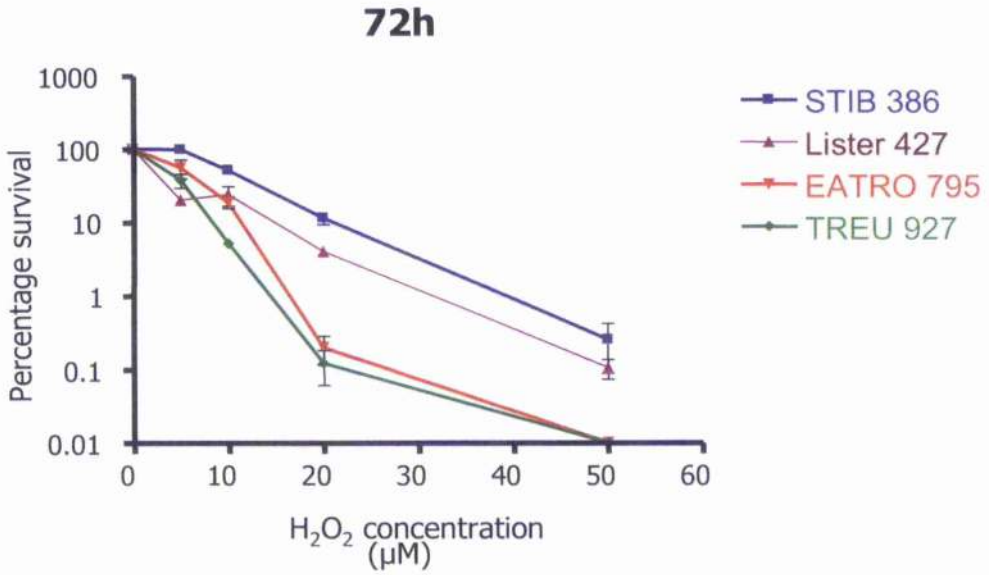


Figure 5.15. Growth in culture of different *T. brucei* strains in the presence of H_2O_2. Procyclic *T. brucei* of strains STIB 386, Lister 427, EATRO 795 and STIB 927 were grown *in vitro* in the presence of different concentrations of H2O2, from a starting density of 5 x 105 cells.ml-1 and cell concentration measured every 24 h. Log of percentage survival compared to the no-H2O2 control are shown against the H_2O_2 concentration (in µM) for 24 h, 48 h and 72 h.

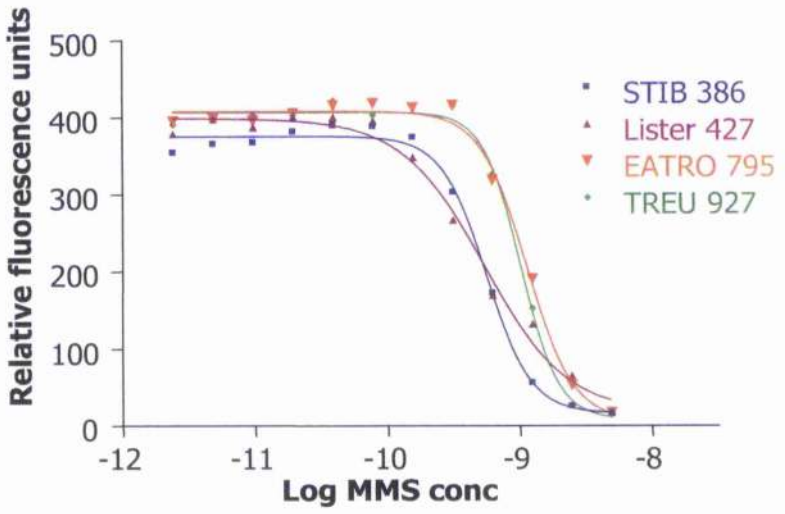
occasions. The percentage survival for each drug concentration was calculated by comparing cell density in the presence of H_2O_2 to the control sample which had not been treated at each timepoint, and the average survival rate is shown in a log-linear plot in figure 5.15.

In all cases, cell viability was impaired by H_2O_2 treatment, since even after 24 h at the lowest H_2O_2 concentration (5 μM), the cell numbers had decreased in comparison to the control culture. However, differences in H_2O_2 sensitivity between the strains were observed. When viewed on a log-linear scale, the strains STIB 386 and Lister 427 had detectably higher survival rates than EATRO 795 and STIB 927 after 24 h at the highest concentration of H_2O_2 (50 μM), and after 48 or 72 h treatment with 20 and 50 μM H_2O_2. These differences follow the same strain groupings as were seen assaying MNNG sensitivity, indicating potentially a common source for both phenotypes. Two-tailed nonparametric t-tests were performed on the data from the highest H_2O_2 concentration (50 μM) at the three timepoints. The P values from comparisons of STIB 386 and Lister 427 strains were 0.9372, 0.4875 and 1, respectively for 24 h, 48h and 72 h, and P values from comparisons of EATRO 795 and TREU 927 were 0.1797, 0.5556 and 1, respectively. This means that the two strains in each group are not significantly different from each other in a 95% confidence interval. One-way nonparametric analysis of variance was also carried out on all four strains, giving P values of 0.0147, 0.0117 and 0.00343, respectively for the three timepoints, implying that all together, the strains are significantly different from each other. Due to the fact that different concentrations of H_2O_2 gave separation of the wild-type and *MSH2-/-* Lister 427 bloodstream stage cells, a direct comparison with these data is not possible.

5.2.2.4 Tolerance of *T. brucei* subspecies and strains to MMS

Methyl methane sulfonate (MMS) is an alkylating agent like MNNG; however, this DNA damaging agent creates predominantly N-methylated bases, leading to blocks in replication (Nowosielska *et al.*, 2006). Homologous recombination is needed to bypass these stalling events and avoid replication fork collapse, creating DSBs. Defects in HR enzymes, including *T. brucei* RAD51 (McCulloch and Barry, 1999), two of the four RAD51 paralogues (Proudfoot and McCulloch, 2005) and BRCA2 (C. Hartley, pers. comm.), are known to cause an increase in MMS sensitivity. Inactivation of MMR is known to result in increased tolerance to alkylating agents such as MNNG and to the compound cisplatin, but to have no reproducible effect on the sensitivity of *mlh1* mutant mammalian cells to other DNA damaging agents, such as ionising radiation, cross-linking agents, or

a)



b)

	Ave IC50 (% x 10^{-4})	St Error (% x 10-4)
STIB 386 PCF	4.55	0.662
Lister 427 PCF	5.633	0.296
EATRO 795 PCF	17.025	2.988
TREU 927 PCF	9.95	0.185

c)

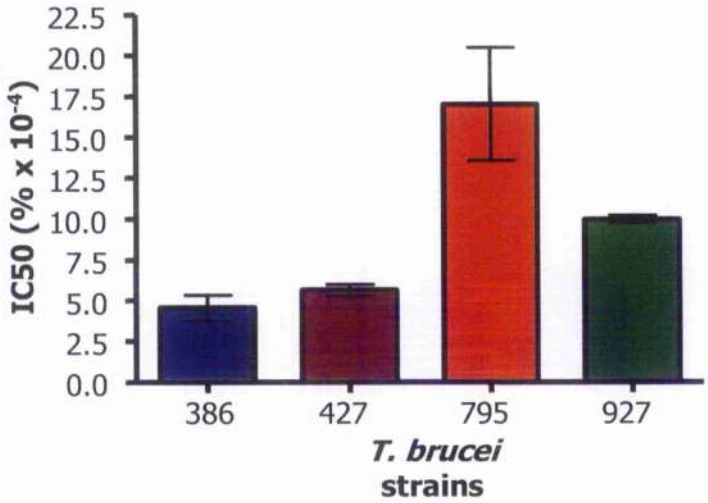


Figure 5.16. An alamar blue assay measuring the MMS tolerance of different *T. brucei* strains. a) Procyclic *T. brucei* of strains STIB 386, Lister 427, EATRO 795 and STIB 927 were grown in 96-well plates in the presence of doubling dilutions of MNNG for 48 h. Alamar blue was added to each well and the cells were grown for a further 24 h before fluorescence was measured. A typical graph for this analysis is shown, in which fluorescence is plotted against the log of MNNG concentration. b) A table showing the average IC50 value for the four procyclic strains, and standard errors calculated from 8 repetitions. c) a bar chart showing the average IC50 values from b).

topoisomerase inhibitors (Cejka *et al.*, 2005). Little has been studied on the effect that MMS has on MMR-deficient cells, but one report indicates that mutation of MLH1, MSH6 or MSH2 causes increased resistance/tolerance to MMS, indicating that MMS cytotoxicity is influenced by MMR (Glaab *et al.*, 1998; Glaab *et al.*, 1999).

T. brucei strains STIB 386, Lister 427, EATRO 795 and TREU 927 were assayed for MMS tolerance. Procyclic stage cells of these four strains, at a starting density of 5×10^5 cells.ml^{-1}, were grown in SDM-79 medium containing doubling dilutions of MMS from 0.005% to 0.00000244%, in 96-well tissue culture plates (200 µl cell culture per well). After 48 h, 20 µl of Alamar Blue was added to each well and fluorescence measured after a further 24 h of growth, as in section 4.2.7.4. This experiment was performed in duplicate twice. An example of a sigmoidal curve with IC50 values plotted against MMS concentrations is shown in figure 5.16 a.

The four strains showed differences in MMS sensitivity. Average IC50 values (see figure 5.16 b) obtained were 0.000455 +/- SE 0.0000662 % for strain STIB 386 and 0.0005633 +/- SE 0.0000296 % for strain Lister 427, compared with 0.001703 +/- SE 0.000299 % for strain EATRO 795, and 0.000995 +/- SE 0.0000185 % for strain TREU 927. The STIB 386 and Lister 427 cells therefore display 2-3-fold greater sensitivity to MMS than TREU 927 and EATRO 795, respectively. Allow this follows broadly the groupings described for the MNNG and H_2O_2 assays, in this analysis EATRO 795 appears more tolerant than TREU 927 to MMS, which contrasts somewhat with previous results as no difference was observed between the two strains in H_2O_2 sensitivity, and TREU 927 was, if anything, more tolerant to MNNG than EATRO 795 (see figure 5.16). That the EATRO 795 strain displays an intermediate phenotype is confirmed by statistical analysis. A two-tailed nonparametric t-test was performed on strains STIB 386 and Lister 427, giving a P value of 0.4857. This means that the two strains in each group are not significantly different from each other in a 95% confidence interval. However, the same analysis for strains EATRO 795 and TREU 927 gave a P value of 0.0286. One-way nonparametric analysis of variance of all four samples gave a P value of 0.0009, and for strains STIB 386, Lister 427 and EATRO 795 gave a P value of 0.021 implying that all together, the strains are significantly different from each other.

5.3 Discussion

1533 bp of DNA, corresponding to the ATPase domain, part of the middle conserved domain and the region of the gene in-between these conserved structural features, was PCR-amplified and sequenced in three segments from the *MSH2* gene of nine different *T. brucei* strains taken from the three subspecies, *T. b. rhodesiense*, *T. b. brucei* and *T. b. rhodesiense*. Multiple alignments of these sequences reveal remarkably low levels of sequence divergence between these strains, with only 9 putative SNPs at the DNA level (summarised in table 5.2), corresponding to 5 putative polymorphisms at the protein level. In fact, this may be an overestimate, as in 3 of these 5 cases the protein sequence of only a single strain was different from the rest, pointing to potential sequencing or PCR errors. Only one of the polymorphisms at the protein level was located within a conserved structural motif (the N-3 nucleotide binding domain of the ATPase domain), although this polymorphism was found in a single strain and requires confirmation from an independent PCR reaction. Irrespective of the validity of these polymorphisms, the level of variation is in contrast to the situation in *T. cruzi*, where analysis of only 829 bp of sequence from different isolates revealed 21 SNPs in the ATPase domain alone, 5 of which cause amino acid changes (Augusto-Pinto *et al.*, 2003). A 715 bp section of the *T. brucei RAD51* ORF was also PCR-amplified, and was shown to contain similarly low levels of polymorphism, with only 3 SNPs and a single amino acid change, again in a single strain. This sequence similarity was not due to contamination of the reagents or primers or of the genomic DNA samples, and appears to indicate that the *T. brucei* subspecies and strains are more closely related than those of *T. cruzi*.

Literature searches do not reveal any other examples of coding sequence having being compared in different *T. brucei* strains; previous work has centred on microsatellites and other polymorphic features of the genome to establish genomic relationships (*e.g.*, Truc and Tibayrenc, 1993). On the other hand, in *T. cruzi* a number of coding sequences have been studied. For example, 46 polymorphic loci are found in 290 bp of sequence from the cytochrome B ORF (de Freitas *et al.*, 2006), and a 1030 bp section of the genome containing the *tcp17* and *tcpgp2* genes contains 83 variable nucleotide positions (Robello *et al.*, 2000). Part of the *RAD51* ORF has also been studied in *T. cruzi*, revealing only 3 SNPs in 359 bp of DNA sequence (Carlos Renato Machado, pers. comm.). Taken together, it seems likely that the levels of polymorphism in the *T. cruzi MSH2* ORF are representative of polymorphism across the genome, whereas the lower levels of divergence in *RAD51* reflect greater functional constraints on this enzyme. *T. cruzi* is a highly

	MSH2-1	MSH2-2	MSH2-3	MSH2-4	MSH2-5	MSH2-6	MSH2-7	MSH2-8	MSH2-9	RAD51-10
A				4	6	3				
C	5	4	2		4		5	5	1	1
G		6		6		4		1		
T	5		8				1		9	9

Table 5.2. A table summarising the SNPs found in the sections of the *MSH2* and *RAD51* ORFs that were sequenced in 10 *T. brucei* strains. The number of strains with each nucleotide at that position is shown.

divergent species, and the genetic distance between subdivisions of the species has been estimated as 4-fold greater than between human and chimpanzee (Tibayrenc, 1995); in fact, there is some debate as to whether it should be classified as a single species at all (Buscaglia and Di Noia, 2003). No such work has allowed an estimate of divergence between *T. brucei* subspecies to be made. Both MSH2 and RAD51 contain a number of well-conserved domains, and future sequence analysis of further genes would be needed to provide a picture of divergence between ORFs in general in *T. brucei*.

The comparison of *MSH2* sequences between *T. brucei* subspecies and strains did not reveal strong evidence for the selection of MMR mutants, in contrast to the findings in *T. cruzi*. In only one case (polymorphism e, in *T. b. gambiense* strain 208) was a single mutation identified that fell within a conserved domain, whilst all other polymorphisms were in non-conserved residues. Further work would be required to assess whether the single changed MSH2 affects MMR. However, despite this lack of clearly selected mutants, differences in drug sensitivity were revealed that may be consistent with alterations in MMR.

MMR efficiency was assayed indirectly in four *T. brucei* strains: STIB 386, Lister 427, EATRO 795 and TREU 927. Tolerance to MNNG has been shown to be increased in *T. brucei* cell lines deficient in either *MSH2* or *MLH1* (Bell *et al.*, 2004), as has been seen in other organisms (see section 4.2.7.4). Although the original analysis of MNNG tolerance used a plating assay and bloodstream stage cells, characterisation of *MSH2-/-* cells using the Alamar Blue assay, as described in section 4.2.7.4, confirms the finding, and demonstrates that this approach can be used to examine the sensitivity of *T. brucei* to drugs such as this. Strains STIB 386 and Lister 427 displayed approximately a 3-fold lower tolerance for MNNG than strains EATRO 795 and TREU 927, and the IC50 values for the former strains (roughly 4 μM) are similar to the values for wild-type bloodstream-stage Lister 427 cells, whereas the IC50 values for the latter (9 μM or higher) are similar to mismatch repair-deficient bloodstream stage *MSH2* mutants. This appears to indicate that drug uptake in bloodstream stage and procyclic form Lister 427 cells is very similar, as is the metabolism of Alamar Blue, which has been used as a measure of cell viability (Wallace *et al.*, 2002). The differences in MNNG tolerance in the different *T. brucei* strains are not simple to reconcile with differences in cellular proliferation, since TREU 927 appeared to divide more rapidly than the other strains, whereas EATRO 795 was very similar to STIB 386 and Lister 427.

The same subdivision of the four strains was also seen when assaying H_2O_2 sensitivity. Whereas mutation in any element of the MMR machinery can lead to MNNG tolerance (Bell *et al.*, 2004), H_2O_2 tolerance is hypothesised to be dependent on MSH2 and MSH6 (see section 5.2.2.3) (but not MSH3; Ni *et al.*, 1999; Berardini *et al.*, 2000) detecting integration of 8-oxoG into DNA. Whatever the exact mechanism of MMR involvement, it is known that H_2O_2 tolerance is decreased in bloodstream-stage *T. brucei* in Lister 427 *MSH2-/-* relative to wild-type cells, and that preliminary results suggest that tolerance remains unchanged in Lister 427 *MLH1-/-* cells (A. Machado da Silva and R. McCulloch, pers. comm.). The higher survival rates of strains STIB 386 and Lister 427 may indicate greater activity or levels of MutS homologues in these strains relative to EATRO 795 and TREU 927. However, H_2O_2 metabolism is also carried out by the mitochondria, a cellular function that could also vary between strains. This cellular activity can be measured using the compound 3[4,5-dimethylthiazol-2-yl]-2,5-diphenyltetrazolium bromide (MTT), a water-soluble tetrazolium salt that is cleaved to formazan by the succinate dehydrogenase system of active mitochondria, causing a colour change from yellow to purple that can be measured by a spectrophotometer (Dreiem *et al.*, 2005). Although some degradation of H_2O_2 is also carried out by other, cytosolic factors (Jezek and Hlavata, 2005), an MTT assay would be valuable to determine how much of this reactive oxygen species is degraded, and therefore how much remains in the cell to cause DNA damage requiring repair by the MMR and BER systems.

Finally, tolerance of the four *T. brucei* strains to the alkylating agent MMS was also measured, and the same differential effects in EATRO 795 and TREU 927 relative to STIB 386 and Lister 427 were found. *T. brucei* cell lines defective in RAD51 (McCulloch and Barry, 1999), two of its four paralogues (Proudfoot and McCulloch, 2005) and BRCA2 (C. Hartley, pers. comm.) suffer from increased MMS sensitivity. In contrast, some work has reported that defects in MMR in human cells result in increased MMS tolerance. The MMS tolerance of *T. brucei* EATRO 795 is approximately 4-fold greater than STIB 386 and Lister 427, whereas the MMS tolerance of TREU 927 is approximately 2-fold greater. MMS tolerance in bloodstream stage Lister 427 *MSH2-/-* and *MLH1-/- T. brucei* has not yet been measured, but this experiment would be vital for the full interpretation of these results. As measurement of MMS sensitivity in other cell lines was carried out by a plating assay, direct comparisons are not possible. The extremely low levels of polymorphism in the *RAD51* gene make it unlikely that the difference in sensitivity is due to differential enzymatic activity in this protein, though variation in other DSB repair genes cannot be excluded. Nevertheless, the increased resistance to MMS in two of the strains, as for the effect seen with MNNG and H_2O_2, may indicate MMR deficiency. Though in human cells

	"wild-type"?			"mutator"?			
	STIB 386 PCF	Lister 427 PCF	Lister 427 BSS	EATRO 795 PCF	TREU 927 PCF	MSH2-/- BSS	MLH-/- BSS
MNNG	sensitive	sensitive	sensitive	tolerant	tolerant	tolerant	tolerant
Hydrogen peroxide	tolerant	tolerant	tolerant	sensitive	sensitive	sensitive	tolerant (?)
MMS	sensitive	sensitive	???	tolerant	tolerant	???	???

Table 5.3. A table summarising the responses of the four procyclic strains, as well as Lister 427 bloodstream stage wild-type and MMR-deficient cell lines, to MNNG, MMS and hydrogen peroxide.

this phenotype appears to be a direct consequence of MMR mutation (Cejka *et al.*, 2005), it is also possibly that the differences in tolerance could result from secondary mutations as a consequence of MMR impairment.

These experiments have shown that *T. brucei* strains show clear separation in their tolerance to the three alkylating agents MNNG, H_2O_2 and MMS, with EATRO 795 and TREU 927 showing statistically significantly different tolerance to EATRO 386 and Lister 427 (with the exception of MMS, where EATRO 795 had an intermediate phenotype). This grouping is compatible with a mutator phenotype in strains EATRO 795 and TREU 927 compared to STIB 386 and Lister 427 (see table 5.2). This potential MMR deficiency is unlikely to be caused by mutations in the *MSH2* ORF, as a significant proportion of the gene, containing the well-conserved ATPase domain, contain very low levels of sequence polymorphism. The only polymorphism identified that could be predicted to affect MSH2 protein function was only found in *T. b. rhodesiense* strain 208 (which is not available in culturable form), but not ILTat 1.2 (which is derived from EATRO 795) or TREU 927. A mutator phenotype could be caused by divergence in the sequence of another MMR protein. H_2O_2 sensitivity has been reported in *S. cerevisiae* to be affected by MSH6, but not MSH3 (Gu *et al.*, 2002; Mazurek *et al.*, 2002a), making MSH6 a more likely candidate to have been mutated in this cases. Similarly, in *S. cerevisiae*, mutation of MMR factor *EXO1* has no effect on MNNG tolerance (Cejka *et al.*, 2005), making this an unlikely that this gene has been mutated. It is also possible that a mutation in the regulatory regions, such as the sequences responsible for *trans* splicing and polyadenylation of the nascent transcript, of *MSH2* or another MMR gene could disrupt expression (Clayton, 2002). Dozens of genes, involved in a range of replication and DNA repair processes, have been identified in *E. coli* as causing mutator phenotype when mutated (Horst *et al.*, 1999). However, the effects of mutations in factors unrelated to MMR on tolerance to these alkylating agents could not be predicted.

Although the data recorded here are consistent with certain strains having a mutator phenotype, this is not the only possible explanation. Differential uptake for all three compounds could lead to different amounts of damage being caused by a constant amount of drug. Although H_2O_2 is a small, uncharged particle with poor reactivity that can pass through membranes easily (Jezek and Hlavata, 2005), MNNG and MMS are bulkier and may require specific uptake into the cell. This is another cellular function that could vary between *T. brucei* subspecies and strains (especially in a mutator strain). However, how this could be studied further without knowledge of the exact mechanism of uptake is hard to imagine. Treatment with alkylating agents such as MNNG also leads to MMR-

dependent cell cycle arrest at G2 phase (O'Brien and Brown, 2006), in a signalling cascade dependent on ATR (Stojic *et al.*, 2004a; Stojic *et al.*, 2004b). A disruption in this pathway could also lead to differences in the survival of different strains.

The question whether this truly represents a mutator phenotype could be answered in three further ways. Firstly, the tolerance of the same *T. brucei* strains to drugs such as phleomycin, or ionising radiation, could be measured in order to assess whether this is a generalised phenomenon, or limited to the alkylating agents studied so far. Secondly, microsatellite instability of the different strains could be compared. Microsatellite instability is a phenotype associated with MMR deficiency; their repetitive nature leads to slippage of DNA polymerase during replication, leading to the formation of insertion-deletion loops (Sia *et al.*, 1997) that are usually repaired by the MMR system (Buermeyer *et al.*, 1999). It is known that deficiencies of either *MSH2* or *MLH1* causes an increase in this mutator phenotype in cloned *T. brucei* bloodstream stage cells (Bell *et al.*, 2004). Thirdly, adaptation of the different strains to stressful conditions, for example sublethal drug pressure, could be studied. This may allow any selective advantage conferred by a potential mutator phenotype to become apparent.

A number of different approaches could be taken to further understand the processes causing the observed differences in drug tolerance. The sequencing of a wider range of MMR and other DNA repair genes would be possible, if time-consuming, as over 15 spontaneous mutator genes have been identified in bacteria (Miller and Michaels, 1996) and the number in eukaryotic cells is likely to be much higher. However, other experimental approaches could be more effective in narrowing down the factors involved. Northern and/or quantitative RT-PCR analysis could be carried out on all MMR factors, as well as other important DNA repair factors such as *RAD51* or *MRE11*, and would identify any changes in expression that could affect DNA repair. Decrease or even increase (Yang *et al.*, 2004) in the expression levels of MMR genes can cause mutator phenotypes. Indeed, modulation of expression levels is adopted as a strategy for a temporary increase in mutation rate in stationary-phase bacteria (see section 5.1.1). Genetic mapping could also be undertaken to understand the basis of these traits. Hybrid cloned trypanosomes and genetic maps are available from crosses of the STIB 247 strain with STIB 386 and TREU 927 (MacLeod *et al.*, 2005b). If it could be established that STIB 247 follows the trend of discrete "sensitive" or "tolerant" phenotypes in response to MNNG or MMS, it should be possible to perform mapping of the phenotype to one or several chromosomal region(s), potentially identifying the cause of the difference (if it has a discrete genetic basis). If this mapping approach, or indeed expression analysis, points to candidate factors for

differential tolerance to the drugs, then sequencing, genetic inactivation and/or complementation analysis would be necessary to confirm this finding.

If the results described in this chapter do reveal that 2/4 *T. brucei* strains do in fact have deficient MMR function, what are the implications for the biology of this parasite? Although the sample size is extremely small, the proportion of putative mutators is extremely high compared to the situation in bacteria; this corresponds to the highest proportion of mutator isolates reported from *Pseudomonas* infections of cystic fibrosis patients (Macia *et al.*, 2005). The groupings observed for the putative mutator phenotype do not depend on the subspecies of the isolate studied: the *T. b. gambiense* strain STIB 386 and the *T. b. brucei* strain Lister 427 had essentially identical phenotypes. This is in contrast to the situation in *T. cruzi*, where all strains can be grouped into clades according to MSH2 sequence polymorphisms (Augusto-Pinto *et al.*, 2003). The advantages to these kinetoplastid parasites that would be afforded by a mutator phenotype are not clear. STIB 386 is a strain of the virulent type 2 *T. b. gambiense* subgroup, and yet appears not to have a mutator phenotype, arguing against a direct link to virulence. It is possible that a mutator phenotype could aid *T. brucei* cells survive in the very different mammalian and tsetse hosts, or facilitate the recombination between divergent sequences that is necessary for antigenic variation. It is not known whether other systems vary in a similar way between *T. brucei* strains. The mechanism of fixation of these mutator phenotypes is equally mysterious; an allelic genetic difference causing a defect in MMR function would have to be spread to the other allele by loss of heterozygosity (in the case of a recessive mutation) before it could spread through a population by second-order selection. However, the population bottlenecks implicit in the antigenic variation system could aid in the fixation of mutations accidentally, reducing the requirement for second-order selection.

To summarise, this work demonstrated extremely low levels of sequence divergence in the *T. brucei MSH2* and *RAD51* ORFs, in contrast to the related kinetoplastid *T. cruzi*, where polymorphisms in the MSH2 ATPase domain are found, and may correlate to differences in MMR efficiency (Augusto-Pinto *et al.*, 2003). In spite of this conservation, differences in tolerance to the alkylating agents MNNG, H_2O_2 and MMS can be seen in different *T. brucei* strains. This could have a number of causes, including polymorphisms in MMR or other repair factors, differential expression of DNA repair proteins, or differential uptake of the DNA damaging agents into the cell. Further work will be needed to resolve these questions.

CHAPTER 6

PERSPECTIVES

6 Perspectives

Prior to this PhD thesis, it had already been shown that the protozoan parasite *T. brucei* possesses a fully functional, standard eukaryotic MMR system (Bell *et al.*, 2004), which, although competent in the regulation of homologous recombination (HR), has no effect on antigenic variation, at least in monomorphic Lister 427 cells (Bell and McCulloch, 2003). The experiments described in this thesis looked at three further roles of *T. brucei* MMR factors in DNA metabolism.

Firstly, the requirements for substrate length and sequence homology in *T. brucei* HR were studied using an assay based on the transformation of linear constructs homologous to a unique site in the *T. brucei* genome. Recombination efficiency was examined over the range 25 bp to 200 bp. This work showed that both substrate length and homology influenced transformation efficiency both in MMR-competent and MMR-deficient cell lines. However, the MMR system only appeared to regulate recombination between moderately (5%) mismatched substrates, with recombination catalysed by perfectly matched and 11% divergent substrates remaining unaffected. The data in this chapter add to previous work by Conway *et al.* (Conway *et al.*, 2002b), and suggest that the presence of two RAD51-independent DNA repair pathways in *T. brucei*. The first is thought to catalyse homologous recombination using shorter stretches of homology than the RAD51-dependent pathway and becomes extremely inefficient below 50 bp, while the microhomology-mediated end-joining (MMEJ) pathway of DSB repair uses only a few base pairs of homology, in the range 5 – 10 bp. Further experiments will be required to test this hypothesis.

The *T. brucei* genome does not contain a Rad52 homologue (El Sayed *et al.*, 2005). In other organisms, this protein binds to Rad51 *via* its N-terminus (Milne and Weaver, 1993), acting as a multimeric ring structure (Singleton *et al.*, 2002), and has roles in facilitating the formation of the Rad51 nucleoprotein filament (Sung, 1997a; Benson *et al.*, 1998). Although Rad52 is found in many eukaryotes (Wu *et al.*, 2006), it has differing importance in HR depending on the organism. In *S. cerevisiae*, Rad52 is extremely important for all pathways of HR, and mutation leads to a 3000-fold reduction in recombination, but it has much less impact on recombination in mammalian cells, and disrupting it has a much smaller effect (Rijkers *et al.*, 1998; Yamaguchi-Iwai *et al.*, 1998b). Fungal species including *S. cerevisiae* (Bai and Symington, 1996) and *Kluyveromyces lactis* (van den *et al.*, 2001) also contain a Rad52 paralogue known as Rad59, which is required for Rad51-

independent recombination and anneals to Rad52 *in vivo* (Davis and Symington, 2001). It is possible that Rad52 is more important in yeast than in mammals because of the relative lack of Rad51 paralogues in these species. *S. cerevisiae* only has two Rad51 paralogues, Rad55 and Rad57, which facilitate nuclear filament formation (Sung, 1997b). Similarly, the fungus *Ustilago maydis* has two Rad51 paralogues (Symington, 2002). On the other hand, mammalian cells contain five Rad51 paralogues (Thompson and Schild, 2001), which appear to have taken on a range of roles in HR (Symington, 2002). In *T. brucei*, in addition to a DMC1 homologue (Proudfoot and McCulloch, 2006), four RAD51 paralogues have been identified. It could be postulated that some of these Rad51 paralogues have taken on the roles of Rad52, causing its relative lack of importance in mammalian cells and rendering it unnecessary in trypanosomatids, where it has been discarded. Only two of the *T. brucei* RAD51 paralogues that have been studied so far, and both appear to act in conjunction with RAD51 (Proudfoot and McCulloch, 2005), although this does not rule out them catalysing other classes of DSB repair as well. The remaining two RAD51 paralogues are being studied at present and this will allow us to have a more complete picture of their role in general DSB repair, and in *VSG* switching.

The genome of *T. brucei*, as well as those of *T. cruzi* and *L. major*, contains orthologues of the meiosis-specific MutS homologues *MSH4* and *MSH5*. Sequence analysis shows that the *T. brucei* genes lack a detectable mismatch interaction domain, in common with those from other organisms. Analysis of *MSH4* and *MSH5* expression showed that transcription of both genes is not restricted to the epimastigote stage where *T. brucei* meiosis is thought to take place (Gibson and Whittington, 1993; Gibson and Bailey, 2003); *MSH5*, though not *MSH4*, RNA was detectable by northern blot in procyclic form and bloodstream stage cells. Attempts to make genetic knockouts of these genes in bloodstream stage cells were not successful, in contrast to other organisms (Ross-Macdonald and Roeder, 1994; Hollingsworth *et al.*, 1995; Zalevsky *et al.*, 1999; Edelmann *et al.*, 1999; de Vries *et al.*, 1999b). Integration of an ectopic copy of the *MSH4* or *MSH5* into the repetitive tubulin array of Lister 427 bloodstream stage cells lead to an increase in *MSH5* mRNA, but *MSH4* RNA remained undetectable, suggesting strong post-transcriptional control for *MSH4* in this life cycle stage. This introduction of the ORFs into an ectopic locus had no effect on MMR, measured indirectly by MNNG sensitivity, or population doubling time.

MSH4 and MSH5 are suggested to be part of a functional meiotic machinery in *T. brucei*. The genomes of the three sequenced kinetoplastids have retained a number of factors involved in different stages of meiosis (see table 4.1). An inventory of meiotic genes (Villeneuve and Hillers, 2001) in the yeast *Candida albicans* (Tzung *et al.*, 2001) validated

this bioinformatic approach when evidence of meiotic recombination was found in this organism (which had previously been thought not to have a sexual cycle; (Tavanti *et al.*, 2004). The retention of meiotic genes in the kinetoplastids and in the protist pathogen *Giardia intestinalis* (Ramesh *et al.*, 2005) points to an ancient origin for the meiotic process. *G. intestinalis* is thought to have adopted the strategy of facultative sex (Dacks and Roger, 1999), with long periods of asexual reproduction punctuated by sexual exchange. Whether or not the same strategy is used in the kinetoplastids is unclear. However, conservation of meiotic factors in all three kinetoplastids may argue against a non-meiotic mechanism of genetic exchangein *T. cruzi*, as suggested by Gaunt *et al.* (Gaunt *et al.*, 2003), and raises the question of whether or not *L. major* is capable of meiosis.

MSH4 and MSH5 are two of many possible meiotic factors found in the *T. brucei* genome. The only other one to have been characterised is the RAD51 homologue DMC1, which can be found in *T. brucei*, *T. cruzi* and *L. major* (Proudfoot and McCulloch, 2006). Like *MSH5*, *T. brucei DMC1* expression can be detected in bloodstream stage trypanosomes by northern analysis (see figure 4.17). This is in contradiction to observations in other organisms, where expression of *DMC1* is limited to meiotic cells and tissues, although it is not known whether the presence of RNA corresponds to the presence of mature protein in either case. However, disruption of the *DMC1* ORF had no effect on MMS sensitivity, transformation efficiency or *VSG* switching (Proudfoot and McCulloch, 2006). While these findings cannot conclusively show meiotic function, this analysis also does not disprove the hypothesis of meiotic recombination in *T. brucei*. Further analysis looking at the putative meiotic life cycle stage of *T. brucei* is necessary to understand the importance of the individual meiotic factors discussed above. Disruption of these genes followed by analysis of genetic exchange in the tsetse fly using drug resistance and fluorescence markers as described by the Gibson group (Bingle *et al.*, 2001; Gibson and Bailey, 2003) would perhaps provide a clearer picture, although the non-essential nature of *T. brucei* genetic exchange would complicate analysis.

All three kinetoplastids appear to encode two copies of *SPO11*. Although most organisms have a single copy of this gene (Grelon *et al.*, 2001), this situation is not unprecedented; in *A. thaliana* there are three SPO11 orthologues (Grelon *et al.*, 2001). These are as different to each other as to SPO11 in other organisms (Hartung and Puchta, 2000), implying that they did not arise from a recent duplication. Both SPO11-1 and SPO11-2 are expressed in somatic tissues (SPO11-1 in an unusually high number of isoforms; (Hartung and Puchta, 2000). Deletion of *SPO11-1* leads to a large reduction in meiotic recombination (Grelon *et*

al., 2001), meaning that the three proteins are not redundant with each other; SPO11-2 and SPO11-3 have been suggested to be involved in recombination in non-meiotic organs (Grelon *et al.*, 2001). In the kinetoplastids, it is tempting to speculate that the presence of two *SPO11* copies could have allowed similar functional diversification. For example, could one of the SPO11 proteins be involved in the creation of DSBs at the initiation stages of *VSG* switching? Although its presence in *T. cruzi* and *L. major* argue against this, experimental analysis is needed to address this question.

The sequence of MSH2 and RAD51 from different *T. brucei* strains and subspecies was studied, and very low levels of divergence were found. This is in contrast to *T. cruzi*, where five polymorphisms were found at the amino acid level in the ATPase domain alone; these polymorphisms have been proposed to modulate MMR function in the same way as bacterial mutator strains (Augusto-Pinto *et al.*, 2003). Despite this, differences were found between strains to the drugs MNNG, H_2O_2 and MMS. This could be due to differences in MMR, another DNA repair pathway, or drug uptake. MMR proteins have been linked to the ATR kinase pathway and apoptotic response to alkylation damage of mammalian DNA (Li, 1999; Yoshioka *et al.*, 2006). MMR proteins are thought to act upstream of the ATR kinase (Wang and Qin, 2003), leading to activation of a number of downstream targets. Both MutS and MutL homologues are required for apoptosis induction following MNNG treatment in human cells (Stojic *et al.*, 2004b). The roles of *T. brucei* MMR proteins in damage sensing and cell cycle control (Stojic *et al.*, 2004a) need to be examined in more detail in order to understand their wider roles in cellular metabolism, in addition to the more classical MMR functions that have been the subject of this thesis.

Appendices

Appendix 1. Oligunucleotide primer sequences used for PCR.

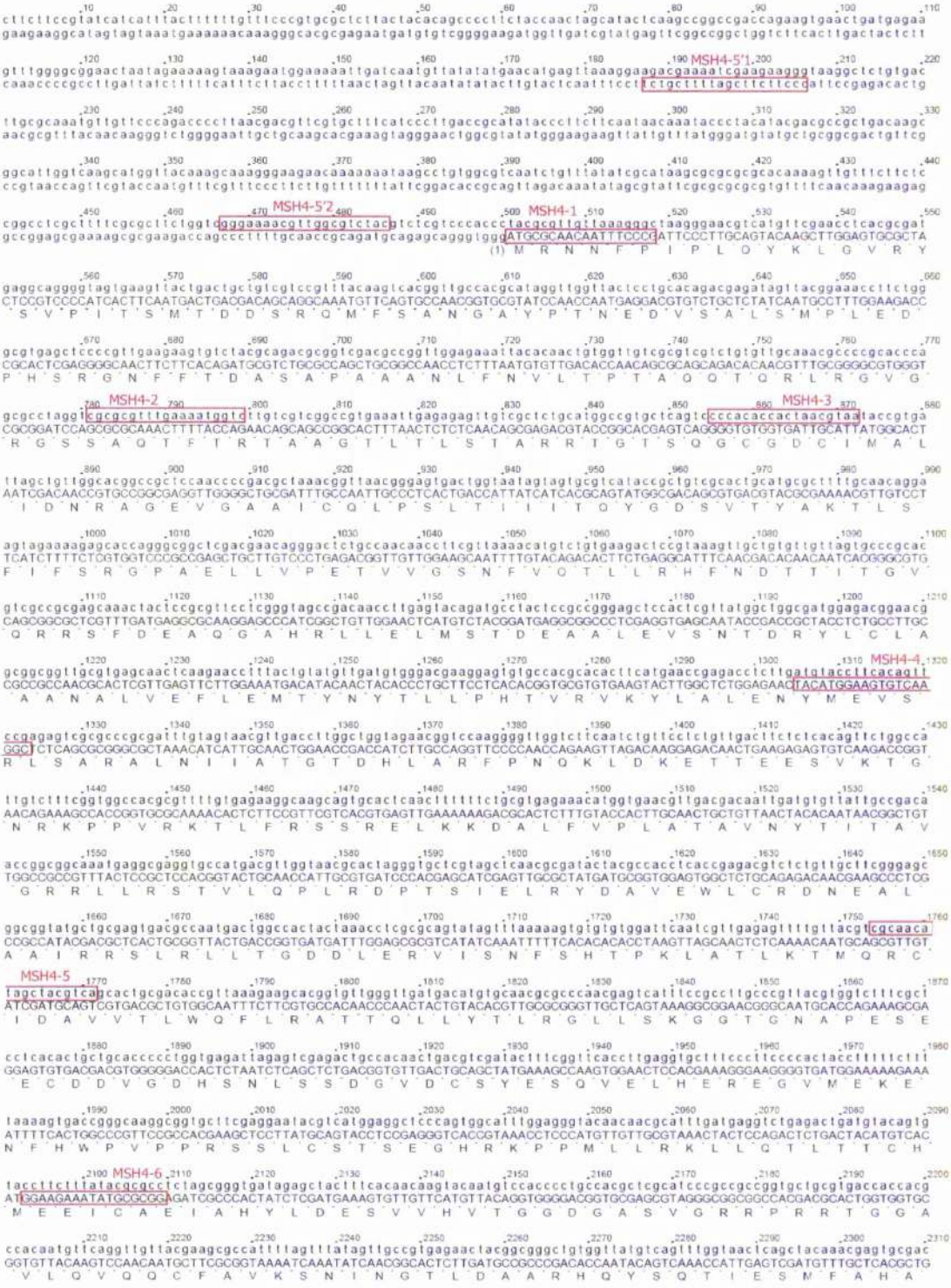
MSH4-7	ccactgtgataaagtgcg
MSH4-8	tattttattgaggcgcgccaccc
MSH4-9	gcgtcgcaaatattgttcgg
MSH4-10	ttgctgacaaggtagttgcgg
MSH4-11	aatagggggttgaacacgc
MSH4-3'1	ccccccggatccttgagaggtatcgtcgaggg
MSH4-3'2	cccccctcgagcctctccaaacaaaagtgcc
MSH4-5'1	cccccctctagatctgcttttagcttcttccc
MSH4-5'2	cccccggatcccatctgcggttgcaaaaggg
MSH5-1	tgtgtggtttgtaaaggggc
MSH5-2	ctctcttcacattagccc
MSH5-3	aatgcaacaaaagcgctgcg
MSH5-4	actaatcttcgacgctccc
MSH5-5	acgcacaaagtagtgtaccc
MSH5-6	atgggagtggaaaatcgg
MSH5-7	aaatgttcgcagagaagcc
MSH5-8	aagcgaacgacatgaatcgg
MSH5-9	acttcgagagcattttgccc
MSH5-10	aacaccgattttccactccc
MSH5-3'1	cccccggatccgtgagtgtgtttgaataccg
MSH5-3'2	cccccctcgaggatgtcgttatgcatcaccc
MSH5-5'1	cccccctctagatcggaaacaacactgcaccg
MSH5-5'2	cccccggatccaaccttcgtccataaagccc
MSH2_ATPase_5	caccccccttgttgaactgcggcagcc
MSH2_ATPase_3	cgccacgtacaatccgtagctgcg
MSH2middle-1	cgctgtactacgggtatggg
MSH2middle-2	ccttcacaagattctgacgc
MSH2middle-3	cgtcagaatcttgtgaaggc
MSH2middle-4	ctgccgcagttcaacaaggg
MSH4overexp3'Mlu	ccccccacgcgttcactcagtagaggaaccg
MSH4overexp5'Mlu	ccccccacgcgtatgcgcaacaatttcccg
MSH5overexp3'Mlu	ccccacgcgttcaaacacactcactgg
MSH5overexp5'Mlu	ccccacgcgtatggacgaaggtttggatg
PolI-3	catgcgcctgtggttcagcatagc
PolI-5	caggaggatcgttcggcaccttggc
PLC-G	caacgacgttggaagagtgtgaac
PLC-H3	ccactgacctttcatttgatcgctttc
Rad51BamHI	ccggatccatgcagcagcaagttggtga
Rad51-U3	tttccaagatgcatctgccg
SL	gtttctgtactatattg

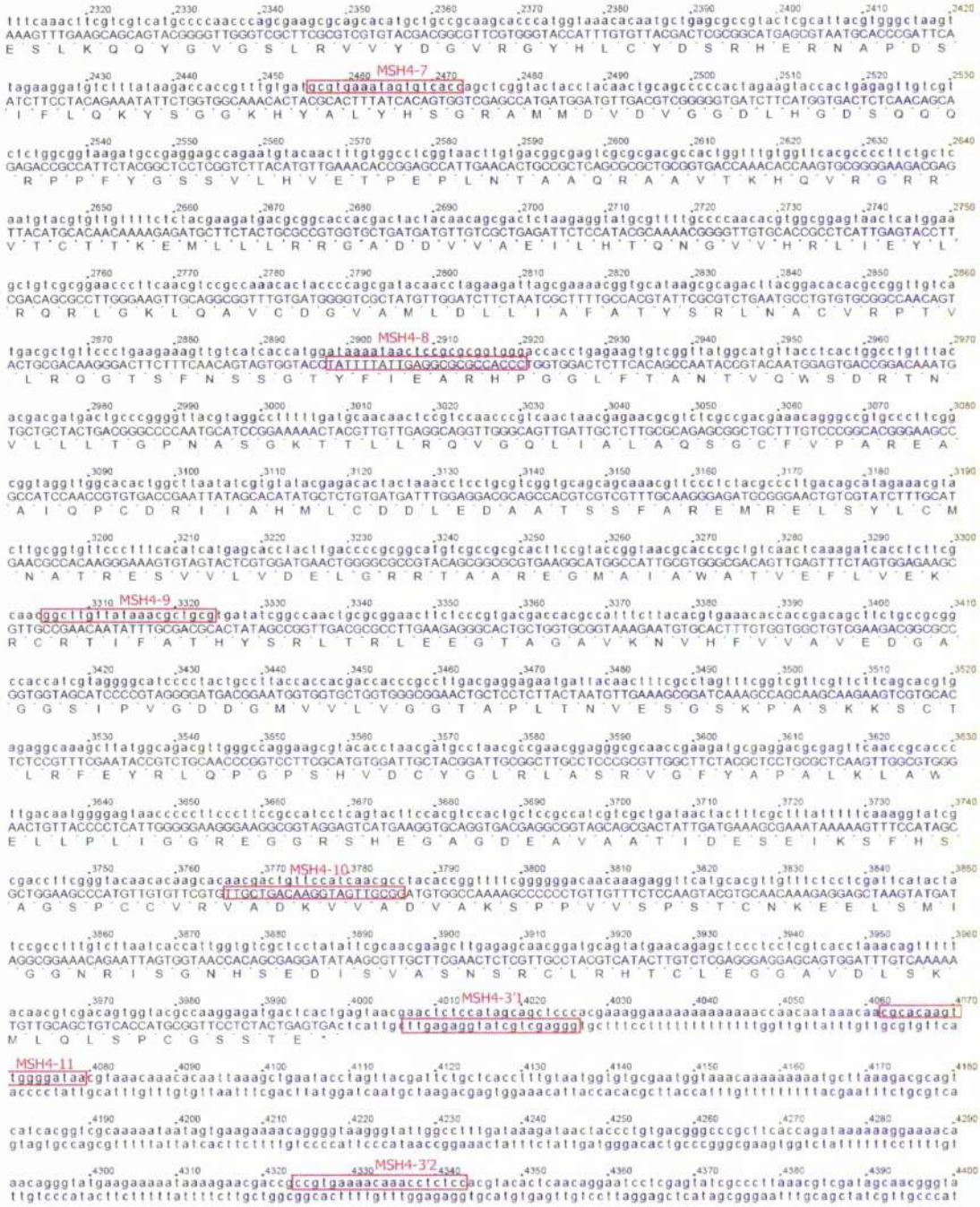
20bp-5'-0%	ggtcaatacactacatggcg
20bp-5'-11%	ggtcgatacactacatggcg
20bp-3'-0%	tgatacacatggggatcagc
20bp-3'-11%	tgatgcacatagggatcagc
50bp-5'-0%	cccattcggaccgcaagg
50bp-5'-11%	cccatccggaccgtaagg
50bp-3'-0%	gtcgtccatcacagtttgcc
50bp-3'-11%	gtcgcccatcgcggtttgcc
100bp-5'-0%	aggccatggatgcgatcgc
100bp-5'-11%	aggccctggatgcgaccgc
100bp-3'-0%	cccaaagcatcagctcatcg
100bp-3'-11%	cccaaggcgtcagctcgccg
150bp-5'-0%	gcaagacctgcctgaaaccg
150bp-5'-11%	gcaagacctgcctgagaccg
150bp-3'-0%	ttgttggagccgaaatccgc
150bp-3'-11%	ttgctggagccggagtccgc
200bp-5'-0%	gcgagagcctgacctattgc
200bp-5'-11%	gcgagggcccgacctactgc
200bp-3'-0%	ctcgctccagtcaatgaccg
200bp-5'-11%	ctcgccccagccagcggccg
Act 3' BamHI	ccggatcctattttatggcagcaacgagacc
βα 5' BamHI	ccggatcctgggtcccattgtttgcctc
βatubulin5'-3'	gccccgacaacttcatctttgga
βatubulin3'-5'	tttcgcatcgaacatctgctgcg
Ble-PstI	ccctgcagaattcatggccaagttgaccagtgcc
Ble-SphI	cccgtacgatatctcagtcctgctcctcggcc
Hyg3'	ctattcctttgccctcggac
Hyg5'	atgaaaaagcctgaactcacc
Hygprobe-3'	tcaagcacttccggaatcgg
JS2-A	gattggcgcaacaactttcacatacg
JS2-B	ctttcttccttggccattgttttactat
Midbleo	tccagaactcgaccgc
MRE11probe-5'	ccctcgatagggtgaagggatgtg
MRE11probe-3'	gccgttaagacagatcaagatgag
MSH4-1	atgcgcaacaatttcccg
MSH4-2	ctggtaaaagtttgcgcgc
MSH4-3	aatgcaatcaccacaccc
MSH4-4	tacatggaagtgtcaaggc
MSH4-5	actgcatcgatacaacgc
MSH4-6	ggaagaaatatgcgcgg

Appendix 2. Accession numbers of genes and proteins used during homology and phylogenetic analysis.

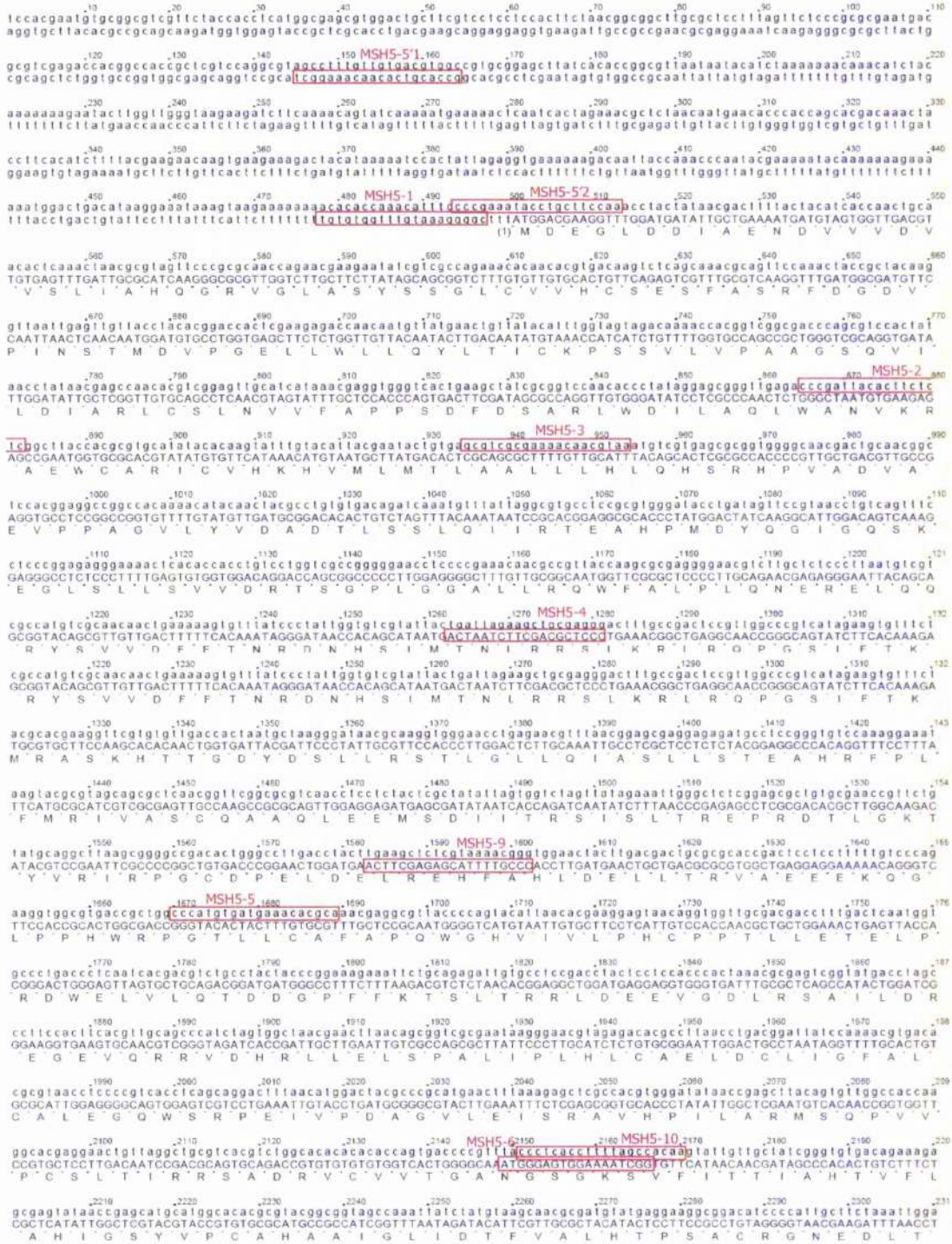
Protein	Species	Accession number
MutS1	*Escherichia coli*	AAL33627
MutS1	*Thermus aquaticus*	QS6215
MutS2	*Haemophilus ducreyi*	AAP6731
MutS2	*Helicobacter pylori*	O24338
MSH2	*Arabidopsis thaliana*	AAB81282
MSH2	*Caenorhabiditis elegans*	AAC78226
MSH2	*Homo sapiens*	NP_000242
MSH2	*Leishmania major*	CAJ05982
MSH2	*Mus musculus*	NP_032654
MSH2	*Saccharomyces cerevisiae*	CAA99102
MSH2	*Trypanosoma brucei*	AAK08648
MSH2	*Trypanosoma cruzi*	AAM1471
MSH3	*Arabidopsis thaliana*	NP_194284
MSH3	*Caenorhabiditis elegans*	AAA80443
MSH3	*Homo sapiens*	AAB47281
MSH3	*Leishmania major*	CAJ03372
MSH3	*Mus musculus*	AAH40784
MSH3	*Saccharomyces cerevisiae*	CAA42247
MSH3	*Trypanosoma brucei*	CAJ03372
MSH3	*Trypanosoma cruzi*	XP_820278
MSH4	*Arabidopsis thaliana*	NP_193469
Him-14	*Caenorhabiditis elegans*	Q23405
MSH4	*Homo sapiens*	AAB72039
MSH4	*Leishmania major*	CAJ04376
MSH4	*Mus musculus*	NP_114076
MSH4	*Saccharomyces cerevisiae*	NP_116652
MSH4	*Trypanosoma brucei*	XP_822385
MSH4	*Trypanosoma cruzi*	XP_815283
MSH5	*Caenorhabiditis elegans*	Q19272
MSH5	*Homo sapiens*	BAB63375
MSH5	*Leishmania major*	XP_847837
MSH5	*Mus musculus*	AAL14462
MSH5	*Saccharomyces cerevisiae*	CAA66337
MSH5	*Trypanosoma brucei*	XP_825491
MSH5	*Trypanosoma cruzi*	XP_809831
MSH6-1	*Arabidopsis thaliana*	O04716
MSH6	*Caenorhabiditis elegans*	AAK95890
MSH6	*Homo sapiens*	AAK21215
MSH8	*Leishmania major*	CAJ09151
MSH6	*Mus musculus*	AAB88445
MSH6	*Saccharomyces cerevisiae*	NP_010382
MSH8	*Trypanosoma brucei*	XP_822883
MSH6	*Trypanosoma cruzi*	XP_821503
MSH6-2	*Arabidopsis thaliana*	AAM13399
Gene	**Species**	**Accession number**
MSH4	*Trypanosoma brucei*	XM_817292
MSH5	*Trypanosoma brucei*	AAX79333

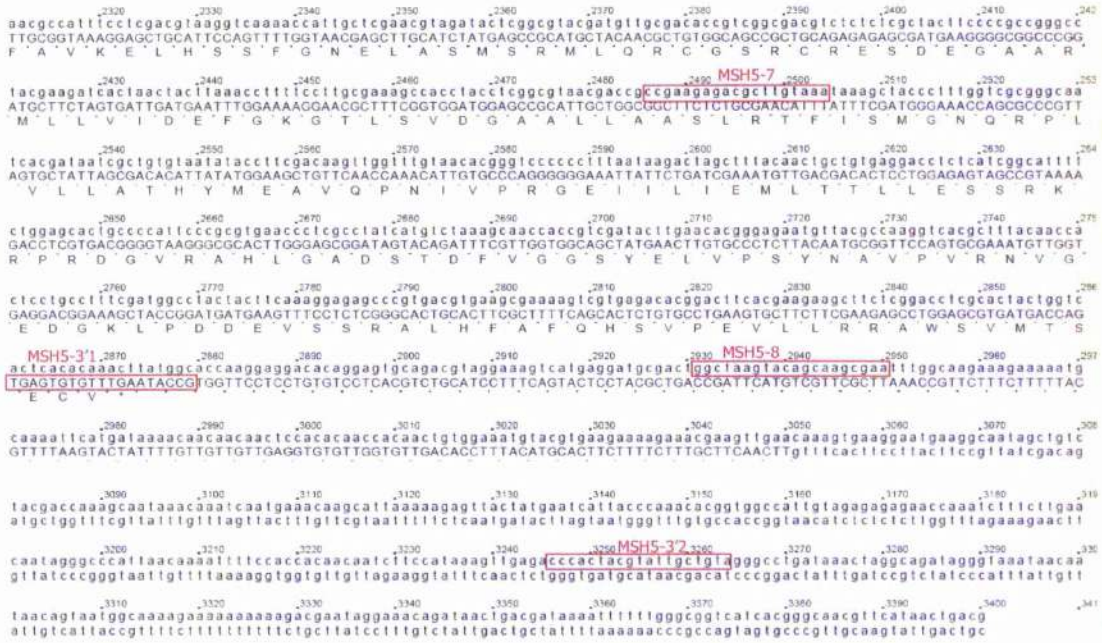
Appendix 3. The gene sequence for *MSH4*. Primers used for sequencing and amplification of targeting flanks are shown in pink.




MSH4-7
MSH4-8
MSH4-9
MSH4-10
MSH4-3'1
MSH4-11
MSH4-3'2

Appendix 4. The gene sequence for *MSH5*. Primers used for sequencing and amplification of targeting flanks are shown in pink.




MSH5-7
MSH5-3'1
MSH5-8
MSH5-3'2

References

Abdulkarim, F. and Hughes, D. (1996). Homologous recombination between the tuf genes of Salmonella typhimurium. *J.Mol.Biol.* **260**, 506-522.

Abdullah, M. F., Hoffmann, E. R., Cotton, V. E. and Borts, R. H. (2004). A role for the MutL homologue MLH2 in controlling heteroduplex formation and in regulating between two different crossover pathways in budding yeast. *Cytogenet.Genome Res.* **107**, 180-190.

Acharya, S., Foster, P. L., Brooks, P. and Fishel, R. (2003). The coordinated functions of the E. coli MutS and MutL proteins in mismatch repair. *Mol.Cell* **12**, 233-246.

Acharya, S., Wilson, T., Gradia, S., Kane, M. F., Guerrette, S., Marsischky, G. T., Kolodner, R. and Fishel, R. (1996). hMSH2 forms specific mispair-binding complexes with hMSH3 and hMSH6. *Proc.Natl.Acad.Sci.U.S.A* **93**, 13629-13634.

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

Adam, R. D. (2000). The Giardia lamblia genome. *Int.J.Parasitol.* **30**, 475-484.

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

Ahmed, S. A., Gogal, R. M., Jr. and Walsh, J. E. (1994). A new rapid and simple non-radioactive assay to monitor and determine the proliferation of lymphocytes: an alternative to [3H]thymidine incorporation assay. *J.Immunol.Methods* **170**, 211-224.

Ahn, B. Y., Dornfeld, K. J., Fagrelius, T. J. and Livingston, D. M. (1988). Effect of limited homology on gene conversion in a Saccharomyces cerevisiae plasmid recombination system. *Mol.Cell Biol.* **8**, 2442-2448.

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

Alani, E., Sokolsky, T., Studamire, B., Miret, J. J. and Lahue, R. S. (1997). Genetic and biochemical analysis of Msh2p-Msh6p: role of ATP hydrolysis and Msh2p-Msh6p subunit interactions in mismatch base pair recognition. *Mol.Cell Biol.* **17**, 2436-2447.

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.

Alexiadis, V. and Kadonaga, J. T. (2002). Strand pairing by Rad54 and Rad51 is enhanced by chromatin. *Genes Dev.* **16**, 2767-2771.

Allen, D. J., Makhov, A., Grilley, M., Taylor, J., Thresher, R., Modrich, P. and Griffith, J. D. (1997). MutS mediates heteroduplex loop formation by a translocation mechanism. *EMBO J.* **16**, 4467-4476.

Allers, T. and Lichten, M. (2001). Differential timing and control of noncrossover and crossover recombination during meiosis. *Cell* **106**, 47-57.

Amiguet-Vercher, A., Perez-Morga, D., Pays, A., Poelvoorde, P., Van Xong, H., Tebabi, P., Vanhamme, L. and Pays, E. (2004). Loss of the mono-allelic control of the VSG expression sites during the development of Trypanosoma brucei in the bloodstream. *Mol.Microbiol.* **51**, 1577-1588.

Amin, N. S., Nguyen, M. N., Oh, S. and Kolodner, R. D. (2001). exo1-Dependent mutator mutations: model system for studying functional interactions in mismatch repair. *Mol.Cell Biol.* **21**, 5142-5155.

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

Antony, E. and Hingorani, M. M. (2004). Asymmetric ATP Binding and Hydrolysis Activity of the Thermus aquaticus MutS Dimer Is Key to Modulation of Its Interactions with Mismatched DNA. *Biochemistry* **43**, 13115-13128.

Argueso, J. L., Kijas, A. W., Sarin, S., Heck, J., Waase, M. and Alani, E. (2003). Systematic mutagenesis of the Saccharomyces cerevisiae MLH1 gene reveals distinct roles for Mlh1p in meiotic crossing over and in vegetative and meiotic mismatch repair. *Mol.Cell Biol.* **23**, 873-886.

Argueso, J. L., Smith, D., Yi, J., Waase, M., Sarin, S. and Alani, E. (2002). Analysis of conditional mutations in the Saccharomyces cerevisiae MLH1 gene in mismatch repair and in meiotic crossing over. *Genetics* **160**, 909-921.

Argueso, J. L., Wanat, J., Gemici, Z. and Alani, E. (2004). Competing crossover pathways act during meiosis in Saccharomyces cerevisiae. *Genetics* **168**, 1805-1816.

Assenmacher, N. and Hopfner, K. P. (2004). MRE11/RAD50/NBS1: complex activities. *Chromosoma* **113**, 157-166.

Au, K. G., Clark, S., Miller, J. H. and Modrich, P. (1989). Escherichia coli mutY gene encodes an adenine glycosylase active on G-A mispairs. *Proc.Natl.Acad.Sci.U.S.A* **86**, 8877-8881.

Au, K. G., Welsh, K. and Modrich, P. (1992). Initiation of methyl-directed mismatch repair. *J.Biol.Chem.* **267**, 12142-12148.

Auffret, C. A. and Turner, M. J. (1981). Variant specific antigens of Trypanosoma brucei exist in solution as glycoprotein dimers. *Biochem.J.* **193**, 647-650.

Augusto-Pinto, L., Bartholomeu, D. C., Teixeira, S. M., Pena, S. D. and Machado, C. R. (2001). Molecular cloning and characterization of the DNA mismatch repair gene class 2 from the Trypanosoma cruzi. *Gene* **272**, 323-333.

Augusto-Pinto, L., Teixeira, S. M., Pena, S. D. and Machado, C. R. (2003). Single-Nucleotide Polymorphisms of the Trypanosoma cruzi MSH2 Gene Support the Existence of Three Phylogenetic Lineages Presenting Differences in Mismatch-Repair Efficiency. *Genetics* **164**, 117-126.

Ayares, D., Chekuri, L., Song, K. Y. and Kucherlapati, R. (1986). Sequence homology requirements for intermolecular recombination in mammalian cells. *Proc.Natl.Acad.Sci.U.S.A* **83**, 5199-5203.

Bai, Y., Davis, A. P. and Symington, L. S. (1999). A novel allele of RAD52 that causes severe DNA repair and recombination deficiencies only in the absence of RAD51 or RAD59. *Genetics* **153**, 1117-1130.

Bai, Y. and Symington, L. S. (1996). A Rad52 homolog is required for RAD51-independent mitotic recombination in Saccharomyces cerevisiae. *Genes Dev.* **10**, 2025-2037.

Bailis, A. M. and Rothstein, R. (1990). A defect in mismatch repair in Saccharomyces cerevisiae stimulates ectopic recombination between homeologous genes by an excision repair dependent process. *Genetics* **126**, 535-547.

Ban, C., Junop, M. and Yang, W. (1999). Transformation of MutL by ATP binding and hydrolysis: a switch in DNA mismatch repair. *Cell* **97**, 85-97.

Ban, C. and Yang, W. (1998). Crystal structure and ATPase activity of MutL: implications for DNA repair and mutagenesis. *Cell* **95**, 541-552.

Baquero, M. R., Galan, J. C., del Carmen, Turrientes M., Canton, R., Coque, T. M., Martinez, J. L. and Baquero, F. (2005). Increased mutation frequencies in Escherichia coli isolates harboring extended-spectrum beta-lactamases. *Antimicrob.Agents Chemother.* **49**, 4754-4756.

Baquero, M. R., Nilsson, A. I., Turrientes, Mdel C., Sandvang, D., Galan, J. C., Martinez, J. L., Frimodt-Moller, N., Baquero, F. and Andersson, D. I. (2004). Polymorphic mutation frequencies in Escherichia coli: emergence of weak mutators in clinical isolates. *J.Bacteriol.* **186**, 5538-5542.

Barbet, A. F. and Kamper, S. M. (1993). The importance of mosaic genes to trypanosome survival. *Parasitol.Today* **9**, 63-66.

Barnes, D. E. and Lindahl, T. (2004). Repair and genetic consequences of endogenous DNA base damage in mammalian cells. *Annu.Rev.Genet.* **38**, 445-476.

Barrett, M. P., Burchmore, R. J., Stich, A., Lazzari, J. O., Frasch, A. C., Cazzulo, J. J. and Krishna, S. (2003). The trypanosomiases. *Lancet* **362**, 1469-1480.

Barry, J. D. (1986). Antigenic variation during Trypanosoma vivax infections of different host species. *Parasitology* **92 (Pt 1)**, 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., Crowe, J. S. and Vickerman, K. (1983). Instability of the Trypanosoma brucei rhodesiense metacyclic variable antigen repertoire. *Nature* **306**, 699-701.

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.

Barry, J. D., Marcello, L., Morrison, L. J., Read, A. F., Lythgoe, K., Jones, N., Carrington, M., Blandin, G., Bohme, U., Caler, E., Hertz-Fowler, C., Renauld, H., El Sayed, N. and Berriman, M. (2005). What the genome sequence is revealing about trypanosome antigenic variation. *Biochem.Soc.Trans.* **33**, 986-989.

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. and McCulloch, R. (2003). Trypanopsome Antigenic Variation - a Heavy Investment in the Evasion of Immunity. In 'Antigenic Variation'. (A. Craig and A. Scherf Eds.) pp. 224-242. (Elsevier Ltd.: London.)

Baur, J. A., Zou, Y., Shay, J. W. and Wright, W. E. (2001). Telomere position effect in human cells. *Science* **292**, 2075-2077.

Bawa, S. and Xiao, W. (2003). A single amino acid substitution in MSH5 results in DNA alkylation tolerance. *Gene* **315**, 177-182.

Bazemore, L. R., Folta-Stogniew, E., Takahashi, M. and Radding, C. M. (1997). RecA tests homology at both pairing and strand exchange. *Proc.Natl.Acad.Sci.U.S.A* **94**, 11863-11868.

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

Bell, J. S., Harvey, T. I., Sims, A. M. and McCulloch, R. (2004). Characterization of components of the mismatch repair machinery in Trypanosoma brucei. *Mol.Microbiol.* **51**, 159-173.

Bell, J. S. and McCulloch, R. (2003). Mismatch repair regulates homologous recombination, but has little influence on antigenic variation, in Trypanosoma brucei. *J.Biol.Chem.* **278**, 45182-45188.

Bellacosa, A., Cicchillitti, L., Schepis, F., Riccio, A., Yeung, A. T., Matsumoto, Y., Golemis, E. A., Genuardi, M. and Neri, G. (1999). MED1, a novel human methyl-CpG-binding endonuclease, interacts with DNA mismatch repair protein MLH1. *Proc.Natl.Acad.Sci.U.S.A* **96**, 3969-3974.

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.

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

Berardini, M., Mazurek, A. and Fishel, R. (2000). The effect of O6-methylguanine DNA adducts on the adenosine nucleotide switch functions of hMSH2-hMSH6 and hMSH2-hMSH3. *J.Biol.Chem.* **275**, 27851-27857.

Berberof, M., Perez-Morga, D. and Pays, E. (2001). A receptor-like flagellar pocket glycoprotein specific to Trypanosoma brucei gambiense. *Mol.Biochem.Parasitol.* **113**, 127-138.

Berriman, M., Ghedin, E., Hertz-Fowler, C., Blandin, G., Renauld, H., Bartholomeu, D. C., Lennard, N. J., Caler, E., Hamlin, N. E., Haas, B., Bohme, U., Hannick, L., Aslett, M. A., Shallom, J., Marcello, L., Hou, L., Wickstead, B., Alsmark, U. C., Arrowsmith, C., Atkin, R. J., Barron, A. J., Bringaud, F., Brooks, K., Carrington, M., Cherevach, I., Chillingworth, T. J., Churcher, C., Clark, L. N., Corton, C. H., Cronin, A., Davies, R. M., Doggett, J., Djikeng, A., Feldblyum, T., Field, M. C., Fraser, A., Goodhead, I., Hance, Z., Harper, D., Harris, B. R., Hauser, H., Hostetler, J., Ivens, A., Jagels, K., Johnson, D., Johnson, J., Jones, K., Kerhornou, A. X., Koo, H., Larke, N., Landfear, S., Larkin, C., Leech, V., Line, A., Lord, A., MacLeod, A., Mooney, P. J., Moule, S., Martin, D. M., Morgan, G. W., Mungall, K., Norbertczak, H., Ormond, D., Pai, G., Peacock, C. S., Peterson, J., Quail, M. A., Rabbinowitsch, E., Rajandream, M. A., Reitter, C., Salzberg, S. L., Sanders, M., Schobel, S., Sharp, S., Simmonds, M., Simpson, A. J., Tallon, L., Turner, C. M., Tait, A., Tivey, A. R., Van Aken, S., Walker, D., Wanless, D., Wang, S., White, B., White, O., Whitehead, S., Woodward, J., Wortman, J., Adams, M. D., Embley, T. M., Gull, K., Ullu, E., Barry, J. D., Fairlamb, A. H., Opperdoes, F., Barrell, B. G., Donelson, J. E., Hall, N., Fraser, C. M., Melville, S. E. and El Sayed, N. M. (2005). The genome of the African trypanosome Trypanosoma brucei. *Science* **309**, 416-422.

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.

Beverley, S. M. and Coburn, C. M. (1990). Recurrent de novo appearance of small linear DNAs in Leishmania major and relationship to extra-chromosomal DNAs in other species. *Mol.Biochem.Parasitol.* **42**, 133-141.

Bhattacharyya, M. K. and Kumar, N. (2003). Identification and molecular characterisation of DNA damaging agent induced expression of Plasmodium falciparum recombination protein PfRad51. *Int.J.Parasitol.* **33**, 1385-1392.

Bhattacharyya, M. K., Norris, D. E. and Kumar, N. (2004). Molecular players of homologous recombination in protozoan parasites: implications for generating antigenic variation. *Infect.Genet.Evol.* **4**, 91-98.

Bingle, L. E., Eastlake, J. L., Bailey, M. and Gibson, W. C. (2001). A novel GFP approach for the analysis of genetic exchange in trypanosomes allowing the in situ detection of mating events. *Microbiology* **147**, 3231-3240.

Biswas, I., Ban, C., Fleming, K. G., Qin, J., Lary, J. W., Yphantis, D. A., Yang, W. and Hsieh, P. (1999). Oligomerization of a MutS mismatch repair protein from Thermus aquaticus. *J.Biol.Chem.* **274**, 23673-23678.

Biswas, I., Obmolova, G., Takahashi, M., Herr, A., Newman, M. A., Yang, W. and Hsieh, P. (2001). Disruption of the helix-u-turn-helix motif of MutS protein: loss of subunit dimerization, mismatch binding and ATP hydrolysis. *J.Mol.Biol.* **305**, 805-816.

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.

Bjedov, I., Tenaillon, O., Gerard, B., Souza, V., Denamur, E., Radman, M., Taddei, F. and Matic, I. (2003). Stress-induced mutagenesis in bacteria. *Science* **300**, 1404-1409.

Bjornson, K. P., Blackwell, L. J., Sage, H., Baitinger, C., Allen, D. and Modrich, P. (2003). Assembly and molecular activities of the MutS tetramer. *J.Biol.Chem.* **278**, 34667-34673.

Blackwell, L. J., Wang, S. and Modrich, P. (2001). DNA chain length dependence of formation and dynamics of hMutSalpha.hMutLalpha.heteroduplex complexes. *J.Biol.Chem.* **276**, 33233-33240.

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.

Blundell, P. A., Rudenko, G. and Borst, P. (1996). Targeting of exogenous DNA into Trypanosoma brucei requires a high degree of homology between donor and target DNA. *Mol.Biochem.Parasitol.* **76**, 215-229.

Bocker, T., Barusevicius, A., Snowden, T., Rasio, D., Guerrette, S., Robbins, D., Schmidt, C., Burczak, J., Croce, C. M., Copeland, T., Kovatich, A. J. and Fishel, R. (1999). hMSH5: a human MutS homologue that forms a novel heterodimer with hMSH4 and is expressed during spermatogenesis. *Cancer Res.* **59**, 816-822.

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.

Boe, L., Danielsen, M., Knudsen, S., Petersen, J. B., Maymann, J. and Jensen, P. R. (2000). The frequency of mutators in populations of Escherichia coli. *Mutat.Res.* **448**, 47-55.

Bogliolo, A. R., Lauria-Pires, L. and Gibson, W. C. (1996). Polymorphisms in Trypanosoma cruzi: evidence of genetic recombination. *Acta Trop.* **61**, 31-40.

Boothroyd, J. C., Cross, G. A., Hoeijmakers, J. H. and Borst, P. (1980). A variant surface glycoprotein of Trypanosoma brucei synthesized with a C-terminal hydrophobic 'tail' absent from purified glycoprotein. *Nature* **288**, 624-626.

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

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

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

Bosher, J. M. and Labouesse, M. (2000). RNA interference: genetic wand and genetic watchdog. *Nat.Cell Biol.* **2**, E31-E36.

Bowers, J., Sokolsky, T., Quach, T. and Alani, E. (1999). A mutation in the MSH6 subunit of the Saccharomyces cerevisiae MSH2-MSH6 complex disrupts mismatch recognition. *J.Biol.Chem.* **274**, 16115-16125.

Bowers, J., Tran, P. T., Joshi, A., Liskay, R. M. and Alani, E. (2001). MSH-MLH complexes formed at a DNA mismatch are disrupted by the PCNA sliding clamp. *J.Mol.Biol.* **306**, 957-968.

Brendel, V., Brocchieri, L., Sandler, S. J., Clark, A. J. and Karlin, S. (1997). Evolutionary comparisons of RecA-like proteins across all major kingdoms of living organisms. *J.Mol.Evol.* **44**, 528-541.

Bridges, B. A. (2001). Hypermutation in bacteria and other cellular systems. *Philos.Trans.R.Soc.Lond B Biol.Sci.* **356**, 29-39.

Bringaud, F., Biteau, N., Donelson, J. E. and Baltz, T. (2001). Conservation of metacyclic variant surface glycoprotein expression sites among different trypanosome isolates. *Mol.Biochem.Parasitol.* **113**, 67-78.

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.

Brown, E. W., LeClerc, J. E., Li, B., Payne, W. L. and Cebula, T. A. (2001). Phylogenetic evidence for horizontal transfer of mutS alleles among naturally occurring Escherichia coli strains. *J.Bacteriol.* **183**, 1631-1644.

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

Bucka, A. and Stasiak, A. (2001). RecA-mediated strand exchange traverses substitutional heterologies more easily than deletions or insertions. *Nucleic Acids Res.* **29**, 2464-2470.

Buermeyer, A. B., Deschenes, S. M., Baker, S. M. and Liskay, R. M. (1999). Mammalian DNA mismatch repair. *Annu.Rev.Genet.* **33**, 533-564.

Bulmer, M. G. (1967). Principles of Statistics. Oliver Boyd, Edinburgh and London.

Buscaglia, C. A. and Di Noia, J. M. (2003). Trypanosoma cruzi clonal diversity and the epidemiology of Chagas' disease. *Microbes.Infect.* **5**, 419-427.

Calmann, M. A. and Marinus, M. G. (2005). Differential effects of cisplatin and MNNG on dna mutants of Escherichia coli. *Mutat.Res.* **578**, 406-416.

Calmann, M. A., Nowosielska, A. and Marinus, M. G. (2005). Separation of mutation avoidance and antirecombination functions in an Escherichia coli mutS mutant. *Nucleic Acids Res.* **33**, 1193-1200.

Carney, J. P., Maser, R. S., Olivares, H., Davis, E. M., Le Beau, M., Yates, J. R., III, Hays, L., Morgan, W. F. and Petrini, J. H. (1998). The hMre11/hRad50 protein complex and Nijmegen breakage syndrome: linkage of double-strand break repair to the cellular DNA damage response. *Cell* **93**, 477-486.

Carrasco, H. J., Frame, I. A., Valente, S. A. and Miles, M. A. (1996). Genetic exchange as a possible source of genomic diversity in sylvatic populations of Trypanosoma cruzi. *Am.J.Trop.Med.Hyg.* **54**, 418-424.

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.

Cejka, P., Mojas, N., Gillet, L., Schar, P. and Jiricny, J. (2005). Homologous recombination rescues mismatch-repair-dependent cytotoxicity of S(N)1-type methylating agents in S. cerevisiae. *Curr.Biol.* **15**, 1395-1400.

Celerin, M., Merino, S. T., Stone, J. E., Menzie, A. M. and Zolan, M. E. (2000). Multiple roles of Spo11 in meiotic chromosome behavior. *EMBO J.* **19**, 2739-2750.

Chappell, C., Hanakahi, L. A., Karimi-Busheri, F., Weinfeld, M. and West, S. C. (2002). Involvement of human polynucleotide kinase in double-strand break repair by non-homologous end joining. *EMBO J.* **21**, 2827-2832.

Chat, L. L., Fons, M. and Taddei, F. (2006). Escherichia coli mutators: selection criteria and migration effect. *Microbiology* **152**, 67-73.

Chattopadhyay, A., Jones, N. G., Nietlispach, D., Nielsen, P. R., Voorheis, H. P., Mott, H. R. and Carrington, M. (2005). Structure of the C-terminal Domain from Trypanosoma brucei Variant Surface Glycoprotein MITat1.2. *J.Biol.Chem.* **280**, 7228-7235.

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, W. and Jinks-Robertson, S. (1998). Mismatch repair proteins regulate heteroduplex formation during mitotic recombination in yeast. *Mol.Cell Biol.* **18**, 6525-6537.

Chen, W. and Jinks-Robertson, S. (1999). The role of the mismatch repair machinery in regulating mitotic and meiotic recombination between diverged sequences in yeast. *Genetics* **151**, 1299-1313.

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.

Chi, N. W. and Kolodner, R. D. (1994a). Purification and characterization of MSH1, a yeast mitochondrial protein that binds to DNA mismatches. *J.Biol.Chem.* **269**, 29984-29992.

Chi, N. W. and Kolodner, R. D. (1994b). The effect of DNA mismatches on the ATPase activity of MSH1, a protein in yeast mitochondria that recognizes DNA mismatches. *J.Biol.Chem.* **269**, 29993-29997.

Chi, P., Van Komen, S., Sehorn, M. G., Sigurdsson, S. and Sung, P. (2006). Roles of ATP binding and ATP hydrolysis in human Rad51 recombinase function. *DNA Repair (Amst)* **5**, 381-391.

Chiurillo, M. A., Cano, I., da Silveira, J. F. and Ramirez, J. L. (1999). Organization of telomeric and sub-telomeric regions of chromosomes from the protozoan parasite Trypanosoma cruzi. *Mol.Biochem.Parasitol.* **100**, 173-183.

Chopra, I., O'Neill, A. J. and Miller, K. (2003). The role of mutators in the emergence of antibiotic-resistant bacteria. *Drug Resist.Updat.* **6**, 137-145.

Chung, M. H., Kasai, H., Jones, D. S., Inoue, H., Ishikawa, H., Ohtsuka, E. and Nishimura, S. (1991). An endonuclease activity of Escherichia coli that specifically removes 8-hydroxyguanine residues from DNA. *Mutat.Res.* **254**, 1-12.

Cibulskis, R.E. (1988). Origins and organization of genetic diversity in natural populations of Trypanosoma brucei. *Parasitology* **96**, 303-322.

Ciofu, O., Riis, B., Pressler, T., Poulsen, H. E. and Hoiby, N. (2005). Occurrence of hypermutable Pseudomonas aeruginosa in cystic fibrosis patients is associated with the oxidative stress caused by chronic lung inflammation. *Antimicrob.Agents Chemother.* **49**, 2276-2282.

Ciotta, C., Ceccotti, S., Aquilina, G., Humbert, O., Palombo, F., Jiricny, J. and Bignami, M. (1998). Increased somatic recombination in methylation tolerant human cells with defective DNA mismatch repair. *J.Mol.Biol.* **276**, 705-719.

Clark, A. B., Valle, F., Drotschmann, K., Gary, R. K. and Kunkel, T. A. (2000). Functional interaction of proliferating cell nuclear antigen with MSH2- MSH6 and MSH2-MSH3 complexes. *J.Biol.Chem.* **275**, 36498-36501.

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

Coban, A. Y., Darka, O., Fisgin, N. T., Cihan, C. C., Bilgin, K., Akgunes, A., Guven, T., Dokuzoguz, B., Birinci, A. and Durupinar, B. (2005). The resazurin microplate method for rapid detection of vancomycin resistance in enterococci. *J.Chemother.* **17**, 361-366.

Colaiacovo, M. P., MacQueen, A. J., Martinez-Perez, E., McDonald, K., Adamo, A., La Volpe, A. and Villeneuve, A. M. (2003). Synaptonemal complex assembly in C. elegans is dispensable for loading strand-exchange proteins but critical for proper completion of recombination. *Dev.Cell* **5**, 463-474.

Colaiacovo, M. P., Paques, F. and Haber, J. E. (1999). Removal of one nonhomologous DNA end during gene conversion by a RAD1- and MSH2-independent pathway. *Genetics* **151**, 1409-1423.

Collins, L. and Franzblau, S. G. (1997). Microplate alamar blue assay versus BACTEC 460 system for high-throughput screening of compounds against Mycobacterium tuberculosis and Mycobacterium avium. *Antimicrob.Agents Chemother.* **41**, 1004-1009.

Colussi, C., Parlanti, E., Degan, P., Aquilina, G., Barnes, D., Macpherson, P., Karran, P., Crescenzi, M., Dogliotti, E. and Bignami, M. (2002). The mammalian mismatch repair pathway removes DNA 8-oxodGMP incorporated from the oxidized dNTP pool. *Curr.Biol.* **12**, 912-918.

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

Constantin, N., Dzantiev, L., Kadyrov, F. A. and Modrich, P. (2005). Human mismatch repair: reconstitution of a nick-directed bidirectional reaction. *J.Biol.Chem.* **280**, 39752-39761.

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, A. B., Lynch, T. W., Zhang, Y., Fortin, G. S., Fung, C. W., Symington, L. S. and Rice, P. A. (2004). Crystal structure of a Rad51 filament. *Nat.Struct.Mol.Biol.*

Conway, C., McCulloch, R., Ginger, M. L., Robinson, N. P., Browitt, A. and Barry, J. D. (2002a). 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. (2002b). Two pathways of homologous recombination in Trypanosoma brucei. *Mol.Microbiol.* **45**, 1687-1700.

Cooper, R., de Jesus, A. R. and Cross, G. A. (1993). Deletion of an immunodominant Trypanosoma cruzi surface glycoprotein disrupts flagellum-cell adhesion. *J.Cell Biol.* **122**, 149-156.

Cooper, V. S. and Lenski, R. E. (2000). The population genetics of ecological specialization in evolving Escherichia coli populations. *Nature* **407**, 736-739.

Courcelle, J., Khodursky, A., Peter, B., Brown, P. O. and Hanawalt, P. C. (2001). Comparative gene expression profiles following UV exposure in wild-type and SOS-deficient Escherichia coli. *Genetics* **158**, 41-64.

Crawford, D. C., Akey, D. P., and Nickerson, D. A. (2005). The patterns of natural variation in human genes. *Annu. Rev. Genomics Hum. Genet.* **6**, 287-312.

Critchlow, S. E. and Jackson, S. P. (1998). DNA end-joining: from yeast to man. *Trends Biochem.Sci.* **23**, 394-398.

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., Taylor, M. C. and Borst, P. (1998). Frequent loss of the active site during variant surface glycoprotein expression site switching in vitro in Trypanosoma brucei. *Mol.Cell Biol.* **18**, 198-205.

Cruz, A. and Beverley, S. M. (1990). Gene replacement in parasitic protozoa. *Nature* **348**, 171-173.

Cruz, A., Coburn, C. M. and Beverley, S. M. (1991). Double targeted gene replacement for creating null mutants. *Proc.Natl.Acad.Sci.U.S.A* **88**, 7170-7174.

Cruz, A. K., Titus, R. and Beverley, S. M. (1993). Plasticity in chromosome number and testing of essential genes in Leishmania by targeting. *Proc.Natl.Acad.Sci.U.S.A* **90**, 1599-1603.

Culligan, K. M. and Hays, J. B. (2000). Arabidopsis MutS homologs-AtMSH2, AtMSH3, AtMSH6, and a novel AtMSH7- form three distinct protein heterodimers with different specificities for mismatched DNA. *Plant Cell* **12**, 991-1002.

Culligan, K. M., Meyer-Gauen, G., Lyons-Weiler, J. and Hays, J. B. (2000). Evolutionary origin, diversification and specialization of eukaryotic MutS homolog mismatch repair proteins. *Nucleic Acids Res.* **28**, 463-471.

Dacks, J. and Roger, A. J. (1999). The first sexual lineage and the relevance of facultative sex. *J.Mol.Evol.* **48**, 779-783.

Dao, V. and Modrich, P. (1998). Mismatch-, MutS-, MutL-, and helicase II-dependent unwinding from the single-strand break of an incised heteroduplex. *J.Biol.Chem.* **273**, 9202-9207.

DasGupta, C. and Radding, C. M. (1982). Polar branch migration promoted by recA protein: effect of mismatched base pairs. *Proc.Natl.Acad.Sci.U.S.A* **79**, 762-766.

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.

Davies, A. A., Masson, J. Y., McIlwraith, M. J., Stasiak, A. Z., Stasiak, A., Venkitaraman, A. R. and West, S. C. (2001). Role of BRCA2 in control of the RAD51 recombination and DNA repair protein. *Mol.Cell* **7**, 273-282.

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.

Davis, A. P. and Symington, L. S. (2003). The Rad52-Rad59 complex interacts with Rad51 and replication protein A. *DNA Repair (Amst)* **2**, 1127-1134.

Davis, A. P. and Symington, L. S. (2004). RAD51-dependent break-induced replication in yeast. *Mol.Cell Biol.* **24**, 2344-2351.

de Freitas, J. M., Augusto-Pinto, L., Pimenta, J. R., Bastos-Rodrigues, L., Goncalves, V. F., Teixeira, S. M., Chiari, E., Junqueira, A. C., Fernandes, O., Macedo, A. M., Machado, C. R. and Pena, S. D. (2006). Ancestral Genomes, Sex, and the Population Structure of Trypanosoma cruzi. *PLoS.Pathog.* **2**, e24.

de Fries, R. and Mitsuhashi, M. (1995). Quantification of mitogen induced human lymphocyte proliferation: comparison of alamarBlue assay to 3H-thymidine incorporation assay. *J.Clin.Lab Anal.* **9**, 89-95.

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

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

de los, Santos T. and Hollingsworth, N. M. (1999). Red1p, a MEK1-dependent phosphoprotein that physically interacts with Hop1p during meiosis in yeast. *J.Biol.Chem.* **274**, 1783-1790.

de los, Santos T., Hunter, N., Lee, C., Larkin, B., Loidl, J. and Hollingsworth, N. M. (2003). The Mus81/Mms4 endonuclease acts independently of double-Holliday junction resolution to promote a distinct subset of crossovers during meiosis in budding yeast. *Genetics* **164**, 81-94.

de Vries, S. S., Baart, E. B., Dekker, M., Siezen, A., de Rooij, D. G., de Boer, P. and te, Riele H. (1999a). Mouse MutS-like protein Msh5 is required for proper chromosome synapsis in male and female meiosis. *Genes Dev.* **13**, 523-531.

de Vries, S. S., Baart, E. B., Dekker, M., Siezen, A., de Rooij, D. G., de Boer, P. and te, Riele H. (1999b). Mouse MutS-like protein Msh5 is required for proper chromosome synapsis in male and female meiosis. *Genes Dev.* **13**, 523-531.

De Wind, N., Dekker, M., Berns, A., Radman, M. and te, Riele H. (1995). Inactivation of the mouse Msh2 gene results in mismatch repair deficiency, methylation tolerance, hyperrecombination, and predisposition to cancer. *Cell* **82**, 321-330.

Degen, R., Pospichal, H., Enyaru, J. and Jenni, L. (1995). Sexual compatibility among Trypanosoma brucei isolates from an epidemic area in southeastern Uganda. *Parasitol.Res.* **81**, 253-257.

Deitsch, K. W. (2005). Malaria virulence genes controlling expression through chromatin modification. *Cell* **121**, 1-2.

Deitsch, K. W., Calderwood, M. S. and Wellems, T. E. (2001). Malaria. Cooperative silencing elements in var genes. *Nature* **412**, 875-876.

Deitsch, K. W. and Hviid, L. (2004). Variant surface antigens, virulence genes and the pathogenesis of malaria. *Trends Parasitol.* **20**, 562-566.

Deitsch, K. W., Moxon, E. R. and Wellems, T. E. (1997). Shared themes of antigenic variation and virulence in bacterial, protozoal, and fungal infections. *Microbiol.Mol.Biol.Rev.* **61**, 281-293.

Della, M., Palmbos, P. L., Tseng, H. M., Tonkin, L. M., Daley, J. M., Topper, L. M., Pitcher, R. S., Tomkinson, A. E., Wilson, T. E. and Doherty, A. J. (2004). Mycobacterial Ku and ligase proteins constitute a two-component NHEJ repair machine. *Science* **306**, 683-685.

Denamur, E., Bonacorsi, S., Giraud, A., Duriez, P., Hilali, F., Amorin, C., Bingen, E., Andremont, A., Picard, B., Taddei, F. and Matic, I. (2002). High frequency of mutator strains among human uropathogenic Escherichia coli isolates. *J.Bacteriol.* **184**, 605-609.

Denamur, E., Lecointre, G., Darlu, P., Tenaillon, O., Acquaviva, C., Sayada, C., Sunjevaric, I., Rothstein, R., Elion, J., Taddei, F., Radman, M. and Matic, I. (2000). Evolutionary implications of the frequent horizontal transfer of mismatch repair genes. *Cell* **103**, 711-721.

Denamur, E. and Matic, I. (2006). Evolution of mutation rates in bacteria. *Mol.Microbiol.* **60**, 820-827.

Denamur, E., Tenaillon, O., Deschamps, C., Skurnik, D., Ronco, E., Gaillard, J. L., Picard, B., Branger, C. and Matic, I. (2005). Intermediate mutation frequencies favor evolution of multidrug resistance in Escherichia coli. *Genetics* **171**, 825-827.

Deng, C. and Capecchi, M. R. (1992). Reexamination of gene targeting frequency as a function of the extent of homology between the targeting vector and the target locus. *Mol.Cell Biol.* **12**, 3365-3371.

Dernburg, A. F., McDonald, K., Moulder, G., Barstead, R., Dresser, M. and Villeneuve, A. M. (1998). Meiotic recombination in C. elegans initiates by a conserved mechanism and is dispensable for homologous chromosome synapsis. *Cell* **94**, 387-398.

DeWeese, T. L., Shipman, J. M., Larrier, N. A., Buckley, N. M., Kidd, L. R., Groopman, J. D., Cutler, R. G., te, Riele H. and Nelson, W. G. (1998). Mouse embryonic stem cells carrying one or two defective Msh2 alleles respond abnormally to oxidative stress inflicted by low-level radiation. *Proc.Natl.Acad.Sci.U.S.A* **95**, 11915-11920.

Di Noia, J. M., Buscaglia, C. A., De Marchi, C. R., Almeida, I. C. and Frasch, A. C. (2002). A Trypanosoma cruzi small surface molecule provides the first immunological evidence that Chagas' disease is due to a single parasite lineage. *J.Exp.Med.* **195**, 401-413.

Dipaolo, C., Kieft, R., Cross, M. and Sabatini, R. (2005). Regulation of Trypanosome DNA Glycosylation by a SWI2/SNF2-like Protein. *Mol.Cell* **17**, 441-451.

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

Dong, C., Whitford, R. and Langridge, P. (2002). A DNA mismatch repair gene links to the Ph2 locus in wheat. *Genome* **45**, 116-124.

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.

Drake, J. W., Charlesworth, B., Charlesworth, D. and Crow, J. F. (1998). Rates of spontaneous mutation. *Genetics* **148**, 1667-1686.

Dreiem, A., Gertz, C. C. and Seegal, R. F. (2005). The effects of methylmercury on mitochondrial function and reactive oxygen species formation in rat striatal synaptosomes are age-dependent. *Toxicol.Sci.* **87**, 156-162.

Drotschmann, K., Yang, W., Brownewell, F. E., Kool, E. T. and Kunkel, T. A. (2001). Asymmetric recognition of DNA local distortion. Structure-based functional studies of eukaryotic Msh2-Msh6. *J.Biol.Chem.* **276**, 46225-46229.

Dujardin, J. C., Banuls, A. L., Llanos-Cuentas, A., Alvarez, E., DeDoncker, S., Jacquet, D., Le Ray, D., Arevalo, J. and Tibayrenc, M. (1995). Putative Leishmania hybrids in the Eastern Andean valley of Huanuco, Peru. *Acta Trop.* **59**, 293-307.

Durant, S. T., Morris, M. M., Illand, M., McKay, H. J., McCormick, C., Hirst, G. L., Borts, R. H. and Brown, R. (1999). Dependence on RAD52 and RAD1 for anticancer drug resistance mediated by inactivation of mismatch repair genes. *Curr.Biol.* **9**, 51-54.

Dutta, R. and Inouye, M. (2000). GHKL, an emergent ATPase/kinase superfamily. *Trends Biochem.Sci.* **25**, 24-28.

Dzantiev, L., Constantin, N., Genschel, J., Iyer, R. R., Burgers, P. M. and Modrich, P. (2004). A defined human system that supports bidirectional mismatch-provoked excision. *Mol.Cell* **15**, 31-41.

Earley, M. C. and Crouse, G. F. (1998). The role of mismatch repair in the prevention of base pair mutations in Saccharomyces cerevisiae. *Proc.Natl.Acad.Sci.U.S.A* **95**, 15487-15491.

Edelmann, W., Cohen, P. E., Kneitz, B., Winand, N., Lia, M., Heyer, J., Kolodner, R., Pollard, J. W. and Kucherlapati, R. (1999). Mammalian MutS homologue 5 is required for chromosome pairing in meiosis. *Nat.Genet.* **21**, 123-127.

Edelmann, W., Umar, A., Yang, K., Heyer, J., Kucherlapati, M., Lia, M., Kneitz, B., Avdievich, E., Fan, K., Wong, E., Crouse, G., Kunkel, T., Lipkin, M., Kolodner, R. D. and Kucherlapati, R. (2000). The DNA mismatch repair genes Msh3 and Msh6 cooperate in intestinal tumor suppression. *Cancer Res.* **60**, 803-807.

Egashira, A., Yamauchi, K., Yoshiyama, K., Kawate, H., Katsuki, M., Sekiguchi, M., Sugimachi, K., Maki, H. and Tsuzuki, T. (2002). Mutational specificity of mice defective in the MTH1 and/or the MSH2 genes. *DNA Repair (Amst)* **1**, 881-893.

Eid, J. and Sollner-Webb, B. (1987). Efficient introduction of plasmid DNA into Trypanosoma brucei and transcription of a transfected chimeric gene. *Proc.Natl.Acad.Sci.U.S.A* **84**, 7812-7816.

Eisen, J. A. (1998). A phylogenomic study of the MutS family of proteins. *Nucleic Acids Res.* **26**, 4291-4300.

Eisen, J. A., Sweder, K. S. and Hanawalt, P. C. (1995). Evolution of the SNF2 family of proteins: subfamilies with distinct sequences and functions. *Nucleic Acids Res.* **23**, 2715-2723.

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

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.

Emery, H. S., Schild, D., Kellogg, D. E. and Mortimer, R. K. (1991). Sequence of RAD54, a Saccharomyces cerevisiae gene involved in recombination and repair. *Gene* **104**, 103-106.

Esashi, F., Christ, N., Gannon, J., Liu, Y., Hunt, T., Jasin, M. and West, S. C. (2005). CDK-dependent phosphorylation of BRCA2 as a regulatory mechanism for recombinational repair. *Nature* **434**, 598-604.

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

Evans, E., Sugawara, N., Haber, J. E. and Alani, E. (2000). The Saccharomyces cerevisiae Msh2 mismatch repair protein localizes to recombination intermediates in vivo. *Mol.Cell* **5**, 789-799.

Fan, W. and Wu, X. (2004). DNA polymerase lambda can elongate on DNA substrates mimicking non-homologous end joining and interact with XRCC4-ligase IV complex. *Biochem.Biophys.Res.Commun.* **323**, 1328-1333.

Farez-Vidal, M. E., Gallego, C., Ruiz-Perez, L. M. and Gonzalez-Pacanowska, D. (2001). Characterization of uracil-DNA glycosylase activity from Trypanosoma cruzi and its stimulation by AP endonuclease. *Nucleic Acids Res.* **29**, 1549-1555.

Felsenstein, J. (1974). The evolutionary advantage of recombination. *Genetics* **78**, 737-756.

Feng, G., Tsui, H. C. and Winkler, M. E. (1996). Depletion of the cellular amounts of the MutS and MutH methyl-directed mismatch repair proteins in stationary-phase Escherichia coli K-12 cells. *J.Bacteriol.* **178**, 2388-2396.

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

Feuerbach, F., Galy, V., Trelles-Sticken, E., Fromont-Racine, M., Jacquier, A., Gilson, E., Olivo-Marin, J. C., Scherthan, H. and Nehrbass, U. (2002). Nuclear architecture and spatial positioning help establish transcriptional states of telomeres in yeast. *Nat.Cell Biol.* **4**, 214-221.

Fishel, R., Lescoe, M. K., Rao, M. R., Copeland, N. G., Jenkins, N. A., Garber, J., Kane, M. and Kolodner, R. (1994). The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer. *Cell* **77**, 167.

Flores-Rozas, H., Clark, D. and Kolodner, R. D. (2000). Proliferating cell nuclear antigen and Msh2p-Msh6p interact to form an active mispair recognition complex. *Nat.Genet.* **26**, 375-378.

Flores-Rozas, H. and Kolodner, R. D. (1998). The Saccharomyces cerevisiae MLH3 gene functions in MSH3-dependent suppression of frameshift mutations. *Proc.Natl.Acad.Sci.U.S.A* **95**, 12404-12409.

Fortini, P., Pascucci, B., Parlanti, E., D'Errico, M., Simonelli, V. and Dogliotti, E. (2003). 8-Oxoguanine DNA damage: at the crossroad of alternative repair pathways. *Mutat.Res.* **531**, 127-139.

Fowler, R. G., White, S. J., Koyama, C., Moore, S. C., Dunn, R. L. and Schaaper, R. M. (2003). Interactions among the Escherichia coli mutT, mutM, and mutY damage prevention pathways. *DNA Repair (Amst)* **2**, 159-173.

Funchain, P., Yeung, A., Stewart, J. L., Lin, R., Slupska, M. M. and Miller, J. H. (2000). The consequences of growth of a mutator strain of Escherichia coli as measured by loss of function among multiple gene targets and loss of fitness. *Genetics* **154**, 959-970.

Galio, L., Bouquet, C. and Brooks, P. (1999). ATP hydrolysis-dependent formation of a dynamic ternary nucleoprotein complex with MutS and MutL. *Nucleic Acids Res.* **27**, 2325-2331.

Galkin, V. E., Esashi, F., Yu, X., Yang, S., West, S. C. and Egelman, E. H. (2005). BRCA2 BRC motifs bind RAD51-DNA filaments. *Proc.Natl.Acad.Sci.U.S.A* **102**, 8537-8542.

Gasior, S. L., Wong, A. K., Kora, Y., Shinohara, A. and Bishop, D. K. (1998). Rad52 associates with RPA and functions with rad55 and rad57 to assemble meiotic recombination complexes. *Genes Dev.* **12**, 2208-2221.

Gaunt, M. and Miles, M. (2000). The ecotopes and evolution of triatomine bugs (triatominae) and their associated trypanosomes. *Mem.Inst.Oswaldo Cruz* **95**, 557-565.

Gaunt, M. W., Yeo, M., Frame, I. A., Stothard, J. R., Carrasco, H. J., Taylor, M. C., Mena, S. S., Veazey, P., Miles, G. A., Acosta, N., de Arias, A. R. and Miles, M. A. (2003). Mechanism of genetic exchange in American trypanosomes. *Nature* **421**, 936-939.

Genest, P. A., ter Riet, B., Dumas, C., Papadopoulou, B., van Luenen, H. G. and Borst, P. (2005). Formation of linear inverted repeat amplicons following targeting of an essential gene in Leishmania. *Nucleic Acids Res.* **33**, 1699-1709.

Genschel, J., Bazemore, L. R. and Modrich, P. (2002). Human exonuclease I is required for 5' and 3' mismatch repair. *J.Biol.Chem.* **277**, 13302-13311.

Ghedin, E., Bringaud, F., Peterson, J., Myler, P., Berriman, M., Ivens, A., Andersson, B., Bontempi, E., Eisen, J., Angiuoli, S., Wanless, D., Von Arx, A., Murphy, L., Lennard, N., Salzberg, S., Adams, M. D., White, O., Hall, N., Stuart, K., Fraser, C. M. and El Sayed, N. M. (2004). Gene synteny and evolution of genome architecture in trypanosomatids. *Mol.Biochem.Parasitol.* **134**, 183-191.

Gibson, W. (2001). Sex and evolution in trypanosomes. *Int.J.Parasitol.* **31**, 643-647.

Gibson, W. (2002). Will the real Trypanosoma brucei rhodesiense please step forward? *Trends Parasitol.* **18**, 486-490.

Gibson, W. and Bailey, M. (2003). The development of Trypanosoma brucei within the tsetse fly midgut observed using green fluorescent trypanosomes. *Kinetoplastid.Biol.Dis.* **2**, 1.

Gibson, W., Crow, M. and Kearns, J. (1997). Kinetoplast DNA minicircles are inherited from both parents in genetic crosses of Trypanosoma brucei. *Parasitol.Res.* **83**, 483-488.

Gibson, W. and Garside, L. (1990). Kinetoplast DNA minicircles are inherited from both parents in genetic hybrids of Trypanosoma brucei. *Mol.Biochem.Parasitol.* **42**, 45-53.

Gibson, W., Garside, L. and Bailey, M. (1992). Trisomy and chromosome size changes in hybrid trypanosomes from a genetic cross between Trypanosoma brucei rhodesiense and T. b. brucei. *Mol.Biochem.Parasitol.* **51**, 189-199.

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

Gibson, W. and Whittington, H. (1993). Genetic exchange in Trypanosoma brucei: selection of hybrid trypanosomes by introduction of genes conferring drug resistance. *Mol.Biochem.Parasitol.* **60**, 19-26.

Gibson, W. C. (1986). Will the real Trypanosoma b. gambiense please stand up. *Parasitol.Today* **2**, 255-257.

Gibson, W. C. (1989). Analysis of a genetic cross between Trypanosoma brucei rhodesiense and T. b. brucei. *Parasitology* **99 Pt 3**, 391-402.

Gibson, W.C. (1995). The significance of genetic exchange in trypanosomes. *Parasitol. Today* **11**, 465-472.

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.

Giraud, A., Matic, I., Tenaillon, O., Clara, A., Radman, M., Fons, M. and Taddei, F. (2001a). Costs and benefits of high mutation rates: adaptive evolution of bacteria in the mouse gut. *Science* **291**, 2606-2608.

Giraud, A., Radman, M., Matic, I. and Taddei, F. (2001b). The rise and fall of mutator bacteria. *Curr.Opin.Microbiol.* **4**, 582-585.

Glaab, W. E., Risinger, J. I., Umar, A., Barrett, J. C., Kunkel, T. A. and Tindall, K. R. (1998). Cellular resistance and hypermutability in mismatch repair-deficient human cancer cell lines following treatment with methyl methanesulfonate. *Mutat.Res.* **398**, 197-207.

Glaab, W. E., Tindall, K. R. and Skopek, T. R. (1999). Specificity of mutations induced by methyl methanesulfonate in mismatch repair-deficient human cancer cell lines. *Mutat.Res.* **427**, 67-78.

Goldmacher, V. S., Cuzick, R. A., Jr. and Thilly, W. G. (1986). Isolation and partial characterization of human cell mutants differing in sensitivity to killing and mutation by methylnitrosourea and N- methyl-N'-nitro-N-nitrosoguanidine. *J.Biol.Chem.* **261**, 12462-12471.

Gommers-Ampt, J. H., Teixeira, A. J., Van de, Werken G., van Dijk, W. J. and Borst, P. (1993). The identification of hydroxymethyluracil in DNA of Trypanosoma brucei. *Nucleic Acids Res.* **21**, 2039-2043.

Gottschling, D. E., Aparicio, O. M., Billington, B. L. and Zakian, V. A. (1990). Position effect at S. cerevisiae telomeres: reversible repression of Pol II transcription. *Cell* **63**, 751-762.

Gradia, S., Acharya, S. and Fishel, R. (1997). The human mismatch recognition complex hMSH2-hMSH6 functions as a novel molecular switch. *Cell* **91**, 995-1005.

Gradia, S., Acharya, S. and Fishel, R. (2000). The role of mismatched nucleotides in activating the hMSH2-hMSH6 molecular switch. *J.Biol.Chem.* **275**, 3922-3930.

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

Grawunder, U., Wilm, M., Wu, X., Kulesza, P., Wilson, T. E., Mann, M. and Lieber, M. R. (1997). Activity of DNA ligase IV stimulated by complex formation with XRCC4 protein in mammalian cells. *Nature* **388**, 492-495.

Greaves, D. R. and Borst, P. (1987). Trypanosoma brucei variant-specific glycoprotein gene chromatin is sensitive to single-strand-specific endonuclease digestion. *J.Mol.Biol.* **197**, 471-483.

Greene, C. N. and Jinks-Robertson, S. (1997). Frameshift intermediates in homopolymer runs are removed efficiently by yeast mismatch repair proteins. *Mol.Cell Biol.* **17**, 2844-2850.

Grelon, M., Vezon, D., Gendrot, G. and Pelletier, G. (2001). AtSPO11-1 is necessary for efficient meiotic recombination in plants. *EMBO J.* **20**, 589-600.

Griffin, S., Branch, P., Xu, Y. Z. and Karran, P. (1994). DNA mismatch binding and incision at modified guanine bases by extracts of mammalian cells: implications for tolerance to DNA methylation damage. *Biochemistry* **33**, 4787-4793.

Grilley, M., Welsh, K. M., Su, S. S. and Modrich, P. (1989). Isolation and characterization of the Escherichia coli mutL gene product. *J.Biol.Chem.* **264**, 1000-1004.

Gu, L., Hong, Y., McCulloch, S., Watanabe, H. and Li, G. M. (1998). ATP-dependent interaction of human mismatch repair proteins and dual role of PCNA in mismatch repair. *Nucleic Acids Res.* **26**, 1173-1178.

Gu, Y., Parker, A., Wilson, T. M., Bai, H., Chang, D. Y. and Lu, A. L. (2002). Human MutY homolog, a DNA glycosylase involved in base excision repair, physically and functionally interacts with mismatch repair proteins human MutS homolog 2/human MutS homolog 6. *J.Biol.Chem.* **277**, 11135-11142.

Guerrette, S., Acharya, S. and Fishel, R. (1999). The interaction of the human MutL homologues in hereditary nonpolyposis colon cancer. *J.Biol.Chem.* **274**, 6336-6341.

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.

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

Habraken, Y., Sung, P., Prakash, L. and Prakash, S. (1998). ATP-dependent assembly of a ternary complex consisting of a DNA mismatch and the yeast MSH2-MSH6 and MLH1-PMS1 protein complexes. *J.Biol.Chem.* **273**, 9837-9841.

Hager, K. M., Pierce, M. A., Moore, D. R., Tytler, E. M., Esko, J. D. and Hajduk, S. L. (1994). Endocytosis of a cytotoxic human high density lipoprotein results in disruption of acidic intracellular vesicles and subsequent killing of African trypanosomes. *J.Cell Biol.* **126**, 155-167.

Hall, B. G. (1995). Adaptive mutations in Escherichia coli as a model for the multiple mutational origins of tumors. *Proc.Natl.Acad.Sci.U.S.A* **92**, 5669-5673.

Harfe, B. D. and Jinks-Robertson, S. (2000). DNA mismatch repair and genetic instability. *Annu.Rev.Genet.* **34**, 359-399.

Harfe, B. D., Minesinger, B. K. and Jinks-Robertson, S. (2000). Discrete in vivo roles for the MutL homologs Mlh2p and Mlh3p in the removal of frameshift intermediates in budding yeast. *Curr.Biol.* **10**, 145-148.

Hariharan, S., Ajioka, J. and Swindle, J. (1993). Stable transformation of Trypanosoma cruzi: inactivation of the PUB12.5 polyubiquitin gene by targeted gene disruption. *Mol.Biochem.Parasitol.* **57**, 15-30.

Harris, R. S., Feng, G., Ross, K. J., Sidhu, R., Thulin, C., Longerich, S., Szigety, S. K., Hastings, P. J., Winkler, M. E. and Rosenberg, S. M. (1999). Mismatch repair is diminished during stationary-phase mutation. *Mutat.Res.* **437**, 51-60.

Harris, S., Rudnicki, K. S. and Haber, J. E. (1993). Gene conversions and crossing over during homologous and homeologous ectopic recombination in Saccharomyces cerevisiae. *Genetics* **135**, 5-16.

Hartung, F. and Puchta, H. (2000). Molecular characterisation of two paralogous SPO11 homologues in Arabidopsis thaliana. *Nucleic Acids Res.* **28**, 1548-1554.

Hasty, P., Rivera-Perez, J. and Bradley, A. (1991). The length of homology required for gene targeting in embryonic stem cells. *Mol.Cell Biol.* **11**, 5586-5591.

Hayashi, H., Tominaga, Y., Hirano, S., McKenna, A. E., Nakabeppu, Y. and Matsumoto, Y. (2002). Replication-associated repair of adenine:8-oxoguanine mispairs by MYH. *Curr.Biol.* **12**, 335-339.

Hays, S. L., Firmenich, A. A. and Berg, P. (1995). Complex formation in yeast double-strand break repair: participation of Rad51, Rad52, Rad55, and Rad57 proteins. *Proc.Natl.Acad.Sci.U.S.A* **92**, 6925-6929.

Hedges, S. B., Blair, J. E., Venturi, M. L. and Shoe, J. L. (2004). A molecular timescale of eukaryote evolution and the rise of complex multicellular life. *BMC.Evol.Biol.* **4**, 2.

Heidenreich, E. and Wintersberger, U. (2001). Adaptive reversions of a frameshift mutation in arrested Saccharomyces cerevisiae cells by simple deletions in mononucleotide repeats. *Mutat.Res.* **473**, 101-107.

Henderson, I. R., Owen, P. and Nataro, J. P. (1999). Molecular switches--the ON and OFF of bacterial phase variation. *Mol.Microbiol.* **33**, 919-932.

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.

Her, C. and Doggett, N. A. (1998). Cloning, structural characterization, and chromosomal localization of the human orthologue of Saccharomyces cerevisiae MSH5 gene. *Genomics* **52**, 50-61.

Her, C., Wu, X., Griswold, M. D. and Zhou, F. (2003). Human MutS homologue MSH4 physically interacts with von Hippel-Lindau tumor suppressor-binding protein 1. *Cancer Res.* **63**, 865-872.

Her, C., Wu, X., Wan, W. and Doggett, N. A. (1999). Identification and characterization of the mouse MutS homolog 5: Msh5. *Mamm.Genome* **10**, 1054-1061.

Heyer, W. D., Ehmsen, K. T. and Solinger, J. A. (2003). Holliday junctions in the eukaryotic nucleus: resolution in sight? *Trends Biochem.Sci.* **28**, 548-557.

Hide, G., Cattand, P., LeRay, D., Barry, J. D. and Tait, A. (1990). The identification of Trypanosoma brucei subspecies using repetitive DNA sequences. *Mol.Biochem.Parasitol.* **39**, 213-225.

Higgins, J. D., Armstrong, S. J., Franklin, F. C. and Jones, G. H. (2004). The Arabidopsis MutS homolog AtMSH4 functions at an early step in recombination: evidence for two classes of recombination in Arabidopsis. *Genes Dev.* **18**, 2557-2570.

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.

Hoffmann, E. R. and Borts, R. H. (2004). Meiotic recombination intermediates and mismatch repair proteins. *Cytogenet.Genome Res.* **107**, 232-248.

Hogardt, M., Schubert, S., Adler, K., Gotzfried, M. and Heesemann, J. (2006). Sequence variability and functional analysis of MutS of hypermutable Pseudomonas aeruginosa cystic fibrosis isolates. *Int.J.Med.Microbiol.*

Hollingsworth, N. M. and Brill, S. J. (2004). The Mus81 solution to resolution: generating meiotic crossovers without Holliday junctions. *Genes Dev.* **18**, 117-125.

Hollingsworth, N. M., Ponte, L. and Halsey, C. (1995). MSH5, a novel MutS homolog, facilitates meiotic reciprocal recombination between homologs in Saccharomyces cerevisiae but not mismatch repair. *Genes Dev.* **9**, 1728-1739.

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.

Hope, M., MacLeod, A., Leech, V., Melville, S., Sasse, J., Tait, A. and Turner, C. M. (1999). Analysis of ploidy (in megabase chromosomes) in Trypanosoma brucei after genetic exchange. *Mol.Biochem.Parasitol.* **104**, 1-9.

Hopfner, K. P., Karcher, A., Craig, L., Woo, T. T., Carney, J. P. and Tainer, J. A. (2001). Structural biochemistry and interaction architecture of the DNA double-strand break repair Mre11 nuclease and Rad50-ATPase. *Cell* **105**, 473-485.

Horn, D. and Barry, J. D. (2005). The central roles of telomeres and subtelomeres in antigenic variation in African trypanosomes. *Chromosome.Res.* **13**, 525-533.

Horn, D. and Cross, G. A. (1995). A developmentally regulated position effect at a telomeric locus in Trypanosoma brucei. *Cell* **83**, 555-561.

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

Horn, D. and Cross, G. A. (1997b). Position-dependent and promoter-specific regulation of gene expression in Trypanosoma brucei. *EMBO J.* **16**, 7422-7431.

Horst, J. P., Wu, T. H. and Marinus, M. G. (1999). Escherichia coli mutator genes. *Trends Microbiol.* **7**, 29-36.

Horwath, M., Kramer, W. and Kunze, R. (2002). Structure and expression of the Zea mays mutS-homologs Mus1 and Mus2. *Theor.Appl.Genet.* **105**, 423-430.

Hughes, A. L., Packer, B., Welch, R., Bergen, A. W., Chanock, S. J., and Yeager, M. (2003). Widespread purifying selection at polymorphic sites in human protein-coding loci. *Proc.Natl.Acad.Sci.U.S.A* **100**, 15754-15757.

Hughes, A. L. and Piontkivska, H. (2003). Phylogeny of Trypanosomatidae and Bodonidae (Kinetoplastida) Based on 18S rRNA: Evidence for Paraphyly of Trypanosoma and Six Other Genera. *Mol.Biol.Evol.* **20**, 644-652.

Humbert, O., Fiumicino, S., Aquilina, G., Branch, P., Oda, S., Zijno, A., Karran, P. and Bignami, M. (1999). Mismatch repair and differential sensitivity of mouse and human cells to methylating agents. *Carcinogenesis* **20**, 205-214.

Humbert, O., Prudhomme, M., Hakenbeck, R., Dowson, C. G. and Claverys, J. P. (1995). Homeologous recombination and mismatch repair during transformation in Streptococcus pneumoniae: saturation of the Hex mismatch repair system. *Proc.Natl.Acad.Sci.U.S.A* **92**, 9052-9056.

Hunter, N., Chambers, S. R., Louis, E. J. and Borts, R. H. (1996). The mismatch repair system contributes to meiotic sterility in an interspecific yeast hybrid. *EMBO J.* **15**, 1726-1733.

Hunter, N. and Kleckner, N. (2001). The single-end invasion: an asymmetric intermediate at the double-strand break to double-holliday junction transition of meiotic recombination. *Cell* **106**, 59-70.

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

Imboden, M. A., Laird, P. W., Affolter, M. and Seebeck, T. (1987). Transcription of the intergenic regions of the tubulin gene cluster of Trypanosoma brucei: evidence for a polycistronic transcription unit in a eukaryote. *Nucleic Acids Res.* **15**, 7357-7368.

Interthal, H. and Heyer, W. D. (2000). MUS81 encodes a novel helix-hairpin-helix protein involved in the response to UV- and methylation-induced DNA damage in Saccharomyces cerevisiae. *Mol.Gen.Genet.* **263**, 812-827.

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.

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.

Ivens, A. C., Peacock, C. S., Worthey, E. A., Murphy, L., Aggarwal, G., Berriman, M., Sisk, E., Rajandream, M. A., Adlem, E., Aert, R., Anupama, A., Apostolou, Z., Attipoe, P., Bason, N., Bauser, C., Beck, A., Beverley, S. M., Bianchettin, G., Borzym, K., Bothe, G., Bruschi, C. V., Collins, M., Cadag, E., Ciarloni, L.,

Clayton, C., Coulson, R. M., Cronin, A., Cruz, A. K., Davies, R. M., De Gaudenzi, J., Dobson, D. E., Duesterhoeft, A., Fazelina, G., Fosker, N., Frasch, A. C., Fraser, A., Fuchs, M., Gabel, C., Goble, A., Goffeau, A., Harris, D., Hertz-Fowler, C., Hilbert, H., Horn, D., Huang, Y., Klages, S., Knights, A., Kube, M., Larke, N., Litvin, L., Lord, A., Louie, T., Marra, M., Masuy, D., Matthews, K., Michaeli, S., Mottram, J. C., Muller-Auer, S., Munden, H., Nelson, S., Norbertczak, H., Oliver, K., O'neil, S., Pentony, M., Pohl, T. M., Price, C., Purnelle, B., Quail, M. A., Rabbinowitsch, E., Reinhardt, R., Rieger, M., Rinta, J., Robben, J., Robertson, L., Ruiz, J. C., Rutter, S., Saunders, D., Schafer, M., Schein, J., Schwartz, D. C., Seeger, K., Seyler, A., Sharp, S., Shin, H., Sivam, D., Squares, R., Squares, S., Tosato, V., Vogt, C., Volckaert, G., Wambutt, R., Warren, T., Wedler, H., Woodward, J., Zhou, S., Zimmermann, W., Smith, D. F., Blackwell, J. M., Stuart, K. D., Barrell, B. and Myler, P. J. (2005). The genome of the kinetoplastid parasite, Leishmania major. *Science* **309**, 436-442.

Jackson, S. P. (2001). Detecting, signalling and repairing DNA double-strand breaks. *Biochem.Soc.Trans.* **29**, 655-661.

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

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

Jezek, P. and Hlavata, L. (2005). Mitochondria in homeostasis of reactive oxygen species in cell, tissues, and organism. *Int.J.Biochem.Cell Biol.* **37**, 2478-2503.

Jinks-Robertson, S., Michelitch, M. and Ramcharan, S. (1993). Substrate length requirements for efficient mitotic recombination in Saccharomyces cerevisiae. *Mol.Cell Biol.* **13**, 3937-3950.

Jiricny, J. (1998). Eukaryotic mismatch repair: an update. *Mutat.Res.* **409**, 107-121.

Jiricny, J. (2006). The multifaceted mismatch-repair system. *Nat.Rev.Mol.Cell Biol.* **7**, 335-346.

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.

Johnson, R. E., Kovvali, G. K., Guzder, S. N., Amin, N. S., Holm, C., Habraken, Y., Sung, P., Prakash, L. and Prakash, S. (1996a). Evidence for involvement of yeast proliferating cell nuclear antigen in DNA mismatch repair. *J.Biol.Chem.* **271**, 27987-27990.

Johnson, R. E., Kovvali, G. K., Prakash, L. and Prakash, S. (1996b). Requirement of the yeast MSH3 and MSH6 genes for MSH2-dependent genomic stability. *J.Biol.Chem.* **271**, 7285-7288.

Jones, M. and Wagner, R. (1981). N-Methyl-N'-nitro-N-nitrosoguanidine sensitivity of E. coli mutants deficient in DNA methylation and mismatch repair. *Mol.Gen.Genet.* **184**, 562-563.

Junop, M. S., Modesti, M., Guarne, A., Ghirlando, R., Gellert, M. and Yang, W. (2000). Crystal structure of the Xrcc4 DNA repair protein and implications for end joining. *EMBO J.* **19**, 5962-5970.

Junop, M. S., Yang, W., Funchain, P., Clendenin, W. and Miller, J. H. (2003). In vitro and in vivo studies of MutS, MutL and MutH mutants: correlation of mismatch repair and DNA recombination. *DNA Repair (Amst)* **2**, 387-405.

Kahn, R., Cunningham, R. P., DasGupta, C. and Radding, C. M.(1981). Polarity of heteroduplex formation promoted by Escherichia coli recA protein. *Proc.Natl.Acad.Sci.U.S.A* **78**, 4786-4790.

Kang, L. E. and Symington, L. S. (2000). Aberrant double-strand break repair in rad51 mutants of Saccharomyces cerevisiae. *Mol.Cell Biol.* **20**, 9162-9172.

Karran, P. and Bignami, M. (1992). Self-destruction and tolerance in resistance of mammalian cells to alkylation damage. *Nucleic Acids Res.* **20**, 2933-2940.

Karran, P. and Marinus, M. G. (1982). Mismatch correction at O6-methylguanine residues in E. coli DNA. *Nature* **296**, 868-869.

Keeney, S., Giroux, C. N. and Kleckner, N. (1997). Meiosis-specific DNA double-strand breaks are catalyzed by Spo11, a member of a widely conserved protein family. *Cell* **88**, 375-384.

Kelly, J. M., Law, J. M., Chapman, C. J., Van Eys, G. J. and Evans, D. A. (1991). Evidence of genetic recombination in Leishmania. *Mol.Biochem.Parasitol.* **46**, 253-263.

Kelly, K. O., Dernburg, A. F., Stanfield, G. M. and Villeneuve, A. M. (2000). Caenorhabditis elegans msh-5 is required for both normal and radiation-induced meiotic crossing over but not for completion of meiosis. *Genetics* **156**, 617-630.

Kennedy, P. G. (2004). Human African trypanosomiasis of the CNS: current issues and challenges. *J.Clin.Invest* **113**, 496-504.

Khanna, K. K. and Jackson, S. P. (2001). DNA double-strand breaks: signaling, repair and the cancer connection. *Nat.Genet.* **27**, 247-254.

Kitten, T. and Barbour, A. G. (1990). Juxtaposition of expressed variable antigen genes with a conserved telomere in the bacterium Borrelia hermsii. *Proc.Natl.Acad.Sci.U.S.A* **87**, 6077-6081.

Kleczkowska, H. E., Marra, G., Lettieri, T. and Jiricny, J. (2001). hMSH3 and hMSH6 interact with PCNA and colocalize with it to replication foci. *Genes Dev.* **15**, 724-736.

Kneitz, B., Cohen, P. E., Avdievich, E., Zhu, L., Kane, M. F., Hou, H., Jr., Kolodner, R. D., Kucherlapati, R., Pollard, J. W. and Edelmann, W. (2000). MutS homolog 4 localization to meiotic chromosomes is required for chromosome pairing during meiosis in male and female mice. *Genes Dev.* **14**, 1085-1097.

Kokoska, R. J., Stefanovic, L., Buermeyer, A. B., Liskay, R. M. and Petes, T. D. (1999). A mutation of the yeast gene encoding PCNA destabilizes both microsatellite and minisatellite DNA sequences. *Genetics* **151**, 511-519.

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

Kondo, E., Horii, A. and Fukushige, S. (2001). The interacting domains of three MutL heterodimers in man: hMLH1 interacts with 36 homologous amino acid residues within hMLH3, hPMS1 and hPMS2. *Nucleic Acids Res.* **29**, 1695-1702.

Kooy, R. F., Hirumi, H., Moloo, S. K., Nantulya, V. M., Dukes, P., Van der Linden, P. M., Duijndam, W. A., Janse, C. J. and Overdulve, J. P. (1989). Evidence for diploidy in metacyclic forms of African trypanosomes. *Proc.Natl.Acad.Sci.U.S.A* **86**, 5469-5472.

Koprowski, P., Fikus, M. U., Mieczkowski, P. and Ciesla, Z. (2002). A dominant mitochondrial mutator phenotype of Saccharomyces cerevisiae conferred by msh1 alleles altered in the sequence encoding the ATP-binding domain. *Mol.Genet.Genomics* **266**, 988-994.

Kowalczykowski, S. C. (2002). Molecular mimicry connects BRCA2 to Rad51 and recombinational DNA repair. *Nat.Struct.Biol.* **9**, 897-899.

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.

Krogh, B. O. and Symington, L. S. (2004). Recombination Proteins in Yeast. *Annu.Rev.Genet.*

Kulasekara, H. D. and Blomfield, I. C. (1999). The molecular basis for the specificity of fimE in the phase variation of type 1 fimbriae of Escherichia coli K-12. *Mol.Microbiol.* **31**, 1171-1181.

Kunkel, T. A. and Erie, D. A. (2005). DNA mismatch repair. *Annu.Rev.Biochem.* **74**, 681-710.

Kyes, S., Horrocks, P. and Newbold, C. (2001). Antigenic variation at the infected red cell surface in malaria. *Annu.Rev.Microbiol.* **55**, 673-707.

Labat, F., Pradillon, O., Garry, L., Peuchmaur, M., Fantin, B. and Denamur, E. (2005). Mutator phenotype confers advantage in Escherichia coli chronic urinary tract infection pathogenesis. *FEMS Immunol.Med.Microbiol.* **44**, 317-321.

Lamers, M. H., Perrakis, A., Enzlin, J. H., Winterwerp, H. H., De Wind, N. and Sixma, T. K. (2000). The crystal structure of DNA mismatch repair protein MutS binding to a G x T mismatch. *Nature* **407**, 711-717.

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.

Langland, G., Kordich, J., Creaney, J., Goss, K. H., Lillard-Wetherell, K., Bebenek, K., Kunkel, T. A. and Groden, J. (2001). The Bloom's syndrome protein (BLM) interacts with MLH1 but is not required for DNA mismatch repair. *J.Biol.Chem.* **276**, 30031-30035.

Langston, L. D. and Symington, L. S. (2005). Opposing roles for DNA structure-specific proteins Rad1, Msh2, Msh3, and Sgs1 in yeast gene targeting. *EMBO J.* **24**, 2214-2223.

las Alas, M. M., de Bruin, R. A., Ten Eyck, L., Los, G. and Howell, S. B. (1998). Prediction-based threading of the hMSH2 DNA mismatch repair protein. *FASEB J.* **12**, 653-663.

Laufer, G., Schaaf, G., Bollgonn, S. and Gunzl, A. (1999). In vitro analysis of alpha-amanitin-resistant transcription from the rRNA, procyclic acidic repetitive protein, and variant surface glycoprotein gene promoters in Trypanosoma brucei. *Mol.Cell Biol.* **19**, 5466-5473.

LeClerc, J. E., Li, B., Payne, W. L. and Cebula, T. A. (1996). High mutation frequencies among Escherichia coli and Salmonella pathogens. *Science* **274**, 1208-1211.

LeClerc, J. E., Payne, W. L., Kupchella, E. and Cebula, T. A. (1998). Detection of mutator subpopulations in Salmonella typhimurium LT2 by reversion of his alleles. *Mutat.Res.* **400**, 89-97.

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, M. G. and Van der Ploeg, L. H. (1997). Transcription of protein-coding genes in trypanosomes by RNA polymerase I. *Annu.Rev.Microbiol.* **51**, 463-489.

Lee, S. D. and Alani, E. (2006). Analysis of interactions between mismatch repair initiation factors and the replication processivity factor PCNA. *J.Mol.Biol.* **355**, 175-184.

Leu, J. Y., Chua, P. R. and Roeder, G. S. (1998). The meiosis-specific Hop2 protein of S. cerevisiae ensures synapsis between homologous chromosomes. *Cell* **94**, 375-386.

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.

Levati, L., Marra, G., Lettieri, T., D'Atri, S., Vernole, P., Tentori, L., Lacal, P. M., Pagani, E., Bonmassar, E., Jiricny, J. and Graziani, G. (1998). Mutation of the mismatch repair gene hMSH2 and hMSH6 in a human T-cell leukemia line tolerant to methylating agents. *Genes Chromosomes.Cancer* **23**, 159-166.

Li, B., Tsui, H. C., LeClerc, J. E., Dey, M., Winkler, M. E. and Cebula, T. A. (2003). Molecular analysis of mutS expression and mutation in natural isolates of pathogenic Escherichia coli. *Microbiology* **149**, 1323-1331.

Li, G. M. (1999). The role of mismatch repair in DNA damage-induced apoptosis. *Oncol.Res.* **11**, 393-400.

Li, L., Jean, M. and Belzile, F. (2006). The impact of sequence divergence and DNA mismatch repair on homeologous recombination in Arabidopsis. *Plant J.* **45**, 908-916.

Liang, F., Han, M., Romanienko, P. J. and Jasin, M. (1998). Homology-directed repair is a major double-strand break repair pathway in mammalian cells. *Proc.Natl.Acad.Sci.U.S.A* **95**, 5172-5177.

Ligtenberg, M. J., Bitter, W., Kieft, R., Steverding, D., Janssen, H., Calafat, J. and Borst, P. (1994). Reconstitution of a surface transferrin binding complex in insect form Trypanosoma brucei. *EMBO J.* **13**, 2565-2573.

Lin, Y. L., Shivji, M. K., Chen, C., Kolodner, R., Wood, R. D. and Dutta, A. (1998). The evolutionarily conserved zinc finger motif in the largest subunit of human replication protein A is required for DNA replication and mismatch repair but not for nucleotide excision repair. *J.Biol.Chem.* **273**, 1453-1461.

Linardopoulou, E. V., Williams, E. M., Fan, Y., Friedman, C., Young, J. M. and Trask, B. J. (2005). Human subtelomeres are hot spots of interchromosomal recombination and segmental duplication. *Nature* **437**, 94-100.

Lipkin, S. M., Moens, P. B., Wang, V., Lenzi, M., Shanmugarajah, D., Gilgeous, A., Thomas, J., Cheng, J., Touchman, J. W., Green, E. D., Schwartzberg, P., Collins, F. S. and Cohen, P. E. (2002). Meiotic arrest and aneuploidy in MLH3-deficient mice. *Nat.Genet.* **31**, 385-390.

Lipkin, S. M., Wang, V., Jacoby, R., Banerjee-Basu, S., Baxevanis, A. D., Lynch, H. T., Elliott, R. M. and Collins, F. S. (2000). MLH3: a DNA mismatch repair gene associated with mammalian microsatellite instability. *Nat.Genet.* **24**, 27-35.

Lisby, M., Mortensen, U. H. and Rothstein, R. (2003). Colocalization of multiple DNA double-strand breaks at a single Rad52 repair centre. *Nat.Cell Biol.*

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, B., Liu, Y., Motyka, S. A., Agbo, E. E. and Englund, P. T. (2005). Fellowship of the rings: the replication of kinetoplast DNA. *Trends Parasitol.* **21**, 363-369.

Liu, Y., Masson, J. Y., Shah, R., O'Regan, P. and West, S. C. (2004). RAD51C is required for Holliday junction processing in mammalian cells. *Science* **303**, 243-246.

Liu, Y. and West, S. C. (2002). Distinct functions of BRCA1 and BRCA2 in double-strand break repair. *Breast Cancer Res.* **4**, 9-13.

Longley, M. J., Pierce, A. J. and Modrich, P. (1997). DNA polymerase delta is required for human mismatch repair in vitro. *J.Biol.Chem.* **272**, 10917-10921.

Lovett, S. T., Hurley, R. L., Sutera, V. A., Jr., Aubuchon, R. H. and Lebedeva, M. A. (2002). Crossing over between regions of limited homology in Escherichia coli. RecA-dependent and RecA-independent pathways. *Genetics* **160**, 851-859.

Lu, A. L., Clark, S. and Modrich, P. (1983). Methyl-directed repair of DNA base-pair mismatches in vitro. *Proc.Natl.Acad.Sci.U.S.A* **80**, 4639-4643.

Lukacsovich, T. and Waldman, A. S. (1999). Suppression of intrachromosomal gene conversion in mammalian cells by small degrees of sequence divergence. *Genetics* **151**, 1559-1568.

Machado, C. A. and Ayala, F. J. (2001). Nucleotide sequences provide evidence of genetic exchange among distantly related lineages of Trypanosoma cruzi. *Proc.Natl.Acad.Sci.U.S.A* **98**, 7396-7401.

Machado, C. R., Augusto-Pinto, L., McCulloch, R. and Teixeira, S. M. (2006). DNA metabolism and genetic diversity in Trypanosomes. *Mutat.Res.* **612**, 40-57.

Macia, M. D., Blanquer, D., Togores, B., Sauleda, J., Perez, J. L. and Oliver, A. (2005). Hypermutation is a key factor in development of multiple-antimicrobial resistance in Pseudomonas aeruginosa strains causing chronic lung infections. *Antimicrob.Agents Chemother.* **49**, 3382-3386.

MacLeod, A., Tweedie, A., Welburn, S.C., Maudlin, I., Turner, C.M., Tait, A. (2000). Minisatellite marker analysis of Trypanosoma brucei: reconciliation of clonal, panmictic, and epidemic population genetic structures. *Proc.Natl.Acad.Sci.U.S.A* **97**, 13442-13447.

MacLeod, A., Tweedie, A., McLellan, S., Taylor, S., Cooper, A., Sweeney, L., Turner, C. M. and Tait, A. (2005a). Allelic segregation and independent assortment in T. brucei crosses: Proof that the genetic system is Mendelian and involves meiosis. *Mol.Biochem.Parasitol.*

MacLeod, A., Tweedie, A., McLellan, S., Taylor, S., Hall, N., Berriman, M., El Sayed, N. M., Hope, M., Turner, C. M. and Tait, A. (2005b). The genetic map and comparative analysis with the physical map of Trypanosoma brucei. *Nucleic Acids Res.* **33**, 6688-6693.

Mahadevaiah, S. K., Turner, J. M., Baudat, F., Rogakou, E. P., de Boer, P., Blanco-Rodriguez, J., Jasin, M., Keeney, S., Bonner, W. M. and Burgoyne, P. S. (2001). Recombinational DNA double-strand breaks in mice precede synapsis. *Nat.Genet.* **27**, 271-276.

Majewski, J. and Cohan, F. M. (1999). DNA sequence similarity requirements for interspecific recombination in Bacillus. *Genetics* **153**, 1525-1533.

Majewski, J., Zawadzki, P., Pickerill, P., Cohan, F. M. and Dowson, C. G. (2000). Barriers to genetic exchange between bacterial species: Streptococcus pneumoniae transformation. *J.Bacteriol.* **182**, 1016-1023.

Maki, H. and Sekiguchi, M. (1992). MutT protein specifically hydrolyses a potent mutagenic substrate for DNA synthesis. *Nature* **355**, 273-275.

Malagon, F. and Aguilera, A. (2001). Yeast spt6-140 mutation, affecting chromatin and transcription, preferentially increases recombination in which Rad51p-mediated strand exchange is dispensable. *Genetics* **158**, 597-611.

Malkov, V. A., Biswas, I., Camerini-Otero, R. D. and Hsieh, P. (1997). Photocross-linking of the NH2-terminal region of Taq MutS protein to the major groove of a heteroduplex DNA. *J.Biol.Chem.* **272**, 23811-23817.

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., Naylor, M. L., Yamaguchi, M., Ira, G. and Haber, J. E. (2005). RAD51-dependent break-induced replication differs in kinetics and checkpoint responses from RAD51-mediated gene conversion. *Mol.Cell Biol.* **25**, 933-944.

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.

Mao, E. F., Lane, L., Lee, J. and Miller, J. H. (1997). Proliferation of mutators in A cell population. *J.Bacteriol.* **179**, 417-422.

Marsischky, G. T., Filosi, N., Kane, M. F. and Kolodner, R. (1996). Redundancy of Saccharomyces cerevisiae MSH3 and MSH6 in MSH2-dependent mismatch repair. *Genes Dev.* **10**, 407-420.

Martin, A., Morcillo, N., Lemus, D., Montoro, E., Telles, M. A., Simboli, N., Pontino, M., Porras, T., Leon, C., Velasco, M., Chacon, L., Barrera, L., Ritacco, V., Portaels, F. and Palomino, J. C. (2005). Multicenter study of MTT and resazurin assays for testing susceptibility to first-line anti-tuberculosis drugs. *Int.J.Tuberc.Lung Dis.* **9**, 901-906.

Masson, J. Y. and West, S. C. (2001). The Rad51 and Dmc1 recombinases: a non-identical twin relationship. *Trends Biochem.Sci.* **26**, 131-136.

Mathieu-Daude, F., Bicart-See, A., Bosseno, M. F., Breniere, S. F. and Tibayrenc, M. (1994). Identification of Trypanosoma brucei gambiense group I by a specific kinetoplast DNA probe. *Am.J.Trop.Med.Hyg.* **50**, 13-19.

Matic, I., Radman, M., Taddei, F., Picard, B., Doit, C., Bingen, E., Denamur, E. and Elion, J. (1997). Highly variable mutation rates in commensal and pathogenic Escherichia coli. *Science* **277**, 1833-1834.

Matic, I., Rayssiguier, C. and Radman, M. (1995). Interspecies gene exchange in bacteria: the role of SOS and mismatch repair systems in evolution of species. *Cell* **80**, 507-515.

Matic, I., Taddei, F. and Radman, M. (2000). Interspecies recombination and mismatch repair. Generation of mosaic genes and genomes. *Methods Mol.Biol.* **152**, 149-157.

Matthews, K. R., Ellis, J. R. and Paterou, A. (2004). Molecular regulation of the life cycle of African trypanosomes. *Trends Parasitol.* **20**, 40-47.

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.

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

Mazin, A. V., Zaitseva, E., Sung, P. and Kowalczykowski, S. C. (2000b). Tailed duplex DNA is the preferred substrate for Rad51 protein-mediated homologous pairing. *EMBO J.* **19**, 1148-1156.

Mazurek, A., Berardini, M. and Fishel, R. (2002b). Activation of human MutS homologs by 8-oxo-guanine DNA damage. *J.Biol.Chem.* **277**, 8260-8266.

Mazurek, A., Berardini, M. and Fishel, R. (2002a). Activation of human MutS homologs by 8-oxo-guanine DNA damage. *J.Biol.Chem.* **277**, 8260-8266.

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

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, S. D., Gu, L. and Li, G. M. (2003). Bi-directional processing of DNA loops by mismatch repair-dependent and -independent pathways in human cells. *J.Biol.Chem.* **278**, 3891-3896.

McEachern, M. J. and Haber, J. E. (2006). Break-Induced Replication and Recombinational Telomere Elongation in Yeast. *Annu.Rev.Biochem.*

McIlwraith, M. J., Hall, D. R., Stasiak, A. Z., Stasiak, A., Wigley, D. B. and West, S. C. (2001). RadA protein from Archaeoglobus fulgidus forms rings, nucleoprotein filaments and catalyses homologous recombination. *Nucleic Acids Res.* **29**, 4509-4517.

McKean, P. G., Keen, J. K., Smith, D. F. and Benson, F. E. (2001). Identification and characterisation of a RAD51 gene from Leishmania major. *Mol.Biochem.Parasitol.* **115**, 209-216.

McKim, K. S. and Hayashi-Hagihara, A. (1998). mei-W68 in Drosophila melanogaster encodes a Spo11 homolog: evidence that the mechanism for initiating meiotic recombination is conserved. *Genes Dev.* **12**, 2932-2942.

Mehr, I. J. and Seifert, H. S. (1998). Differential roles of homologous recombination pathways in Neisseria gonorrhoeae pilin antigenic variation, DNA transformation and DNA repair. *Mol.Microbiol.* **30**, 697-710.

Mello, J. A., Acharya, S., Fishel, R. and Essigmann, J. M. (1996). The mismatch-repair protein hMSH2 binds selectively to DNA adducts of the anticancer drug cisplatin. *Chem.Biol.* **3**, 579-589.

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., Leech, V., Navarro, M. and Cross, G. A. (2000). The molecular karyotype of the megabase chromosomes of Trypanosoma brucei stock 427. *Mol.Biochem.Parasitol.* **111**, 261-273.

Mezard, C. and Nicolas, A. (1994). Homologous, homeologous, and illegitimate repair of double-strand breaks during transformation of a wild-type strain and a rad52 mutant strain of Saccharomyces cerevisiae. *Mol.Cell Biol.* **14**, 1278-1292.

Michel, B., Grompone, G., Flores, M. J. and Bidnenko, V. (2004). Multiple pathways process stalled replication forks. *Proc.Natl.Acad.Sci.U.S.A* **101**, 12783-12788.

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.

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.

Miller, J. H. and Michaels, M. (1996). Finding new mutator strains of Escherichia coli--a review. *Gene* **179**, 129-132.

Miller, K., O'Neill, A. J. and Chopra, I. (2002). Response of Escherichia coli hypermutators to selection pressure with antimicrobial agents from different classes. *J.Antimicrob.Chemother.* **49**, 925-934.

Milne, G. T. and Weaver, D. T. (1993). Dominant negative alleles of RAD52 reveal a DNA repair/recombination complex including Rad51 and Rad52. *Genes Dev.* **7**, 1755-1765.

Mitra, S., Pal, B. C. and Foote, R. S. (1982). O6-methylguanine-DNA methyltransferase in wild-type and ada mutants of Escherichia coli. *J.Bacteriol.* **152**, 534-537.

Modrich, P. (1991). Mechanisms and biological effects of mismatch repair. *Annu.Rev.Genet.* **25**, 229-253.

Modrich, P. and Lahue, R. (1996). Mismatch repair in replication fidelity, genetic recombination, and cancer biology. *Annu.Rev.Biochem.* **65**, 101-133.

Moens, P. B., Kolas, N. K., Tarsounas, M., Marcon, E., Cohen, P. E. and Spyropoulos, B. (2002). The time course and chromosomal localization of recombination-related proteins at meiosis in the mouse are compatible with models that can resolve the early DNA-DNA interactions without reciprocal recombination. *J.Cell Sci.* **115**, 1611-1622.

Mookerjee, S. A., Lyon, H. D. and Sia, E. A. (2005). Analysis of the functional domains of the mismatch repair homologue Msh1p and its role in mitochondrial genome maintenance. *Curr.Genet.* **47**, 84-99.

Moreland, N. J., Illand, M., Kim, Y. T., Paul, J. and Brown, R. (1999). Modulation of drug resistance mediated by loss of mismatch repair by the DNA polymerase inhibitor aphidicolin. *Cancer Res.* **59**, 2102-2106.

Morrison, L. J., Majiwa, P., Read, A. F. and Barry, J. D. (2005). Probabilistic order in antigenic variation of Trypanosoma brucei. *Int.J.Parasitol.* **35**, 961-972.

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.

Mowatt, M. R., Aggarwal, A. and Nash, T. E. (1991). Carboxy-terminal sequence conservation among variant-specific surface proteins of Giardia lamblia. *Mol.Biochem.Parasitol.* **49**, 215-227.

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.

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

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.

Nakagawa, T., Datta, A. and Kolodner, R. D. (1999). Multiple functions of MutS- and MutL-related heterocomplexes. *Proc.Natl.Acad.Sci.U.S.A* **96**, 14186-14188.

Nash, T. E. (2002). Surface antigenic variation in Giardia lamblia. *Mol.Microbiol.* **45**, 585-590.

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.

Neale, M. J., Pan, J. and Keeney, S. (2005). Endonucleolytic processing of covalent protein-linked DNA double-strand breaks. *Nature* **436**, 1053-1057.

Negritto, M. T., Wu, X., Kuo, T., Chu, S. and Bailis, A. M. (1997). Influence of DNA sequence identity on efficiency of targeted gene replacement. *Mol.Cell Biol.* **17**, 278-286.

Neyton, S., Lespinasse, F., Moens, P. B., Paul, R., Gaudray, P., Paquis-Flucklinger, V. and Santucci-Darmanin, S. (2004). Association between MSH4 (MutS homologue 4) and the DNA strand-exchange RAD51 and DMC1 proteins during mammalian meiosis. *Mol.Hum.Reprod* **10**, 917-924.

Ni, T. T., Marsischky, G. T. and Kolodner, R. D. (1999). MSH2 and MSH6 are required for removal of adenine misincorporated opposite 8-oxo-guanine in S. cerevisiae. *Mol.Cell* **4**, 439-444.

Nicholson, A., Hendrix, M., Jinks-Robertson, S. and Crouse, G. F. (2000). Regulation of mitotic homeologous recombination in yeast. Functions of mismatch repair and nucleotide excision repair genes. *Genetics* **154**, 133-146.

Nilsson, A. I., Kugelberg, E., Berg, O. G. and Andersson, D. I. (2004). Experimental adaptation of Salmonella typhimurium to mice. *Genetics* **168**, 1119-1130.

Njiokou, F., Nkinin, S.W., Grebaut, P., Penchenier, L., Barnabe, C., Tibayrenc, M., and Herder, S. (2004). An isoenzyme survey of Trypanosoma brucei s.l. from the Central African subregion: population structure, taxonomic and epidemiological considerations. *Parasitology* **128**, 645-653.

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

Novak, J. E., Ross-Macdonald, P. B. and Roeder, G. S. (2001). The budding yeast Msh4 protein functions in chromosome synapsis and the regulation of crossover distribution. *Genetics* **158**, 1013-1025.

Nowosielska, A., Smith, S. A., Engelward, B. P. and Marinus, M. G. (2006). Homologous recombination prevents methylation-induced toxicity in Escherichia coli. *Nucleic Acids Res.* **34**, 2258-2268.

O'Brien, J., Wilson, I., Orton, T. and Pognan, F. (2000). Investigation of the Alamar Blue (resazurin) fluorescent dye for the assessment of mammalian cell cytotoxicity. *Eur.J.Biochem.* **267**, 5421-5426.

O'Brien, V. and Brown, R. (2006). Signalling cell cycle arrest and cell death through the MMR System. *Carcinogenesis* **27**, 682-692.

Obmolova, G., Ban, C., Hsieh, P. and Yang, W. (2000). Crystal structures of mismatch repair protein MutS and its complex with a substrate DNA. *Nature* **407**, 703-710.

Oda, S., Zhao, Y., and Maehara, Y. (2005). Microsatellite instability in gastrointestinal tract cancers: a brief update. Surg. Today **35**, 1005-1015.

Oliver, A., Canton, R., Campo, P., Baquero, F. and Blazquez, J. (2000). High frequency of hypermutable Pseudomonas aeruginosa in cystic fibrosis lung infection. *Science* **288**, 1251-1254.

Onyango, J. D., Burri, C. and Brun, R. (2000). An automated biological assay to determine levels of the trypanocidal drug melarsoprol in biological fluids. *Acta Trop.* **74**, 95-100.

Opperman, R., Emmanuel, E. and Levy, A. A. (2004). The effect of sequence divergence on recombination between direct repeats in Arabidopsis. *Genetics* **168**, 2207-2215.

Osman, F., Dixon, J., Doe, C. L. and Whitby, M. C. (2003). Generating crossovers by resolution of nicked Holliday junctions: a role for Mus81-Emel in meiosis. *Mol.Cell* **12**, 761-774.

Otsu, K., Donelson, J. E. and Kirchhoff, L. V. (1993). Interruption of a Trypanosoma cruzi gene encoding a protein containing 14-amino acid repeats by targeted insertion of the neomycin phosphotransferase gene. *Mol.Biochem.Parasitol.* **57**, 317-330.

Overath, P. and Engstler, M. (2004). Endocytosis, membrane recycling and sorting of GPI-anchored proteins: Trypanosoma brucei as a model system. *Mol.Microbiol.* **53**, 735-744.

Overath, P., Steverding, D., Chaudhri, M., Stierhof, Y. D. and Ziegelbauer, K. (1994). Structure and function of GPI-anchored surface proteins of Trypanosoma brucei. *Braz.J.Med.Biol.Res.* **27**, 343-347.

Padmore, R., Cao, L. and Kleckner, N. (1991). Temporal comparison of recombination and synaptonemal complex formation during meiosis in S. cerevisiae. *Cell* **66**, 1239-1256.

Page, S. L. and Hawley, R. S. (2003). Chromosome choreography: the meiotic ballet. *Science* **301**, 785-789.

Page, S. L. and Hawley, R. S. (2004). The genetics and molecular biology of the synaptonemal complex. *Annu.Rev.Cell Dev.Biol.* **20**, 525-558.

Paindavoine, P., Zampetti-Bosseler, F., Pays, E., Schweizer, J., Guyaux, M., Jenni, L. and Steinert, M. (1986). Trypanosome hybrids generated in tsetse flies by nuclear fusion. *EMBO J.* **5**, 3631-3636.

Pang, Q., Prolla, T. A. and Liskay, R. M. (1997). Functional domains of the Saccharomyces cerevisiae Mlh1p and Pms1p DNA mismatch repair proteins and their relevance to human hereditary nonpolyposis colorectal cancer-associated mutations. *Mol.Cell Biol.* **17**, 4465-4473.

Panton, L. J., Tesh, R. B., Nadeau, K. C. and Beverley, S. M. (1991). A test for genetic exchange in mixed infections of Leishmania major in the sand fly Phlebotomus papatasi. *J.Protozool.* **38**, 224-228.

Papadopoulou, B. and Dumas, C. (1997). Parameters controlling the rate of gene targeting frequency in the protozoan parasite Leishmania. *Nucleic Acids Res.* **25**, 4278-4286.

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.

Paquis-Flucklinger, V., Santucci-Darmanin, S., Paul, R., Saunieres, A., Turc-Carel, C. and Desnuelle, C. (1997). Cloning and expression analysis of a meiosis-specific MutS homolog: the human MSH4 gene. *Genomics* **44**, 188-194.

Parker, B. O. and Marinus, M. G. (1992). Repair of DNA heteroduplexes containing small heterologous sequences in Escherichia coli. *Proc.Natl.Acad.Sci.U.S.A* **89**, 1730-1734.

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.

Paull, T. T. and Gellert, M. (1999). Nbs1 potentiates ATP-driven DNA unwinding and endonuclease cleavage by the Mre11/Rad50 complex. *Genes Dev.* **13**, 1276-1288.

Pavlov, Y. I., Mian, I. M. and Kunkel, T. A. (2003). Evidence for preferential mismatch repair of lagging strand DNA replication errors in yeast. *Curr.Biol.* **13**, 744-748.

Pavlov, Y. I., Newlon, C. S. and Kunkel, T. A. (2002). Yeast origins establish a strand bias for replicational mutagenesis. *Mol.Cell* **10**, 207-213.

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

Pays, E., Guyaux, M., Aerts, D., Van Meirvenne, N. and Steinert, M. (1985). Telomeric reciprocal recombination as a possible mechanism for antigenic variation in trypanosomes. *Nature* **316**, 562-564.

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.

Pays, E., Van Meirvenne, N., Le Ray, D. and Steinert, M. (1981). Gene duplication and transposition linked to antigenic variation in Trypanosoma brucei. *Proc.Natl.Acad.Sci.U.S.A* **78**, 2673-2677.

Pays, E., Vanhamme, L. and Berberof, M. (1994). Genetic controls for the expression of surface antigens in African trypanosomes. *Annu.Rev.Microbiol.* **48**, 25-52.

Pays, E., Vanhamme, L. and Perez-Morga, D. (2004). Antigenic variation in Trypanosoma brucei: facts, challenges and mysteries. *Curr.Opin.Microbiol.* **7**, 369-374.

Pays, E., Vanhollebeke, B., Vanhamme, L., Paturiaux-Hanocq, F., Nolan, D. P. and Perez-Morga, D. (2006). The trypanolytic factor of human serum. *Nat.Rev.Microbiol.* **4**, 477-486.

Pedrazzi, G., Perrera, C., Blaser, H., Kuster, P., Marra, G., Davies, S. L., Ryu, G. H., Freire, R., Hickson, I. D., Jiricny, J. and Stagljar, I. (2001). Direct association of Bloom's syndrome gene product with the human mismatch repair protein MLH1. *Nucleic Acids Res.* **29**, 4378-4386.

Pena-Diaz, J., Akbari, M., Sundheim, O., Farez-Vidal, M. E., Andersen, S., Sneve, R., Gonzalez-Pacanowska, D., Krokan, H. E. and Slupphaug, G. (2004). Trypanosoma cruzi contains a single detectable uracil-DNA glycosylase and repairs uracil exclusively via short patch base excision repair. *J.Mol.Biol.* **342**, 787-799.

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

Petit, M. A., Dimpfl, J., Radman, M. and Echols, H. (1991). Control of large chromosomal duplications in Escherichia coli by the mismatch repair system. *Genetics* **129**, 327-332.

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. V., Pezza, R. J., Vanevski, F., Ploquin, M., Masson, J. Y. and Camerini-Otero, R. D. (2005). The Hop2 and Mnd1 proteins act in concert with Rad51 and Dmc1 in meiotic recombination. *Nat.Struct.Mol.Biol.* **12**, 449-453.

Petukhova, G. V., Romanienko, P. J. and Camerini-Otero, R. D. (2003). The Hop2 protein has a direct role in promoting interhomolog interactions during mouse meiosis. *Dev.Cell* **5**, 927-936.

Picard, B., Duriez, P., Gouriou, S., Matic, I., Denamur, E. and Taddei, F. (2001). Mutator natural Escherichia coli isolates have an unusual virulence phenotype. *Infect.Immun.* **69**, 9-14.

Piontkivska, H. and Hughes, A. L. (2005). Environmental kinetoplastid-like 18S rRNA sequences and phylogenetic relationships among Trypanosomatidae: paraphyly of the genus Trypanosoma. *Mol.Biochem.Parasitol.* **144**, 94-99.

Pochart, P., Woltering, D. and Hollingsworth, N. M. (1997). Conserved properties between functionally distinct MutS homologs in yeast. *J.Biol.Chem.* **272**, 30345-30349.

Polach, K. J. and Widom, J. (1995). Mechanism of protein access to specific DNA sequences in chromatin: a dynamic equilibrium model for gene regulation. *J.Mol.Biol.* **254**, 130-149.

Priebe, S. D., Westmoreland, J., Nilsson-Tillgren, T. and Resnick, M. A. (1994). Induction of recombination between homologous and diverged DNAs by double-strand gaps and breaks and role of mismatch repair. *Mol.Cell Biol.* **14**, 4802-4814.

Proudfoot, C. and McCulloch, R. (2005). Distinct roles for two RAD51-related genes in Trypanosoma brucei antigenic variation. *Nucleic Acids Res.* **33**, 6906-6919.

Proudfoot, C. and McCulloch, R. (2006). Trypanosoma brucei DMC1 does not act in DNA recombination, repair or antigenic variation in bloodstream stage cells. *Mol.Biochem.Parasitol.* **145**, 245-253.

Prunier, A. L. and Leclercq, R. (2005). Role of mutS and mutL genes in hypermutability and recombination in Staphylococcus aureus. *J.Bacteriol.* **187**, 3455-3464.

Radding, C. M. (1991). Helical interactions in homologous pairing and strand exchange driven by RecA protein. *J.Biol.Chem.* **266**, 5355-5358.

Radman, M., Matic, I., Halliday, J. A. and Taddei, F. (1995). Editing DNA replication and recombination by mismatch repair: from bacterial genetics to mechanisms of predisposition to cancer in humans. *Philos.Trans.R.Soc.Lond B Biol.Sci.* **347**, 97-103.

Ramesh, M. A., Malik, S. B. and Logsdon, J. M., Jr. (2005). A phylogenomic inventory of meiotic genes; evidence for sex in Giardia and an early eukaryotic origin of meiosis. *Curr.Biol.* **15**, 185-191.

Ramsden, D. A. and Gellert, M. (1998). Ku protein stimulates DNA end joining by mammalian DNA ligases: a direct role for Ku in repair of DNA double-strand breaks. *EMBO J.* **17**, 609-614.

Rattray, A. J. and Symington, L. S. (1994). Use of a chromosomal inverted repeat to demonstrate that the RAD51 and RAD52 genes of Saccharomyces cerevisiae have different roles in mitotic recombination. *Genetics* **138**, 587-595.

Rayssiguier, C., Thaler, D. S. and Radman, M. (1989). The barrier to recombination between Escherichia coli and Salmonella typhimurium is disrupted in mismatch-repair mutants. *Nature* **342**, 396-401.

Raz, B., Iten, M., Grether-Buhler, Y., Kaminsky, R. and Brun, R. (1997). The Alamar Blue assay to determine drug sensitivity of African trypanosomes (T.b. rhodesiense and T.b. gambiense) in vitro. *Acta Trop.* **68**, 139-147.

Reenan, R. A. and Kolodner, R. D. (1992). Characterization of insertion mutations in the Saccharomyces cerevisiae MSH1 and MSH2 genes: evidence for separate mitochondrial and nuclear functions. *Genetics* **132**, 975-985.

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.

Ricchetti, M., Dujon, B. and Fairhead, C. (2003). Distance from the chromosome end determines the efficiency of double strand break repair in subtelomeres of haploid yeast. *J.Mol.Biol.* **328**, 847-862.

Richardson, A. R., Yu, Z., Popovic, T. and Stojiljkovic, I. (2002). Mutator clones of Neisseria meningitidis in epidemic serogroup A disease. *Proc.Natl.Acad.Sci.U.S.A* **99**, 6103-6107.

Richardson, C., Horikoshi, N. and Pandita, T. K. (2004). The role of the DNA double-strand break response network in meiosis. *DNA Repair (Amst)* **3**, 1149-1164.

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.

Ristic, D., Modesti, M., Kanaar, R. and Wyman, C. (2003). Rad52 and Ku bind to different DNA structures produced early in double-strand break repair. *Nucleic Acids Res.* **31**, 5229-5237.

Robello, C., Gamarro, F., Castanys, S. and Alvarez-Valin, F. (2000). Evolutionary relationships in Trypanosoma cruzi: molecular phylogenetics supports the existence of a new major lineage of strains. *Gene* **246**, 331-338.

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.

Roditi, I. and Liniger, M. (2002). Dressed for success: the surface coats of insect-borne protozoan parasites. *Trends Microbiol.* **10**, 128-134.

Romanienko, P. J. and Camerini-Otero, R. D. (2000). The mouse Spo11 gene is required for meiotic chromosome synapsis. *Mol.Cell* **6**, 975-987.

Rosenberg, S. M. (2001). Evolving responsively: adaptive mutation. *Nat.Rev.Genet.* **2**, 504-515.

Ross-Macdonald, P. and Roeder, G. S. (1994). Mutation of a meiosis-specific MutS homolog decreases crossing over but not mismatch correction. *Cell* **79**, 1069-1080.

Rubnitz, J. and Subramani, S. (1984). The minimum amount of homology required for homologous recombination in mammalian cells. *Mol.Cell Biol.* **4**, 2253-2258.

Rudenko, G., Blundell, P. A., Dirks-Mulder, A., Kieft, R. and Borst, P. (1995). A ribosomal DNA promoter replacing the promoter of a telomeric VSG gene expression site can be efficiently switched on and off in T. brucei. *Cell* **83**, 547-553.

Rudenko, G., Chaves, I., Dirks-Mulder, A. and Borst, P. (1998). Selection for activation of a new variant surface glycoprotein gene expression site in Trypanosoma brucei can result in deletion of the old one. *Mol.Biochem.Parasitol.* **95**, 97-109.

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.

Russo, M. T., Blasi, M. F., Chiera, F., Fortini, P., Degan, P., Macpherson, P., Furuichi, M., Nakabeppu, Y., Karran, P., Aquilina, G. and Bignami, M. (2004). The oxidized deoxynucleoside triphosphate pool is a significant contributor to genetic instability in mismatch repair-deficient cells. *Mol.Cell Biol.* **24**, 465-474.

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

Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989). Molecular cloning: a laboratory manual. (Cold Spring Harbor: N.Y.)

Santucci-Darmanin, S., Neyton, S., Lespinasse, F., Saunieres, A., Gaudray, P. and Paquis-Flucklinger, V. (2002). The DNA mismatch-repair MLH3 protein interacts with MSH4 in meiotic cells, supporting a role for this MutL homolog in mammalian meiotic recombination. *Hum.Mol.Genet.* **11**, 1697-1706.

Santucci-Darmanin, S., Paul, R., Michiels, J. F., Saunieres, A., Desnuelle, C. and Paquis-Flucklinger, V. (1999). Alternative splicing of hMSH4: two isoforms in testis and abnormal transcripts in somatic tissues. *Mamm.Genome* **10**, 423-427.

Santucci-Darmanin, S., Vidal, F., Scimeca, J. C., Turc-Carel, C. and Paquis-Flucklinger, V. (2001). Family of SRY/Sox proteins is involved in the regulation of the mouse Msh4 (MutS Homolog 4) gene expression. *Mol.Reprod.Dev.* **60**, 172-180.

Santucci-Darmanin, S., Walpita, D., Lespinasse, F., Desnuelle, C., Ashley, T. and Paquis-Flucklinger, V. (2000). MSH4 acts in conjunction with MLH1 during mammalian meiosis. *FASEB J.* **14**, 1539-1547.

Saparbaev, M., Prakash, L. and Prakash, S. (1996). Requirement of mismatch repair genes MSH2 and MSH3 in the RAD1-RAD10 pathway of mitotic recombination in Saccharomyces cerevisiae. *Genetics* **142**, 727-736.

Saunders, N. J., Peden, J. F., Hood, D. W. and Moxon, E. R. (1998). Simple sequence repeats in the Helicobacter pylori genome. *Mol.Microbiol.* **27**, 1091-1098.

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

Scherf, A., Figueiredo, L. M. and Freitas-Junior, L. H. (2001). Plasmodium telomeres: a pathogen's perspective. *Curr.Opin.Microbiol.* **4**, 409-414.

Schlacher, K., Pham, P., Cox, M. M. and Goodman, M. F. (2006). Roles of DNA polymerase V and RecA protein in SOS damage-induced mutation. *Chem.Rev.* **106**, 406-419.

Schmutte, C., Marinescu, R. C., Sadoff, M. M., Guerrette, S., Overhauser, J. and Fishel, R. (1998). Human exonuclease I interacts with the mismatch repair protein hMSH2. *Cancer Res.* **58**, 4537-4542.

Schofield, M. J., Brownewell, F. E., Nayak, S., Du, C., Kool, E. T. and Hsieh, P. (2001). The Phe-X-Glu DNA binding motif of MutS. The role of hydrogen bonding in mismatch recognition. *J.Biol.Chem.* **276**, 45505-45508.

Schreer, A., Tinson, C., Sherry, J. P. and Schirmer, K. (2005). Application of Alamar blue/5-carboxyfluorescein diacetate acetoxymethyl ester as a noninvasive cell viability assay in primary hepatocytes from rainbow trout. *Anal.Biochem.* **344**, 76-85.

Schwacha, A. and Kleckner, N. (1997). Interhomolog bias during meiotic recombination: meiotic functions promote a highly differentiated interhomolog-only pathway. *Cell* **90**, 1123-1135.

Schweizer, J., Pospichal, H. and Jenni, L. (1991). Hybrid formation between African trypanosomes in vitro. *Acta Trop.* **49**, 237-240.

Seebeck, T., Whittaker, P. A., Imboden, M. A., Hardman, N. and Braun, R. (1983). Tubulin genes of Trypanosoma brucei: a tightly clustered family of alternating genes. *Proc.Natl.Acad.Sci.U.S.A* **80**, 4634-4638.

Seitz, E. M., Brockman, J. P., Sandler, S. J., Clark, A. J. and Kowalczykowski, S. C. (1998). RadA protein is an archaeal RecA protein homolog that catalyzes DNA strand exchange. *Genes Dev.* **12**, 1248-1253.

Selva, E. M., New, L., Crouse, G. F. and Lahue, R. S. (1995). Mismatch correction acts as a barrier to homeologous recombination in Saccharomyces cerevisiae. *Genetics* **139**, 1175-1188.

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.

Shaver, A. C., Dombrowski, P. G., Sweeney, J. Y., Treis, T., Zappala, R. M. and Sniegowski, P. D. (2002). Fitness evolution and the rise of mutator alleles in experimental Escherichia coli populations. *Genetics* **162**, 557-566.

Shea, C., Glass, D. J., Parangi, S. and Van der Ploeg, L. H. (1986). Variant surface glycoprotein gene expression site switches in Trypanosoma brucei. *J.Biol.Chem.* **261**, 6056-6063.

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.

Sheader, K., te, Vruchte D. and Rudenko, G. (2004). Bloodstream form-specific up-regulation of silent vsg expression sites and procyclin in Trypanosoma brucei after inhibition of DNA synthesis or DNA damage. *J.Biol.Chem.* **279**, 13363-13374.

Shen, P. and Huang, H. V. (1986). Homologous recombination in Escherichia coli: dependence on substrate length and homology. *Genetics* **112**, 441-457.

Shen, S., Arhin, G. K., Ullu, E. and Tschudi, C. (2001). In vivo epitope tagging of Trypanosoma brucei genes using a one step PCR-based strategy. *Mol.Biochem.Parasitol.* **113**, 171-173.

Shi, B. S., Cai, Z. N., Yang, J. and Yu, Y. N. (2004). N-methyl-N'-nitro-N-nitrosoguanidine sensitivity, mutator phenotype and sequence specificity of spontaneous mutagenesis in FEN-1-deficient cells. *Mutat.Res.* **556**, 1-9.

Shiflett, A. M., Bishop, J. R., Pahwa, A. and Hajduk, S. L. (2005). Human high density lipoproteins are platforms for the assembly of multi-component innate immune complexes. *J.Biol.Chem.* **280**, 32578-32585.

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.

Sia, E. A., Jinks-Robertson, S. and Petes, T. D. (1997). Genetic control of microsatellite stability. *Mutat.Res.* **383**, 61-70.

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.

Sigurdsson, S., Van Komen, S., Bussen, W., Schild, D., Albala, J. S. and Sung, P. (2001). Mediator function of the human Rad51B-Rad51C complex in Rad51/RPA-catalyzed DNA strand exchange. *Genes Dev.* **15**, 3308-3318.

Sigurdsson, S., Van Komen, S., Petukhova, G. and Sung, P. (2002). Homologous DNA pairing by human recombination factors Rad51 and Rad54. *J.Biol.Chem.* **277**, 42790-42794.

Singleton, M. R., Wentzell, L. M., Liu, Y., West, S. C. and Wigley, D. B. (2002). Structure of the single-strand annealing domain of human RAD52 protein. *Proc.Natl.Acad.Sci.U.S.A* **99**, 13492-13497.

Slupphaug, G., Kavli, B. and Krokan, H. E. (2003). The interacting pathways for prevention and repair of oxidative DNA damage. *Mutat.Res.* **531**, 231-251.

Smith, G. C. and Jackson, S. P. (1999). The DNA-dependent protein kinase. *Genes Dev.* **13**, 916-934.

Smith, G. R., Boddy, M. N., Shanahan, P. and Russell, P. (2003). Fission yeast Mus81.Eme1 Holliday junction resolvase is required for meiotic crossing over but not for gene conversion. *Genetics* **165**, 2289-2293.

Sniegowski, P. D., Gerrish, P. J. and Lenski, R. E. (1997). Evolution of high mutation rates in experimental populations of E. coli. *Nature* **387**, 703-705.

Snowden, T., Acharya, S., Butz, C., Berardini, M. and Fishel, R. (2004). hMSH4-hMSH5 recognizes Holliday Junctions and forms a meiosis-specific sliding clamp that embraces homologous chromosomes. *Mol.Cell* **15**, 437-451.

Spampinato, C. and Modrich, P. (2000). The MutL ATPase is required for mismatch repair. *J.Biol.Chem.* **275**, 9863-9869.

Spell, R. M. and Jinks-Robertson, S. (2003). Role of mismatch repair in the fidelity of R. *Genetics* **165**, 1733-1744.

Spell, R. M. and Jinks-Robertson, S. (2004). Determination of mitotic recombination rates by fluctuation analysis in Saccharomyces cerevisiae. *Methods Mol.Biol.* **262**, 3-12.

Springer, B., Sander, P., Sedlacek, L., Hardt, W. D., Mizrahi, V., Schar, P. and Bottger, E. C. (2004). Lack of mismatch correction facilitates genome evolution in mycobacteria. *Mol.Microbiol.* **53**, 1601-1609.

Stahl, F. W., Foss, H. M., Young, L. S., Borts, R. H., Abdullah, M. F. and Copenhaver, G. P. (2004). Does crossover interference count in Saccharomyces cerevisiae? *Genetics* **168**, 35-48.

Stark, J. M., Hu, P., Pierce, A. J., Moynahan, M. E., Ellis, N. and Jasin, M. (2002). ATP hydrolysis by mammalian RAD51 has a key role during homology-directed DNA repair. *J.Biol.Chem.* **277**, 20185-20194.

Sternberg, J., Turner, C. M., Wells, J. M., Ranford-Cartwright, L. C., Le Page, R. W. and Tait, A. (1989). Gene exchange in African trypanosomes: frequency and allelic segregation. *Mol.Biochem.Parasitol.* **34**, 269-279.

Sternberg, J. M. (2004). Human African trypanosomiasis: clinical presentation and immune response. *Parasite Immunol.* **26**, 469-476.

Stojic, L., Brun, R. and Jiricny, J. (2004a). Mismatch repair and DNA damage signalling. *DNA Repair (Amst)* **3**, 1091-1101.

Stojic, L., Mojas, N., Cejka, P., Di Pietro, M., Ferrari, S., Marra, G. and Jiricny, J. (2004b). Mismatch repair-dependent G2 checkpoint induced by low doses of SN1 type methylating agents requires the ATR kinase. *Genes Dev.* **18**, 1331-1344.

Stracker, T. H., Theunissen, J. W., Morales, M. and Petrini, J. H. (2004). The Mre11 complex and the metabolism of chromosome breaks: the importance of communicating and holding things together. *DNA Repair (Amst)* **3**, 845-854.

Su, S. S. and Modrich, P. (1986). Escherichia coli mutS-encoded protein binds to mismatched DNA base pairs. *Proc.Natl.Acad.Sci.U.S.A* **83**, 5057-5061.

Sugawara, N., Ira, G. and Haber, J. E. (2000). DNA length dependence of the single-strand annealing pathway and the role of Saccharomyces cerevisiae RAD59 in double-strand break repair. *Mol.Cell Biol.* **20**, 5300-5309.

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.

Sun, H., Treco, D. and Szostak, J. W. (1991). Extensive 3'-overhanging, single-stranded DNA associated with the meiosis-specific double-strand breaks at the ARG4 recombination initiation site. *Cell* **64**, 1155-1161.

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

Sung, P. (1997b). Yeast Rad55 and Rad57 proteins form a heterodimer that functions with replication protein A to promote DNA strand exchange by Rad51 recombinase. *Genes Dev.* **11**, 1111-1121.

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.

Svetlanov, A. and Cohen, P. E. (2004). Mismatch repair proteins, meiosis, and mice: understanding the complexities of mammalian meiosis. *Exp.Cell Res.* **296**, 71-79.

Symington, L. S. (2002). Role of RAD52 epistasis group genes in homologous recombination and double-strand break repair. *Microbiol.Mol.Biol.Rev.* **66**, 630-70, table.

Taddei, F., Vulic, M., Radman, M. and Matic, I. (1997). Genetic variability and adaptation to stress. *EXS* **83**, 271-290.

Tait, A. (1980). Evidence for diploidy and mating in trypanosomes. *Nature* **287**, 536-538.

Tait, A., Buchanan, N., Hide, G., Turner, M. (1993). Genetic recombination and karyotype inheritance in *T. brucei* spp. In: Mozaria S.P, editor. Genome Analysis of Protozoan Parasites. Nairobi: ILRAD, p 93-107.

Tait, A., Buchanan, N., Hide, G. and Turner, C. M. (1996). Self-fertilisation in Trypanosoma brucei. *Mol.Biochem.Parasitol.* **76**, 31-42.

Tait, A., Turner, C. M., Le Page, R. W. and Wells, J. M. (1989). Genetic evidence that metacyclic forms of Trypanosoma brucei are diploid. *Mol.Biochem.Parasitol.* **37**, 247-255.

Takata, M., Sasaki, M. S., Sonoda, E., Morrison, C., Hashimoto, M., Utsumi, H., Yamaguchi-Iwai, Y., Shinohara, A. and Takeda, S. (1998). Homologous recombination and non-homologous end-joining pathways of DNA double-strand break repair have overlapping roles in the maintenance of chromosomal integrity in vertebrate cells. *EMBO J.* **17**, 5497-5508.

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.

Tanaka, M. M., Bergstrom, C. T. and Levin, B. R. (2003). The evolution of mutator genes in bacterial populations: the roles of environmental change and timing. *Genetics* **164**, 843-854.

Tarsounas, M., Davies, D. and West, S. C. (2003). BRCA2-dependent and independent formation of RAD51 nuclear foci. *Oncogene* **22**, 1115-1123.

Tarsounas, M., Munoz, P., Claas, A., Smiraldo, P. G., Pittman, D. L., Blasco, M. A. and West, S. C. (2004). Telomere maintenance requires the RAD51D recombination/repair protein. *Cell* **117**, 337-347.

Tavanti, A., Gow, N. A., Maiden, M. C., Odds, F. C. and Shaw, D. J. (2004). Genetic evidence for recombination in Candida albicans based on haplotype analysis. *Fungal.Genet.Biol.* **41**, 553-562.

te, Riele H., Maandag, E. R. and Berns, A. (1992). Highly efficient gene targeting in embryonic stem cells through homologous recombination with isogenic DNA constructs. *Proc.Natl.Acad.Sci.U.S.A* **89**, 5128-5132.

Ten Asbroek, A. L., Ouellette, M. and Borst, P. (1990). Targeted insertion of the neomycin phosphotransferase gene into the tubulin gene cluster of Trypanosoma brucei. *Nature* **348**, 174-175.

Tenaillon, O., Le Nagard, H., Godelle, B. and Taddei, F. (2000). Mutators and sex in bacteria: conflict between adaptive strategies. *Proc.Natl.Acad.Sci.U.S.A* **97**, 10465-10470.

Tenaillon, O., Taddei, F., Radman, M. and Matic, I. (2001). Second-order selection in bacterial evolution: selection acting on mutation and recombination rates in the course of adaptation. *Res.Microbiol.* **152**, 11-16.

Tenaillon, O., Toupance, B., Le Nagard, H., Taddei, F. and Godelle, B. (1999). Mutators, population size, adaptive landscape and the adaptation of asexual populations of bacteria. *Genetics* **152**, 485-493.

Tetley, L., Vickerman, K. and Moloo, S. K. (1981). Absence of a surface coat from metacyclic Trypanosoma vivax: possible implications for vaccination against vivax trypanosomiasis. *Trans.R.Soc.Trop.Med.Hyg.* **75**, 409-414.

Thomas, D., Scot, A. D., Barbey, R., Padula, M. and Boiteux, S. (1997). Inactivation of OGG1 increases the incidence of G . C-->T . A transversions in Saccharomyces cerevisiae: evidence for endogenous oxidative damage to DNA in eukaryotic cells. *Mol.Gen.Genet.* **254**, 171-178.

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. and Higgins, D. G. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25**, 4876-4882.

Thompson, L. H. and Schild, D. (2001). Homologous recombinational repair of DNA ensures mammalian chromosome stability. *Mutat.Res.* **477**, 131-153.

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.

Thorne, L. W. and Byers, B. (1993). Stage-specific effects of X-irradiation on yeast meiosis. *Genetics* **134**, 29-42.

Tibayrenc, M. (1995). Population genetics of parasitic protozoa and other microorganisms. *Adv.Parasitol.* **36**, 47-115.

Tibayrenc, M. and Ayala, F. J. (2002). The clonal theory of parasitic protozoa: 12 years on. *Trends Parasitol.* **18**, 405-410.

Tibayrenc, M., Ward, P., Moya, A. and Ayala, F. J. (1986). Natural populations of Trypanosoma cruzi, the agent of Chagas disease, have a complex multiclonal structure. *Proc.Natl.Acad.Sci.U.S.A* **83**, 115-119.

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.

Tishkoff, D. X., Boerger, A. L., Bertrand, P., Filosi, N., Gaida, G. M., Kane, M. F. and Kolodner, R. D. (1997). Identification and characterization of Saccharomyces cerevisiae EXO1, a gene encoding an exonuclease that interacts with MSH2. *Proc.Natl.Acad.Sci.U.S.A* **94**, 7487-7492.

Tran, H. T., Degtyareva, N. P., Gordenin, D. A. and Resnick, M. A. (1999). Genetic factors affecting the impact of DNA polymerase delta proofreading activity on mutation avoidance in yeast. *Genetics* **152**, 47-59.

Trouiller, B., Schaefer, D. G., Charlot, F. and Nogue, F. (2006). MSH2 is essential for the preservation of genome integrity and prevents homeologous recombination in the moss Physcomitrella patens. *Nucleic Acids Res.* **34**, 232-242.

Truc, P. and Tibayrenc, M. (1993). Population genetics of Trypanosoma brucei in central Africa: taxonomic and epidemiological significance. *Parasitology* **106 (Pt 2)**, 137-149.

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.

Tschudi, C. and Ullu, E. (1988). Polygene transcripts are precursors to calmodulin mRNAs in trypanosomes. *EMBO J.* **7**, 455-463.

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.

Tsukamoto, Y., Mitsuoka, C., Terasawa, M., Ogawa, H. and Ogawa, T. (2005). Xrs2p regulates Mre11p translocation to the nucleus and plays a role in telomere elongation and meiotic recombination. *Mol.Biol.Cell* **16**, 597-608.

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.

Turner, C. M., Aslam, N. and Dye, C. (1995). Replication, differentiation, growth and the virulence of Trypanosoma brucei infections. *Parasitology* **111 (Pt 3)**, 289-300.

Turner, C. M., Aslam, N., Smith, E., Buchanan, N. and Tait, A. (1991). The effects of genetic exchange on variable antigen expression in Trypanosoma brucei. *Parasitology* **103 Pt 3**, 379-386.

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., 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., Sternberg, J., Buchanan, N., Smith, E., Hide, G. and Tait, A. (1990). Evidence that the mechanism of gene exchange in Trypanosoma brucei involves meiosis and syngamy. *Parasitology* **101 Pt 3**, 377-386.

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.

Tzung, K. W., Williams, R. M., Scherer, S., Federspiel, N., Jones, T., Hansen, N., Bivolarevic, V., Huizar, L., Komp, C., Surzycki, R., Tamse, R., Davis, R. W. and Agabian, N. (2001). Genomic evidence for a complete sexual cycle in Candida albicans. *Proc.Natl.Acad.Sci.U.S.A* **98**, 3249-3253.

Uetz, P., Giot, L., Cagney, G., Mansfield, T. A., Judson, R. S., Knight, J. R., Lockshon, D., Narayan, V., Srinivasan, M., Pochart, P., Qureshi-Emili, A., Li, Y., Godwin, B., Conover, D., Kalbfleisch, T., Vijayadamodar, G., Yang, M., Johnston, M., Fields, S. and Rothberg, J. M. (2000). A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. *Nature* **403**, 623-627.

Ullu, E., Matthews, K. R. and Tschudi, C. (1993). Temporal order of RNA-processing reactions in trypanosomes: rapid trans splicing precedes polyadenylation of newly synthesized tubulin transcripts. *Mol.Cell Biol.* **13**, 720-725.

Umar, A., Buermeyer, A. B., Simon, J. A., Thomas, D. C., Clark, A. B., Liskay, R. M. and Kunkel, T. A. (1996). Requirement for PCNA in DNA mismatch repair at a step preceding DNA resynthesis. *Cell* **87**, 65-73.

Umezu, K., Sugawara, N., Chen, C., Haber, J. E. and Kolodner, R. D. (1998). Genetic analysis of yeast RPA1 reveals its multiple functions in DNA metabolism. *Genetics* **148**, 989-1005.

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, Bosch M., Zonneveld, J. B., Lohman, P. H. and Pastink, A. (2001). Isolation and characterization of the RAD59 homologue of Kluyveromyces lactis. *Curr.Genet.* **39**, 305-310.

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 Dyck, E., Stasiak, A. Z., Stasiak, A. and West, S. C. (1999). Binding of double-strand breaks in DNA by human Rad52 protein. *Nature* **398**, 728-731.

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., Kieft, R., Cross, M. and Borst, P. (2000). Tandemly repeated DNA is a target for the partial replacement of thymine by beta-D-glucosyl-hydroxymethyluracil in Trypanosoma brucei. *Mol.Biochem.Parasitol.* **109**, 133-145.

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 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.

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.

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.

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.

Venkatesan, R. N., Bielas, J. H. and Loeb, L. A. (2006). Generation of mutator mutants during carcinogenesis. *DNA Repair (Amst)* **5**, 294-302.

Victoir, K. and Dujardin, J. C. (2002). How to succeed in parasitic life without sex? Asking Leishmania. *Trends Parasitol.* **18**, 81-85.

Villeneuve, A. M. and Hillers, K. J. (2001). Whence meiosis? *Cell* **106**, 647-650.

Viswanathan, M., Burdett, V., Baitinger, C., Modrich, P. and Lovett, S. T. (2001). Redundant exonuclease involvement in Escherichia coli methyl-directed mismatch repair. *J.Biol.Chem.* **276**, 31053-31058.

Vo, A. T., Zhu, F., Wu, X., Yuan, F., Gao, Y., Gu, L., Li, G. M., Lee, T. H. and Her, C. (2005). hMRE11 deficiency leads to microsatellite instability and defective DNA mismatch repair. *EMBO Rep.* **6**, 438-444.

Vulic, M., Dionisio, F., Taddei, F. and Radman, M. (1997). Molecular keys to speciation: DNA polymorphism and the control of genetic exchange in enterobacteria. *Proc.Natl.Acad.Sci.U.S.A* **94**, 9763-9767.

Vulic, M., Lenski, R. E. and Radman, M. (1999). Mutation, recombination, and incipient speciation of bacteria in the laboratory. *Proc.Natl.Acad.Sci.U.S.A* **96**, 7348-7351.

Wada, M. and Nakamura, Y. (1996). Unique telomeric expression site of major-surface-glycoprotein genes of Pneumocystis carinii. *DNA Res.* **3**, 55-64.

Waldman, A. S. and Liskay, R. M. (1988). Dependence of intrachromosomal recombination in mammalian cells on uninterrupted homology. *Mol.Cell Biol.* **8**, 5350-5357.

Walker, J. R., Corpina, R. A. and Goldberg, J. (2001). Structure of the Ku heterodimer bound to DNA and its implications for double-strand break repair. *Nature* **412**, 607-614.

Wallace, L. J., Candlish, D. and De Koning, H. P. (2002). Different substrate recognition motifs of human and trypanosome nucleobase transporters. Selective uptake of purine antimetabolites. *J.Biol.Chem.* **277**, 26149-26156.

Wan, L., de los, Santos T., Zhang, C., Shokat, K. and Hollingsworth, N. M. (2004). Mek1 kinase activity functions downstream of RED1 in the regulation of meiotic double strand break repair in budding yeast. *Mol.Biol.Cell* **15**, 11-23.

Wang, T. F., Kleckner, N. and Hunter, N. (1999). Functional specificity of MutL homologs in yeast: evidence for three Mlh1-based heterocomplexes with distinct roles during meiosis in recombination and mismatch correction. *Proc.Natl.Acad.Sci.U.S.A* **96**, 13914-13919.

Wang, T. F. and Kung, W. M. (2002). Supercomplex formation between Mlh1-Mlh3 and Sgs1-Top3 heterocomplexes in meiotic yeast cells. *Biochem.Biophys.Res.Commun.* **296**, 949-953.

Wang, X. and Haber, J. E. (2004). Role of Saccharomyces single-stranded DNA-binding protein RPA in the strand invasion step of double-strand break repair. *PLoS.Biol.* **2**, E21.

Wang, Y. and Qin, J. (2003). MSH2 and ATR form a signaling module and regulate two branches of the damage response to DNA methylation. *Proc.Natl.Acad.Sci.U.S.A* **100**, 15387-15392.

Welz-Voegele, C., Stone, J. E., Tran, P. T., Kearney, H. M., Liskay, R. M., Petes, T. D. and Jinks-Robertson, S. (2002). Alleles of the yeast Pms1 mismatch-repair gene that differentially affect recombination- and replication-related processes. *Genetics* **162**, 1131-1145.

Westmoreland, J., Porter, G., Radman, M. and Resnick, M. A. (1997). Highly mismatched molecules resembling recombination intermediates efficiently transform mismatch repair proficient Escherichia coli. *Genetics* **145**, 29-38.

Weterings, E. and van Gent, D. C. (2004). The mechanism of non-homologous end-joining: a synopsis of synapsis. *DNA Repair (Amst)* **3**, 1425-1435.

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

Wickstead, B., Ersfeld, K. and Gull, K. (2004). The small chromosomes of Trypanosoma brucei involved in antigenic variation are constructed around repetitive palindromes. *Genome Res.* **14**, 1014-1024.

Wickstead, Bill, Ersfeld, Klaus and Gull, Keith (2003). The frequency of gene targeting in Trypanosoma brucei is independent of target site copy number. *Nucleic Acids Research* **31**, 3993.

Wilson, J. H., Leung, W. Y., Bosco, G., Dieu, D. and Haber, J. E. (1994). The frequency of gene targeting in yeast depends on the number of target copies. *Proc.Natl.Acad.Sci.U.S.A* **91**, 177-181.

Wilson, T. E., Grawunder, U. and Lieber, M. R. (1997). Yeast DNA ligase IV mediates non-homologous DNA end joining. *Nature* **388**, 495-498.

Winand, N. J., Panzer, J. A. and Kolodner, R. D. (1998). Cloning and characterization of the human and Caenorhabditis elegans homologs of the Saccharomyces cerevisiae MSH5 gene. *Genomics* **53**, 69-80.

Wiseman, H. and Halliwell, B. (1996). Damage to DNA by reactive oxygen and nitrogen species: role in inflammatory disease and progression to cancer. *Biochem.J.* **313 (Pt 1)**, 17-29.

Wolner, B. and Peterson, C. L. (2005). ATP-dependent and ATP-independent roles for the Rad54 chromatin remodeling enzyme during recombinational repair of a DNA double strand break. *J.Biol.Chem.* **280**, 10855-10860.

Woltering, D., Baumgartner, B., Bagchi, S., Larkin, B., Loidl, J., de los, Santos T. and Hollingsworth, N. M. (2000). Meiotic segregation, synapsis, and recombination checkpoint functions require physical interaction between the chromosomal proteins Red1p and Hop1p. *Mol.Cell Biol.* **20**, 6646-6658.

Worth, L., Jr., Bader, T., Yang, J. and Clark, S. (1998). Role of MutS ATPase activity in MutS,L-dependent block of in vitro strand transfer. *J.Biol.Chem.* **273**, 23176-23182.

Worth, L., Jr., Clark, S., Radman, M. and Modrich, P. (1994). Mismatch repair proteins MutS and MutL inhibit RecA-catalyzed strand transfer between diverged DNAs. *Proc.Natl.Acad.Sci.U.S.A* **91**, 3238-3241.

Wu, L. and Hickson, I. D. (2001). RecQ helicases and topoisomerases: components of a conserved complex for the regulation of genetic recombination. *Cell Mol.Life Sci.* **58**, 894-901.

Wu, S. Y., Culligan, K., Lamers, M. and Hays, J. (2003). Dissimilar mispair-recognition spectra of Arabidopsis DNA-mismatch-repair proteins MSH2*MSH6 (MutSalpha) and MSH2*MSH7 (MutSgamma). *Nucleic Acids Res.* **31**, 6027-6034.

Wu, T. H. and Marinus, M. G. (1994). Dominant negative mutator mutations in the mutS gene of Escherichia coli. *J.Bacteriol.* **176**, 5393-5400.

Wu, Y., Sugiyama, T. and Kowalczykowski, S. C. (2006). DNA annealing mediated by rad52 and rad59 proteins. *J.Biol.Chem.*

Wyrzykowski, J. and Volkert, M. R. (2003). The Escherichia coli methyl-directed mismatch repair system repairs base pairs containing oxidative lesions. *J.Bacteriol.* **185**, 1701-1704.

Xong, H. V., Vanhamme, L., Chamekh, M., Chimfwembe, C. E., Van den, Abbeele J., Pays, A., Van Meirvenne, N., Hamers, R., De Baetselier, P. and Pays, E. (1998). A VSG expression site-associated gene confers resistance to human serum in Trypanosoma rhodesiense. *Cell* **95**, 839-846.

Xu, Z., Fulop, Z., Zhong, Y., Evinger, A. J., III, Zan, H. and Casali, P. (2005). DNA lesions and repair in immunoglobulin class switch recombination and somatic hypermutation. *Ann.N.Y.Acad.Sci.* **1050**, 146-162.

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

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

Yang, H., Li, Q., Fan, J., Holloman, W. K. and Pavletich, N. P. (2005). The BRCA2 homologue Brh2 nucleates RAD51 filament formation at a dsDNA-ssDNA junction. *Nature* **433**, 653-657.

Yang, H., Wolff, E., Kim, M., Diep, A. and Miller, J. H. (2004). Identification of mutator genes and mutational pathways in Escherichia coli using a multicopy cloning approach. *Mol.Microbiol.* **53**, 283-295.

Yokoyama, H., Kurumizaka, H., Ikawa, S., Yokoyama, S. and Shibata, T. (2003). Holliday junction binding activity of the human Rad51B protein. *J.Biol.Chem.* **278**, 2767-2772.

Yoshioka, K., Yoshioka, Y. and Hsieh, P. (2006). ATR kinase activation mediated by MutSalpha and MutLalpha in response to cytotoxic O6-methylguanine adducts. *Mol.Cell* **22**, 501-510.

Yu, D. S., Sonoda, E., Takeda, S., Huang, C. L., Pellegrini, L., Blundell, T. L. and Venkitaraman, A. R. (2003). Dynamic control of Rad51 recombinase by self-association and interaction with BRCA2. *Mol.Cell* **12**, 1029-1041.

Yu, H. G. and Koshland, D. E. (2003). Meiotic condensin is required for proper chromosome compaction, SC assembly, and resolution of recombination-dependent chromosome linkages. *J.Cell Biol.* **163**, 937-947.

Zahrt, T. C., Buchmeier, N. and Maloy, S. (1999). Effect of mutS and recD mutations on Salmonella virulence. *Infect.Immun.* **67**, 6168-6172.

Zahrt, T. C. and Maloy, S. (1997). Barriers to recombination between closely related bacteria: MutS and RecBCD inhibit recombination between Salmonella typhimurium and Salmonella typhi. *Proc.Natl.Acad.Sci.U.S.A* **94**, 9786-9791.

Zahrt, T. C., Mora, G. C. and Maloy, S. (1994). Inactivation of mismatch repair overcomes the barrier to transduction between Salmonella typhimurium and Salmonella typhi. *J.Bacteriol.* **176**, 1527-1529.

Zalevsky, J., MacQueen, A. J., Duffy, J. B., Kemphues, K. J. and Villeneuve, A. M. (1999). Crossing over during Caenorhabditis elegans meiosis requires a conserved MutS-based pathway that is partially dispensable in budding yeast. *Genetics* **153**, 1271-1283.

Zampetti-Bosseler, F., Schweizer, J., Pays, E., Jenni, L. and Steinert, M. (1986). Evidence for haploidy in metacyclic forms of Trypanosoma brucei. *Proc.Natl.Acad.Sci.U.S.A* **83**, 6063-6064.

Zawadzki, P., Roberts, M. S. and Cohan, F. M. (1995). The log-linear relationship between sexual isolation and sequence divergence in Bacillus transformation is robust. *Genetics* **140**, 917-932.

Zhang, H., Tombline, G. and Weber, B. L. (1998). BRCA1, BRCA2, and DNA damage response: collision or collusion? *Cell* **92**, 433-436.

Zhang, Y., Yuan, F., Presnell, S. R., Tian, K., Gao, Y., Tomkinson, A. E., Gu, L. and Li, G. M. (2005). Reconstitution of 5'-directed human mismatch repair in a purified system. *Cell* **122**, 693-705.

Zhao, Z., Fu, Y. X., Hewett-Emmett, D., Boerwinkle, E.(2003). Investigating single nucleotide polymorphism (SNP) density in the human genome and its implications for molecular evolution. *Gene* **312**, 207-213.

Zierhut, C., Berlinger, M., Rupp, C., Shinohara, A. and Klein, F. (2004). Mnd1 is required for meiotic interhomolog repair. *Curr.Biol.* **14**, 752-762.

Zomerdijk, J. C., Kieft, R., Duyndam, M., Shiels, P. G. and Borst, P. (1991). Antigenic variation in Trypanosoma brucei: a telomeric expression site for variant-specific surface glycoprotein genes with novel features. *Nucleic Acids Res.* **19**, 1359-1368.

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.