



<https://theses.gla.ac.uk/>

Theses Digitisation:

<https://www.gla.ac.uk/myglasgow/research/enlighten/theses/digitisation/>

This is a digitised version of the original print thesis.

Copyright and moral rights for this work are retained by the author

A copy can be downloaded for personal non-commercial research or study,  
without prior permission or charge

This work cannot be reproduced or quoted extensively from without first  
obtaining permission in writing from the author

The content must not be changed in any way or sold commercially in any  
format or medium without the formal permission of the author

When referring to this work, full bibliographic details including the author,  
title, awarding institution and date of the thesis must be given

Enlighten: Theses

<https://theses.gla.ac.uk/>  
[research-enlighten@glasgow.ac.uk](mailto:research-enlighten@glasgow.ac.uk)

**Mismatch Repair in  
DNA Recombination and Antigenic  
Variation in *Trypanosoma brucei***

**Joanna Sharlene Bell**

**Wellcome Centre for Molecular Parasitology  
The Anderson College  
University of Glasgow  
56 Dumbarton Road  
Glasgow G11 6NU**

**Submitted for the degree of Doctor of Philosophy  
May 2002**

ProQuest Number: 10644168

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



ProQuest 10644168

Published by ProQuest LLC (2017). Copyright of the Dissertation is held by the Author.

All rights reserved.

This work is protected against unauthorized copying under Title 17, United States Code  
Microform Edition © ProQuest LLC.

ProQuest LLC.  
789 East Eisenhower Parkway  
P.O. Box 1346  
Ann Arbor, MI 48106 – 1346

## ACKNOWLEDGEMENTS

My greatest thanks obviously go to Richard McCulloch, my supervisor, for his guidance, encouragement, support and patience during the course of this project. Warm thanks also to Dave Barry, Andy Tait and Chris Oura for their advice, help and reassurance. I'm especially grateful too to all the WUMPies whose assistance and friendship I've enjoyed over the past 3½ years; including Affy, Alex, Ali B, Ally L, Caroline, Chris P, Colin, Des, Fiona, Kat, Katie, Liam (Legend), Mags, Mary, Michael (Ginge), Michael (Fothers), Miriam, Nick (Robbo), Pat, Pauline, Pete, Rui, Sandra, Simon, Tansy and Willie. Special thanks particularly to Annette, Alison and Marianne for allowing me to pester them on an almost daily basis regarding microsatellites, and Tim for helping me plough through the workload which went towards Chapter 3. It was a pleasure working with everyone.

Many thanks also to Bob Brown and Keith Gull for useful suggestions, and Mike Cross for his gift of antibiotic resistance cassettes. I'm also in debt to a large number of genome sequencing initiatives, especially at Sanger and TIGR.

Big hugs a plenty to Mum, Dad, Sam and Steve for their constant support and generous donations to the Jo Bell PhD fund, but especially for listening to my endless grumblings without complaint. Sorry for phoning so late all the time. Thanks also to Cathy and Cath for making Glasgow feel like home – the Christmas decorations were overdoing it though. Drinks all round for my friends who kept 'forgetting' it was my turn to get the beers in, and for not mentioning the 'T' word in the pub. A glass of wine also to the ladies who lunch for giving me something to look forward to every Friday.

Thanks everyone for keeping my chin up!

# TABLE OF CONTENTS

<b>TITLE PAGE</b> .....	1
<b>DECLARATION</b> .....	2
<b>ACKNOWLEDGEMENTS</b> .....	4
<b>TABLE OF CONTENTS</b> .....	5
<b>LIST OF FIGURES</b> .....	10
<b>LIST OF TABLES</b> .....	15
<b>LIST OF APPENDICES</b> .....	17
<b>ABBREVIATIONS</b> .....	18
<b>ABSTRACT</b> .....	22
<b>Chapter 1: Introduction</b> .....	23
1.1 General introduction.....	24
1.2 Antigenic variation .....	28
1.3 Antigenic variation in <i>T. brucei</i> .....	31
1.4 DNA recombination and antigenic variation .....	44
1.5 Mismatch repair in eukaryotes.....	51
1.6 Mismatch repair in DNA recombination .....	65
1.7 Aims of this thesis.....	71
<b>Chapter 2: Materials and Methods</b> .....	72
2.1 Trypanosome strains, cell culture and transformation .....	73
2.2 Standard molecular biology techniques .....	74
2.3 Genomic DNA isolation and Southern analysis .....	78
2.4 RNA isolation and reverse transcription PCR .....	79
2.5 Manufacture of radiolabelled probes and DNA hybridisation.....	80
2.6 Screening of the ILTat 1.2 genomic lambda library and mapping of clones.....	81
2.7 Sequencing .....	83
2.8 Data mining the National Center for Biotechnology Information (NCBI) database, the <i>T. brucei</i> and other sequencing databases, multiple alignment of polypeptide sequences and gap comparisons.....	84
2.9 Isolation of the complete coding sequences of three MutS homologues from <i>T. brucei</i> .....	86
2.10 Isolation of the complete coding sequences of two MutL homologues from <i>T. brucei</i> .....	88
2.11 Generation and analysis of <i>T. brucei</i> mismatch repair knock-out cell lines .....	90

2.12	Re-expressing a copy of the wild-type <i>MSH2</i> allele in the homozygous knock-outs .....	92
2.13	Measuring the growth rates of trypanosome mutant cell lines .....	92
2.14	Measuring cell survival following treatment with <i>N</i> -methyl- <i>N</i> '-nitro- <i>N</i> -nitrosoguanidine (MNNG) .....	93
2.15	Generation and analysis of trypanosomes that have undergone antigenic variation .....	93
2.16	Assay for microsatellite instability in the <i>T. brucei</i> mismatch repair mutants .....	94
2.17	Assay for the frequency of integration of homologous and homeologous cassettes into trypanosome mismatch repair mutants .....	96
<b>Chapter 3: Identification of three putative MutS homologues and two putative MutL homologues from <i>T. brucei</i> .....</b>		<b>105</b>
3.1	Cloning and characterisation of a putative MutS Homologue 2 ( <i>MSH2</i> ) orthologue in <i>T. brucei</i> .....	106
3.1.1	Introduction .....	106
3.1.2	Identification of a putative <i>T. brucei MSH2</i> orthologue by degenerate PCR .....	106
3.1.3	Cloning of the complete ORF of the putative <i>T. brucei MSH2</i> gene from a genomic lambda library .....	107
3.1.4	Southern analysis of the putative <i>T. brucei MSH2</i> locus .....	107
3.1.5	Analysis of the expression of the putative <i>T. brucei MSH2</i> gene .....	108
3.1.6	Pairwise comparison of the putative <i>T. brucei MSH2</i> polypeptide with other MutS homologues .....	108
3.1.7	Multiple alignment of the putative <i>T. brucei MSH2</i> polypeptide with eukaryotic <i>MSH2</i> orthologues .....	108
3.1.8	The genomic environment of the putative <i>T. brucei MSH2</i> ORF in TREU 927/4 .....	109
3.2	Identification of two further putative MutS homologues, <i>MSH3</i> and <i>MSH8</i> , in <i>T. brucei</i> .....	123
3.2.1	Introduction .....	123
3.2.2	Identification of a second putative MutS homologue by searching the <i>T. brucei</i> genome sequencing databases .....	123
3.2.3	Cloning and sequencing of the complete ORF of the putative <i>T. brucei MSH8</i> gene .....	124
3.2.4	Southern analysis of the putative <i>T. brucei MSH8</i> locus .....	125
3.2.5	Analysis of the expression of the putative <i>T. brucei MSH8</i> gene .....	125

3.2.6	Identification of a third putative MutS homologue by searching of the <i>T. brucei</i> genome sequencing databases .....	125
3.2.7	Pairwise comparison of the putative <i>T. brucei</i> MSH3 and MSH8 polypeptides with other MutS homologues .....	126
3.2.8	Investigation of the 5' splicing of the putative <i>T. brucei</i> MSH8 mRNA... ..	127
3.2.9	Multiple alignment of the putative <i>T. brucei</i> MSH3 and MSH8 polypeptides with eukaryotic MSH3, MSH6 and MSH7 homologues.....	127
3.2.10	The genomic environments of the putative <i>T. brucei</i> MSH8 and MSH3 genes in TREU 927/4.....	129
3.3	Cloning and characterisation of a putative MutL Homologue 1 (MLH1) orthologue in <i>T. brucei</i> .....	152
3.3.1	Introduction.....	152
3.3.2	Identification of the putative <i>T. brucei</i> MLH1 gene by searching the <i>T. brucei</i> genomic sequencing databases .....	152
3.3.3	Cloning a complete ORF from a genomic lambda library putatively encoding the <i>T. brucei</i> MLH1 gene.....	153
3.3.4	Southern analysis of the putative <i>T. brucei</i> MLH1 locus.....	154
3.3.5	Analysis of the expression of the putative <i>T. brucei</i> MLH1 gene.....	154
3.3.6	Pairwise comparison of the putative <i>T. brucei</i> MLH1 polypeptide with other MutL homologues.....	154
3.3.7	Multiple alignment of the putative <i>T. brucei</i> MLH1 polypeptide with eukaryotic MLH1 orthologues.....	155
3.4	Cloning and characterisation of a <i>T. brucei</i> gene encoding a putative orthologue of Post-Meiotic Segregation I (PMS1) from <i>S. cerevisiae</i> .....	167
3.4.1	Introduction.....	167
3.4.2	Identification of a second putative MutL homologue by searching of the <i>T. brucei</i> genome sequencing databases .....	167
3.4.3	Cloning of the complete ORF of the putative <i>T. brucei</i> PMS1 gene from a genomic lambda library .....	168
3.4.4	Southern analysis of the putative <i>T. brucei</i> PMS1 locus .....	168
3.4.5	Analysis of the expression of the putative <i>T. brucei</i> PMS1 gene .....	168
3.4.6	Pairwise comparison of the putative <i>T. brucei</i> PMS1 polypeptide with other MutL homologues.....	169
3.4.7	Multiple alignment of the putative <i>T. brucei</i> PMS1 polypeptide with orthologues of PMS1 from <i>S. cerevisiae</i> .....	169

3.4.8	The genomic environment of the putative <i>T. brucei</i> <i>PMS1</i> gene in TREU 927/4.....	170
3.5	Discussion.....	182
<b>Chapter 4: Generation and analysis of homozygous <i>MSH2</i> and <i>MLH1</i> knockouts in <i>T. brucei</i>.....</b>		<b>188</b>
4.1	Introduction.....	189
4.2	Generation and analysis of homozygous <i>MSH2</i> knockout trypanosomes.....	191
4.2.1	Introduction.....	191
4.2.2	Design and generation of <i>MSH2</i> knockout constructs.....	192
4.2.3	Design and generation of <i>MSH2</i> re-expression construct.....	192
4.2.4	Generation of <i>MSH2</i> mutant cell lines in <i>T. brucei</i> .....	193
4.2.5	Growth of <i>MSH2</i> knockout trypanosomes <i>in vitro</i> .....	195
4.2.6	Survival of <i>MSH2</i> mutant trypanosomes in the presence of the alkylating agent, <i>N</i> -methyl- <i>N'</i> -nitro- <i>N</i> -nitrosoguanidine (MNNG).....	195
4.2.7	Microsatellite instability in <i>MSH2</i> knockout cell lines.....	197
4.2.8	Antigenic variation in <i>MSH2</i> knockout trypanosomes.....	203
4.3	Generation and analysis of homozygous <i>MLH1</i> knockout trypanosomes.....	247
4.3.1	Introduction.....	247
4.3.2	Design and generation of <i>MLH1</i> knockout constructs.....	248
4.3.3	Generation of <i>MLH1</i> mutant cell lines in <i>T. brucei</i> .....	248
4.3.4	Growth of <i>MLH1</i> knockout trypanosomes <i>in vitro</i> .....	250
4.3.5	Survival of <i>MLH1</i> mutant trypanosomes in the presence of the alkylating agent, MNNG.....	250
4.3.6	Microsatellite instability in <i>MLH1</i> knockout cell lines.....	251
4.2.7	Antigenic variation in <i>MLH1</i> knockout trypanosomes.....	254
4.4	Discussion.....	282
<b>Chapter 5: An assay to determine the influence of mismatch repair on trypanosome homeologous recombination.....</b>		<b>291</b>
5.1	Introduction.....	292
5.2	Design of an assay to determine the frequency of integration of homologous and homeologous constructs into MMR-deficient trypanosomes.....	293
5.3	Generation of the transgenic HTUB strain of <i>T. brucei</i> .....	295
5.4	Generation of <i>MSH2</i> mutants in the HTUB trypanosome strain.....	296
5.5	Generation of homologous and homeologous <i>HYG</i> targeting constructs.....	298

5.6	Assay to determine the effect of <i>MSH2</i> mutation on the frequency of integration of homologous and homeologous <i>HYG</i> targeting constructs into the HTUB cell line.....	300
5.7	Discussion.....	301
<b>Chapter 6: Final Discussion.....</b>		<b>322</b>
6.1	Final discussion.....	323
<b>APPENDICES.....</b>		<b>334</b>
<b>LIST OF REFERENCES.....</b>		<b>371</b>

## LIST OF FIGURES

Figure 1.1	The life cycle of <i>Trypanosoma brucei</i> .....	27
Figure 1.2	<i>VSG</i> expression sites in trypanosomes.....	38
Figure 1.3	Expression linked copy (ELC) formation.....	39
Figure 1.4	Telomere conversion.....	40
Figure 1.5	Telomere reciprocal recombination.....	41
Figure 1.6	Mosaic gene formation.....	42
Figure 1.7	<i>In situ</i> transcriptional switching.....	43
Figure 1.8	Pathways of double-strand break (DSB) repair.....	50
Figure 1.9	The proteins involved in the post-replicative mismatch repair (MMR) pathway.....	61
Figure 1.10	The mechanism of <i>E. coli</i> methyl-directed MMR.....	62
Figure 1.11	Interactions between MutS- and MutL-related proteins during mismatch repair in <i>Saccharomyces cerevisiae</i> .....	63
Figure 1.12	A scheme for eukaryotic mismatch repair.....	64
Figure 1.13	Activities of MutS-related proteins during homologous recombination in <i>Saccharomyces cerevisiae</i> .....	69
Figure 1.14	A scheme for DNA recombination and the roles of MutS and MutL homologues in <i>Saccharomyces cerevisiae</i> .....	70
Figure 3.1	Design of degenerate PCR primers to amplify an <i>MSH2</i> orthologue from <i>T. brucei</i> .....	110
Figure 3.2	Degenerate PCR of <i>MSH2</i> .....	111
Figure 3.3	Restriction mapping of Plaque 1 - single digestions.....	112
Figure 3.4	Restriction mapping of Plaque 1 - double digestions.....	113
Figure 3.5	Restriction map of the putative <i>T. brucei MSH2</i> lambda clone Plaque 1..	114
Figure 3.6	Genomic Southern blot probed for the putative <i>T. brucei MSH2</i> gene.....	115
Figure 3.7	Analysis of the expression of the putative <i>T. brucei MSH2</i> gene.....	116
Figure 3.8	A graphical representation of the similarity between the putative <i>T. brucei MSH2</i> polypeptide and other MutS-related proteins.....	118
Figure 3.9	Global multiple alignment of the putative <i>T. brucei MSH2</i> polypeptide with a range of <i>MSH2</i> orthologues.....	119
Figure 3.10	Analysis of the genomic environment of the putative <i>T. brucei MSH2</i> gene.....	122
Figure 3.11	Strategy for PCR amplification of a fragment of the putative <i>T. brucei</i> <i>MSH2</i> ORF.....	131

Figure 3.12	PCR of a 2.0 kb region of the putative <i>T. brucei</i> <i>MSH8</i> gene .....	132
Figure 3.13	Strategy for PCR amplification of the putative <i>T. brucei</i> <i>MSH8</i> ORF .....	133
Figure 3.14	PCR of the upstream and downstream regions of the putative <i>T. brucei</i> <i>MSH8</i> gene.....	135
Figure 3.15	PCR of <i>T. brucei</i> <i>MSH8</i> probe sequence .....	136
Figure 3.16	Genomic Southern blot probed for the putative <i>T. brucei</i> <i>MSH8</i> gene .....	137
Figure 3.17	Analysis of the expression of the putative <i>T. brucei</i> <i>MSH8</i> gene.....	138
Figure 3.18	A graphical representation of the similarity between the putative <i>T. brucei</i> <i>MSH8</i> polypeptide and other MutS-related proteins.....	140
Figure 3.19	A graphical representation of the similarity between the putative <i>T. brucei</i> <i>MSH3</i> polypeptide and other MutS-related proteins.....	141
Figure 3.20	Analysis of the 5' splicing of the putative <i>T. brucei</i> <i>MSH8</i> gene.....	142
Figure 3.21	Global multiple alignment of the putative <i>T. brucei</i> <i>MSH3</i> and <i>MSH8</i> polypeptides with a range of <i>MSH3</i> , <i>MSH6</i> and <i>MSH7</i> orthologues.....	143
Figure 3.22	Alignment of the conserved PCNA binding motifs present in the <i>MSH3/MSH6</i> sub-family of MutS homologues.....	149
Figure 3.23	Analysis of the genomic environment of the putative <i>T. brucei</i> <i>MSH8</i> gene.....	150
Figure 3.24	Analysis of the genomic environment of the putative <i>T. brucei</i> <i>MSH3</i> gene.....	151
Figure 3.25	Strategy for PCR amplification of the putative <i>T. brucei</i> <i>MLH1</i> ORF.....	156
Figure 3.26	PCR of a 2.9 kb region of the putative <i>T. brucei</i> <i>MLH1</i> gene.....	157
Figure 3.27	PCR of the putative <i>T. brucei</i> <i>MLH1</i> probe sequence .....	158
Figure 3.28	Genomic Southern blot probed for the putative <i>T. brucei</i> <i>MLH1</i> gene .....	159
Figure 3.29	Genomic Southern blot probed for the putative <i>T. brucei</i> <i>MLH1</i> gene .....	160
Figure 3.30	Analysis of the expression of the putative <i>T. brucei</i> <i>MLH1</i> gene.....	161
Figure 3.31	A graphical representation of the similarity between the putative <i>T. brucei</i> <i>MLH1</i> polypeptide and other MutL-related proteins.....	163
Figure 3.32	Global multiple alignment of the putative <i>T. brucei</i> <i>MLH1</i> polypeptide with a range of <i>MLH1</i> orthologues.....	164
Figure 3.33	Strategy for PCR amplification of the putative <i>T. brucei</i> <i>PMS1</i> ORF.....	171
Figure 3.34	PCR of a 2.7 kb region of the putative <i>T. brucei</i> <i>PMS1</i> gene.....	172
Figure 3.35	PCR of <i>T. brucei</i> <i>PMS1</i> probe sequence.....	173
Figure 3.36	Genomic Southern blot probed for the putative <i>T. brucei</i> <i>PMS1</i> gene.....	174
Figure 3.37	Analysis of the expression of the putative <i>T. brucei</i> <i>PMS1</i> gene .....	175

Figure 3.38	A graphical representation of the similarity between the putative <i>T. brucei</i> PMS1 polypeptide and other MutL-related proteins.....	177
Figure 3.39	Global multiple alignment of the putative <i>T. brucei</i> PMS1 polypeptide with a range of PMS1 orthologues.....	178
Figure 3.40	Analysis of the genomic environment of the putative <i>T. brucei</i> PMS1 gene.....	181
Figure 4.1	A strategy for the generation of <i>T. brucei</i> MSH2 knockout mutants.....	206
Figure 4.2	PCR amplification of MSH2 knockout construct flanks.....	207
Figure 4.3	MSH2 re-expression construct.....	208
Figure 4.4	Southern analysis of the putative 3174 MSH2 mutant cell lines .....	209
Figure 4.5	PCR analysis of the putative 3174 MSH2 mutant cell lines .....	210
Figure 4.6	Analysis of the expression of MSH2 in the putative 3174 MSH2 mutant cell lines.....	211
Figure 4.7	Growth of MSH2 knockout trypanosomes.....	212
Figure 4.8	Survival of MSH2 knockout trypanosomes in the presence of MNNG.....	214
Figure 4.9	Model of microsatellite mutation by replication slippage .....	215
Figure 4.10	Sequence of microsatellite locus JS-2 .....	216
Figure 4.11	PCR amplification of microsatellite JS-2 in 3174 MSH2 mutant cell lines.....	217
Figure 4.12	Analysis of microsatellite JS-2 in MSH2 knockout trypanosomes.....	219
Figure 4.13	Genescan analysis of fluorescently-labelled PCR products derived from four independent microsatellite loci from 3174 transgenic trypanosomes .....	220
Figure 4.14	Microsatellite loci on Chromosome I .....	221
Figure 4.15	Microsatellite loci on Chromosome II.....	222
Figure 4.16	Analysis of microsatellite ChrI-7 in MSH2 knockout trypanosomes.....	226
Figure 4.17	Spectrum of mutations observed at microsatellite ChrI-7 in MSH2 homozygous mutants .....	227
Figure 4.18	Analysis of microsatellite ChrI-15 in MSH2 knockout trypanosomes.....	231
Figure 4.19	Spectrum of mutations observed at microsatellite ChrI-15 in MSH2 homozygous mutants .....	232
Figure 4.20	Analysis of microsatellite PLC in MSH2 knockout trypanosomes.....	236
Figure 4.21	Spectrum of mutations observed at microsatellite PLC in MSH2 homozygous mutants .....	237
Figure 4.22	Use of the 3174 transgenic trypanosomes strain to characterise VSG switching events.....	242

Figure 4.23	Effect of <i>MSH2</i> mutation on the frequency of VSG switching in the 3174 trypanosome strain.....	244
Figure 4.24	VSG switching mechanisms used by <i>MSH2</i> knockout trypanosomes.....	246
Figure 4.25	A strategy for the generation of <i>T. brucei MLH1</i> knockout mutants.....	256
Figure 4.26	PCR amplification of <i>MLH1</i> knockout construct flanks.....	257
Figure 4.27	Southern analysis of 3174 <i>MLH1</i> knockout cell lines.....	258
Figure 4.28	PCR analysis of the 3174 <i>MLH1</i> knockout cell lines.....	259
Figure 4.29	Analysis of the expression of <i>MLH1</i> in the putative 3174 <i>MLH1</i> mutant cell lines.....	260
Figure 4.30	Growth of <i>MLH1</i> knockout trypanosomes.....	261
Figure 4.31	Survival of <i>MLH1</i> knockout trypanosomes in the presence of MNNG.....	263
Figure 4.32	PCR amplification of microsatellite JS-2 in 3174 <i>MLH1</i> mutant cell lines.....	264
Figure 4.33	Analysis of microsatellite JS-2 in <i>MLH1</i> knockout trypanosomes.....	266
Figure 4.34	Analysis of microsatellite ChrI-7 in <i>MLH1</i> knockout trypanosomes.....	270
Figure 4.35	Spectrum of mutations observed at microsatellite ChrI-7 in <i>MLH1</i> homozygous mutants.....	271
Figure 4.36	Analysis of microsatellite ChrI-15 in <i>MLH1</i> knockout trypanosomes.....	275
Figure 4.37	Spectrum of mutations observed at microsatellite ChrI-15 in <i>MLH1</i> homozygous mutants.....	276
Figure 4.38	Effect of <i>MLH1</i> mutation on the frequency of VSG switching in the 3174 trypanosome strain.....	279
Figure 4.39	VSG switching mechanisms used by <i>MLH1</i> knockout trypanosomes.....	281
Figure 5.1	Overview of experimental approach.....	306
Figure 5.2	Targeting a hygromycin phosphotransferase ORF into the tubulin array..	308
Figure 5.3	Design of constructs with homeologous integration flanks to target the <i>HYG</i> ORF present in the HTUB transgenic cell line.....	309
Figure 5.4	PCR amplification of homologous <i>HYG</i> targeting flanks and <i>HYG</i> probe sequence.....	310
Figure 5.5	Southern analysis of the IITUB transgenic cell line.....	311
Figure 5.6	Southern analysis of HTUB <i>MSH2</i> knockout cell lines.....	312
Figure 5.7	Southern analysis of HTUB <i>MSH2</i> knockout cell lines.....	313
Figure 5.8	PCR analysis of the HTUB <i>MSH2</i> knockout cell lines.....	314
Figure 5.9	Analysis of the expression of <i>MSH2</i> in the putative HTUB <i>MSH2</i> mutant cell lines.....	315
Figure 5.10	PCR amplification of homeologous <i>HYG</i> targeting flanks.....	316

Figure 5.11	Multiple alignment of the <i>pHYGWT</i> , <i>pHYG01</i> and <i>pHYG03</i> sequences...317
Figure 5.12	Estimation of the concentration of the linearised <i>HYGWT::BLE</i> , <i>HYG01::BLE</i> , and <i>HYG03::BLE</i> constructs.....319
Figure 5.13	The frequency of integration of constructs containing increasing numbers of mismatches into HTUB <i>MSH2</i> knockout cell lines..... 321
Figure 6.1	Scheme of MutS/MSII evolution.....332

## LIST OF TABLES

Table 2.1	Oligonucleotides used in this study .....	98
Table 2.2	Unicellular parasite genome sequencing initiatives.....	103
Table 3.1	Pairwise comparison of the putative <i>T. brucei</i> MSH2 polypeptide with a range of MutS homologues .....	117
Table 3.2	Pairwise comparison of the putative <i>T. brucei</i> MSH8 and MSH3 polypeptides with a range of MutS homologues.....	139
Table 3.3	Pairwise comparison of the putative <i>T. brucei</i> MLH1 polypeptide with a range of MutL homologues.....	162
Table 3.4	Pairwise comparison of the putative <i>T. brucei</i> PMS1 polypeptide with a range of MutL homologues.....	176
Table 4.1	Survival of <i>MSH2</i> knockout trypanosomes in the presence of MNNG.....	213
Table 4.2	Synopsis of results for microsatellite JS-2 in 3174 <i>MSH2</i> knockout trypanosomes .....	218
Table 4.3	Genescan analysis of microsatellite ChrI-7 in 3174 <i>MSH2</i> mutant cell lines.....	223
Table 4.4	Synopsis of results for microsatellite ChrI-7 in 3174 <i>MSH2</i> knockout trypanosomes .....	225
Table 4.5	Genescan analysis of microsatellite ChrI-15 in 3174 <i>MSH2</i> mutant cell lines .....	228
Table 4.6	Synopsis of results for microsatellite ChrI-15 in 3174 <i>MSH2</i> knockout trypanosomes .....	230
Table 4.7	Genescan analysis of microsatellite PLC in 3174 <i>MSH2</i> mutant cell lines .....	233
Table 4.8	Synopsis of results for microsatellite PLC in 3174 <i>MSH2</i> knockout trypanosomes .....	235
Table 4.9	Genescan analysis of microsatellite ChrII-6 in 3174 <i>MSH2</i> mutant cell lines.....	238
Table 4.10	Synopsis of results for microsatellite ChrII-6 in 3174 <i>MSH2</i> knockout trypanosomes .....	240
Table 4.11	Mutational spectrum of <i>MSH2</i> <sup>-/-</sup> mutants analysed by GeneScan in <i>T. brucei</i> .....	241
Table 4.12	Effect of <i>MSH2</i> mutation on the frequency of VSG switching in the 3174 trypanosome strain.....	243

Table 4.13	<i>VSG</i> switching mechanisms used by <i>MSH2</i> wild-type and <i>MSH2</i> mutant trypanosomes .....	245
Table 4.14	Survival of <i>MLH1</i> knockout trypanosomes in the presence of MNNG.....	262
Table 4.15	Synopsis of results for microsatellite JS-2 in 3174 <i>MLH1</i> knockout trypanosomes .....	265
Table 4.16	Genescan analysis of microsatellite ChrI-7 in 3174 <i>MLH1</i> mutant cell lines .....	267
Table 4.17	Synopsis of results for microsatellite ChrI-7 in 3174 <i>MLH1</i> knockout trypanosomes .....	269
Table 4.18	Genescan analysis of microsatellite ChrI-15 in 3174 <i>MLH1</i> mutant cell lines .....	272
Table 4.19	Synopsis of results for microsatellite ChrI-15 in 3174 <i>MLH1</i> knockout trypanosomes .....	274
Table 4.20	Mutational spectrum of <i>MLH1</i> <sup>-/-</sup> mutants analysed by GeneScan in <i>T. brucei</i> .....	277
Table 4.21	Effect of <i>MLH1</i> mutation on the frequency of VSG switching in the 3174 trypanosome strain.....	278
Table 4.22	<i>VSG</i> switching mechanisms used by <i>MLH1</i> wild-type and <i>MLH1</i> mutant trypanosomes .....	280
Table 5.1	Effect of <i>MSH2</i> mutation on the frequency of integration of constructs containing increasing numbers of mismatches into the HTUB trypanosome strain.....	320
Table 6.1	MutS and MutL homologues in unicellular eukaryotic parasites.....	333

## LIST OF APPENDICES

Appendix 1	Sequence map showing the <i>T. brucei</i> <i>MSH2</i> ORF .....	335
Appendix 2	Sequence map showing the putative <i>T. brucei</i> <i>MSH8</i> ORF.....	342
Appendix 3	Sequence map showing the putative <i>T. brucei</i> <i>MSH3</i> ORF.....	348
Appendix 4	Sequence map showing the <i>T. brucei</i> <i>MLH1</i> ORF .....	353
Appendix 5	Sequence map showing the putative <i>T. brucei</i> <i>PMS1</i> ORF .....	360
Appendix 6	Multiple alignment of the pHygro-Tub, p <i>HYGWT</i> , p <i>HYG01</i> , p <i>HYG02</i> , p <i>HYG03</i> , p <i>HYG05</i> , p <i>IYG07</i> , p <i>HYG09</i> and p <i>HYG11</i> .....	368

## ABBREVIATIONS

6-FAM	6-carboxyfluorescein
8-oxo-dGTP	2'-Deoxy-8-hydroxyguanosine-5'-triphosphate, tetrasodium salt
BC	basic copy
BIR	break-induced replication
<i>BES</i>	bloodstream expression site
<i>BLE</i>	bleomycin resistance protein gene
bp	base-pairs
BrdU	5-bromodeoxyuridine
<i>BSD</i>	blasticidin-S-deaminase gene
cDNA	complementary DNA
CIP	alkaline phosphatase, calf intestinal
Cisplatin	<i>cis</i> -platinum(II) diammine dichloride
DEPC	diethyl pyrocarbonate used at 0.1% to remove RNase
DMSO	dimethyl sulphoxide
dNTP	deoxynucleoside triphosphate
dPTP	6-(2-deoxy- $\beta$ -D-erythropentofuranosyl)-3,4-dihydro-8H-pyrimido-[4,5C][1,2]oxazine-7-one-5'-triphosphate, triethylammonium salt

EATRO	East African Trypanosomiasis Research Organisation
ELC	expression linked copy
ESAG	expression site associated gene
ESB	expression site body
EtBr	ethidium bromide
FISH	fluorescent <i>in situ</i> hybridisation
gDNA	genomic DNA
HIV1	human immunodeficiency virus 1
HR	homologous recombination
<i>HYG</i>	hygromycin phosphotransferase gene
IDL	insertion/deletion loop
ILTat	International Laboratory for research on animal diseases, Trypanozoon antigen type
IPTG	isopropyl- $\beta$ -D-thiogalactopyranoside
kb	kilobase-pairs
Mb	megabase-pairs
MGMT	O <sup>6</sup> -methylguanine DNA methyltransferase
MITat	Moltino Institute – <i>Trypanozoan</i> antigen type
<i>MLH1</i> <sup>+/-</sup>	<i>MLH1</i> heterozygous mutant

<i>MLH1</i> <sup>-/-</sup>	<i>MLH1</i> homozygous mutant
MMR	mismatch repair
MMS	methyl methanesulphonate
MNNG	<i>N</i> -methyl- <i>N</i> '-nitro- <i>N</i> -nitrosoguanidine
<i>MSH2</i> <sup>+/-</sup>	<i>MSH2</i> heterozygous mutant
<i>MSH2</i> <sup>-/-</sup>	<i>MSH2</i> homozygous mutant
<i>MSH2</i> <sup>+/+</sup>	<i>MSH2</i> re-expressor
MSI	microsatellite instability
NER	nucleotide excision repair
NHEJ	non-homologous end joining
ORF	open reading frame
PCNA	proliferating cell nuclear antigen
PCR	polymerase chain reaction
<i>PUR</i>	puromycin-N-acetyltransferase gene
RNAi	RNA interference
RT	reverse transcriptase
RT-PCR	reverse transcription polymerase chain reaction
SDS	sodium dodecyl sulphate

SDSA	synthesis-dependent strand annealing
<i>SRA</i>	serum resistance-associated gene
SSA	single-strand annealing
SSC	sodium chloride / sodium citrate solution (1x): 0.15 M NaCl, 0.015 M Na <sub>3</sub> C <sub>6</sub> H <sub>5</sub> O <sub>7</sub> ·2H <sub>2</sub> O
TAE	TRIS / acetate / EDTA buffer (1x): 0.04 M TRIS base, 0.04 M glacial acetic acid, 1 mM EDTA
TAMRA	N, N, N', N'-tetramethyl-6-carboxyrhodamine
TBE	Tris / borate / EDTA buffer (1x): 0.089 M Tris base, 0.089 M ortho-boric acid, 2 mM EDTA
TE	10 mM Tris.Cl, 1 mM EDTA
TLF	trypanosome lytic factor
TREU	Trypanosomiasis Research Edinburgh University
UTR	untranslated region
VAT	variable antigen type
VSG	variant surface glycoprotein
X-Gal	5-bromo-4-chloro-3-indolyl-β-D-galactoside
ZMG	Zimmerman postfusion medium supplemented with glucose 132 mM NaCl, 8 mM Na <sub>2</sub> HPO <sub>4</sub> , 1.5 mM KH <sub>2</sub> PO <sub>4</sub> , 0.5 mM magnesium acetate, 0.09 mM calcium acetate [pH 7.0], 1% glucose

## ABSTRACT

African trypanosomes evade the mammalian host immune system for prolonged periods by periodically switching the protective Variant Surface Glycoprotein (VSG) coat expressed on their cell surface. Recombination reactions, which introduce novel *VSG* gene copies into specialised expression sites, constitute the principal mechanism of this antigenic variation process. Since *VSG* genes occupy cassettes with common upstream and downstream sequences it is hypothesised that *VSG* switching reactions occur primarily by homologous recombination. It has previously been shown that at least some of the *VSG* switching reactions are either catalysed or regulated by RAD51, a fundamental component of the homologous recombination machinery in eukaryotes.

In order to investigate further the factors regulating antigenic variation we have characterised several components of the highly conserved post-replicative mismatch repair (MMR) system in *T. brucei*. The nuclear MMR system recognises mispaired bases in heteroduplex recombination intermediates preventing recombination between divergent DNA molecules, and thereby confining such reactions to highly homologous sequences. In yeast and mammals this system is composed of three homologues of the bacterial MutS enzyme and three bacterial MutL-related enzymes. We have identified five single-copy genes putatively encoding three MutS homologues and two MutL homologues, suggesting that most of the components of eukaryotic MMR are conserved in trypanosomes.

Knockout mutants have been generated in two of the genes, *MSH2* and *MLH1*, to investigate the activities of MMR in this organism. Both *MSH2*-deficient and *MLH1*-deficient trypanosomes exhibit microsatellite instability, a mutator phenotype, and tolerance to the methylating agent MNNG, demonstrating that these genes encode components of an active MMR system in *T. brucei*. Moreover, deletion of *MSH2* increases the frequency of recombination between 3% divergent sequences, revealing that MMR is also involved in the regulation of recombination in this organism. Despite this, no effect on the frequency or mechanisms used in antigenic variation was observed in either *MSH2* or *MLH1* mutants, suggesting that MMR has little or no influence on this process, and that perhaps the homologous recombination reactions underlying antigenic variation proceed by a mechanism distinct from general recombination.

# **CHAPTER 1**

## **Introduction**

## 1.1 General introduction

The *Kinetoplastida* are an order of unicellular flagellated protozoa that diverged early in eukaryotic evolution (Sogin *et al.*, 1986; Sogin, 1989). They possess many unique genetic and metabolic features, including an organelle containing kinetoplast (mitochondrial) DNA, after which they are named. However, they also retain many of the characteristics conserved throughout the eukaryotic lineage. Within this order are the *Trypanosomatidae*, a family that boasts many vertebrate parasites (Maslov *et al.*, 2001). *Trypanosoma brucei* belongs to a group named the African Trypanosomes, which are mostly found in sub-Saharan and equatorial Africa. They are extracellular parasites that proliferate in the lymphatic and vascular systems of their mammalian host, causing interstitial inflammation and necrosis within the capillaries of major organs (Vickerman, 1985). The resulting parasitaemia typically produces sub-clinical infections within the reservoir of African game. However, in imported breeds of domestic cattle severe malnutrition and possibly death, in the form of a disease called nagana, occurs. In humans, the equivalent disease is called sleeping sickness, because the parasites also cause disorders of the central nervous system. Several morphologically indistinguishable sub-species of differing infectivity exist. *T. b. brucei* is believed to be non-infective to man because of a non-immune high-density lipoprotein called trypanosome lytic factor (TLF) found in human serum which lyses the parasites (Hajduk *et al.*, 1992; Smith *et al.*, 1995; Smith and Hajduk, 1995). However, *T. b. rhodesiense* and *T. b. gambiense* are extremely virulent in humans, causing the acute and chronic forms of sleeping sickness respectively, and are generally fatal if left untreated. The factor which allows *T. b. rhodesiense* to avoid killing by TLF is the product of the serum resistance-associated (*SRA*) gene, although its mode of action has yet to be determined (Pays *et al.*, 2001). The mechanism conferring TLF resistance to *T. b. gambiense* is unknown.

The developmental cycle of *T. brucei*, shown in Figure 1.1, requires both a mammalian host and an insect vector of the genus *Glossina*, or tsetse fly (Vickerman *et al.*, 1988). During this cycle, the parasite undergoes a number of highly co-ordinated and functionally organised differentiation events, and alternates between proliferative and non-dividing stages. The replicative stages proliferate by repeated mitotic division and are responsible for establishing and maintaining infection in their environment, whereas the non-dividing stages are pre-adapted to transmit the infection to the next host or vector (Shapiro *et al.*, 1984; Matthews and Gull, 1997).

Non-dividing metacyclic trypomastigotes, which develop in the salivary gland of the tsetse fly, are introduced into the mammal when an infective tsetse fly feeds (Barry and McCulloch, 2001). There they differentiate into the rapidly proliferating long slender bloodstream form, which is adapted for growth within the mammal, scavenging glucose from the host's blood serum and relying on aerobic glycolysis within specialised organelles called glycosomes (Priest and Hajduk, 1994). The resultant parasitaemia does not, however, normally kill the host as it is attenuated in two main ways. Firstly, the host immune response clears the majority of the parasites, and secondly, the long slender form ceases to divide at high cell densities (Pays *et al.*, 2001). Cessation of the mitotic cell cycle and differentiation into the non-proliferative short stumpy bloodstream form is induced by a low molecular weight factor, secreted by the long slender stage, which operates through a cyclic adenosine monophosphate (cAMP) signal transduction pathway (Hamm *et al.*, 1990; Vassella *et al.*, 1997). The short stumpy bloodstream cells are pre-adapted to the conditions within the tsetse fly gut and show signs of mitochondrial reactivation (Vickerman, 1985).

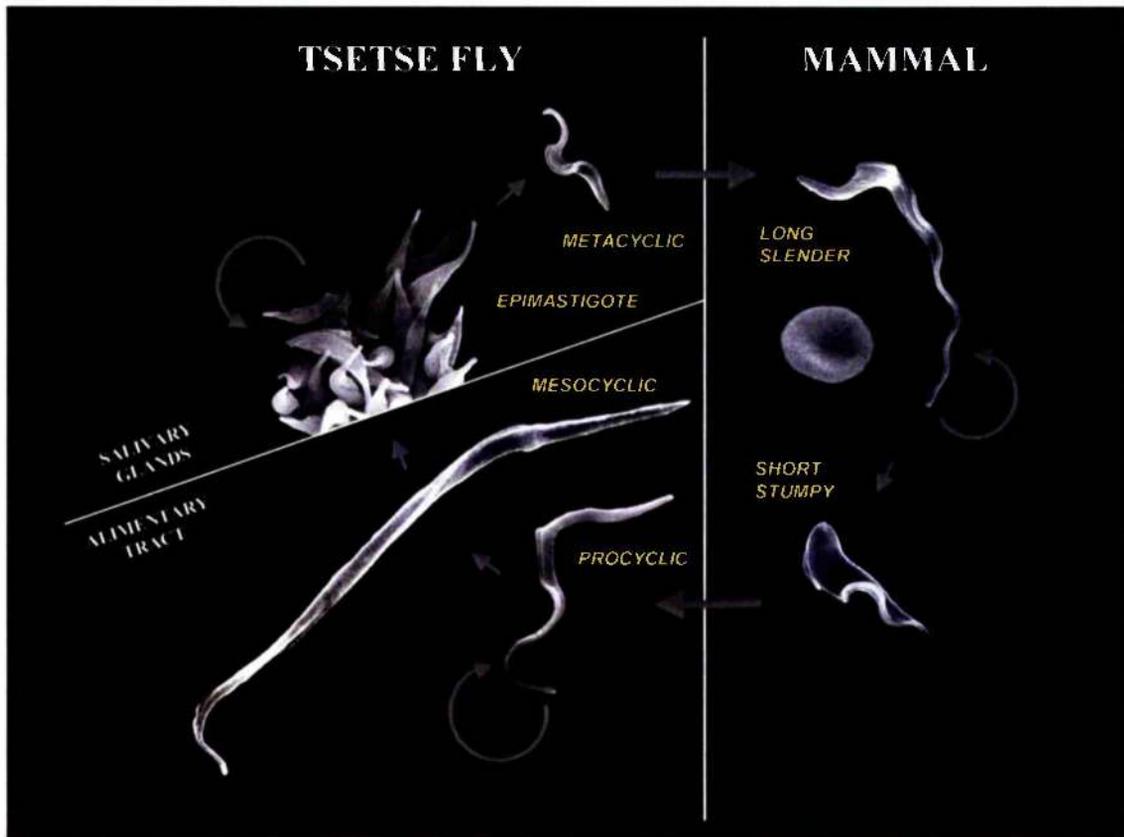
Metacyclic cells and both bloodstream forms possess a densely packed surface coat composed of Variant Surface Glycoprotein (VSG) (Cross, 1975). In excess of  $10^7$  copies of a single VSG species are expressed at the cell surface (Cross, 1996), and these molecules, which are present as dimers, define the variable antigen type (VAT) (Barry, 1997a; Barry and McCulloch, 2001). While VSG molecules are highly immunogenic, they function to protect the underlying invariant antigens from immune recognition and prevent the action of non-specific immune pathways through steric hindrance (Overath *et al.*, 1994). The ability of the bloodstream forms to evade the host immune system therefore relies on their capacity to periodically change this coat by a process known as antigenic variation. Each VSG coat is encoded by a distinct *VSG* gene and studies have suggested that *T. brucei* possesses a repertoire of approximately 1000 different *VSGs* (Van der Ploeg *et al.*, 1982). Immune evasion is dependent upon the parasite's ability to express a novel *VSG*, thereby switching the VSG coat exposed at the cell surface. *VSG* sequences are extremely variable so that the host immune system is likely to be unfamiliar with all the exposed epitopes in the new VSG coat (Blum *et al.*, 1993). Once infected, a host is often infected for life, although there are reported cases where the parasite eventually exhausts its *VSG* repertoire, and under these conditions the host can clear the infection in a phenomenon known as 'self-cure' (Nantulya *et al.*, 1984; Barry, 1986b; Nantulya *et al.*, 1986).

When a tsetse fly takes a blood meal from the infected host, only the short stumpy bloodstream forms survive as they are pre-adapted to the environment within the fly (Vassella *et al.*, 2000). These differentiate into procyclic trypomastigotes in the posterior midgut of the fly and undergo a switch to mitochondrial respiration using proline as the main energy source (Vickerman *et al.*, 1988; Priest and Hajduk, 1994). The VSG coat is shed and replaced by an invariant coat called procyclin, which is carbohydrate-rich (Roditi and Pearson, 1990). The procyclin coat comprises two distinct proteins, EP procyclin and GPEET procyclin, which are named after their characteristic internal repeat motifs (using the single letter amino acid code; Ruepp *et al.*, 1997). These proteins are induced early in differentiation, although EP expression is maintained while GPEET expression is repressed as established procyclic forms develop (Vassella *et al.*, 2000). Knockout studies have revealed that the EP and GPEET procyclins are not functionally equivalent, although the role of the procyclin coat is still under debate (Ruepp *et al.*, 1997). It is possible, however, that these proteins are required to avoid the proteolytic activities prominent in the fly's gut (Barry, 1997a).

The procyclic cells differentiate through a series of developmental stages, including the non-dividing mesocyclic form, while migrating from the midgut into the foregut, then through the oesophagus until they reach the salivary glands (Van Den Abbeele *et al.*, 1999). At this point, they differentiate into the epimastigote form which attaches to the microvilli of the salivary gland. Here they develop into nascent metacyclic cells, at which point they cease to divide, regain the VSG coat, the glycosomes enlarge and mitochondrial functions are repressed (Vickerman *et al.*, 1988). The cells detach from the microvilli to become free metacyclic trypomastigotes in the salivary gland space before entering the mammal when the tsetse fly feeds, and so beginning the life cycle again.

**Figure 1.1** The life cycle of *Trypanosoma brucei*

The different life cycle stages shown are scanning electron micrographs, reproduced to scale. Stages capable of replication are denoted by curved arrows, while straight arrows indicate differentiation and progression through the life cycle. An erythrocyte is shown next to the long slender bloodstream form for comparison. Epimastigote stage cells are shown as a cluster on the salivary gland epithelium where they grow as a dense monolayer (After Barry and McCulloch, 2001).



## 1.2 Antigenic variation

Enhanced phenotypic variation is a collective term encompassing several distinct mechanisms, including antigenic variation, which have evolved independently in a broad variety of pathogenic microorganisms. These phenomena enhance propagation or prolong survival of the micropathogens in the host, and ultimately promote transmission to new hosts (Barry and McCulloch, 2001). Although these events are generally random, the genomes of many of the organisms undergoing these processes are organised so that these events are more likely to occur, often at high frequencies (greater than  $10^{-5}$  events per generation; Henderson *et al.*, 1999). Phenotypic variation thereby generates a heterogeneous population capable of exploiting more diverse opportunities.

For viruses, evasion of the host immune system is the primary advantage of phenotypic variation, with the best characterised examples being those of the influenza virus and human immunodeficiency virus 1 (HIV1). Phenotypic variation in influenza virus alters the capsid haemagglutinin protein which is required for binding host cell receptors, but which also elicits the host immune response. Antigenic drift, which involves the accumulation of point mutations and therefore amino acid replacements in the haemagglutinin protein, allows the virus to initiate new rounds of infection in previously exposed populations (Gorman *et al.*, 1992). However, influenza pandemics arise due to a more fundamental change termed antigenic shift, which occurs less frequently and involves genetic reassortment between two viruses infecting the same host, leading to the development of a virus expressing a serologically distinct haemagglutinin coat (Fanning and Taubenberger, 1999).

In HIV1, the two main conserved regions required for viral attachment to CD4<sup>+</sup> host cells are the CD4 receptor and the chemokine co-receptor, which are both present on the homotrimeric gp120 envelope glycoprotein (Kwong *et al.*, 1998; Wyatt *et al.*, 1998). The CD4 binding site of gp120 is located in a deeply recessed groove that renders it inaccessible to antibodies. The chemokine co-receptor, however, is protected by three variable loops which undergo antigenic variation, thereby hiding the receptor site from the immune system while allowing host cell binding. This can be considered a form of intramolecular steric hindrance where variable sequences shield the underlying sequences from immune recognition.

Phenotypic variation in bacteria is achieved by a host of different mechanisms, and is not limited solely to pathogenic species. Most commonly affected by these processes are

surface structures such as fimbriae, flagella, lipopolysaccharide (LPS) and outer membrane proteins, which alter how the organism interacts with its environment or host. Two broad categories of phenotypic variation can be distinguished in bacteria. Firstly, processes which cause expression of a particular factor to be switched on or off are termed phase variation. For example, while shrouded by a surface capsule, meningococci are able to evade the host immune system. However, upon switching to the alternative state, they become able to invade host cells (Bucci *et al.*, 1999). The second, more complex form of phenotypic variation found in bacteria is known as antigenic variation, a phenomenon which allows the organism to switch between the expression of a number of alternative surface phenotypes. This strategy allows the pathogen to evade the specific response of the host immune system and prolong infection over long periods. *Borrelia hermsii* is a tick-borne spirochaete which causes relapsing fever, a disease characterised by a cyclic rise and fall in body temperature. Each relapse population of spirochaetes expresses an antigenically distinct surface lipoprotein that is encoded by a separate gene, and at least 25 of these genes can be expressed alternately from a single expression site (Barbour, 2002). Thus, although the host immune response clears the cell population expressing one lipoprotein, another population expressing a different lipoprotein continues to multiply. A similar system for the antigenic variation of surface lipoproteins has been described in the Lyme-disease spirochaete, *B. burgdorferi* (Zhang *et al.*, 1997).

A number of strategies for antigenic variation are also successfully utilised by unicellular eukaryotic parasites. African trypanosomes, including *T. brucei*, are able to repeatedly replace the VSG coat expressed at their cell surface using both recombinational and transcriptional switching mechanisms (see section 1.3). In fact, the similarities between the systems executing antigenic variation in trypanosomes and *Borrelia* spp. are striking, at both the genetic and phenotypic levels. Antigenic variation has also been described in the opportunistic fungal pathogen *Pneumocystis carinii*, which encodes a large number of major surface glycoprotein (MSG) genes that can be introduced into the expression site by recombination (Stringer and Keely, 2001). Furthermore, the intestinal parasite, *Giardia lamblia*, also possesses a considerable potential to alter the variant-specific surface protein (VSP) coat which covers its entire cell surface, using a system which resembles the transcriptional switching mechanism of *T. brucei* (Svärd *et al.*, 1998).

Antigenic variation is not limited solely to extracellular pathogens whose surface is in direct contact with the host immune system, as the intraerythrocytic parasites *Plasmodium falciparum* and *Babesia bovis* have also evolved systems which enable them to vary the antigens they express on the surface of their host cells (Barbour and Restrepo, 2000).

However, the functions of these proteins are more complex than the simple role in immune evasion required of the surface antigens of extracellular parasites. For instance, members of the PfEMP1 protein family, encoded by the *var* genes of *P. falciparum*, are receptors which concentrate into knob-like protrusions on the surface of the parasitised cell. These receptors bind host molecules on endothelial cells, thereby sequestering infected erythrocytes from immune surveillance at sites such as the spleen (Gardner *et al.*, 1996; Wahlgren *et al.*, 1999). Two other protein families, encoded by the *rif* and *stevor* genes, are also thought to undergo antigenic variation in *P. falciparum*, although the functions of these antigens have yet to be fully elucidated (Beeson and Brown, 2002).

### 1.3 Antigenic variation in *T. brucei*

Since the advent of molecular biology, researchers have been studying the mechanisms involved in antigenic variation (reviewed in Donelson, 1995; Cross, 1996; Barry, 1997b; Borst *et al.*, 1998; Pays and Nolan, 1998; Barry and McCulloch, 2001). Until recently, most of this research was conducted using monomorphic lines of trypanosomes developed from a number of stocks with differing genetic backgrounds by prolonged, often rapid syringe passing of *T. brucei* through laboratory animals (Barry and McCulloch, 2001). These cell lines are unable to differentiate from long slender to short stumpy forms, at least in the case of strain MITat 1.2a S427 because they have lost the ability to signal a response to the stumpy induction factor (Vassella *et al.*, 1997). Consequently, they reach high parasitaemias *in vivo*, and can readily grown *in vitro*. Non-adapted, or pleomorphic, trypanosomes continue to generate short stumpy forms both *in vitro* and *in vivo* and are readily able to establish tsetse fly infections. Monomorphic lines have been shown to switch at a low frequency of  $10^{-6}$  to  $10^{-7}$  switches/cell/generation (Lamont *et al.*, 1986). In contrast, in pleomorphic lines the switch rate is at least  $10^{-5}$  switches/cell/generation and can be as high as  $10^{-2}$  switches/cell/generation (Turner and Barry, 1989; Turner, 1997). The basis for this switching difference is unknown, but may be connected to changes in cell cycle control, or could indicate that monomorphic cells are deficient in a switch-promoting function, such as an endonuclease that promotes homologous recombination (Barry, 1997b). One advantage of the switch deficiency in monomorphic cells is it has allowed researchers to elucidate that several mechanisms enable *T. brucei* to express a novel *VSG*, and thereby a novel surface coat (see below). Such studies would have been seriously hampered by the inherent instability of *VSG* gene expression in pleomorphic cells. However, recent data have suggested that monomorphic and pleomorphic cells differ in their usage of the described switching mechanisms, highlighting the need for further research into the genetic controls of antigenic variation in both these cell lines (Barry, 1997b; Robinson *et al.*, 1999). It should also be noted that the rates of underlying *VSG* activation are probably underestimated by methodologies recording phenotypic switching only, since a trypanosome population can contain individual cells expressing the same *VSG* which arose via multiple independent gene rearrangement events (Timmers *et al.*, 1987).

*VSG* genes are only expressed when present in specialised telomeric locations called *VSG* expression sites (Figure 1.2; Pays *et al.*, 2001). In a given cell only a single expression site is transcriptionally active at a time, ensuring that a homogeneous *VSG* coat is expressed at the cell surface. Expression sites comprise stage-specific transcription units which are thought to be transcribed by RNA polymerase I, the enzyme also responsible for

transcribing the ribosomal DNA and the procyclin loci (Barry and McCulloch, 2001; Navarro and Gull, 2001). Two types of expression site have been described. These are known as metacyclic expression sites (*MESs*), so named because they are expressed within the metacyclic developmental stage, and bloodstream expression sites (*BESs*), which are expressed in both long slender and short stumpy cells within the mammal. All expression sites are telomeric, and the *VSG* gene is invariably found directly adjacent to the telomeric repeats ((CCCTAA)<sub>n</sub>; Van der Ploeg *et al.*, 1984). Both the *MESs* and *BESs* are present on all size classes of chromosome except the minichromosomes. *MESs* are generally about 4-6 kb in length and there are an estimated 27 such loci within a cell (Figure 1.2b; Turner *et al.*, 1988). These loci are the only known protein-coding monocistrons in trypanosomes, the *VSG* being the single transcribed gene (Barry and McCulloch, 2001). Other features characteristic of *MESs* include a poorly conserved promoter region (Bringaud *et al.*, 2001) and a small number of 70-bp repeats (typically 1 or 2 repeat units) found upstream of the *VSG*, although some *MESs* lack these repeats (Lenardo *et al.*, 1984). The metacyclic population of a tsetse fly salivary gland is heterogeneous, consisting of a mixture of variable antigenic types (*VATs*), thereby creating a diversity which should maximise infection of new hosts and circumvent the problem of existing anti-*VSG* antibodies in previously infected hosts (Barry *et al.*, 1998). This is achieved by each metacyclic cell randomly selecting and activating a single *MES*.

Metacyclic cells introduced into a mammalian host quickly re-enter the cell cycle and differentiate into the long slender proliferative bloodstream form. Within a few days these cells switch their *VSG* coat and initiate transcription from one of approximately 20 *BESs* (Figure 1.2a; Pays *et al.*, 2001; Vanhamme *et al.*, 2001). These loci are large polycistronic transcription units, up to 65 kb in size, and the overall structure, and indeed sequence, of these sites appears to be largely conserved. *BESs* are demarcated at their 5' end by a large array of 50-bp repeats located upstream of the *BES* promoter, while a number of non-*VSG* expression site-associated genes (*ESAGs*) are located downstream of the promoter. The functions of a number of the *ESAGs* have been elucidated; for instance, *ESAG6* and *ESAG7* encode a heterodimeric receptor for the uptake of host transferrin (Salmon *et al.*, 1994; Steverding *et al.*, 1994), while *ESAG4* encodes an adenylate cyclase (Paindavoine *et al.*, 1992). However, many of the *ESAG* gene products have yet to be assigned a function, nor is it known why it is necessary that these genes be co-expressed with the *VSG*. Separating the final *ESAG* and the *VSG* is a large 'barren' region composed of degenerate 70-bp repeats, believed to facilitate recombinational *VSG* switching (see below).

Since the VSG coat present on the trypanosome cell surface is highly immunogenic, the parasite must alter the active VSG gene to express a novel surface coat in order to avoid the host immune system and maintain infection. For this reason the *T. brucei* genome encodes an estimated 1000 VSG genes, although many of these may be damaged or incomplete (Van der Ploeg *et al.*, 1982). The majority of these genes are thought to reside in large arrays within chromosome-internal locations, and are termed basic-copy (BC) genes, while most of the remaining VSGs are found in telomeric regions, especially those of the 100 or so mini-chromosomes (Gull *et al.*, 1998). The remaining VSGs are located within the MESs and BESs. VSGs can be considered as occupying cassettes with common upstream and downstream sequences. At the 5' end, the majority of VSGs are flanked by runs of imperfect 70-bp repeats (Liu *et al.*, 1983; Aline *et al.*, 1985). BC genes possess a small number of such repeats, between 1 and 10, while mini-chromosomal telomeric VSGs have been shown to have more. As described above, within BESs the 'barren' region may contain hundreds of these repeats spanning several kilobases upstream of the VSG. The ESAGs also provide many kilobases of homology between BESs (Thi *et al.*, 1991; Xong *et al.*, 1998). Downstream, at the 3' end of all VSGs, are the highly conserved carboxyterminal coding regions followed by the 3' untranslated regions. For mini-chromosomal telomeric and expression site VSGs the telomeric hexanucleotide repeats also provide several kilobases of homology (Aline *et al.*, 1985; Aline and Stuart, 1989). These cassettes possess the features of functional units designed to undergo homologous recombination, and it is hypothesised that the homologous recombination machinery underlies a number of the reported VSG switching mechanisms (see below and section 1.4).

Currently, there are five main mechanisms reported for antigenic variation in *T. brucei* (Figures 1.3, 1.4, 1.5, 1.6 and 1.7). The first mechanism is termed expression-linked copy (ELC) formation, and comprises the first of two forms of duplicative transposition (Figure 1.3). During this mechanism a chromosome-internal VSG gene is copied and inserted into the active BES. The VSG that previously occupied the BES is deleted, and the new VSG copy, or ELC, is expressed (Michels *et al.*, 1983; Pays *et al.*, 1983b). Since this mechanism shares similarities with gene conversion reactions found in all organisms, which are mediated by cellular DNA recombination enzymes (see section 1.4), it is possible these same enzymes assist in ELC formation. This is the only mechanism allowing the activation of BC genes, which account for the majority of silent VSGs and which cannot be expressed unless transposed into the active BES. Numerically, this must be an important mechanism for antigenic variation. ELC formation occurs readily in both monomorphic and pleomorphic lines of trypanosomes, and is mediated by a recombination

reaction that relies upon the regions of homology which flank the *VSGs* and form the gene cassette. In most cases, imperfect 70-bp repeats delineate the upstream limit for recombination (Campbell *et al.*, 1984; De Lange *et al.*, 1985; Florent *et al.*, 1987; Lee and Van der Ploeg, 1987), as is true for every reported case in pleomorphic lines so far (Matthews *et al.*, 1990). However, in monomorphic lines, the region of transposition can extend to other homologous sequences (Michiels *et al.*, 1983; Pays *et al.*, 1983b; Lee and Van der Ploeg, 1987) and deletion of the 70-bp repeats from the active ES have been shown not to affect ELC formation (McCulloch *et al.*, 1997). These data have led to the suggestion that ELC formation occurs by general homologous recombination in monomorphic trypanosomes and that these cell lines are deficient in another, dedicated system for duplicative transposition which operates in pleomorphic cells (Barry, 1997b).

The second form of duplicative transposition is called telomere conversion, and occurs when a *VSG* is copied from the telomere of a chromosome and inserted into the active ES (Figure 1.4). Again, the *VSG* originally found in the *BES* is deleted. While exhibiting many similarities with ELC formation, telomere conversion differs because the transposed region can terminate downstream of the end of the *VSG* gene cassette, within subtelomeric regions of homology (De Lange *et al.*, 1983; Bernards *et al.*, 1984) or even the hexanucleotide telomeric repeats at the end of the chromosome (De Lange *et al.*, 1983). In fact, some telomere conversions may simply copy the entire donor telomere into the *BES* (De Lange *et al.*, 1983). This mechanism is highly significant in antigenic variation, as it is probably the main route for mini-chromosomal *VSG* gene activation, both in monomorphic and pleomorphic trypanosomes.

Telomere reciprocal recombination is the third form of recombination-dependent antigenic variation in trypanosomes (Figure 1.5). This mechanism involves the exchange of telomeric regions containing *VSG* genes between chromosomes, and is mediated by a simple genetic cross-over event (Barry, 1997b). It does not involve gene conversion, and the *VSG* within the active site is not deleted. To effect an antigenic switch this genetic exchange must introduce a novel *VSG* gene into the active *BES*, and the previously expressed *VSG* is moved to a silent telomere. For this process, the position of genetic exchange may be within the 70-bp repeats (Pays *et al.*, 1985a) or further upstream within the *BES* (Pays *et al.*, 1983a; Shea *et al.*, 1986). Until recently only two cases of this mechanism had been reported. However, Rudenko *et al.* (1996) found more examples, and suggested that this may be a major mechanism of antigenic switching in monomorphic lines. Nevertheless, since this mechanism has not so far been observed in pleomorphic

cells, it is likely that this type of antigenic variation is simply the result of random mitotic cross-over events.

The final recombination-dependent *VSG* switching mechanism is mosaic gene formation (Figure 1.6). Using this process, which is also called segmental gene conversion, novel *VSGs* can be generated by chance recombination between two or more *VSG* genes (reviewed in Barbet and Kamper, 1993). Segmental gene conversion was originally reported to involve the retention of the 3' coding sequence of the *VSG* initially in the *BES*, while the upstream region was replaced by a fragment of a novel gene (Pays *et al.*, 1983a). Later, mosaic gene formation was reported between members of the same *VSG* gene family (Pays *et al.*, 1985b), and more recently it has been shown that many mosaic genes are constructed from fragments of pseudogenes (Roth *et al.*, 1986; Thon *et al.*, 1990). The nature of this mechanism suggests that such events are likely to be rare. However, if a chronically infected host eventually develops immunity against all the readily expressed *VSGs* in the repertoire, no other VATs can survive and the products of this mechanism may then be detected. It seems likely that this mechanism is also unregulated, and relies upon chance homologies within the recombining *VSG* genes.

The fifth mechanism allowing the expression of a novel *VSG* in *T. brucei* is termed *in situ* transcriptional switching (Figure 1.7). This type of antigenic variation involves the transcriptional activation of one of the silent *BESs* along with the concomitant silencing of the active *BES*, thereby switching between the two *VSGs* that occupy those particular sites and enabling the cell to express a new surface coat. In monomorphic cells *in situ* switching is often the most commonly observed mechanism of antigenic variation (McCulloch *et al.*, 1997; McCulloch and Barry, 1999). However, polymorphic cells rarely seem to utilise this mechanism, perhaps because it can only be used to activate the *VSGs* which happen to occupy the *BESs* (Robinson *et al.*, 1999). It has been known for many years that only a single expression site can be active at a time, and many theories have been advanced over the years to explain this phenomenon (reviewed in Barry and McCulloch, 2001). However, the recent discovery of a subnuclear structure associated with expression of the active *BES* by RNA polymerase I appears to have solved the problem (Navarro and Gull, 2001). This structure, termed the expression site body (ESB), is present only in the bloodstream form, and appears to be a coherent structure inside which the active *BES* is located and transcribed. Inactive *BESs* are not associated with this transcriptional body. Since the ESB recruits RNA polymerase I for transcription, it is likely to share some features with the nucleolus but is nevertheless entirely separate (Chaves *et al.*, 1998; Navarro and Gull, 2001). These results suggest a model in which

*BES* recruitment and activation relies on interaction with the ESB, and where inactive *BESs* are excluded from the structure while it is occupied by the active *BES*. This is supported by evidence showing that two *BESs* cannot be fully active at the same time, and that attempts to force cells to express two sites simultaneously, using antibiotic selection, appears to cause the cells to switch rapidly back and forth between the two sites (Chaves *et al.*, 1999). Although the mechanism ensuring that only one *BES* occupies the ESB at a time remains unknown, it may rely on some conserved feature of the expression site, since the sequences of *BESs*, outwith the *VSGs*, have been shown to be around 90% identical (Xong *et al.*, 1998; Navarro and Gull, 2001). Furthermore, this model suggests that *in situ* switching would proceed by a process which caused the interaction between the active *BES* and ESB to become unstable, and an inactive *BES* could then move to occupy the ESB, resulting in displacement of the active *BES* by an inactive one. How this is co-ordinated, and whether it is related to DNA replication, recombination or chromatin remodelling, remains unclear (Navarro and Gull, 2001).

A sixth mechanism which has been proposed involves the accumulation of point mutations during the generation of a copy of a previously expressed *VSG* gene sequence (Donelson, 1995). However, all the exposed epitopes encoded by the *VSG* would have to be altered from the progenitor sequence for the cell to avoid destruction by the host immune system, and it is difficult to see how such a high rate of point mutation could be achieved.

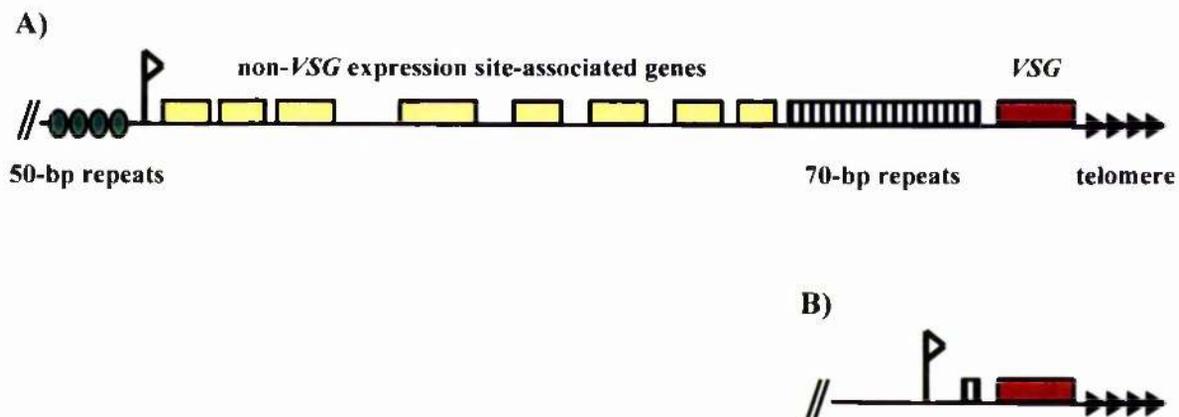
Explanations for many aspects of antigenic variation in *T. brucei* remain elusive, and almost nothing is known about the genes in trypanosomes that regulate it. The reduction in switch rate in monomorphic lines has prompted the suggestion that an active mechanism for antigenic variation exists that is diminished or absent in these cell lines (Barry, 1997b). Since the switch rate is comparable to that expected by background mutation alone, it is possible switching occurs in monomorphic cells by random homologous recombination, and that pleomorphic cells possess a dedicated pathway for *VSG* recombination.

Trypanosomes do possess pathways to ensure *VSGs* are expressed in a loose hierarchy, leading to the appearance of VATs in a semi-definable order (Gray, 1965; Capbern *et al.*, 1977; Barry, 1986a; Thon *et al.*, 1990), so preventing the exhaustion of the whole *VSG* repertoire within the early waves of parasitaemia. This hierarchy appears to be based on the probability of activation of *VSG* genes in a variety of locations and under a number of circumstances (Barry and McCulloch, 2001). For instance, *VSGs* are more frequently activated when present in telomeric rather than chromosome-internal locations (Liu *et al.*, 1985; Robinson *et al.*, 1999). In pleomorphic trypanosomes, duplication is by far the predominant switching mechanism, while *in situ* switching events are rare (Robinson *et al.*,

1999). It follows that telomere reciprocal recombination and mosaic gene formation events would be less frequent still. At the head of the hierarchy are the telomeric *VSGs*, perhaps because the flanking regions of these genes are likely to share longer stretches of homology with the active *BES*, thereby increasing the probability of successful gene conversion.

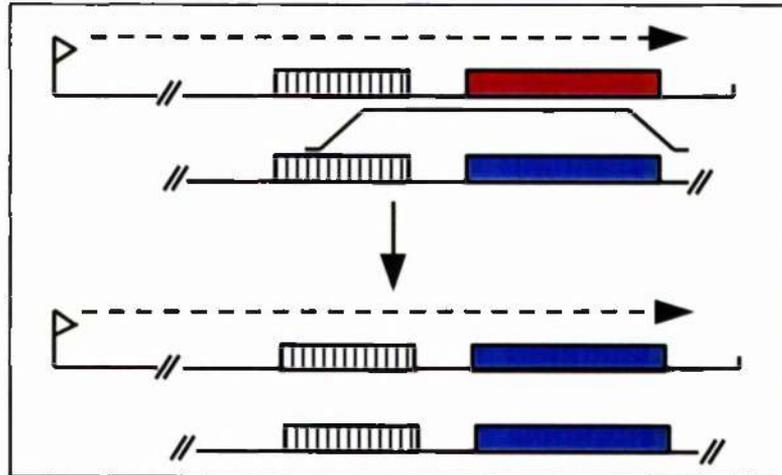
## Figure 1.2 *VSG* expression sites in trypanosomes

(A) Bloodstream expression sites (*BESs*) are large polycistronic transcription units up to about 60 kb in length (Pays *et al.*, 2001). The expression site promoter (shown as a flag) drives transcription of up to a dozen non-*VSG* expression site-associated genes (*ESAGs*; shown in yellow), as well as the *VSG* gene (shown in red) which is found at the terminus of the *BES*, proximal to the telomeric repeats (shown as black arrowheads). Separating the *ESAGs* and *VSG* gene are a large array of 70-bp repeats (shown as vertical black and white stripes), while the 5' boundary of the *BES* is delimited by an array of 50-bp repeats (shown as green ovals). (B) Metacyclic expression sites (*MESs*) are monocistronic transcription units approximately 4-6 kb in length which contain only a *VSG* gene and a small number of 70-bp repeats upstream of the telomeric repeats (Barry and McCulloch, 2001).



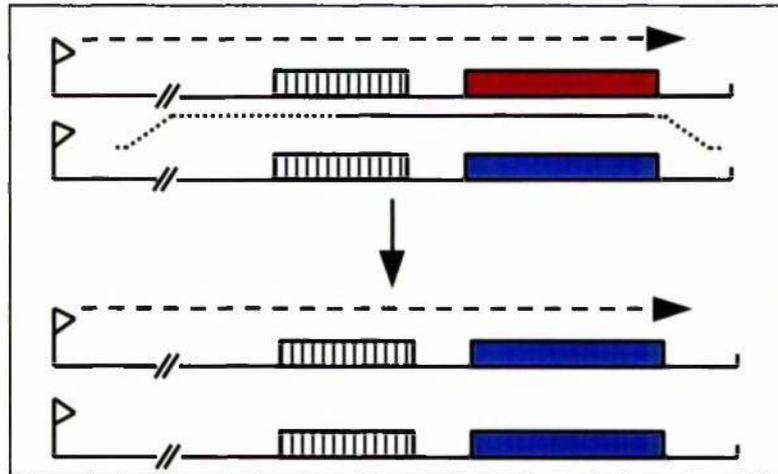
### Figure 1.3 Expression linked copy (ELC) formation

A simplified scheme showing the replacement of the *VSG* (shown in red) in the active *BES* (indicated by a dashed arrow) by a novel inactive chromosome-internal *VSG* (shown in blue) using homologous sequences present in the 70-bp repeats (shown as vertical black and white stripes) and downstream of the *VSG*s. The *BES* promoter is depicted as a flag. The expressed *VSG* is destroyed and replaced by a copy of the inactive *VSG* (ELC) which is then transcribed.



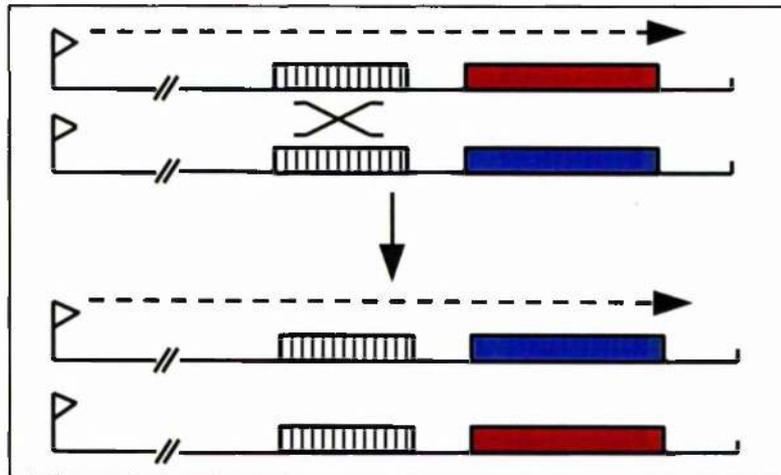
### Figure 1.4 Telomere conversion

A simplified scheme showing the replacement of the *VSG* (shown in red) in the active *BES* (indicated by a dashed arrow) by a novel inactive telomeric *VSG* (shown in blue) using 5' homologous sequences present in the 70-bp repeats (shown as vertical black and white stripes) or between the upstream regions of two expression sites; the 3' conversion limit may lie in homologous sequences downstream of the *VSG*s, or conversion may copy the entire donor telomere into the *BES*. The *BES* promoter is depicted as a flag. The expressed *VSG* is destroyed and replaced by a copy of the inactive *VSG* which is then transcribed.



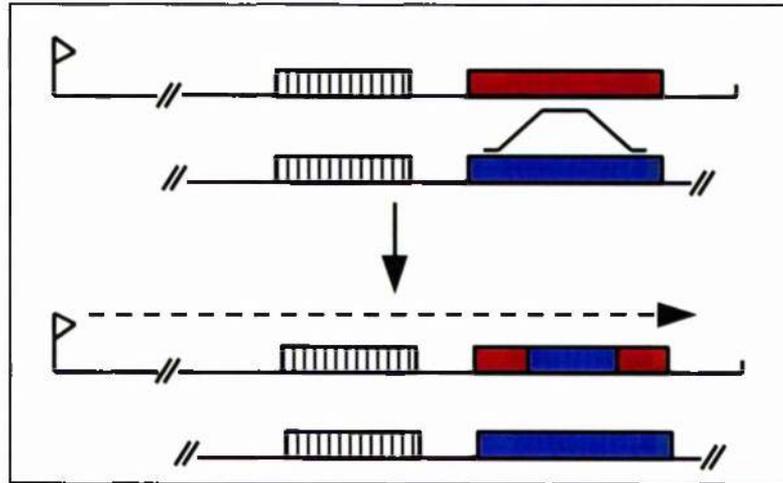
### Figure 1.5 Telomere reciprocal recombination

A simplified scheme showing crossing-over between the active *BES* (indicated by a dashed arrow) and an inactive *BES*. During this process the expressed *VSG* (shown in red) is exchanged for a novel inactive *VSG* (shown in blue) which is then transcribed. These events utilise homologous sequences upstream of the *VSG*s, either within the 70-bp repeats (shown as vertical black and white stripes) or between the upstream regions of two *BES*s. The *BES* promoter is depicted as a flag.



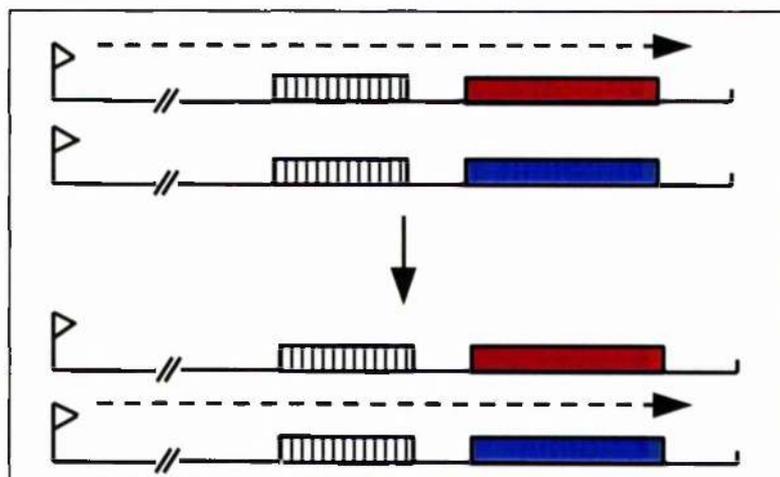
### Figure 1.6 Mosaic gene formation

A simplified scheme showing recombination between homologous sequences present within an inactive or damaged *VSG* within an inactive *BES* (shown in red) with another inactive or damaged *VSG* (shown in blue). This process may result in the generation of a novel *VSG* (shown in red/blue) within the inactive *BES*. Activation of this *BES* (indicated by a dashed arrow) results in the expression of a novel *VSG*. The *BES* promoter is depicted as a flag.



**Figure 1.7** *In situ* transcriptional switching

A simplified scheme showing the inactivation of the active *BES* (indicated by a dashed arrow) containing the expressed *VSG* (shown in red), along with the simultaneous activation of an inactive *BES* containing a novel *VSG* (shown in blue). The *BES* promoters are shown as flags and the 70-bp repeats as vertical black and white stripes.



#### 1.4 DNA recombination and antigenic variation

DNA recombination has been recognised as one of the primary forces underlying antigenic variation in *T. brucei* (McCulloch and Barry, 1999). At present, very little is known about the recombination machinery possessed by trypanosomes, although recent work suggests that multiple pathways of recombination and DNA break repair may be present (see below), mirroring the situation in other eukaryotes.

Extensive research into the recombination and DNA break repair pathways of yeast and mammals has revealed that DNA recombination is usually initiated by the presence of a double-strand break (DSB) or free double-strand end (free end). These lesions can arise by the direct action of ionising radiation and some chemicals, by the action of endonucleases during events such as meiosis or mating type switching in *Saccharomyces cerevisiae*, or perhaps most importantly, during DNA replication when a replication fork encounters a single-strand break (a nick) or collides with a protein that terminates replication (Haber, 1999; Karran, 2000; Cromie *et al.*, 2001). If such lesions remain unrepaired they may lead to chromosome loss and ultimately cell death, while imprecise repair can result in mutations and chromosomal rearrangements (van Gent *et al.*, 2001). For these reasons, eukaryotes have evolved an impressive array of DSB-repair pathways.

Two general types of DSB repair can be distinguished in eukaryotes: non-homologous end joining (NHEJ) in which the broken ends are directly rejoined to one another; and homologous recombination (HR), where an intact double-strand DNA molecule acts as a template to repair the broken molecule. HR can in fact be sub-divided into several pathways, including gene conversion and break-induced replication (BIR), in which an intact homologous chromosome or sister chromatid is used as a template for repair, and single-strand annealing (SSA) where complementary regions flanking both ends of the DSB are joined to repair the lesion (Figure 1.8; Pâques and Haber, 1999). These mechanisms have been conserved throughout evolution, and adapted to accommodate the demands of increasingly complex organisms and their environments. Many of the proteins central to these processes in *S. cerevisiae* have homologues in other eukaryotes, although the relative contribution of NHEJ and HR varies between lower and higher eukaryotes. Moreover, within an organism the choice of repair pathway may also be influenced by the phase of the cell cycle, the nature of the lesion to be repaired and, in multicellular organisms, the cell type and stage of development (reviewed in Pâques and Haber, 1999; Pastink and Lohman, 1999; Haber, 2000; Karran, 2000; Cromie *et al.*, 2001). Supplementary to their role in genome maintenance, these repair pathways are also utilised

during processes which generate genetic diversity; for instance, genetic exchange during meiosis (Pâques and Haber, 1999), or V(D)J recombination during the generation of B and T cell lineages (Featherstone and Jackson, 1999).

NHEJ is the most prevalent form of non-homologous, or illegitimate, DSB repair, and requires little or no base-pairing to join the DNA ends (Figure 1.8a; Moore and Haber, 1996; Jeggo, 1998). In mammalian cells, this process begins with the assembly of the multiprotein complex, DNA-dependent protein kinase (DNA-PK). Both of the DNA ends are bound by heterodimers composed of the KU70 and KU80 proteins, which recruit and activate the catalytic subunit of DNA-PK (DNA-PK<sub>cs</sub>). The DNA-PK complex then recruits DNA ligase IV and XRCC4 (X-ray cross-complementing 4), which together accomplish the ligation step. The same core proteins are required during NHEJ in *S. cerevisiae*, although no homologue of DNA-PK<sub>cs</sub> has been identified in this organism. A further difference lies in the involvement of the RAD50-MRE11-containing complex in NHEJ in *S. cerevisiae* (Moore and Haber, 1996), while the absence of MRE11 has little effect on the vertebrate system (Yamaguchi-Iwai *et al.*, 1999). Whilst KU70/80-dependent NHEJ is the best characterised reaction, genetic evidence in yeast has suggested that forms of NHEJ can occur in KU70/80 mutants, albeit less efficiently, and with different reaction requirements (Pâques and Haber, 1999; Manolis *et al.*, 2001).

In *S. cerevisiae*, DSB repair by HR relies upon the RAD52 epistasis group proteins, including RAD50, RAD51, RAD52, RAD54, RAD57, RAD59, MRE11 and XRS2, and homologues of the majority of these proteins have been identified in mammalian cells (Pâques and Haber, 1999). A number of distinct mechanisms of HR have been described in eukaryotes to explain experimental observations gathered using a variety of assay methods, and to provide a framework for further study. However, *in vivo* it is unlikely that these repair pathways are as clearly delineated. Within the cell, molecular cross-talk between the pathways may generate flexibility, allowing HR to repair a broad spectrum of lesions using different combinations of the RAD52-group proteins.

The simplest HR pathway, SSA, is dependent on the presence of direct repeats on both sides of the DSB (Figure 1.8b; Pâques and Haber, 1999). In *S. cerevisiae*, after 5' to 3' resection of both ends of the break by the RAD50-MRE11-XRS2 complex, or perhaps by other nuclease activities, these regions of homology are exposed and can anneal together to form a joint molecule. This process appears to be facilitated by RAD52, since it is capable of promoting the annealing of complementary single strands *in vitro* (Mortensen *et al.*, 1996; Shinohara *et al.*, 1998), but other RAD52-group members in *S. cerevisiae* do not

seem to be required (Pastink and Lohman, 1999). The single-stranded non-homologous tails are then removed by the RAD1-RAD10 endonuclease working in conjunction with the mismatch repair heterodimer MSH2-MSH3 (see section 1.6). Repair DNA synthesis is followed by ligation to complete the repair. The SSA repair process results in deletion of the sequence between the two repeats along with one of the repeat regions. SSA has been observed between molecules containing as little as 30 bp of complementary sequence, although the efficiency of the process is greatly enhanced where 200-400 bp of homologous sequence are available (Sugawara *et al.*, 2000), and repair events resulting in the deletion of as much as 15 kb have been reported (Sugawara and Haber, 1992). SSA also occurs in mammalian cells, and although little is known about the genetic requirements of this process, it is assumed that the mammalian counterparts of the RAD50-MRE11-XRS2 nuclease and RAD1-RAD10 endonuclease complexes are involved (Karran, 2000).

Gene conversion, or the nonreciprocal transfer of DNA from one molecule to its homologue, is the most extensively studied form of HR, and is thought to occur by two pathways: gene conversion involving Holliday junctions (Figure 1.8d; Szostak *et al.*, 1983) and synthesis-dependent strand annealing (SDSA; Figure 1.8e; Nassif *et al.*, 1994). The pathways share many mechanistic features, and the early stages of these repair events appear to be identical (Pâques and Haber, 1999). As in SSA, the ends of the DSB are resected, probably by the RAD50-MRE11-XRS2 complex, to generate 3' single-stranded tails. RAD51 protein polymerises onto the single-stranded tails to form a nucleoprotein filament that searches for homologous duplex DNA. Upon successful completion of this search, DNA strand exchange forms a joint molecule between the homologous intact duplex DNA and the damaged DNA. The formation of RAD51 nucleoprotein filaments is facilitated by the co-ordinated action of RAD52, the RAD55-RAD57 heterodimer and replication protein A (RPA; Sung, 1994; Sung, 1997; Shinohara and Ogawa, 1998); the strand exchange reaction is promoted by RAD54 (Petukhova *et al.*, 1998). Repair DNA synthesis requires DNA polymerase and accessory factors, but the identity of the polymerases involved remains unclear. The two models differ mainly in the methods used to resolve the strand exchange recombination intermediates, prior to repair completion by DNA ligase. In gene conversion involving Holliday junctions, the 3' ends are thought to invade the homologous template and initiate new DNA synthesis, leading to the formation of two four-stranded branched structures called Holliday junctions. Theoretically, these Holliday junctions should be cleaved by resolvases, analogous to RuvC or RusA in eubacteria (Connolly *et al.*, 1991; Chan *et al.*, 1997). So far, only one candidate resolvase, active in the nuclei of eukaryotes, has been identified. This protein, called MUS81,

cleaves Holliday junctions *in vitro* (Chen *et al.*, 2001), but is thought to use a novel enzymatic strategy, dissimilar from those of the eubacterial enzymes, and it does not seem to possess all the expected properties of a eukaryotic Holliday junction resolvase (Haber and Heyer, 2001). A complex of unknown composition which can cleave Holliday junctions in a manner similar to that of RuvC has also been identified in mammalian cell extracts (Constantinou *et al.*, 2001). Other Holliday junction resolvases have also been identified in the mitochondria of budding yeast and fission yeast (Kleff *et al.*, 1992; White and Lilley, 1997). Gene conversion is not associated with crossing-over if both Holliday junctions are cleaved in the same direction. Alternatively, if the crossed strands of one Holliday junction are cleaved, while the non-crossed strands of the other Holliday junction are cut, then gene conversion is associated with the crossing-over of flanking markers. The second gene conversion model, SDSA, does not involve Holliday junctions. Instead, the newly synthesised DNA strands are returned to the broken molecule after displacement from the template, after which the two newly synthesised strands can anneal to each other. DNA synthesis then fills any gaps remaining in the broken molecule to complete the repair. This process is not associated with the crossing-over of markers flanking the DSB.

The final HR pathway, BIR, has been postulated to explain the observation of very long gene conversion tracts, in some cases up to 400kb in length (Figure 1.8c; Esposito, 1978; Pâques and Haber, 1999; Kraus *et al.*, 2001). Very little is known about the mechanism or the components involved in this type of repair, although evidence implicates a number of the RAD52-group proteins along with the replication machinery. BIR appears to be initiated in a similar manner to gene conversion, with the 3' end of the single-stranded tail invading intact homologous duplex and forming a joint molecule. The 3' end then probably initiates both leading- and lagging-strand synthesis to form a true replication fork. As in replication, this fork is able to synthesise many kilobases of DNA, explaining the large conversion tracts observed as a result of this repair mechanism. This recombination-induced replication process may also result in DNA synthesis extending the length of whole chromosome arms, with synthesis proceeding to the end of the telomere.

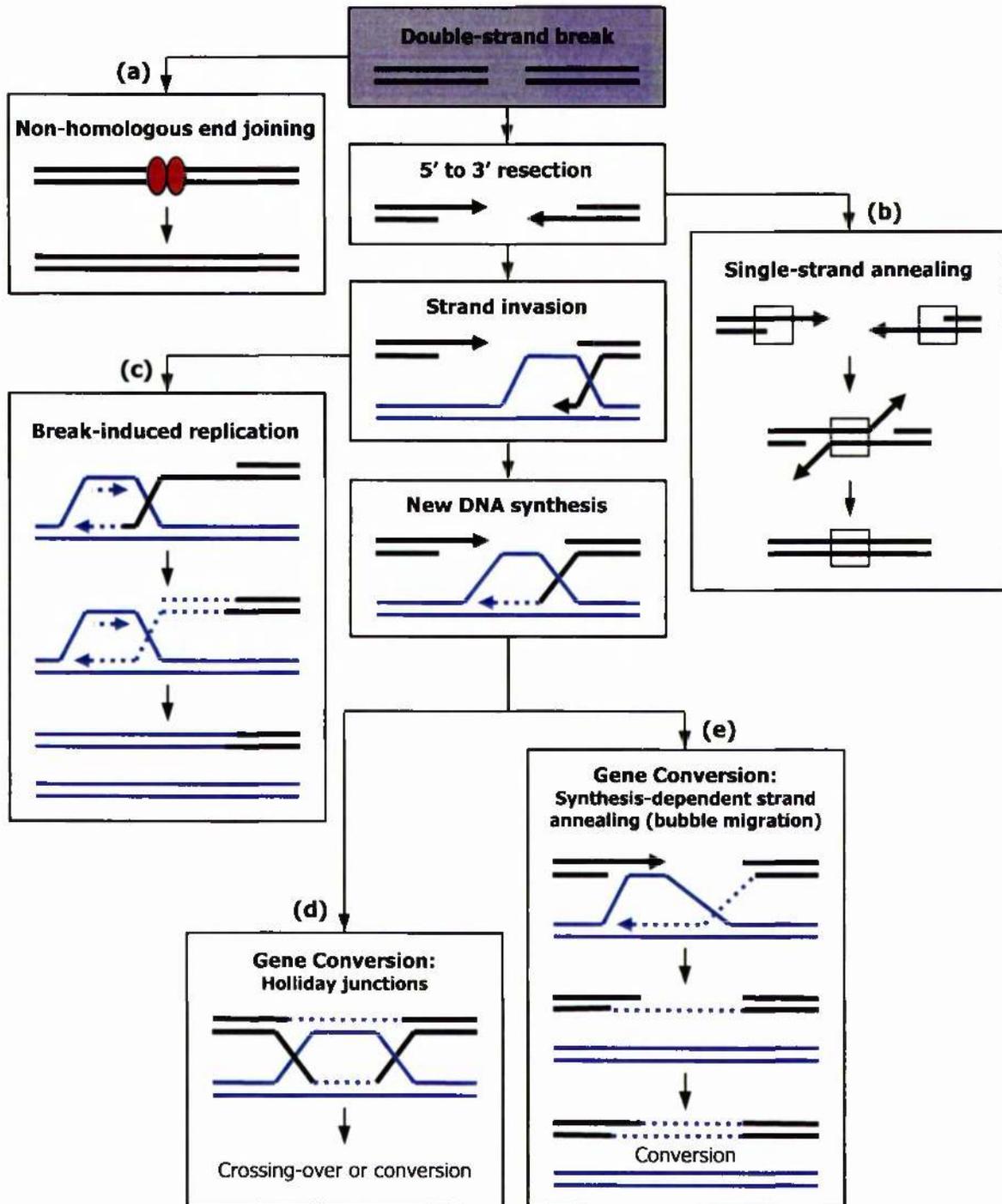
The majority of *VSG* switching events in *T. brucei* involve DNA rearrangements, and as discussed in section 1.3, *VSG* genes occupy cassettes with common upstream and downstream sequences, suggesting that homologous recombination underlies these events. In recent years, homologues of a number of proteins central to the recombination pathways of other eukaryotes have been identified in trypanosomes, and functional analyses of these genes has begun to shed light on the recombination mechanisms available to this organism. The first recombinase gene characterised in *T. brucei* encodes a homologue of the highly

conserved RAD51 protein (McCulloch and Barry, 1999). RAD51 proteins bind both single- and double-stranded DNA to form nucleoprotein filaments in the presence of ATP (Benson *et al.*, 1994). These proteins display ATPase activity in the presence of single-stranded DNA, and upon binding, RAD51 can catalyse the transfer of a single-stranded DNA tail into intact duplex DNA, to generate homology-dependent strand-exchange intermediates in which the single-stranded DNA is annealed with the complementary strand of the duplex DNA, and the non-complementary strand is displaced (Sung, 1994; Baumann *et al.*, 1996). Homozygous *T. brucei* *RAD51* mutants exhibited an increased sensitivity to the radiomimetic alkylating agent methyl methanesulphonate (MMS) and a reduced efficiency of integration of DNA constructs following transformation, both phenotypes consistent with the loss of a component of the HR machinery. Analysis of the role of *RAD51* in VSG switching revealed a complex phenotype: the mutants switched their VSG coats at a reduced overall frequency relative to wild-type cells, but interestingly this was due to down-regulation of both recombinational switching and *in situ* transcriptional switching. Switching events of both types were still detectable in the homozygous mutants, however, indicating that *RAD51*-independent pathways of switching, and therefore DNA recombination, exist in *T. brucei*. Recent investigations into this residual recombination activity, using DNA transformation, have revealed that trypanosomes contain at least two, and perhaps three, pathways of homologous recombination (Conway *et al.*, 2002b).

More recently, the roles of the *KU70* and *MRE11* genes have been investigated in trypanosomes. *KU70* associates with *KU80* to form a heterodimer which is essential for efficient and accurate NHEJ (Featherstone and Jackson, 1999). The primary function of the *KU70/80* heterodimer is to bind DNA ends and recruit factors, such as DNA-PK<sub>cs</sub>, which repair the DSB (Featherstone and Jackson, 1999). *MRE11* forms a complex with *RAD50*, and either *XRS2* in yeast, or *NBS1* in mammals (D'Amours and Jackson, 2002). This complex is required for the nucleolytic processing of DSBs in HR and possibly NHEJ, and is thought to play a structural role which may facilitate both strand exchange and end-joining reactions (Furuse *et al.*, 1998; Paull and Gellert, 1998; Paull and Gellert, 2000; Hopfner *et al.*, 2002). The *MRE11* complex has also been associated with DNA damage sensing and signalling activities (D'Amours and Jackson, 2002; Hopfner *et al.*, 2002). *KU70* mutant trypanosomes showed no increase in sensitivity to the cytotoxic effects of MMS compared with wild-type cells, indicating that NHEJ plays a minor role in DSB repair in the parasite (Conway *et al.*, 2002a). Similarly, *MRE11* mutant trypanosomes showed no increased sensitivity to MMS compared to wild-type cells, but did display an increase in sensitivity to the radiomimetic drug bleomycin, which

predominantly causes single- and double-strand DNA breaks, suggesting that *MRE11* is involved in DSB repair (Robinson *et al.*, 2002). However, given that *MRE11* only appears to be involved in the repair of lesions created after treatment with bleomycin, but not MMS, it appears that *MRE11* is only involved in the repair of a sub-set of DNA damage, whereas *RAD51* has a more widespread function. Large scale genome rearrangements were also detected in *MRE11* mutants after prolonged passage *in vitro*. Furthermore, *MRE11* mutation has been shown to have a direct effect on recombination, as a 3- to 4-fold reduction in the frequency of integration of homologous targeting constructs was observed in *MRE11* mutants compared with wild-type cells. This reduction in integration efficiency is approximately half that seen in *RAD51* mutant trypanosomes (McCulloch and Barry, 1999). In addition, *MRE11* mutants were seen to integrate the DNA constructs into two available homologous locations, a different pattern of recombination compared with wild-type cells, but comparable to *RAD51* mutant cell lines (McCulloch and Barry, 1999; Robinson *et al.*, 2002). A possible explanation for these observations is that resection of the ends of DSBs occurs much less efficiently in *MRE11* mutant trypanosomes, so the 3' single-stranded tails generated by this process are shorter than in wild-type cells, perhaps forcing *MRE11* mutant cells to repair the break using the RAD51-independent homologous recombination pathway (Conway *et al.*, 2002b). Despite these alterations to recombination in trypanosomes, *MRE11* mutants, like *KU70* mutants, displayed no changes in either the frequency of, or mechanisms used, during *VSG* switching (Conway *et al.*, 2002a; Robinson *et al.*, 2002). It remains unclear what this tells us about either the initiating event in *VSG* switching, or the recombination pathway that is used.

Figure 1.8 Pathways of double-strand break (DSB) repair



## 1.5 Mismatch repair in eukaryotes

Given the conflicting evidence of RAD51 and MRE11 involvement in *VSG* switching, it remains unclear exactly what form of *VSG* recombination reaction(s) are used. The involvement of RAD51 is compatible with the molecular data suggesting that homologous recombination is involved. Nevertheless, the MRE11 data illustrate that we still need to understand the genetic controls that operate on this process. The post-replicative mismatch repair system is an important metabolic pathway first described in prokaryotes (Cox, 1976). This mechanism functions mainly to recognise and correct base-base mispairs and misalignments that arise as a result of DNA damage or replication errors, but significantly it also regulates homologous recombination, at least between divergent DNA sequences. This makes it an attractive system to examine in the context of *T. brucei* *VSG* switching.

Although the best-researched mismatch repair (MMR) systems are the MTHLS-dependent pathways of *Escherichia coli* and *Salmonella typhimurium* (Modrich, 1991), work in eukaryotes has revealed extensive sequence conservation between the functional homologues of the principal polypeptides involved (Eisen, 1998; Culligan *et al.*, 2000). Such homology suggests the fundamental components of MMR have been conserved throughout evolution and adapted to the requirements of increasingly complex genomes (Fishel and Kolodner, 1995; Fishel and Wilson, 1997). These major components fall into two main categories: MutS homologues (MSH) and MutL homologues (MLH), which are related to the bacterial MutS and MutL proteins respectively (Figure 1.9). MutS family members typically exhibit 20-30% amino acid identity and the family has been divided into two main lineages on the basis of phylogenetic analysis of the highly conserved ATP binding domain: MutS-I, which encompasses the proteins involved in MMR; and MutS-II, where the eukaryotic proteins are required during meiotic crossing over and chromosome segregation (the bacterial homologues have yet to be assigned a function; Eisen, 1998). MutS-I homologues generally possess four conserved subdomains: a putative ATP binding site (Gorbalenya and Koonin, 1990) and a helix-turn-helix motif towards the C-terminal end (Biswas *et al.*, 2001), as well as a middle conserved domain (Culligan *et al.*, 2000) and an N-terminal mismatch interacting domain (Malkov *et al.*, 1997; Culligan *et al.*, 2000; Schofield *et al.*, 2001). MutS-II proteins also possess the ATP-binding site and helix-turn-helix motif, and some possess the middle conserved domain; they lack the N-terminal domain, however (Eisen, 1998; Culligan *et al.*, 2000). MutL-related proteins also possess a putative ATP binding site found in a conserved region of approximately 300 residues at their amino-terminus (Bergerat *et al.*, 1997; Mushegian *et al.*, 1997; Ban and Yang, 1998;

Dutta and Inouye, 2000), while the rather diverse carboxy-terminal region is responsible for dimer formation (Pang *et al.*, 1997; Schär *et al.*, 1997; Kondo *et al.*, 2001).

The MutHLS or MMR system in *Escherichia coli* has been reviewed by Modrich, 1991; Fishel and Kolodner, 1995; Kolodner, 1996; Modrich and Lahue, 1996; Jiricny, 1998; Buermeyer *et al.*, 1999; Harfe and Jinks-Robertson, 2000 (Figure 1.10). In this system, the MutS homodimer recognises and binds base-base mismatches and small insertion-deletion loops (IDLs). The MutS-DNA complex is then bound by a homodimer of MutL. MutL is thought to function as a molecular matchmaker, interacting with and modulating the activities of other proteins involved in MMR. MutL recruits MutH, an endonuclease which nicks the unmethylated strand at a hemimethylated GATC site. Excision from the nick past the mismatch and resynthesis of the unmethylated strand to complete the repair can be bi-directional, and requires the co-ordinated activities of any of a number of single-strand-specific exonucleases including ExoI, ExoVII, ExoX or RecJ, as well as UvrD (MutU) helicase II, single-strand binding protein (SSB), DNA polymerase III holoenzyme and DNA ligase. The use of *dam* methylation to distinguish between parental and newly replicated DNA appears to be unique to gram negative bacteria, and since no homologues of MutH have been identified outside this group, other organisms must possess different mechanisms of signal transduction and strand discrimination (Jiricny, 1998; Buermeyer *et al.*, 1999; Kolodner and Marsischky, 1999; Harfe and Jinks-Robertson, 2000).

Although the model for MMR in eukaryotes is far from complete, it is thought to have evolved from an ancestral system also related to the MutHLS system in bacteria (Eisen, 1998; Culligan *et al.*, 2000). In *S. cerevisiae*, six MSH genes and four MLH genes have been described (Kolodner and Marsischky, 1999). Furthermore, there has been extensive progress in identifying and understanding the proteins involved in mammalian MMR since the discovery that hereditary nonpolyposis colorectal carcinoma (HNPCC), a common form of inherited cancer predisposition syndrome, is caused by mutations in some of the genes encoding components of the MMR system (Peltomäki *et al.*, 1997).

Of the six MSH genes found in *S. cerevisiae*, MSH2-6 encode proteins that act on nuclear DNA, and homologues of each have been found in a large array of eukaryotes. However, MSH1, which is required for the stability of the mitochondrial genome, has so far only been identified in *S. cerevisiae*, *Schizosaccharomyces pombe* and *Arabidopsis thaliana* (Reenan and Kolodner, 1992a; Reenan and Kolodner, 1992b; Culligan and Hays, 2000). Of the five nuclear homologues, three (MSH2, MSH3 and MSH6) are involved in mismatch repair (Johnson *et al.*, 1996b; Marsischky *et al.*, 1996), whereas the remaining two (MSH4

and MSH5) facilitate crossing-over during meiotic recombination (Ross-Macdonald and Roeder, 1994; Hollingsworth *et al.*, 1995). MSH1, MSH2, MSH3 and MSH6 are members of the MutS-I lineage whereas MSH4 and MSH5 have been classified as MutS-II proteins (Eisen, 1998). Orthologues of MSH2, MSH3 and MSH6 have been described in both mammals and plants, and in fact, plants possess an extra MutS homologue, MSH7, which also appears to function in MMR (Culligan and Hays, 2000; Dong *et al.*, 2002).

In contrast to bacteria, eukaryotic MutS homologues have evolved to form heterodimers (with the exception of MSH1; Nakagawa *et al.*, 1999). An explanation for this may have been uncovered by structural studies of MutS homodimers in *E. coli* and *Thermus aquaticus* which revealed that while the protein subunits are identical, together they form structural heterodimers by adopting different conformations upon mismatch binding (Lamers *et al.*, 2000; Obmolova *et al.*, 2000). In eukaryotes, the possession of multiple MutS-I homologues allowed the evolution of distinct heterodimers which may have broadened the range of mismatches the cell is able to detect and repair. In yeast and mammals, MSH2 forms complexes with MSH3 and MSH6 (Figure 1.11; Acharya *et al.*, 1996; Iaccarino *et al.*, 1996; Marsischky *et al.*, 1996). These heterodimers are also present in plants, and are joined by a third as MSH2 also interacts with MSH7 (Culligan and Hays, 2000). Interactions between MSH4 and MSH5 have also been documented in yeast (Pochart *et al.*, 1997), but discussion of the activities of this heterodimer is outwith the scope of this work.

Evidence from *in vitro* binding assays and mutation spectra suggest that MSH2 is necessary for the repair of all types of mismatches. The MSH2-MSH6 heterodimer (also known as the MutS $\alpha$  complex in mammals and plants) is believed to recognise and bind base-base mispairs preferentially, while the MSH2-MSH3 complex (MutS $\beta$  in mammals and plants) is primarily involved in the repair of IDLs (Acharya *et al.*, 1996; Habraken *et al.*, 1996; Johnson *et al.*, 1996b; Marsischky *et al.*, 1996; Sia *et al.*, 1997b; Earley and Crouse, 1998; Genschel *et al.*, 1998). In plants, the MSH7 protein is closely related to MSH6 and forms the heterodimer MSH2-MSH7, or MutS $\gamma$  (Culligan and Hays, 2000). *In vitro* studies indicate that the spectrum of mismatches recognised by MSH2-MSH7 may be similar to that of MSH2-MSH6. However, the affinity of the MSH2-MSH7 heterodimer for the mismatches tested was lower than that of the MSH2-MSH6 heterodimer for the same mispairs. This may suggest that MSH2-MSH7 is required for the repair of specialised DNA lesions or mispairs in specialised contexts (Culligan and Hays, 2000). The *in vivo* activities of MSH2-MSH7 have yet to be investigated. MSH2-MSH6

and MSH2-MSH3 appear to be partially redundant in function as they compete for the repair of small 1- to 2-nucleotide loops (Acharya *et al.*, 1996; Johnson *et al.*, 1996b; Marsischky *et al.*, 1996; Greene and Jinks-Robertson, 1997; Genschel *et al.*, 1998; Culligan and Hays, 2000). Furthermore, MSH2-MSH6 is considered to be the predominant complex during post-replicative mismatch repair, as studies of the protein levels within MMR-proficient human cells revealed it to be in excess over MSH2-MSH3 (Drummond *et al.*, 1997; Genschel *et al.*, 1998; Marra *et al.*, 1998). Such studies have yet to be performed in yeast and plants.

As previously stated, the eukaryotic MutS homologues involved in nuclear MMR possess four conserved domains. The best studied of these is the highly conserved ATP-binding domain which contains the four motifs characteristic of the ABC ATPase superfamily (I, II, III and IV; Gorbalenya and Koonin, 1990; Holland and Blight, 1999). In this superfamily, motif I corresponds to the 'Walker A' consensus sequence which is required to bind the phosphoryl moiety of ATP or GTP (also known as the P-loop), and motif III corresponds to the 'Walker B' site which is required for Mg<sup>2+</sup> binding (Walker *et al.*, 1982; Gorbalenya and Koonin, 1990). Purified bacterial, yeast and human MutS homologues have been shown to possess an intrinsic low level ATP hydrolysis (ATPase) activity (Haber and Walker, 1991; Chi and Kolodner, 1994; Alani *et al.*, 1997b; Gradia *et al.*, 1997) which is essential for the correct functioning of these proteins in MMR. Genetic analyses of *mutS*-related genes from bacteria and *S. cerevisiae* revealed that mutations within the P-loop resulted in decreased *in vitro* ATPase activity, and over-expression of the mutant alleles led to dominant negative mutator phenotypes (Haber and Walker, 1991; Alani *et al.*, 1997b; Iaccarino *et al.*, 1998; Studamire *et al.*, 1998). In *in vitro* DNA binding assays, addition of ATP to reactions containing mismatched substrates abolished MSH2-MSH6 mismatch binding (Drummond *et al.*, 1995; Gradia *et al.*, 1997). This ATP-dependent mismatch release probably results from the translocation of the heterodimer away from the mismatch and off the free ends of the bound DNA molecule. In similar experiments, where the ends of the DNA substrates were physically blocked by biotin-streptavidin complexes, or indeed where the DNA substrates were circular, this ATP-induced release of MSH2-MSH6 was prevented (Gradia *et al.*, 1999). It has been suggested that MSH2-MSH6 functions as a molecular switch similar to G protein signalling pathways (Gradia *et al.*, 1997; Fishel, 1998; Gradia *et al.*, 1999; Fishel *et al.*, 2000). In this model, the MSH2-MSH6 heterodimer hydrolyses ATP to form the mismatch-binding competent ADP-bound form. Recognition of a mismatch induces ADP-ATP exchange which results in the formation of a freely diffusible clamp which is stably associated with DNA. The clamp dissociates from the mismatch and is capable of

rapidly diffusing away in either a 5' or 3' direction, allowing a further MSH2-MSH6 complex to bind the mismatch. This hypothesis accounts both for the bidirectionality of MMR, and affords a possible mechanism of signal transduction to activate the repair machinery. The target of the MSH2-MSH6 sliding clamp is unknown, but both proliferating cell nuclear antigen (PCNA) and exonuclease 1 (EXO1) are attractive candidates (see below; Gradia *et al.*, 1999). It is likely that the other eukaryotic MutS-related heterodimers involved in MMR function in a similar manner given the high level of conservation of their ATP-binding domains.

The second conserved domain found towards the C-terminus is the helix-turn-helix motif. This region is required for dimerisation of the MutS homologues as revealed by structural studies of bacterial MutS (Lamers *et al.*, 2000; Obmolova *et al.*, 2000). Mutation of the helix-turn-helix motif of both MutS and MSH2 not only leads to loss of dimerisation but also loss of mismatch binding and ATP hydrolysis (Alani, 1996; Alani *et al.*, 1997b; Biswas *et al.*, 2001). This can be explained by the structural data which reveals that both the ATP-binding site and the DNA binding site of MutS are composed of domains derived from both subunits (Lamers *et al.*, 2000; Obmolova *et al.*, 2000). If this is also the case in eukaryotes, disruption of dimerisation would also affect both the ATPase and mismatch recognition activities of the MSH2-containing heterodimers.

The middle conserved domain is the third region of conservation found in the MutS homologues (Culligan *et al.*, 2000). Three dimensional structural analysis of *Homo sapiens* MSH2 indicates that this domain lies on the surface of the molecule (de Las Alas *et al.*, 1998), and a dominant negative phenotype is observed upon mutation of a highly conserved arginine in this region of *E. coli* MutS (R305H) (Wu and Marinus, 1994). While the middle domain is probably not directly involved in heterodimer formation (Alani, 1996; Guerrette *et al.*, 1998), it may have a role in mediating interactions with other components of the MMR system, for instance members of the MutL family of proteins.

The final conserved region found in MutS homologues involved in MMR is the N-terminal mismatch interacting domain. Within the N-terminal domain of *E. coli* MutS and *T. aquaticus* MutS, there is a conserved Phe-X-Glu DNA binding motif, where the glutamic acid residue hydrogen bonds with either an unpaired thymidine or the thymidine of a G-T mismatch (Lamers *et al.*, 2000; Schofield *et al.*, 2001). This motif is also present in MSH6 and MSH7 homologues from *A. thaliana*, *H. sapiens* and *S. cerevisiae*, suggesting this motif is common to proteins binding base-base mismatches and small IDLs

(Acharya *et al.*, 1996; Marsischky *et al.*, 1996; Culligan and Hays, 2000; Culligan *et al.*, 2000). MSH3 orthologues, which are involved in the repair of larger IDLs do not possess the conserved Phc-X-Glu motif, although they do show sequence conservation within this region (Culligan *et al.*, 2000). A similar situation is observed within the MSH2 orthologues.

Analysis of genetic and protein-protein interactions suggests that eukaryotic MMR is fundamentally similar to the bacterial system, and that a heterodimer of MutL homologues interacts with an MSH2-containing heterodimeric complex bound to mismatched DNA (Figure 1.12) (Jiricny, 1998; Buermeyer *et al.*, 1999; Kolodner and Marsischky, 1999; Nakagawa *et al.*, 1999; Harfe and Jinks-Robertson, 2000). Four genes possessing the regions of homology characteristic of MutL-family members have been cloned in *S. cerevisiae*; these are designated MLH1, MLH2, MLH3 and PMS1 (named for the phenotype of post-meiotic segregation). Four MutL homologues have so far been described in mammals: MLH1, MLH3, PMS1 and PMS2. The closest mammalian homologue of *S. cerevisiae* PMS1 is PMS2, but it is unclear as to whether mammalian PMS1 is more closely related to *S. cerevisiae* PMS1 or MLH2 (Lipkin *et al.*, 2000). Currently only MLH1 has been described in *A. thaliana*, although studies of MMR in plants are still in their infancy (Jean *et al.*, 1999).

MLH1 has been identified as the common subunit of three MutL-related complexes in yeast, associating with PMS1, MLH2 and MLH3 (see Figure 1.11) (Wang *et al.*, 1999). In yeast, *MLH1* mutants, *MLH1 PMS1* double mutants and *MSH2* mutants all exhibit a similar level of phenotypic deficiency in assays examining the repair of replication errors, indicating that the MLH1 and PMS1 proteins are necessary during the repair of such errors (Strand *et al.*, 1993; Prolla *et al.*, 1994a; Greene and Jinks-Robertson, 1997). MLH1-PMS1 (MLH1-PMS2 in mammals, where it is also called MutL $\alpha$ ) is known to form mismatch-dependent ternary complexes with both MSH2-MSH6 and MSH2-MSH3 and so plays an important role in the repair of both base-base mispairs and IDLs (Habracken *et al.*, 1997; Habracken *et al.*, 1998; Blackwell *et al.*, 2001). The MLH1-MLH3 heterodimer is reported to interact with MSH2-MSH3 only, probably to promote the repair of specific insertion-deletion mispairs (Flores-Rozas and Kolodner, 1998), but it is also thought to complex with the MSH4-MSH5 heterodimer to facilitate crossing-over during meiosis (Wang *et al.*, 1999). The MLH1-MLH2 complex has only recently been described, and *MLH2* mutants are resistant to cisplatin and related anticancer drugs thereby implying a role in processing some types of DNA damage (Durant *et al.*, 1999).

As for MutL in *E. coli*, the MLH-related heterodimers in eukaryotes are believed to act as "coupling" factors in the assembly of higher order complexes during MMR, or indeed during meiotic crossing-over (Buermeier *et al.*, 1999). Consistent with this idea are the findings that MLH1 interacts with proteins which are thought to be involved in the downstream processing and repair of mismatches, including PCNA and EXO1 (see below). It is also probable that the MutL homologues enhance mismatch recognition, as it is known that MutL increases the efficiency of mismatch recognition by MutS (Drotschmann *et al.*, 1998) and similar results have been reported for yeast MLH1-PMS1 (Habraken *et al.*, 1997).

As described above, MutL-related proteins possess two characteristic conserved domains, an ATP-binding domain and a dimerisation domain. The ATP-binding domain is highly conserved between the MutL homologues and is a member of the recently described GHKL ATPase/kinase superfamily (Bergerat *et al.*, 1997; Mushegian *et al.*, 1997; Ban and Yang, 1998; Dutta and Inouye, 2000). The best researched member of the MLH family, *E. coli* MutL, possesses a weak ATPase activity which is required to transduce the signal from MutS to factors which effect the repair of the mismatch (Aronshtam and Marinus, 1996; Ban and Yang, 1998; Ban *et al.*, 1999; Spampinato and Modrich, 2000). Weak ATPase activities have also been demonstrated in *S. cerevisiae* MLH1 and PMS1 and amino acid substitutions within the conserved ATP-binding motifs of these proteins result in a reduction in ATP-binding and hydrolysis and lead to mutator phenotypes *in vivo* (Pang *et al.*, 1997; Tran and Liskay, 2000; Hall *et al.*, 2002). Further to this, it has been demonstrated that *S. cerevisiae* MLH1 binds ATP more efficiently than PMS1, suggesting that MLH1 binds ATP first, resulting in a conformational change allowing PMS1 to bind ATP, and allowing the heterodimer to modulate interactions with other MMR components (Hall *et al.*, 2002).

The second domain present in MutL homologues is required for dimerisation and is found in the C-terminal portion of these proteins (Pang *et al.*, 1997; Guerrette *et al.*, 1999; Kondo *et al.*, 2001). This domain is poorly conserved between the paralogous MutL-related protein sub-families, but well-conserved regions can be identified between orthologues (Pang *et al.*, 1997; Lipkin *et al.*, 2000). For instance, yeast PMS1 shares the sequence PWNCPHGRPTMRH with its orthologue human PMS2, but not with its paralogues, yeast MLH1 or human PMS1 (Pang *et al.*, 1997). Deletion studies show that proteins which lack this domain cannot interact with their normal MutL-related binding partners (Pang *et al.*, 1997; Guerrette *et al.*, 1999; Kondo *et al.*, 2001). Furthermore, mutational studies of the dimerisation domains of human MutL homologues revealed that a number of mutations

reported in HNPCC kindreds caused a strong reduction in heterodimerisation, explaining the MMR-deficiency in these patients (Guerrette *et al.*, 1999).

A final conserved motif, which is present only in orthologues of MLH1, has been termed the carboxy-terminal homology (CTH) motif (Pang *et al.*, 1997). This domain is essential for mismatch repair function in *S. cerevisiae* MLH1, and mutations in this motif resulted in strong mutator phenotypes, but did not affect interaction with PMS1 (Pang *et al.*, 1997). However, the exact role of this motif in the functioning of MLH1 has not yet been elucidated.

The proteins involved in DNA strand discrimination, and excision and resynthesis of the nascent strand have yet to be fully characterised in eukaryotes (Figure 1.12), but it is expected that many of the activities required to complete MMR in bacteria will also be necessary in eukaryotes. As described above, in *E. coli* strand discrimination is achieved by the endonuclease MutH, which introduces a nick in the nascent strand at hemi-methylated GATC sequence (Modrich and Lahue, 1996). Eukaryotes do not possess *dam* methylation, but it is hypothesised that single-strand breaks and strand discontinuities created during replication are utilised to distinguish the nascent strand, since *in vitro* repair of mismatched plasmid substrates containing a single-strand break occurs on the nicked strand (Holmes *et al.*, 1990; Thomas *et al.*, 1991; Buermeier *et al.*, 1999). PCNA, an essential factor in DNA replication, forms a ring shaped trimeric complex which encircles DNA and tethers DNA polymerase to the template (Warbrick, 2000). It also functions as a sliding platform which can mediate interactions between several different proteins at once, as well as with DNA (Jonsson and Hubscher, 1997; Kelman, 1997; Kelman and Hurwitz, 1998). There is mounting evidence to suggest that PCNA may also mediate interactions between the mismatch recognition complex and the replication machinery, thereby facilitating strand discrimination (Figure 1.12). Consistent with this suggestion is the finding that human MSH3 and MSH6 colocalise with PCNA to replication foci (Kleczkowska *et al.*, 2001). Moreover, both yeast and human MSH3 and MSH6 have been shown to interact with PCNA as part of a heterodimer with MSH2 (Johnson *et al.*, 1996a; Clark *et al.*, 2000; Flores-Rozas *et al.*, 2000; Kleczkowska *et al.*, 2001), and conserved PCNA-binding sites have been found at the extreme N-termini of MSH3 and MSH6 homologues from yeast and humans (Clark *et al.*, 2000; Warbrick, 2000). This motif is also present in MSH3, MSH6 and MSH7 from *A. thaliana*, although it is absent from the MSH2, MLH1 and PMS1 orthologues in all these species (Kleczkowska *et al.*, 2001). It has also been shown that interaction with PCNA is strongly reduced where the PCNA binding sites of MSH3 and MSH6 contain mutations (Clark *et al.*, 2000;

Kleczkowska *et al.*, 2001), and that interaction with this replication factor enhances the mispair binding of the MSH2-MSH6 heterodimer (Flores-Rozas *et al.*, 2000). The importance of these interactions for functional mismatch repair is underlined by *in vivo* studies of the yeast MSH3 and MSH6 proteins which revealed that mutations in the PCNA binding site led to a mutator phenotype (Clark *et al.*, 2000; Flores-Rozas *et al.*, 2000). It has been suggested that these phenotypes arise due to defects in the strand discrimination process, an idea supported by the finding that some mutations in *S. cerevisiae pol30* (the gene encoding PCNA in this organism) cause mutator phenotypes epistatic to MMR-inactivating mutations (Chen *et al.*, 1999). Interaction with PCNA is not limited to the MutS homologues, however, as MLH1 has also been shown to bind PCNA, although the region of the MLH1 protein involved in this interaction has not been identified (Umar *et al.*, 1996). Furthermore, a complex containing MSH2, MLH1, PMS2 and PCNA was co-immunoprecipitated in the presence of DNA and ATP from nuclear extracts of MMR-proficient human cells, but not from MLH1-deficient human cells (Gu *et al.*, 1998). Recently, it has been shown that ternary complexes composed of MSH2-MSH6 and MLH1-PMS1, while stable in the presence of mismatched DNA and ATP, are disrupted by the addition of PCNA, indicating that this protein not only interacts with these complexes, but may induce translocation of the complex away from the mismatch in order to recruit downstream factors required for mismatch correction (Bowers *et al.*, 2001). In addition to the role of PCNA during the early stages of MMR, Gu *et al.* (1998) suggested that this replication factor might also be required during the DNA resynthesis step, as unrepaired molecules and repair intermediates that were blocked at the step of DNA resynthesis by limiting amounts of PCNA were converted to full-length products in the presence of excess PCNA. While these data suggest a central role for PCNA in MMR, evidence indicates that other replication factors such as replication protein A (RPA) and replication factor C (RFC) are also required (Lin *et al.*, 1998; Xie *et al.*, 1999). Indeed, recently it has been suggested that RPA may be required to protect single-stranded regions of the template DNA strand from nuclease degradation (Ramilo *et al.*, 2002), thereby ensuring that the template strand is suitable for repair DNA synthesis and echoing one of the roles of single-strand binding protein (SSB) in *E. coli* (Meyer and Laine, 1990).

As in bacteria, a number of exonucleases have also been implicated in eukaryotic MMR. EXO1 has been reported to interact with MSH2, MSH3 and MLH1 *in vitro*, and studies show that EXO1 and the MutL $\alpha$  complex can be co-immunoprecipitated from human extracts (Tishkoff *et al.*, 1997; Schmutte *et al.*, 1998; Schmutte *et al.*, 2001). Mutations in EXO1 caused a weak mutator phenotype in both budding and fission yeast and analysis

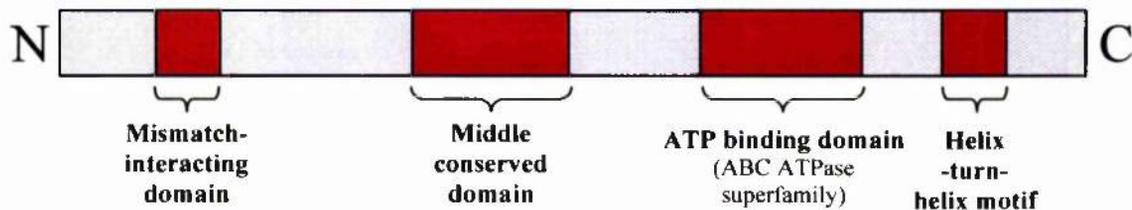
showed that MSH2 was epistatic over EXO1 (Tishkoff *et al.*, 1997; Rudolph *et al.*, 1998). It has also been suggested that the 3'-5' exonuclease functions of both DNA polymerases delta (Pol  $\delta$ ) and epsilon (Pol  $\epsilon$ ) are partially redundant with the 5'-3' activity of EXO1, as combined inactivation of these three exonuclease activities give rise to a mutation rate similar to that seen in MMR-deficient yeast cells (Tran *et al.*, 1999). Further to this, human extracts depleted of Pol  $\delta$  are defective in the repair DNA synthesis step of mismatch correction, but this defect is alleviated by addition of this enzyme, implicating this DNA polymerase in the reaction (Longley *et al.*, 1997). It is also noteworthy that Pol  $\delta$  has been reported to interact with PCNA (Prelich *et al.*, 1987), again suggesting an association between the MMR system and the replication machinery.

Our understanding of the eukaryotic MMR system still remains incomplete (Figure 1.12), and, as yet, no eukaryotic helicase(s) or DNA ligase(s) have been implicated in MMR. It is possible that a number of helicases can function in eukaryotic MMR, and that this functional redundancy has masked the identities of these components. Another possibility is that eukaryotic MMR may involve exonucleases which work on double-stranded DNA (for instance EXO1), thereby removing the need for helicase activity (Harfe and Jinks-Robertson, 2000). It is also possible that functional redundancy has prevented the identification of the ligase which completes the MMR reaction, although the most attractive candidate is DNA ligase I (Marra and Schär, 1999).

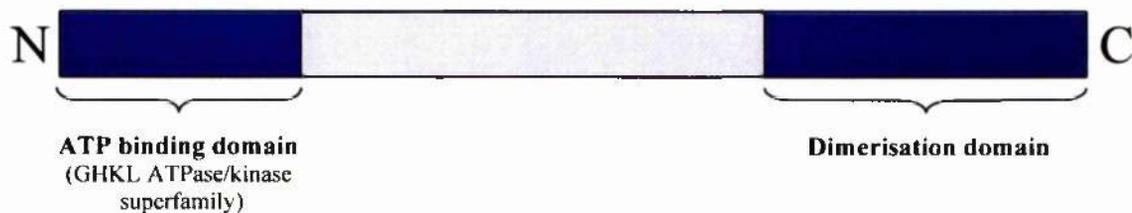
**Figure 1.9 The proteins involved in the post-replicative mismatch repair (MMR) pathway**

a) The MutS homologues involved in MMR possess four conserved subdomains (shown in red): a mismatch interacting domain (Malkov *et al.*, 1997; Schofield *et al.*, 2001), a middle conserved domain (Culligan *et al.*, 2000), an ATP binding domain (Gorbalenya and Koonin, 1990) and a helix-turn-helix motif (Biswas *et al.*, 2001). b) MutL-related peptides contain a highly conserved ATP binding domain (Bergerat *et al.*, 1997; Mushegian *et al.*, 1997; Ban and Yang, 1998; Dutta and Inouye, 2000) and a more diffuse region of conservation required for dimer formation (shown in blue; Pang *et al.*, 1997; Schär *et al.*, 1997; Kondo *et al.*, 2001). The amino- and carboxy-termini are designated N and C respectively.

**a) MutS homologues**

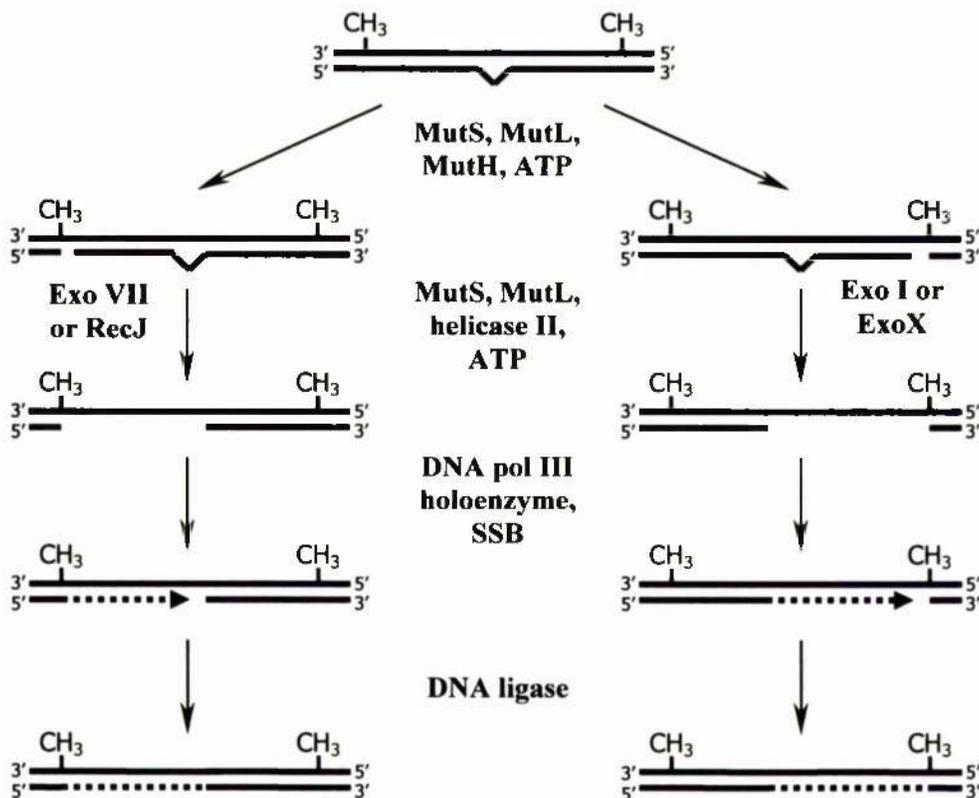


**b) MutL homologues**



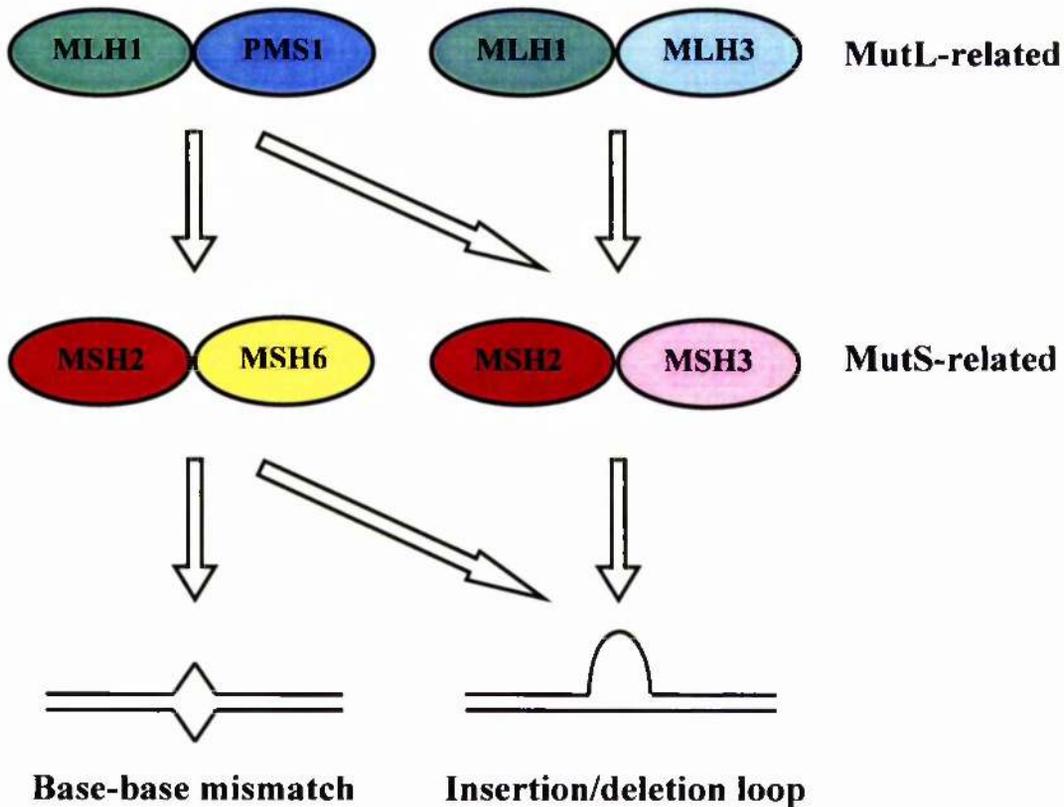
**Figure 1.10 The mechanism of *E. coli* methyl-directed MMR**

MMR is initiated when a MutS homodimer binds a mismatch. The subsequent binding of the MutL homodimer is required to recruit the MutH endonuclease, which nicks the newly-replicated unmethylated strand at a hemi-methylated GATC site, located either 5' or 3' of the mismatch. This is followed by an excision step which requires the activity of helicase II to displace the newly-replicated DNA strand, as well as one of a number of single-stranded exonucleases to digest the displaced strand leaving a gap in the DNA. Excision from the nick to the mismatch may occur in either a 5' to 3' direction, performed by ExoVII or RecJ, or a 3' to 5' direction, using ExoI or ExoX. Resynthesis to fill in the resulting gap requires the DNA polymerase III holoenzyme complex, SSB and DNA ligase. In this system, MMR is coupled to DNA replication, so that mismatches formed during DNA replication can be repaired using the methylated parental strands as templates, thereby reducing misincorporation errors. (Adapted from Grilley *et al.*, 1993)



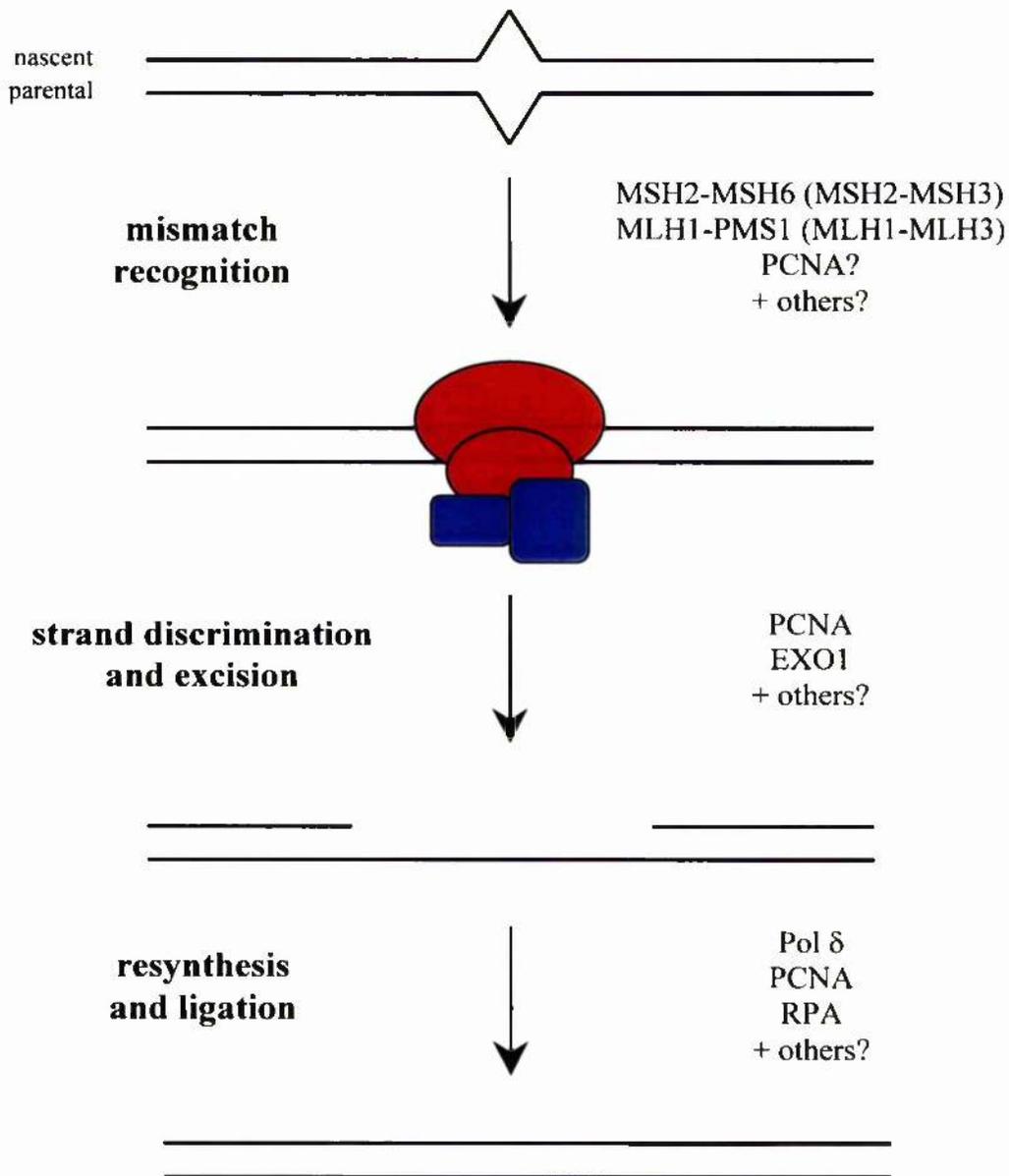
**Figure 1.11 Interactions between MutS- and MutL-related proteins during mismatch repair in *Saccharomyces cerevisiae***

Depending on the type of mismatch, the distorted heteroduplex is bound by either MSH2-MSH6, in the case of a base-base mispair or small insertion/deletion loop (IDL), or MSH2-MSH3 where a larger IDL is detected. The MSH2-containing heterodimer/DNA complex is then bound by a MutL-related heterodimer; in yeast this is usually MLH1-PMS1, however, some specific IDLs may induce binding of MLH1-MLH3 to the MSH2-MSH3 complex. (Adapted from Kolodner and Marsischky, 1999)



**Figure 1.12 A scheme for eukaryotic mismatch repair**

Initially, a mismatch is recognised by an MSH2-containing heterodimer (shown in red), which, in turn, is bound by an MLH1-containing heterodimer (shown in blue). Formation of this ternary-complex is believed to enable the recruitment of factors required for strand discrimination and excision of the mismatch, followed by resynthesis and ligation of the nascent strand. A number of the components required during this process have yet to be identified. (Adapted from Buermeyer *et al.*, 1999)



## 1.6 Mismatch repair in DNA recombination

Recombination between divergent (homeologous) DNA sequences has been found to occur at a reduced frequency when compared with recombination between homologous sequences (Evans and Alani, 2000; Harfe and Jinks-Robertson, 2000). During the formation of heteroduplex joints between independent but complementary single-stranded parental DNA molecules, mispaired bases arise where the two single strands differ in sequence. Where very high levels of sequence divergence are encountered, stably base paired strand exchange intermediates seem to be unable to form, thereby preventing recombination. However, even a single mismatched base-pair serves to inhibit recombination, indicating that an active anti-recombination system also exists (Datta *et al.*, 1996; Datta *et al.*, 1997). MutS- and MutL-related proteins, interacting with mispaired bases in strand exchange intermediates, form the basis of this activity, since the frequency of recombination between homeologous substrates is often greatly elevated in MMR deficient cell lines (Shen and Huang, 1989; Alani *et al.*, 1994; Selva *et al.*, 1995; Datta *et al.*, 1996; Datta *et al.*, 1997; Evans and Alani, 2000). During heteroduplex formation mismatches are processed by components of the MutHLS system in bacteria, and evidence from yeast and mammalian cells suggests that the barrier to recombination imposed by mismatch repair is conserved throughout evolution (Rayssiguier *et al.*, 1989; de Wind *et al.*, 1995; Selva *et al.*, 1995; Datta *et al.*, 1996).

One function of the MMR anti-recombination activity is likely in speciation, acting as a genetic barrier that prevents recombination between the genomes of closely related species (Rayssiguier *et al.*, 1989; Hunter *et al.*, 1996; Štambuk and Radman, 1998; Vulic *et al.*, 1999). For example, the bacterial species, *E. coli* and *Salmonella typhimurium*, cannot normally undergo successful mating. If MMR mutants are crossed, however, this barrier is alleviated (Rayssiguier *et al.*, 1989). In fact, over 1% of cells in wild populations of pathogenic *E. coli* or *Salmonella enterica* may be MMR-deficient, posing the question whether these cells enhance transmission of antibiotic resistance and virulence genes between bacterial populations through interspecies mating (LeClerc *et al.*, 1996). The anti-recombination activities of MMR may also enhance genome stability by inhibiting gene amplification, recombination between diverged repetitive sequences, and gross chromosomal rearrangements, thereby protecting against tumour formation in mammals (Harfe and Jinks-Robertson, 2000; Lin *et al.*, 2001; Myung *et al.*, 2001; Pastink *et al.*, 2001).

Two general models have been proposed to explain how components of the MMR machinery act to down-regulate homeologous recombination (Rayssiguier *et al.*, 1989). The heteroduplex destruction model suggests that MMR-directed excision leads to the generation of multiple single-strand breaks which destroy the recombination intermediate. In contrast, the non-destructive heteroduplex rejection model suggests that the MMR proteins act to reverse the formation of mismatch-containing heteroduplex DNA. The finding that mitotic recombination between 94%-identical substrates generated gene conversion tracts which were significantly longer in MMR-defective yeast strains when compared with wild-type strains is consistent with either of these models (Chen and Jinks-Robertson, 1998; Nicholson *et al.*, 2000). However, it has been reported that MMR activity does not greatly affect the transformation efficiency of plasmids containing a preformed heteroduplex region into *E. coli*, arguing against the creation of DNA breaks by the heteroduplex destruction model (Westmoreland *et al.*, 1997). Further evidence supporting the heteroduplex rejection model includes the finding that *E. coli* MutS and MutL proteins reduce both the rate and extent of RecA-mediated branch migration *in vitro* when the recombining heteroduplexes contain mismatches (Worth *et al.*, 1994; Worth *et al.*, 1998). These experiments also indicated that inhibition of homeologous recombination required the ATPase activity of the MutS protein (Worth *et al.*, 1998). More recently, MutS and MutL have also been shown to inhibit RuvAB-dependent stimulation of RecA-promoted branch migration through regions containing sequence heterologies (Fabisiewicz and Worth, 2001).

The mechanism of heteroduplex rejection is unknown, although several explanations have been suggested. Firstly, either reverse branch migration or immediate resolution of the recombination intermediate may result from a direct interaction between MMR proteins and the recombination machinery upon detection of mismatch-containing heteroduplex DNA (Alani *et al.*, 1994). Alternatively, rejection of the donor strand could result from unwinding of mismatched heteroduplex molecules by an MMR-associated helicase activity (Claverys and Lacks, 1986; Zahrt and Maloy, 1997). This hypothesis is supported by evidence indicating that MMR regulates homeologous recombination in *E. coli* at two stages, including an early stage which requires the UvrD helicase and the RecBCD nuclease, and a late MutH-dependent stage (Štambuk and Radman, 1998).

As stated above, a single mismatch is sufficient to activate the MMR-dependent anti-recombination mechanism, and further mismatches have a cumulative negative effect on the frequency of recombination (Datta *et al.*, 1997; Chen and Jinks-Robertson, 1999). As the level of sequence heterology between the substrate molecules increases, the

probability of the recombination intermediate avoiding detection by the MMR system decreases, and where the sequences are diverged by several percent the presence of additional mismatches fails to induce further anti-recombination activity. Furthermore, the overall rates of recombination between identical DNA sequences are also increased in MMR-deficient cells, although exactly why this should be so is not clear (Datta *et al.*, 1997). One explanation might be that mispairs are formed during the early steps of homology searching, and these are resolved as longer strand exchange intermediates are formed.

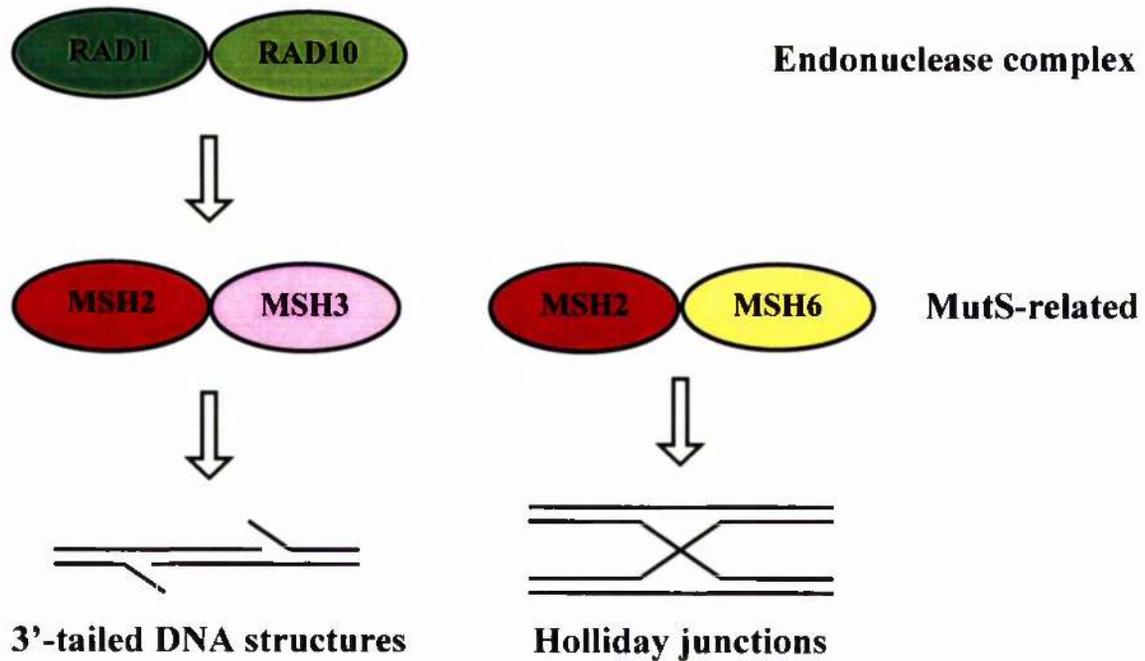
Studies of mitotic recombination in yeast have revealed that MSH2, MSH3, MSH6, MLH1, PMS1, RAD1, RAD10 and EXO1 all inhibit recombination in the presence of mismatches (Chen and Jinks-Robertson, 1999; Nicholson *et al.*, 2000). In general, the MSH2-MSH3 and MSH2-MSH6 heterodimers appear to recognise a similar spectrum of mismatches during both recombination and replication (Nicholson *et al.*, 2000). The greatest increases in homeologous recombination rates, relative to wild-type strains, are seen in *MSH2* mutants. *MSH6* mutants seem unable to inhibit recombination between sequences containing base-base mispairs and 1-nucleotide loops, while in *MSH3* mutants recombination rates increase between sequences containing IDLs of up to 18-nucleotides, but also, surprisingly, some types of base-base mispair. Disruption of *MLH1*, *PMS1* or *MLH1 PMS1* all lead to similar increases in homeologous recombination rates, suggesting that the peptides operate as a heterodimer (Nicholson *et al.*, 2000). However, recombination rates are lower in these backgrounds than in *MSH2* mutant strains, indicating that MSH2 is involved in blocking recombination by a pathway independent of these proteins, and that mismatch binding alone is sufficient to elicit some anti-recombination activity (Chen and Jinks-Robertson, 1999; Nicholson *et al.*, 2000). It should be noted, however, that the roles of MLH2 and MLH3 in repressing recombination between divergent sequences have yet to be assessed. Further evidence indicates that disruption of either *RAD1* or *RAD10* also increases homeologous recombination to rates similar to those seen in *MSH3* mutants (Nicholson *et al.*, 2000). Since other nucleotide excision repair (NER) proteins do not seem to affect recombination between divergent sequences, it is thought that the RAD1-RAD10 endonuclease is acting outside its role in NER. Finally, EXO1 has also been reported to play a minor role in the inhibition of homeologous recombination (Nicholson *et al.*, 2000). Epistasis analysis indicates that PMS1, RAD1 and EXO1 act in different pathways or complexes to inhibit homeologous recombination. Furthermore, while the effects of *MSH6* and *PMS1* deletion mutations were epistatic, the effects of *MSH3* and *PMS1* deletion mutations were not, suggesting that the anti-recombination activity of MSH3 does not rely on interaction with the

MLH1-PMS1 heterodimer and may involve other factors such as the RAD1-RAD10 endonuclease or EXO1 (Nicholson *et al.*, 2000). Current evidence suggests that the roles of MMR proteins in prevention of homeologous recombination are more complex than in post-replicative mismatch repair, and further research should elucidate the differences between these two mechanisms, leading to a more complete model for inhibition of homeologous recombination in eukaryotes.

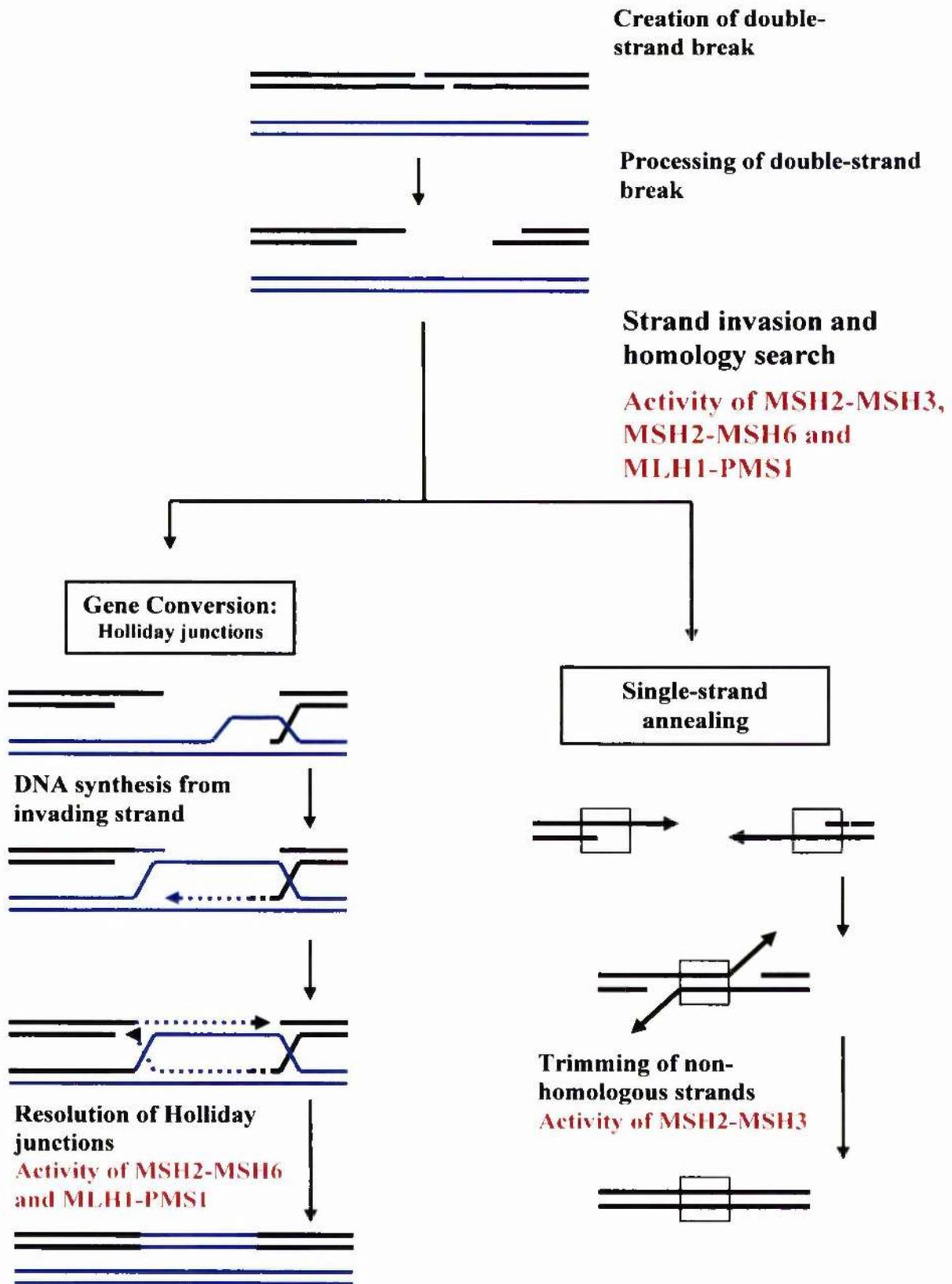
In addition to their roles in mismatch correction and homeologous recombination, the MMR proteins are also required to facilitate the processing of recombination intermediates during homologous recombination. Evidence shows that MSH2 can bind a variety of DNA structures, including Holliday junctions and Y-structures as well as a diverse array of mismatches (Alani *et al.*, 1997a; Sugawara *et al.*, 1997; Marsischky *et al.*, 1999). Moreover, when MSH2 is complexed with either MSH6 or MSH3 another level of specificity is added allowing different heterodimers to act upon different types of recombination intermediate (Figures 1.13 and 1.14). In *S. cerevisiae* the MSH2-MSH3 heterodimer, along with the RAD1-RAD10 endonuclease, is known to be required for the processing of non-homologous ends during gene conversion and single-strand annealing (Figure 1.13 and 1.14); MSH6, MLH1 and PMS1 are not required during these events (Saparbaev *et al.*, 1996; Sugawara *et al.*, 1997). Indeed, the yeast MSH2 protein has been reported to localise to sites near double-strand breaks possessing non-homologous ends *in vivo* (Evans *et al.*, 2000). Additional studies have suggested that MSH2 and RAD10 interact with each other directly (Bertrand *et al.*, 1998). It has been hypothesised that during this process MSH2-MSH3 binds to the non-homologous single-stranded DNA at the end of the annealed region, thus stabilising the recombination intermediate, and allowing the RAD1-RAD10 endonuclease to locate and cleave the non-homologous tail (Sugawara *et al.*, 1997). Finally, MSH2-MSH6 has been found to bind Holliday junctions with an affinity and specificity that is at least as high as it has for mispaired bases, as well as enhancing their cleavage by phage resolution enzymes (Figure 1.13 and 1.14; Marsischky *et al.*, 1999). The structural features which cause MSH2-MSH6 to recognise and bind Holliday junctions have yet to be elucidated.

**Figure 1.13** Activities of MutS-related proteins during homologous recombination in *Saccharomyces cerevisiae*

Different recombination intermediates have been found to interact with different mismatch repair components. The MSH2-MSH6 heterodimer has been shown to associate with Holliday junctions (Marsischky *et al.*, 1999). 3'-tailed DNA structures (created during single-strand annealing and gene conversion events) are believed to interact with the MSH2-MSH3 heterodimer which then recruits the RAD1-RAD10 endonuclease complex acting outside its role in the nucleotide excision repair pathway (Sugawara *et al.*, 1997).



**Figure 1.14** A scheme for DNA recombination and the roles of MutS and MutL homologues in *Saccharomyces cerevisiae*



## 1.7 Aims of this thesis

At the outset of this thesis it was unknown whether *T. brucei* possessed a functional MMR system, or whether MMR proteins regulated recombination between divergent sequences in this organism. However, it was hypothesised that, if present, the components of MMR might be involved in the regulation of antigenic variation in this organism. The primary aim of this thesis was to establish whether the *T. brucei* genome encodes components of a functional post-replicative nuclear MMR system, and to determine the roles of this system in both DNA recombination and antigenic variation in trypanosomes.

Initially we attempted to identify any highly conserved *mutS*- and *mutL*-related genes present in *T. brucei* using both molecular biology techniques, such as PCR amplification with degenerate primers, and a computational approach involving searching of the genome sequencing databases using the protein sequences of MutS and MutL family members from other organisms. Putative MMR components were identified only as gene fragments, and so cloning methods were needed to recover the complete genes in order to assign identities based on the similarity of the predicted polypeptide sequences to MutS- and MutL-related proteins from other eukaryotes.

Those genes encoding putative homologues of MSH2 and MLH1, which are, respectively, the primary components of the MutS- and MutL-related heterodimers in eukaryotes, were chosen for functional analysis. To this end, homozygous knockout mutants were generated for each gene, and their roles in MMR confirmed by examining their sensitivities to the alkylating agent *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine (MNNNG), and their ability to correct replication slippage events which result in a mutator phenotype termed microsatellite instability. To assess the roles of the putative *MSH2* and *MLH1* genes in trypanosome antigenic variation, a *VSG* switching assay described in McCulloch *et al.* (1997) was adopted. Finally, an assay to elucidate the role of MMR in trypanosome homologous and homeologous recombination was developed, and tested in both *MSH2* mutants and wild-type *T. brucei*.

# **CHAPTER 2**

## **Materials and Methods**

## **2.1 Trypanosome strains, cell culture and transformation**

### **2.1.1 Trypanosome strains**

Three *Trypanosoma brucei* cell lines were utilised in this study. Clones derived from *Trypanosoma brucei* MITat (Moltino Institute – *Trypanozoan* antigen type) 1.2a S427 were used in the studies on monomorphic bloodstream form trypanosomes. The ILTat (I.L.R.A.D. – International Laboratory for Research in Animal Diseases – *Trypanozoan* antigen type) 1.2 and TREU (Trypanosomiasis Research Edinburgh University) 927/4 cell lines were used solely in molecular studies.

### **2.1.2 Routine culture of trypanosomes**

*In vitro* cell culture of bloodstream form trypanosomes was performed using HMI-9 culture medium (Hirumi and Hirumi, 1989) at 37°C.

### **2.1.3 Transformation and cloning of trypanosomes**

Bloodstream form trypanosomes were transformed by electroporation of cells with approximately 5 µg of plasmid DNA that had been digested with the relevant restriction enzymes, phenol-chloroform extracted and then ethanol precipitated. Electroporation of  $5 \times 10^7$  cells was performed in 0.5 ml of Zimmerman postfusion medium (132 mM NaCl, 8 mM Na<sub>2</sub>HPO<sub>4</sub>, 1.5 mM KH<sub>2</sub>PO<sub>4</sub>, 0.5 mM magnesium acetate, 0.09 mM calcium acetate [pH 7.0]) supplemented with 1% glucose (ZMG). All electroporation of cells was performed using a Bio-Rad Gene Pulser II set at 1.4 kV and 25-µF capacitance (giving a delivered voltage of 1.40 to 1.60 kV and a time constant of 0.40 to 0.60 ms). Cells were allowed to recover for 18 h in HMI-9 before transformants were selected using the relevant antibiotic. Transformants were either selected on semi-solid agarose plates (Carruthers and Cross, 1992), where  $1 \times 10^7$  cells were spread on plates containing the relevant antibiotic, or  $1 \times 10^7$  or  $5 \times 10^6$  cells were diluted into 18 ml HMI-9, containing the relevant antibiotic(s) and spread in 1.5 ml samples over 12 wells of a 24-well tissue culture plate. Typically, transformants grew out after 7 to 14 days.

Bloodstream form trypanosomes were cloned in one of two ways. Firstly, by plating 100 cells on semi-solid agarose plates (Carruthers and Cross, 1992), containing the relevant antibiotic(s). Clones typically grew out after 7 to 14 days. Secondly by diluting 20 to 30

cells in 20 ml of HMI-9 containing the relevant antibiotic(s), and plating 200  $\mu$ l samples over a 96-well culture dish. Clones typically grew out after 5-7 days.

#### 2.1.4 Stabilate manufacture and growth

Cryogenically frozen stocks of trypanosomes ('stabilates') were prepared from bloodstream form cultures grown to approximately  $2.5 \times 10^6$  cells.ml<sup>-1</sup> by mixing with 100% glycerol, to give a 1 ml sample containing 10% glycerol. Stabilates were frozen to -70°C for 24 h before being transferred to liquid nitrogen for storage. When a stabilate was required for culture it was thawed rapidly at 37°C and transferred to a culture flask containing 10 ml of culture medium.

## 2.2 Standard molecular biology techniques

### 2.2.1 General PCR

Amplification of template using *Taq* DNA polymerase was performed in 1 x buffer IV with separate MgCl<sub>2</sub> (all reagents supplied by Advanced Biotechnologies), producing products with single 3' deoxyadenosine residues at either end owing to the nontemplate-dependent terminal transferase activity of the enzyme. *Pfu* polymerase was used to amplify high fidelity PCR products with blunt termini in 1 x *Pfu* buffer (Stratagene). Herculase Enhanced DNA polymerase is based on *Pfu* polymerase with a small amount of *Taq* polymerase added (Stratagene); it therefore yields high fidelity products with a mixture of blunt termini and termini with single 3' deoxyadenosine overhangs and was used in 1 x Herculase buffer. An extension time of 1 min.kb<sup>-1</sup> was allowed for both *Taq* polymerase and Herculase polymerase, whilst 2 min.kb<sup>-1</sup> was allowed for *Pfu* polymerase. Each primer was added to a final concentration of 200 nM (Sigma Genosys), and 5-20 ng of template DNA was used. Deoxynucleotides (dATP, dCTP, dGTP and dTTP) were added to a final concentration of 400  $\mu$ M in total (100  $\mu$ M each) (Amersham Pharmacia Biotech). Reactions were performed in thin-walled 200  $\mu$ l microeppendorf tubes (Perkin-Elmer), and were overlaid with a drop of mineral oil (Sigma) when thermal cycling was performed in a Robocycler 96 or Robocycler Gradient 96 (Stratagene), but this was unnecessary when using a PCR Sprint (Hybaid). The reaction conditions for each individual PCR are described in the relevant sections and include the following details: the primers and polymerase used, the template and Mg<sup>2+</sup> concentrations, the cycling conditions, and the final reaction volume.

### 2.2.2 Restriction digestion

Restriction digestion reactions were usually performed in a final volume of 20  $\mu$ l containing 1 x reaction buffer and 20-40 U enzyme (all reagents supplied by New England Biolabs). For double digestions the most compatible reaction buffer was chosen and equal amounts of the two enzymes were added. Reactions were incubated for 2 h at the optimal temperature for the enzyme. Each reaction contained 200-600 ng DNA. Where greater amounts of DNA needed to be digested several reactions were set up, and normally combined either by gel purification and pooling or by phenol/chloroform extraction, ethanol precipitation and pooling.

### 2.2.3 General gel electrophoresis

Samples of DNA were typically electrophoresed on 0.6 to 1.0% agarose gels (Gibco BRL, Life Technologies) made with either 1 x TAE buffer or 0.5 x TBE buffer containing 0.2  $\mu$ g.ml<sup>-1</sup> EtBr at 20 to 100 kV for 1 to 18 h, using a commercial 1 kb ladder as a size marker (Gibco BRL, Life Technologies). Gels were then visualised under UV.

### 2.2.4 Ligation of PCR products

PCR products with blunt termini were purified using the Strataprep PCR purification kit and cloned into PCR-Script Amp SK(+) plasmids (all reagents supplied by Stratagene, following the manufacturer's protocols). In most cases the inserts were amplified using *Pfu* polymerase or Herculase polymerase. However, in some instances *Taq* polymerase-generated PCR products, which possess 3' terminal adenine residues, were 'polished' by the addition of *Pfu* polymerase after purification (using reagents supplied with the PCR-Script Amp Cloning Kit, Stratagene, following the manufacturer's protocol). Ligations were performed at room temperature for 1 h. For each cloning attempt a range of different insert:vector ratios were prepared to maximise the likelihood of a successful ligation. The PCR Script system relies on the addition of *SrfI*, which is an infrequently cleaving restriction enzyme which cuts in the plasmid multiple cloning site to generate blunt termini, into the ligation mix. Any plasmid which religates without an insert will be restricted by *SrfI*, thereby enhancing the overall yield of correct ligations.

PCR products generated using *Taq* polymerase were cloned directly into pCR2.1-TOPO (Invitrogen, according to the manufacturer's instructions). Ligations were performed at room temperature for 5 to 30 min. The TOPO TA cloning system relies on vector with an

open multiple cloning site possessing single 3'-thymidine overhangs at both termini which are covalently bound to Topoisomerase I by phospho-tyrosyl bonds between the DNA and the enzyme. When an insert possessing 3'-adenine overhangs (i.e. a *Taq* PCR product) is introduced, the 5'-hydroxyl groups of the insert molecule can attack the phospho-tyrosyl bonds, thereby releasing the topoisomerase and ligating the insert to the vector.

### 2.2.5 Ligation of restricted DNA fragments

The treatment of DNA fragments after restriction digestion depended on the type of cloning involved. For ligations involving compatible restriction sites in the vector, 20 U of Calf Intestinal Phosphatase (CIP; Roche) was added to restriction digestion reactions of vector DNA and incubated for 5 min at room temperature to remove the 5' terminal phosphates and prevent religation of the vector prior to purification. Where 5' overhanging restriction sites needed to be made blunt for ligating non-compatible ends, the 20  $\mu$ l restriction digestion reaction was supplemented with 0.5  $\mu$ l 10 x digestion buffer, 1  $\mu$ l of 10mM dNTPs (containing dATP, dCTP, dGTP and dTTP at 2.5 mM each), 5 U DNA polymerase I large-fragment (Klenow, New England Biolabs) and 2.5  $\mu$ l distilled H<sub>2</sub>O to bring the volume of the restriction digestion up to 25  $\mu$ l, before incubation at 37°C for 30 min. Subsequently, 20 U of CIP was added to reactions containing vector DNA and incubated at room temperature for 5 min before purification, while insert DNA was purified immediately after digestion without further modification.

After restriction digestion and treatment with modifying enzymes, the DNA fragment of interest was, in most cases, separated on a 0.6% agarose gel, excised using a scalpel and purified using the QIAquick gel extraction kit (QIAGEN, following the manufacturer's protocol). However, since restriction digestion of lambda clones produces a large number of fragments, and gel extraction results in the loss of approximately 50% of the sample, sub-cloning of lambda fragments relied on the 'shotgun' approach, whereby all the fragments generated in the restriction digestion were added to the ligation mix, and plasmids containing the fragment of interest were retrieved after transformation of the ligation mix. Therefore, restriction digestions of lambda clones were phenol-chloroform extracted and then chloroform extracted (in both cases, the extractions were centrifuged at 10 000 g for 10 min). Afterwards 1/20 volume of 3 M sodium acetate was added to the aqueous phase and the DNA was ethanol precipitated by adding 2 volumes of 99% ethanol, incubating at -20°C for at least 30 min before centrifugation at 10 000 g for 30 min to pellet the DNA. The ethanol was then aspirated, and the pellet washed with 1 volume of

70% ethanol before a similar centrifugation for 5 min. Finally the 70% ethanol was removed and the pellet air-dried for 5 min before resuspension in an appropriate volume of distilled H<sub>2</sub>O.

For each cloning attempt a range of different insert:vector ratios were prepared to increase the probability of a successful ligation. Both cohesive-end and blunt-end ligations were performed in a final volume of 10 µl, using 400 U T4 DNA ligase (New England Biolabs). Ligations were incubated at room temperature for 3-4 h.

### **2.2.6 Transformation and plasmid retrieval**

After ligation, 5 µl of the ligation mix was added to a 50 µl aliquot of competent cells (previously thawed on ice), mixed gently, and incubated on ice for 30 min. The cells were then heat shocked at 42°C for 47 s, and chilled immediately on ice for 2 min. After cooling, 150 µl of L-broth was added before the cells were incubated at 37°C for 1 h. After this 'recovery step', the cells were spread on selective L-agar plates containing 50 µg.ml<sup>-1</sup> ampicillin (Sigma), which were incubated for 16-18 h at 37°C. For ligations where constructs with and without inserts could be distinguished by blue/white colony colour due to active/inactive β-galactosidase, the plates were also supplemented with 0.125 mM IPTG (Boehringer Mannheim) and 0.05 mM X-Gal (Boehringer Mannheim). Plasmids were transformed into either XL1-Blue MRF' supercompetent cells (Stratagene), XL10-Gold Kan ultracompetent cells (Stratagene), or TOP10F' competent cells (Invitrogen).

Single transformed colonies were inoculated into 5 ml L-broth (supplemented with ampicillin to a final concentration of 50 µg.ml<sup>-1</sup>) and grown at 37°C for 16-18 h. The plasmids were then retrieved from 3 ml of this culture using the QIAprep miniprep kit (QIAGEN, following the manufacturer's protocol). The insert size and orientation were then confirmed by restriction digestion of 250 ng plasmid using enzymes chosen according to their position in the predicted restriction map for the plasmid. Where large amounts of a plasmid were required for later manipulations, a single colony was inoculated into 200 ml of L-broth supplemented with 50 µg.ml<sup>-1</sup> ampicillin, and incubated at 37°C for 16-18 h. The plasmids were then retrieved from 100 ml of TOP10F' cell culture or 200 ml of XL1-Blue MRF' or XL10-Gold Kan cell culture using the QIAGEN plasmid maxiprep kit (according to the manufacturer's protocol).

## **2.3 Genomic DNA isolation and Southern analysis**

### **2.3.1 Standard genomic DNA preparations**

A maximum of  $1 \times 10^8$  trypanosomes were centrifuged at 1620 g and the supernatant decanted. The cells were resuspended in 0.5 ml digestion buffer (100 mM NaCl, 1 mM EDTA, 50 mM Tris.HCl (pH 8.0)) supplemented with SDS to 1% and Proteinase K to  $10 \mu\text{g}.\text{ml}^{-1}$  and incubated at 37° for 18 h. The genomic DNA was phenol/chloroform extracted twice; the phases were separated by centrifugation at 10,000 g for 10 min. To ensure minimal shearing, wide-bore disposable plastic pipettes were used to aspire the aqueous phase. Afterwards, the DNA was ethanol precipitated by gentle mixing with 2 volumes of 99% ethanol. The suspension was allowed to stand at room temperature for 30 min before centrifugation at 10,000 g for 30 min. The pellet was drained and washed with 70% ethanol to remove salts, then drained and air-dried before resuspension in TE buffer to a final concentration to  $2 \mu\text{g}.\mu\text{l}^{-1}$ .

### **2.3.2 Digestion of genomic DNA**

Usually 2  $\mu\text{g}$  of genomic DNA prepared as described above was digested in a 20  $\mu\text{l}$  reaction containing 1 x reaction buffer and 20-40 U of restriction enzyme.

### **2.3.3 Southern blotting**

Restriction digestions of genomic DNA were electrophoresed on 0.6% agarose gels made with either 1 x TAE buffer or 0.5 x TBE buffer at between 15-30 kV for 18 h, post-stained with EtBr and visualised. The gels were then immersed in 0.125 M HCl for 30 min, rinsed in H<sub>2</sub>O, immersed in denaturing solution (0.5 M NaOH, 1.5 M NaCl) for 30 mins, rinsed in H<sub>2</sub>O, immersed in neutralisation solution (1 M Tris-HCl pH 8.0, 1.5 M NaCl) for 30 min and finally rinsed with H<sub>2</sub>O. The DNA was then transferred to a positively charged nylon membrane (Hybond-N or Hybond-XL, Amersham Pharmacia Biotech; or Roche) by capillary blotting using 20 x SSC as the transfer buffer (Sambrook and Russell, 2001), typically for 18 h. The DNA was then crosslinked to the nylon membrane by exposure to 1200J UV light using a UV Stratalinker 2400 (Stratagene).

## 2.4 RNA isolation and reverse transcription PCR

### 2.4.1 RNA isolation

Total RNA was isolated from between  $4 \times 10^7$  to  $1 \times 10^8$  cells using 1 ml TRIzol reagent (Gibco BRL, Life Technologies). Cells were pelleted by centrifugation at 1620 g, and lysed in TRIzol reagent by repetitive pipetting. The lysed samples were incubated at room temperature for 5 min before 200  $\mu$ l of chloroform was added. The tubes were shaken vigorously for 15 s and then incubated at room temperature for 2-3 min, before the samples were centrifuged at 12 000 g for 15 min at 4°C. The colourless upper aqueous phase was removed to a fresh tube, and the RNA precipitated by the addition of 500  $\mu$ l of iso-propanol. Samples were incubated at room temperature for 10 min, before the RNA was pelleted by centrifugation at 12 000 g for 10 min at 4°C. The supernatant was removed and the pellet washed once with 1 ml 75% ethanol, air-dried and dissolved in 20  $\mu$ l DEPC-treated H<sub>2</sub>O.

### 2.4.2 Reverse transcription and cDNA synthesis

Total RNA was treated with DNase I (amplification grade; Invitrogen Life Technologies, according to the manufacturer's protocol) to remove genomic DNA contamination. Afterwards, cDNA was generated by reverse transcription using Superscript II Reverse Transcriptase, and the Superscript first-strand synthesis system for RT-PCR with the random hexamer primers provided for generating the first-strand of DNA (Invitrogen Life Technologies, according to the manufacturer's protocol). Target cDNA was then amplified using 5 U *Taq* polymerase with 2 mM MgCl<sub>2</sub>, 400  $\mu$ M dNTPs (100 $\mu$ M each), 200 nM sense primer, 200 nM antisense primer, and 2  $\mu$ l of cDNA (from the first-strand reaction) in a final volume of 50  $\mu$ l, under the following cycling conditions: 94°C for 5 min, 30 cycles of 94°C for 1 min, 55°C for 1 min and 72°C for 1 min, with a final step of 72°C for 10 min. Amplification of the internal fragment of *MSH2* used the *MSH2D5* and *MSH2U2* primers (all primer sequences shown in Table 2.1). The primers *MSH8D3* and *MSH8U1* or *MSH8D6* and *MSH8U7* were used to amplify the internal fragment of *MSH8*, while the primers *EcoR1S1(1)* and *MSH8U7* or *MSH8U8* were used to confirm 5' splicing of this gene. *MLH1* was amplified using the *MLH1U1* and *MLH1D6* primers. *PMS1* was amplified using the following primers: *PMS1U1* and *PMS1D1*. A control amplification of the 5' end of the gene encoding the large subunit of RNA polymerase I using the primer pair PolI 5' and PolI 3' was performed to assess the integrity of the cDNA (Rudenko *et al.*,

1996). The presence or absence of mRNA encoding a given protein in wild-type trypanosomes was confirmed using cDNA generated from both MITat 1.2 bloodstream forms and procyclic cells.

## **2.5 Manufacture of radiolabelled probes and DNA hybridisation**

### **2.5.1 Radiolabelled probes**

The majority of the probes used in this study were PCR amplified from MITat 1.2 gDNA or from plasmids containing the antibiotic resistance genes. These fragments were then gel purified using the QIAGEN gel extraction kit (following the manufacturer's protocol). Radiolabelling was performed using the Prime-It II Random Primer Labelling Kit (Stratagene). Initially, 10  $\mu$ l of random 9-mer oligonucleotides (27 ODunits.ml<sup>-1</sup>) were mixed with 200 ng of template DNA and sterile, distilled water to a total reaction volume of 36  $\mu$ l. The DNA was denatured by heating to 95°C for 5 min, and cooled to room temperature before brief centrifugation. Subsequently, 10  $\mu$ l 5 x primer buffer, 3  $\mu$ l  $\alpha$ -<sup>32</sup>P labelled dCTP and 1  $\mu$ l Klenow (5 U. $\mu$ l<sup>-1</sup>) were added in order, mixed, and incubated at 37°C for 10 min. This mixture was then briefly centrifuged before purification of the probes from the unincorporated nucleotides using MicroSpin S-200 HR columns (Amersham Pharmacia Biotech, following the manufacturer's protocol). After purification, the probes were denatured at 95°C for 5 min immediately prior to use.

### **2.5.2 Hybridisation**

Nylon filters were hybridised to radioactive probes in glass hybridisation tubes. Where several nylon filters were to undergo hybridisation simultaneously in the same tube they were placed between two sheets of hybridisation mesh (Amersham, Life Technologies), before being transferred to a glass hybridisation tube. Approximately 50 ml of Church-Gilbert solution (0.342 M Na<sub>2</sub>HPO<sub>4</sub>, 0.158 M NaH<sub>2</sub>PO<sub>4</sub>.2H<sub>2</sub>O), 0.257 M SDS and 1 mM EDTA) was added and the filters were pre-hybridised by rotation at 65°C for a minimum of 3 h in a hybridisation oven. After addition of the purified, denatured probe the hybridisation was incubated at 65°C in the oven for at least 18 h. Afterwards, the filters were washed to high stringency in the oven at 65°C with the following series of solutions: 100 ml 2 x SSC, 0.1% SDS (twice – initially for 5 min then again for 30 min); and 100 ml 0.2 x SSC, 0.1% SDS (for 30 min). The filters were then heat-sealed in plastic and visualised in one of two ways: filters were either placed next to medical photographic film

(Konica Medical Corporation) in an autoradiography cassette at  $-70^{\circ}\text{C}$  for 15 min to 7 days, or next to a phosphoimaging plate (Fujifilm) in an autoradiography cassette for 15 min to 24 hours, exposure time depending of the expected strength of the signal in both cases. The photographic film was developed using an M35-M X-OMAT Processor (Kodak), and phosphoimages were read using a Typhoon 8600 Variable Mode Imager (Molecular Dynamics).

### 2.5.3 Stripping of nylon filters

Nylon filters were stripped with a boiling solution of 0.4 M NaOH, 0.1% SDS. The solution was poured onto the filters in a heat resistant container and allowed to cool to room temperature. The filters were then rinsed in 2 x SSC before re-use. Following stripping the filters were placed next to a phosphoimaging plate for 24 h to ensure that the procedure had been successful.

## 2.6 Screening of the ILTat 1.2 genomic lambda library and mapping of clones

### 2.6.1 Preparation of phage-lambda plating bacteria

A single colony of *E. coli* LE392 (Promega) was inoculated into 10 ml L-broth supplemented with 0.2% maltose and incubated at  $37^{\circ}\text{C}$  for 18 h. The  $\text{OD}_{600}$  of a 1/10 dilution of the culture was measured and the amount of culture which would give a final concentration of 2.0  $\text{OD}_{600}$  in 10 ml was centrifuged at 4000 g for 10 min at room temperature. The supernatant was aspirated and the cell pellet resuspended in 10 ml of 10 mM  $\text{MgSO}_4$ . This suspension of plating bacteria was stored at  $4^{\circ}\text{C}$  for up to 3 weeks.

### 2.6.2 Plating of phage-lambda

For the primary screen, an ILTat 1.2 genomic bacteriophage-lambda library was diluted to  $1.5 \times 10^5$  pfu.ml<sup>-1</sup> (a kind gift of Nils Burman; the library was generated using ILTat 1.2 genomic DNA which had undergone a partial restriction digestion with *Sau3A* which was cloned into the *XhoI* site of the LambdaGEM-12 vector). Phage suspension and plating bacteria were mixed in a 1:1 ratio (300  $\mu\text{l}$  each for a primary screen or 100  $\mu\text{l}$  each for later screens), and incubated at  $37^{\circ}\text{C}$  for 30 min. Next, top agarose (6.5 ml for a primary screen or 2.5 ml for a later screen), was mixed with the phage-bacteria suspension and poured onto a bottom agar plate (14 cm diameter plate for a primary screen or 8 cm diameter plate for a later screen). The top agarose was allowed to solidify for 5 min, before the plate was

inverted and incubated at 37°C for 18 h. For this protocol, bottom agar plates were dried and pre-warmed to 37°C, while top agarose was pre-heated to 48°C. After incubation phage plates were stored at 4°C. For secondary and tertiary and quaternary screens, the same procedure was followed using tenfold serial dilutions ( $1 \times 10^3$ ,  $1 \times 10^4$  and  $1 \times 10^5$ ) of isolated phage plaques.

### **2.6.3 Transfer of phage to nylon filters**

Phage plates were placed at 4°C for at least 1 h to allow the agarose to solidify before the transfer. A nylon filter (Hybond-N, Amersham Pharmacia Biotech) was placed onto the surface of the agar for 1 min. During this time, the filter was stabbed asymmetrically in 3 places with a needle and syringe containing electrophoresis loading dye, leaving a mark in the agar where the hole was made and allowing later orientation. The filter was removed and a replica filter was placed on the agar surface for 3 min, while the needle was used to puncture the filter in the same places as the original filter. The position of the holes was also marked on the underside of the petri dish. The filters were allowed to dry, plaque side up, for 20 min. The filters were then incubated, plaque side up, for 2 min in the following solutions in order: Southern denaturation solution, Southern neutralisation solution, and 2 x SSC (see section 2.2.3). The filters were then allowed to dry, and the DNA was crosslinked to the nylon filters by exposure to 1200J UV light using a UV Stratalinker 2400.

### **2.6.4 Isolation of target phage-lambda clones**

The nylon filters were then hybridised with radiolabelled probes as described in section 2.5. The filters were visualised by autoradiography, and then the duplicate filters were examined for hybridisation. Where the probe had hybridised to the same place on both filters, the autoradiograph was aligned with the phage plate and the target plaque identified. A plug of agar encompassing such a plaque was then transferred to 250 µl of phage suspension buffer. This suspension was then used as the inoculum for a further round of phage plating as in section 2.6.2. The screening process was repeated until every plaque on the agar plate hybridised to the probe showing the suspension to be clonal.

### **2.6.5 Liquid lysis of phage-lambda clones**

A well-isolated phage plaque from the final screening plate was transferred to 100 µl phage suspension buffer and left for 1 h at 37°C for use as the inoculum for liquid lysis of

the clone. A single colony of LE392 was inoculated into 5 ml L-broth supplemented with 10 mM MgSO<sub>4</sub> and 0.2% Maltose and incubated at 37°C for 18 h. 500 µl of this culture was added to the tube containing the clonal target phage plug and incubated for 20 min at 37°C. The phage-bacteria suspension was then used to inoculate 100 ml of L-broth supplemented with 10 mM MgSO<sub>4</sub> and pre-warmed to 37°C. The culture was shaken at 37°C until lysis occurred (usually 6-7 h), then 500µl of chloroform was added and the culture shaken for a further 15 min. After lysis the culture was centrifuged at 8 000 g for 10 min to remove the cellular debris. The supernatant was transferred to a sterile tube, 200 µl of chloroform was added and the suspension was stored at 4°C for 18 h. The target phage were isolated from the supernatant using LambdaSorb Phage Adsorbant (Promega), and the DNA was purified and resuspended in TE buffer (5 µl of TE buffer per 10 ml supernatant, according to Promega's protocol).

### **2.6.6 Restriction mapping of phage-lambda clones**

Approximately 600 ng of phage DNA was added to each restriction digestion performed as described in section 2.2.2. For each gene, a range of single and double digestions were performed on each lambda clone, using the enzymes suggested by the available gene sequence data and the results of the genomic Southern. The restriction digestion reactions were separated on a 0.6% agarose gel in TAE buffer, visualised and Southern blotted. The results of the Southern analyses, the sequence of the lambda arms and the available sequence data for the gene in question allowed restriction maps of the lambda clones to be developed. Examination of these maps revealed restriction fragments which were likely to contain the entire ORF of the gene as well as processing flanks. These restriction digestions were then repeated and the fragment of interest cloned as described in section 2.2.5.

### **2.7 Sequencing**

In a thin-walled microeppendorf tube, 500 ng of plasmid DNA (prepared using the QIAprep miniprep kit) was mixed with 3.2 pmoles of primer and 8 µl ABI PRISM dGTP BigDye terminator ready reaction solution (Applied Biosystems) in a final volume of 20 µl. Thermal cycling was then performed with a hot start in a PCR Sprint at 96°C for 4 min followed by 25 cycles of 96°C for 10 s, 50°C for 5 s and 60°C for 4 min. Afterwards, the solution was transferred to a 0.5 ml eppendorf tube and mixed with 6 µl of 3 M sodium acetate and 150 µl of 95% ethanol. After incubation at room temperature for

15 min, the DNA was pelleted by centrifugation at 10 000 g for 20 min. The supernatant was aspirated and the pellet washed with 250 µl 70% ethanol and dried in a Robocycler 96 at 90°C for 1 min. Automated sequencing was then performed by either the Molecular Biology Support Unit (MBSU) of the University of Glasgow, or Oswel Research Products at the University of Southampton.

## **2.8 Data mining the National Center for Biotechnology Information (NCBI) database, the *T. brucei* and other sequencing databases, multiple alignment of polypeptide sequences and gap comparisons**

### **2.8.1 NCBI data retrieval**

The MutS/MSH and MutL/MLH protein sequences were obtained from the protein database of the National Center for Biotechnology Information (NCBI) (<http://www.ncbi.nlm.nih.gov/>). The mismatch repair protein homologues used during this study (and their accession numbers) are as follows: MutS from *Escherichia coli* (AAC75775); MSH1 (NP\_011988) from *Saccharomyces cerevisiae*; homologues of the *S. cerevisiae* MSH2 protein (CAA99102) including *Arabidopsis thaliana* (O24617), *Drosophila melanogaster* (SPEL1 – P43248), *Encephalitozoon cuniculi* (CAD26200), *Homo sapiens* (NP\_000242), *Mus musculus* (AAA75027), *Neurospora crassa* (AAB84225), *Schizosaccharomyces pombe* (CAA07342), *Xenopus laevis* (S53609), *T. brucei* (AAK08648), *Trypanosoma cruzi* (AAG00261) and *Zea mays* (Q9XGC9); homologues of the *S. cerevisiae* MSH3 protein (CAA42247) from *A. thaliana* (CAA07684), *H. sapiens* (NP\_002430), *S. pombe* (SWI4 - CAB52164), and *T. brucei* (no accession number); MSH4 (BAA09235) and MSH5 (CAA98728) from *S. cerevisiae*; homologues of the *S. cerevisiae* MSH6 protein (CAA87671) from *A. thaliana* (CAB53337), *D. melanogaster* (Q9VUM0), *E. cuniculi* (CAD25790) and *H. sapiens* (NP\_000170); MSH7 from *A. thaliana* (AAF06013) and *Z. mays* (MUS2 – CAB42555); MSH8 from *T. brucei* (AAK51796); MutL from *E. coli* (AAC77127); homologues of the *S. cerevisiae* MLH1 protein (CAA89803) from *A. thaliana* (NP\_192653), *D. melanogaster* (AAF59117), *E. cuniculi* (CAD26547), *H. sapiens* (NP\_000240) and *T. brucei* (AAK29067); homologues of the *S. cerevisiae* PMS1 protein (CAA95956) from *A. thaliana* (AAL01156), *D. melanogaster* (PMS2 - AAF58142), *E. cuniculi* (CAD26036), *H. sapiens* (PMS2 - NP\_000526) and *T. brucei* (AAK08649); the *S. cerevisiae* MLH2 protein (CAA97559); the *S. cerevisiae* MLH3 protein (CAA97869); and the PMS1 (XP\_010779) and MLH3 (NP\_055196) proteins from *H. sapiens*.

### 2.8.2 Multiple alignment of polypeptide sequences

Multiple alignment of polypeptide sequences was performed using several programs including: the PileUp program from the GCG Wisconsin Package (Wisconsin Package Version 10.0, Genetics Computer Group (GCG), Madison, Wisc.); CLUSTAL X (version 1.8; Thompson *et al.*, 1997); and Multalin available on the server of the Institut National de la Recherche Agronomique (<http://prodes.toulouse.inra.fr/multalin/multalin.html>) (Corpet, 1988).

### 2.8.3 Pairwise comparisons

Pairwise comparisons were performed using the Gap pairwise comparison program of the GCG Wisconsin Package to assess the percentage similarities and identities between two sequences. The full length protein sequences were compared using the program's default settings.

### 2.8.4 Data mining the *T. brucei* sequencing databases

Preliminary multiple alignments of the MutS, MSH2, MSH3, MSH6, MutL, MLH1 and PMS1 sequences (see section 2.8.1) were performed using PileUp. The regions of conservation between the orthologous sequences included in each alignment were identified using output from the BOXSHADE server ([http://www.ch.embnet.org/software/BOX\\_form.html](http://www.ch.embnet.org/software/BOX_form.html)) using the RTF\_new output format and the fraction of sequences (that must agree for shading) set between 0.5 to 1.0.

BLAST searches of the TIGR ([http://www.tigr.org/tdb/mdb/tbdb/seq\\_search.html](http://www.tigr.org/tdb/mdb/tbdb/seq_search.html)) and Sanger ([http://www.sanger.ac.uk/Projects/T\\_brucei/Toolkit/blast\\_server.shtml](http://www.sanger.ac.uk/Projects/T_brucei/Toolkit/blast_server.shtml)) *T. brucei* genome sequencing databases using the MutS- and MutL-related sequences from *E. coli*, *S. cerevisiae*, *H. sapiens* and *A. thaliana* allowed identification of clones encoding homologous sequences in *T. brucei*. Data retrieved from the *T. brucei* databases were then used in further BLAST searches of the TIGR and Sanger servers as described above to find any overlapping sequences. This process was repeated with each sequence until no further overlapping sequences could be identified. All sequences retrieved from the *T. brucei* databases were then aligned into contigs using the Fragment Assembly series of programs provided by the GCG Wisconsin Package. The contiguous sequences were then used to BLAST search the NCBI database (<http://www.ncbi.nlm.nih.gov/BLAST/>) to assess their similarities with known proteins.

Large contiguous sequences retrieved from the Sanger database were analysed using the Nucleotide Identification (NIX) server (<http://www.hgmp.mrc.ac.uk/Registered/webapp/nix/>) provided by the UK Human Genome Mapping Project Resource Centre. This server is an integrated interface allowing the user to run a number of programs which predict sequence features (eg. exons) and perform BLAST searches, with an output which allows the user to easily compare the results of the different programs against each other. The *T. brucei* sequences analysed were input into the NIX server as genomic bacterial sequences (because *T. brucei* transcribes polycistronic mRNA and lacks introns). The NIX outputs presented in this study only show the results of selected programs: the BLAST search outputs from EMBL, TrEMBL and SwissProt; the gene prediction program HMMGene; and the exon prediction program Fex.

### **2.8.5 Data mining the sequencing databases of other unicellular parasitic organisms**

BLAST searching of genome sequencing databases (Table 2.2) available for *Babesia bovis*, *Encephalitozoon cuniculi*, *Entamoeba histolytica*, *Giardia lamblia*, *Leishmania major*, *Plasmodium falciparum*, *Theileria annulata*, *Toxoplasma gondii*, and *Trypanosoma cruzi* was performed using the MutS- and MutL-related sequences from *A. thaliana*, *S. cerevisiae* and *T. brucei* to identify similar sequences in these organisms. Data retrieved from these databases were then used to BLAST search the NCBI database to assess their similarities with known proteins, and putative protein identities were assigned solely on this basis. Also, the NCBI protein database (<http://www.ncbi.nlm.nih.gov/>) was searched for published MutS and MutL homologous sequences in these organisms.

## **2.9 Isolation of the complete coding sequences of three MutS homologues from *T. brucei***

### **2.9.1 MSH2**

Degenerate PCR primers were designed against the ATP-binding motif of MSH2 (*MSH2* 5'(1) and *MSH2* 3'(1); all primer sequences shown in Table 2.1). PCR was performed using 20 ng of MITat 1.2 gDNA, 5 U *Taq* polymerase, 2 mM MgCl<sub>2</sub> and 2 μM of each primer in a final volume of 50 μl, with an initial 5 min denaturation at 94°C, followed by 30 cycles of 94°C for 1 min, 48°C for 2 min, and 72°C for 1 min, followed a final 5 min extension at 72°C. A 315 bp product was generated and cloned into pBluescript II KS (Stratagene), which upon sequencing was predicted to encode the expected region of *T. brucei* MSH2. This plasmid was restriction-digested using *EcoRI*

and *Bam*HI, and the fragment encoding the ATP-binding domain of *MSH2* was purified as described in section 2.2.5. This fragment was then used as a probe for a MITat 1.2 and ILTat 1.2 genomic Southern and also as a probe to screen an ILTat 1.2 genomic lambda library. Several lambda clones were identified, restriction-digested and Southern blotted. A 4.5kb *Eco*RI fragment that contained the complete *MSH2* ORF was cloned into pBluescript II KS to create the plasmid pJB100 and the sequence was determined on both strands using the following primers: T3, T7, *MSH2D1*, *MSH2D2*, *MSH2D3*, *MSH2D4*, *MSH2D5*, *MSH2F1*, *MSH2F2*, *MSH2F3*, *MSH2F4*, *MSH2J1*, *MSH2J2*, *MSH2J3*, *MSH2J4*, *MSH2U1*, *MSH2U2*, *MSH2U3*, *MSH2U4*, *MSH2U5*, and *MSH2U6*. The sequence predicts a polypeptide 951 amino acids in length, as well as the 5' and 3' untranslated regions of the gene (accession number AY026905).

### 2.9.2 *MSH8*

PCR primers were designed from two contiguous sequences identified by the procedure described in section 2.8.4 (*MSH8* 5' and *MSH8* 3'; all primer sequence shown in Table 2.1). PCR was performed with an initial 5 min denaturation at 94°C followed by 30 cycles of 94°C for 1 min, 55°C for 1 min, and 72°C for 5 min, followed a final 10 min extension at 72°C, using 20 ng MITat 1.2 gDNA, 5 U *Taq* polymerase and 2 mM MgCl<sub>2</sub> in a final volume of 50 µl. The 2.0 kb product was 'polished' with *Pfu* polymerase and cloned into pPCR-Script Amp SK(+), creating the plasmids pJB301.1 and pJB301.2, and the sequence of the two independently derived clones was determined on both strands using custom designed primers. From this sequence, two oligonucleotides (*MSH8D3* and *MSH8U1*) were designed to PCR amplify a 540 bp fragment for use as a probe for a MITat 1.2 and ILTat 1.2 genomic Southern. The reaction conditions were similar to above except that thermal cycling was performed for 5 min at 94°C, 30 cycles of 94°C for 1 min, 55°C for 1 min and 72°C for 2 min, followed by a final extension step of 10 min at 72°C.

Several months later, further searching of the *T. brucei* sequencing databases was performed using the sequence derived from pJB301 revealing further contiguous sequence. Oligonucleotides directed against these sequences allowed PCR using 2.5 U Herculase polymerase with otherwise similar conditions to those used to generate the PCR product cloned to give pJB301. Two products were generated: the first 2.9 kb in length (using primers *MSH8* UP(1) and *MSH8U3*) and the second 2.2 kb in length (using primers *MSH8D3* and *MSH8* DN(1)). Two independent PCR reactions were cloned into pPCR-Script Amp SK(+) for each product, generating the plasmids pJB302 and pJB303 respectively, and the sequence of each determined as above. The 2.9 kb and 2.2 kb

products encoded the N-terminal and C-terminal regions of MSH8 respectively.

Sequencing of the locus encoding *MSH8* required the following oligonucleotides: T3, T7, M13 For, M13 Rev, *MSH8D1*, *MSH8D2*, *MSH8D3*, *MSH8D4*, *MSH8D5*, *MSH8D6*, *MSH8D7*, *MSH8D8*, *MSH8U1*, *MSH8U2*, *MSH8U3*, *MSH8U4*, *MSH8U5*, *MSH8U6*, *MSH8U7*, *MSH8U8*, and *MSH8U9*. Contiguous sequence derived from all three *MSH8* sequences revealed an ORF encoding a polypeptide 997 amino acids in length (accession number AF350880).

Expression of the gene and the 5'-splicing of the gene were confirmed by RT-PCR as described in section 2.4.

### 2.9.3 *MSH3*

A recent BLAST search of the Sanger *T. brucei* genome sequencing database using the *T. brucei* MSH8 polypeptide sequence identified a large contig (TRYP9.0.000864) encoding another MutS homologue from *T. brucei*. The hypothetical ORF (TRYP9.0.000864\_4) encoding this polypeptide was used to BLAST search the NCBI database in order to assess its similarity to known sequences (see section 2.8.4). This approach suggested the ORF encoded an MSH3 homologue.

## 2.10 Isolation of the complete coding sequences of two MutL homologues from *T. brucei*

### 2.10.1 *MLH1*

PCR primers were designed from three contiguous sequences identified by the procedure described in section 2.8.4 (*MLH1* 5'(1) and *MLH1* 3'(1); all primer sequences shown in Table 2.1). PCR was performed with an initial 5 min denaturation step of 94°C, followed by 30 cycles of 94°C for 1 min, 50°C for 1 min and 72°C for 5 min, with a final step of 72°C for 10 min using 20 ng MITat 1.2 gDNA, 5 U *Taq* polymerase and 2 mM MgCl<sub>2</sub> in a final volume of 50 µl. A 2.9 kb product was generated using these primers showing that the three contigs composed parts of the same gene. A further set of oligonucleotides (*MLHIU1* and *MLIID6*) were designed from the contig sequence in order to amplify a 427 bp product, predicted to encode the ATP binding domain of *MLH1*, for use as a probe for both a MITat 1.2 and ILTat 1.2 genomic Southern and an ILTat 1.2 genomic lambda library. The reaction conditions were similar to those used above except that thermal cycling was performed for 5 min at 94°C, with 30 cycles of 94°C for 1 min, 55°C for 1 min

and 72°C for 2 min, followed by a final extension step of 10 min at 72°C. Several lambda clones were isolated and a 4.0 kb *Acc65I* (*KpnI*) fragment that contained the complete *MLH1* ORF was cloned into pBluescript II KS, generating the plasmid pJB200. The sequence was generously determined by T. I. Harvey on both strands by sequencing using the following primers: T3, T7, *MLH1D1*, *MLH1D2*, *MLH1D3*, *MLH1D4*, *MLH1D5*, *MLH1D6*, *MLH1D7*, *MLH1D8*, *MLH1U1*, *MLH1U2*, *MLH1U3*, *MLH1U4*, *MLH1U5*, *MLH1U6*, and *MLH1U7*. The sequence predicts a polypeptide 887 amino acids in length, along with 5' and 3' untranslated regions (accession number AF346620).

### 2.10.2 *PMS1*

PCR primers were designed from three contiguous sequences identified by the procedure described in section 2.8.4 (*PMS1* 3'(2) and *PMS1* 5'(2) or *PMS1U2*; all primer sequences shown in Table 2.1). PCR using primers *PMS1* 5'(2) and *PMS1* 3'(2) was performed with an initial 5 min denaturation at 94°C, followed by 30 cycles of 94°C for 1 min, 55°C for 1 min, and 72°C for 4 min, followed a final 10 min extension at 72°C, using 20 ng MITat 1.2 gDNA, 5 U *Taq* polymerase and 2 mM MgCl<sub>2</sub> in a final volume of 50 µl. PCR using these primers generated a 2.7 kb product confirming that the contigs were part of the same gene. PCR using the primers *PMS1U2* and *PMS1* 3'(2) generated a 460 bp product for use as a probe for both a MITat 1.2 and ILTat 1.2 genomic Southern and an ILTat 1.2 genomic lambda library. The reaction conditions were similar to those used above except that the thermal cycling was performed using an initial 5 min denaturation at 94°C, followed by 30 cycles of 94°C for 1 min, 50°C for 1 min, and 72°C for 1 min, followed a final 10 min extension at 72°C. Several lambda clones were isolated and a 5.5kb *EcoRI* fragment that contained the complete *PMS1* ORF was cloned into pBluescript II KS, generating the plasmid pJB400. The library screening and sub-cloning and sequencing were generously performed by T. I. Harvey. The sequence was determined on both strands by sequencing with the following custom designed primers: T3, T7, *PMS1D1*, *PMS1D2*, *PMS1D3*, *PMS1D4*, *PMS1D5*, *PMS1D6*, *PMS1D7*, *PMS1U1*, *PMS1U2*, *PMS1U3*, *PMS1U4*, *PMS1U5*, and *PMS1U6*. The sequence encodes a polypeptide 788 amino acids in length (accession number AY026906).

## 2.11 Generation and analysis of *T. brucei* mismatch repair knock-out cell lines

### 2.11.1 *MSH2*

*T. brucei* *MSH2* knockout constructs were designed to contain 430bp of 5' sequence and 370bp of 3' sequence, which act as targeting sequences for replacement of the entire *MSH2* ORF. The *MSH2* flanking sequences were amplified by PCR from 20 ng MITat 1.2 gDNA with 1.25 U *Pfu* polymerase supplemented with 2 mM MgSO<sub>4</sub> in a final volume of 50 µl, with an initial 5 min denaturation at 94°C followed by 30 cycles of 94°C for 1 min, 50°C for 1 min, and 72° for 1 min, followed a final 10 min extension at 72°C. The 5' flank was amplified using the primers *MSH2* 5'(A) and *MSH2* 5'(B), while for the 3' flank the primers *MSH2* 3'(A) and *MSH2* 3'(B) were used (primer sequences shown in Table 2.1). Both flanks were cloned simultaneously into pBluescript II KS. These targeting flanks were separated by one of two antibiotic resistance cassettes, containing either the 600 bp puromycin-N-acetyltransferase ORF (*PUR*) or the 400 bp blasticidin-S-deaminase ORF (*BSD*) (both gifts of M. Cross), flanked by 230 bp of β-α tubulin and 330 bp of α-β tubulin processing signals. After integration, the resistance genes are expressed by endogenous *MSH2* transcription, since the knockout cassettes contain no promoter elements. Cloning of the resistance cassettes between the *MSH2* targeting sequences generated the constructs  $\Delta$ *MSH2*::*PUR* and  $\Delta$ *MSH2*::*BSD* respectively. The plasmids were digested with *Xho*I and *Nof*I prior to being transformed into *T. brucei* as described in section 2.1.3. Transformants were selected with 1.0 µg.ml<sup>-1</sup> puromycin dihydrochloride (Calbiochem) or 2.5 µg.ml<sup>-1</sup> blasticidin S hydrochloride (Calbiochem).

Correct integration of the constructs into the *MSH2* locus was determined in three ways. Firstly, gDNA from the transformants was digested with *Sac*I, Southern blotted and probed with the 430 bp *MSH2* 5' integration flank described above, to determine that the *MSH2* wild-type allele was removed from the genome in the blasticidin and puromycin double resistant clones. In both digestions a 5.8 kb fragment was detected in the wild-type and heterozygous mutants, corresponding to the intact wild-type *MSH2* locus. This fragment was absent in the homozygous knockout clones. The probe also hybridised to 2.85 kb and 2.65 kb fragments corresponding to the  $\Delta$ *MSH2*::*PUR* and  $\Delta$ *MSH2*::*BSD* alleles respectively. Secondly, to show that no copies of the *MSH2* ORF existed with the *MSH2* homozygous mutant cell lines, PCR was performed on 20 ng gDNA isolated from the transformants using internal primers complementary to the *MSH2* ORF (*MSH2D5* and *MSH2U2*; sequences shown in Table 2.1), using 5 U *Taq* polymerase and 2 mM MgCl<sub>2</sub> in

a final volume of 50  $\mu$ l with an initial 5 min denaturation at 94°C followed by 30 cycles of 94°C for 1 min, 55°C for 1 min, and 72° for 1 min, followed a final 10 min extension at 72°C in a PCR Sprint. A control amplification of the 5' end of the gene encoding the large subunit of RNA polymerase I using the primer pair PolI 5' and PolI 3' was performed to assess the integrity of the gDNA (Rudenko *et al.*, 1996). Finally, total RNA from the *MSH2* knockout mutants was prepared with TRIzol and cDNA was synthesised as described in section 2.4. RT-PCR was performed to determine whether or not intact *MSH2* mRNA was present in the transformants.

### 2.11.2 *MLH1*

The *MLH1* knock-out constructs were generated and utilised as described for the *MSH2* knockout constructs, with the following modifications. The *T. brucei MLH1* targeting sequences were designed to replace the entire *MLH1* ORF using 413 bp of 5' flank and 520 bp of 3' flank. These were amplified from 20 ng of MITat 1.2 gDNA by PCR using 2.5 U of Herculase polymerase in a final volume of 50  $\mu$ l with the following cycling conditions: 94°C for 5 min, 30 cycles of 94°C for 45 s, 55°C for 45 s and 72°C for 30 s, followed by 72°C for 10 min. The 5' flank was amplified using the primers *MLH1* 5'(A) and *MLH1* 5'(B), while the 3' flank was amplified with the primers *MLH1* 3'(A) and *MLH1* 3'(B) (sequences shown in Table 2.1). Cloning of the resistance cassettes between the *MLH1* targeting sequences generated the  $\Delta MLH1::PUR$  and  $\Delta MLH1::BSD$  constructs respectively.

Correct integration of the constructs into the *MLH1* locus was determined in three ways. Firstly, in order to determine that the *MLH1* wild-type allele was removed from the genome in the blasticidin and puromycin double resistant clones, gDNA from the transformants was digested with *Kpn*I, Southern blotted and probed with the 415 bp *MLH1* 5' integration flank described above. In both digestions a 4.5 kb fragment corresponding to the intact wild-type *MLH1* locus was detected in the wild-type and heterozygous mutants. This fragment was absent in the homozygous knockout clones. The probe also hybridised to 1.3 kb and 1.1 kb fragments corresponding to the  $\Delta MLH1::PUR$  and  $\Delta MLH1::BSD$  alleles respectively. Secondly, PCR was performed on 20 ng gDNA isolated from the transformants using internal primers complementary to the *MLH1* ORF (*MLH1U1* and *MLH1D6*; sequences shown in Table 2.1), using 5 U *Taq* polymerase and 2 mM  $MgCl_2$  in a final volume of 50  $\mu$ l with an initial 5 min denaturation at 94°C followed by 30 cycles of 94°C for 1 min, 55°C for 1 min, and 72° for 1 min, followed a final 10 min

extension at 72°C in a PCR Sprint, to show that no copies of the *MLH1* ORF existed with the *MLH1* homozygous mutant cell lines. In order to assess the integrity of the gDNA, a control amplification of the 5' end of the gene encoding the large subunit of RNA polymerase I using the primer pair PolI 5' and PolI 3' was performed (Rudenko *et al.*, 1996). Finally, total RNA from the *MLH1* knockout mutants was prepared with TRIzol and cDNA was synthesised as described in section 2.4. RT-PCR was performed to determine whether or not intact *MLH1* mRNA was present in the transformants.

## 2.12 Re-expressing a copy of the wild-type *MSH2* allele in the homozygous knock-outs

The *T. brucei* *MSH2* re-expression construct was based on the *MSH2* lambda sub-clone pJB100. An antibiotic resistance cassette containing the 400 bp bleomycin resistance protein (*BLE*) ORF, flanked at the 5' end by 400 bp of processing flank from the actin locus and at the 3' end by 330 bp of  $\alpha$ - $\beta$  tubulin intergenic sequence, was cloned into the multiple cloning site downstream of the 3' end of the *MSH2* 3' processing flank in pJB100. This cloning generated the construct  $\Delta$ *MSH2::MSH2-BLE*. The targeting flanks used by this construct to replace either the *MSH2::PUR* or *MSH2::BLE* alleles are the *MSH2* 5' flank and the  $\alpha$ - $\beta$  tubulin intergenic sequence. As this re-expression cassette contains no promoter element, both *MSH2* and the resistance gene will be expressed by endogenous *MSH2* transcription, thereby returning *MSH2* to a normal level of expression.

The plasmid was digested with *Hind*III prior to being transformed into *T. brucei* as described in section 2.1.3. Transformants were selected with 2.5  $\mu$ g.ml<sup>-1</sup> phleomycin (Cayla). Correct integration of the construct into the *MSH2* locus was determined as described in section 2.11.1. Where correct integration had occurred Southern analysis detected a 6.2 kb band corresponding to the *MSH2::MSH2-BLE* allele.

## 2.13 Measuring the growth rates of trypanosome mutant cell lines

*In vitro* growth rates were measured for each 3174 *MSH2* or *MLH1* transformant by diluting mid-log phase trypanosomes to a concentration of 1 x 10<sup>5</sup> cells.ml<sup>-1</sup> in 2.5 ml of culture medium in 6-well culture dishes. Cell concentrations were measured at 24 h intervals thereafter with a haemocytometer (Sigma).

## 2.14 Measuring cell survival following treatment with *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine (MNNG)

Recently cloned bloodstream form trypanosomes (see section 2.1.3) were grown to mid-log phase before being diluted to  $10 \text{ cells.ml}^{-1}$  and plated across 96-well plates at  $200 \mu\text{l}$  per well. For each of three repetitions, one culture dish for each 3174 *MSH2* or *MLH1* transformant was plated at each drug concentration. MNNG (Aldrich Chemical Company) was dissolved to  $5 \text{ mg.ml}^{-1}$  in 100% DMSO before dilution into HMI-9. The following drug concentrations were used independently: 0, 0.25, 0.5, 0.75 and  $1.0 \mu\text{g.ml}^{-1}$  in HMI-9. Plates were incubated for 12 days before wells were scored for cell survival. The results have been depicted as an average percentage survival without standard error bars as the three repetitions for each cell line were performed simultaneously using the same mid-log phase cell culture (see sections 4.2.6 and 4.3.5 and Figures 4.8 and 4.31).

## 2.15 Generation and analysis of trypanosomes that have undergone antigenic variation

The isolation of VSG-switched variants utilised the approach described in McCulloch *et al.* (1997) with some modifications. A number of ICR mice were intraperitoneally injected with between  $2 \times 10^5$  and  $4 \times 10^5$  3174 trypanosomes (which are wild-type at both the *MSH2* and *MLH1* loci; McCulloch *et al.*, 1997) growing on hygromycin B ( $5.0 \mu\text{g.ml}^{-1}$ ; Roche) and G418 ( $2.5 \mu\text{g.ml}^{-1}$ ; Sigma Cell Culture). The infections were cured 3-4 days later by injection with cymelarsan (Rhone Merieux;  $5 \text{ mg.kg}^{-1}$  body weight) to generate mice immunised against the VSG221 coat. The immunised mice were used between 7 and 35 days post-curing. To select for switched variants, the 3174 *MSH2* or *MLH1* transformants were passaged from medium containing hygromycin and G418 into non-selective medium at a density of  $2 \times 10^5 \text{ cells.ml}^{-1}$  and grown for nine generations (approximately 72 h) until they reached densities of  $2 \times 10^6$  to  $3 \times 10^6 \text{ cells.ml}^{-1}$ . The trypanosomes were counted, and  $4 \times 10^7$  cells of each transformant were centrifuged at 580 g, resuspended in  $300 \mu\text{l}$  of HMI-9 and then intraperitoneally injected into an immunised mouse. Twenty-four hours later the mice were bled by cardiac puncture. Switched variants were isolated from 0.4 ml of blood as described by Rudenko *et al.* (1996) and were immediately cloned after dilution into 20ml HMI-9; this volume was spread over a 96-well culture dish. Typically, a single trypanosome would grow out to a visible population after 5 to 7 days. We have assumed that all the cell populations double

every 8 hours in order to calculate the VSG switching frequencies from the numbers of clones that grew out.

To assay for the drug sensitivity of the switched variant clones, approximately  $2 \times 10^4$  cells were passaged from the 96-well dish into 1 ml of HMI-9 medium alone, or supplemented with either hygromycin B ( $5.0 \mu\text{g} \cdot \text{ml}^{-1}$ ) or G418 ( $2.5 \mu\text{g} \cdot \text{ml}^{-1}$ ). Growth was scored 7 days later. Small-scale genomic DNA preparations for the PCR assays were made from  $3 \times 10^5$  to  $5 \times 10^5$  cells as described in section 2.3.1. The presence or absence of the hygromycin phosphotransferase, neomycin phosphotransferase, and *VSG221* genes was assessed by PCR using the following primer pairs: Hygro 5' and Hygro 3', Neo 5' and Neo 3', VSG221 5' and VSG221 3' (sequences shown in Table 2.1). The integrity of the DNA was assessed by a control amplification of the 5' end of the gene encoding the large subunit of RNA polymerase I using the primer pair ( PolI 5' and PolI 3' – for sequences see Table 2.1; Rudenko *et al.*, 1996) as described in section 2.4.2. PCR amplification in a final volume of 25  $\mu\text{l}$  was performed using 5 U of *Taq* polymerase with 2 mM  $\text{MgCl}_2$  in a Robocycler 96; the reaction used 1/50 of the total DNA prepared as above and involved an initial 5 min denaturation at 94°C followed by 30 cycles of 94°C for 1 min, 55°C for 1 min and 72°C for 1 min and finally 10 min at 72°C.

## 2.16 Assay for microsatellite instability in the *T. brucei* mismatch repair mutants

Trypanosomes were cloned by the 96-well plate method described in section 2.1.3. For each 3174 *MSH2* or *MLH1* transformant 10 clones (200  $\mu\text{l}$  culture) were passaged into 25 ml HMI-9 each and grown to a density of  $1.5 \times 10^6$  to  $4.5 \times 10^6$  cells. $\text{ml}^{-1}$ . Genomic DNA was prepared from each culture as detailed in section 2.3.1, and diluted in distilled  $\text{H}_2\text{O}$  to 5 ng. $\mu\text{l}^{-1}$  for use as template to PCR amplify the chosen microsatellite loci. PCR amplification was performed using a Robocycler 96, using the following reagents in a final volume of 20  $\mu\text{l}$ : either 0.5 U Herculase polymerase in 1 x Herculase buffer with 5 ng template DNA, 400 nM of each primer (all primer sequences shown in Table 2.1), and 125  $\mu\text{M}$  each of dATP, dCTP, dGTP and dTTP; or 0.5 U *Taq* polymerase in 45mM Tris.HCl pH 8.8, 11 mM  $(\text{NH}_4)_2\text{SO}_4$ , 4.5 mM  $\text{MgCl}_2$ , 6.7 mM 2-mercaptoethanol, 4.4  $\mu\text{M}$  EDTA, 113  $\mu\text{g} \cdot \text{ml}^{-1}$  BSA, with 1 mM each of dATP, dCTP, dGTP, and dTTP, 5 ng template DNA, and 400 nM of each primer. Microsatellite JS-2 (Hope *et al.*, 1999) was amplified by 28 cycles of 96°C for 50 s, 58°C for 50 s and 70°C for 50 s using the primers JS-2A and JS-2B with Herculase polymerase for the *MSH2* transformant samples and *Taq* polymerase for the *MLH1* transformant samples, before analysis by electrophoresis on a

3% NuSieve GTG agarose gel (FMC BioProducts). All other microsatellite loci were amplified using *Taq* polymerase and separated on 4% NuSieve gels to estimate the dilution factor for later analysis. Electrophoresis was performed in 0.5 x TBE buffer containing  $0.2 \mu\text{g}\cdot\text{ml}^{-1}$  EtBr at 150 kV for 2.5-4.0 h, using Superladder-Low 20 bp ladder as a size marker (Advanced Biotechnologies). Microsatellite PLC (A. MacLeod, S. McLellan, C.M.R. Turner and A. Tait, personal communication) was amplified by 27 cycles of 95°C for 50 s, 58°C for 50 s and 70°C for 50 s using the primers PLCG and PLCH3. Microsatellite ChrII-6 (A. MacLeod, A. Tweedie and A. Tait, personal communication) was amplified by 30 cycles of 95°C for 50 s, 52°C for 50 s and 72°C for 2 min using the primers ChrII-6A and ChrII-6B. Microsatellites ChrI-7 and ChrI-15 (A. MacLeod, A. Tweedie, A. Tait, personal communication) were amplified by 28 cycles of 95°C for 50 s, 54°C for 50 s and 68°C for 50 s using the primers ChrI-7A and ChrI-7B, and ChrI-15A and ChrI-15B, respectively. The wild-type PCR products were cloned into pCR2.1-TOPO for microsatellites PLC, ChrII-6, ChrI-7 and ChrI-15 and sequenced on both strands. One primer from each pair for each microsatellite was labelled with the fluorescent dye 6-carboxyfluorescein (6-FAM; Sigma Genosys) for detection with an automated DNA sequencer. Microsatellites PLC, ChrII-6, ChrI-7, and ChrI-15 were analysed by GeneScan analysis as follows. The PCR products were diluted 1/5 to 1/50 in H<sub>2</sub>O and 1  $\mu\text{l}$  of each was run on a 36 cm acrylamide-urea gel (5% polyacrylamide, 6 M urea and 1 x TBE buffer for 3 h at 3 000 V and 51°C). N, N, N', N'-tetramethyl-6-carboxyrhodamine (TAMRA)-labelled GeneScan size standard (Applied Biosystems) was loaded into each well along with the PCR products. Signals were read with an automatic sequencer (ABI 377; Applied Biosystems), and the data were stored and analysed with GeneScan software (version 3.1.2; Applied Biosystems) by the local Southern sizing method. The acrylamide gel electrophoresis and reading of the fluorescent signals were performed by the Molecular Biology Support Unit (MBSU) of the University of Glasgow. Where possible, the PCR products generated from the microsatellite loci given above were cloned into pCR2.1-TOPO (Invitrogen; see section 2.2.4), and sequenced in order to determine the sequence of the alleles present in the 3174 *MSH2* and *MLH1* wild-type trypanosomes. The results have been tabulated as a qualitative assessment of the presence or absence of microsatellite length changes (see sections 4.2.7 and 4.3.6).

## 2.17 Assay for the frequency of integration of homologous and homeologous cassettes into trypanosome mismatch repair mutants

### 2.17.1 Generation of phleomycin resistance cassettes with homologous and homeologous integration flanks derived from the hygromycin phosphotransferase ORF

The hygromycin phosphotransferase (*HYG*) ORF present in the plasmid pHygro-Tub, was used as a template for the amplification of targeting flanks derived from the wild-type *HYG* gene by PCR using the primers *HYG*For and *HYG*Rev (all primer sequences shown in Table 2.1). All reactions were performed in a PCR Sprint, using 2 ng of template and the following cycling conditions: 94°C for 10 min, 30 cycles of 94°C for 1 min, 65°C for 1 min and 72°C for 1 min, followed by a final step of 72°C for 10 min. High fidelity amplification of the targeting flanks was performed using 2.5 U Herculase polymerase in a final volume of 50 µl to generate flanks which were perfectly homologous to the *HYG* template. A random mutagenesis approach was utilised to generate flanks homeologous (non-identical) to the *HYG* template. Two nucleotide analogues, dPTP and 8-oxo-dGTP (both supplied by Amersham Pharmacia Biotech), were added to a standard *Taq* polymerase amplification reaction (25 µl, containing 5 U *Taq* and 2 mM MgCl<sub>2</sub>) to a final concentration of 100 µM each. *Taq* polymerase incorporates dPTP in place of dTTP, and with a lower efficiency dCTP, while 8-oxo-dGTP is incorporated in place of dTTP only (Zaccolo *et al.*, 1996). At the end of cycles 5, 10, 15, 20, 25, and 30, 1 µl of the reaction mix was removed and used as a template for a standard *Taq* polymerase amplification, required to remove the mismatches present in the PCR products prior to cloning. The six resulting *Taq* PCR products and the high-fidelity Herculase PCR product were cloned into pCR2.1-TOPO, and clones were restriction-digested with *Asc*I and *Nde*I to ensure the presence of these sites within the plasmid. A number of clones derived from each PCR product were sequenced on both strands with the following primers: M13 For, M13 Rev, *HYGD*1, *HYGD*2, and *HYGU*1. Furthermore, pHygro-Tub was sequenced using the primers *HYGD*1, *HYGD*2, *HYGU*1, ResFor and ResRev. Selection of clones was governed by the level of mutation which they had incorporated. The clones chosen were as follows: pHYG01 (1%) and pHYG03 (3%) (all percentages rounded to the nearest integer) and pHYGWT, which sequencing revealed to be 100% identical to the *HYG* template.

The *HYG* integration flanks were separated by an antibiotic resistance cassette containing the *BLE* ORF, flanked at the 5' end by 400 bp of processing flank from the actin locus and

at the 3' end by 400 bp of processing flank from the calmodulin locus. The resistance gene will be expressed by endogenous tubulin transcription since the cassette contains no promoter element. Cloning of the resistance cassette between the *HYG* targeting sequences generated the constructs *HYGWT::BLE*, *HYG01::BLE* and *HYG03::BLE*. The plasmids were digested with *AscI* and resuspended to  $1 \mu\text{g} \cdot \mu\text{l}^{-1}$  in distilled  $\text{H}_2\text{O}$  prior to transformation into *T. brucei*.

### **2.17.2 Generation of a trypanosome cell line containing a single copy of the *HYG* ORF**

The plasmid pHygro-Tub which contains the *HYG* ORF flanked by 230 bp of  $\beta$ - $\alpha$  tubulin and 300 bp of  $\alpha$ - $\beta$  tubulin targeting flanks was digested with *XhoI* and *XbaI* prior to being transformed into *T. brucei* MITat 1.2a S427 as described into section 2.1.3. Integration of this cassette replaces an  $\alpha$  tubulin gene. As the resistance cassette does not contain a promoter element the *HYG* gene will be expressed by endogenous tubulin transcription. Transformants were selected at  $5 \mu\text{g} \cdot \text{ml}^{-1}$  hygromycin B.

Correct integration of the cassette into the tubulin locus was determined by Southern analysis. Genomic DNA from the transformants was digested with *KpnI*, Southern blotted and probed with the *HYG* ORF generated by PCR with Herculase polymerase in section 2.17.1. In the *KpnI* digestion a 2.9 kb fragment was detected where the cassette had integrated correctly. The resulting cell line, HTUB, was then used to generate both heterozygous and homozygous *MSH2* knock-out mutants as described in section 2.11.1.

### **2.17.3 Assay to determine the frequency of construct integration**

Electroporation of  $2.5 \times 10^7$  cells of each HTUB *MSH2* transformant was performed in 0.5 ml ZMG medium with  $6 \mu\text{g}$  of *HYG::BLE* construct as described in section 2.1.3. For each transformation  $1.5 \times 10^7$  cells were diluted into 36 ml HMI-9, containing  $2.5 \mu\text{g} \cdot \text{ml}^{-1}$  phleomycin and spread in 1.5 ml samples over a 24-well tissue culture plate. Three repetitions were performed for each of the constructs, *HYGWT::BLE*, *HYG01::BLE* and *HYG03::BLE*, in each of the HTUB cell lines. Plates were incubated at  $37^\circ\text{C}$  for 7 days before wells were scored for cell growth.

**Table 2.1 Oligonucleotides used in this study**

<b>Primer Name</b>	<b>Primer Sequence</b>
<i>BLE</i> For	5'-GGACCTCGTGGAGGACG-3'
ChrI-7A	5'-CCTGTTGGCCGCATTATTCGATGC-3'; labelled with 6-FAM
ChrI-7B	5'-TGAGGAAGTGAGGGGAGACGGAAG-3'
ChrI-15A	5'-TGGCTAGTTACACTGTAGTTCTCC-3'; labelled with 6-FAM
ChrI-15B	5'-CCACAACCACTCCTTGATATTCAC-3'
ChrII-6A	5'-GTTGTGTGAGCAACGCAAGGGCGG-3'; labelled with 6-FAM
ChrII-6B	5'-TATAACAACAGGTGAAGGAGAGGG-3'
<i>Eco</i> RISL(1)	5'- <u>CCGAATTC</u> GTCTTCTGTACTATATTG-3'; <i>Eco</i> RI sequence underlined
<i>HYGD</i> 1	5'-CAAGACCTGCCTGAAACC-3'
<i>HYGD</i> 2	5'-TTCCCAATACGAGGTCGC-3'
<i>HYG</i> For	5'- <u>AAGGCGCGCC</u> AGCCTGAACTCACCGCGACG-3'; <i>Asc</i> I sequence underlined
<i>HYG</i> Rev	5'-ATGGCGCGCCCTCCGGATCGGACGATTGCG-3'; <i>Asc</i> I sequence underlined
Hygro 5'	5'-ATGAAAAGCCTGAACTCACC-3'
Hygro 3'	5'-CTATTCCTTTGCCCTCGGAC-3'
<i>HYGU</i> 1	5'-CATCAGCTCATCGAGAGC-3'
JS-2A	5'-GATTGGCGCAACAACCTTTCACATACG-3'
JS-2B	5'-CCCTTTCTTCCTTGGCCATTGTTTTACTAT-3'
M13 For	5'-GTAAAACGACGGCCAG-3'
M13 Rev	5'-GGAAACAGCTATGAC-3'
<i>MLHI</i> 3'(1)	5'- <u>AAGGATCC</u> GCTCAGCAACGCTCGAAGACC-3'; <i>Bam</i> HI sequence underlined
<i>MLHI</i> 3'(A)	5'-GG <u>ICTAGAGATATC</u> CTCTTCTTTCTGTCGTGG-3'; <i>Xba</i> I sequence underlined and <i>Eco</i> RV sequence double underlined
<i>MLHI</i> 3'(B)	5'- <u>AACTCGAGAGG</u> AAACAAGCCACAAACACGCCG-3'; <i>Xho</i> I sequence underlined

Table 2.1 continued

Primer Name	Primer Sequence
MLHI 5'(1)	5'- <u>CCGAATTCAATTCGGACATAATGTTTCGC</u> -3'; <i>EcoRI</i> sequence underlined
MLHI 5'(A)	5'- <u>GGAATTCTCTTAGATATGGGTATGC</u> -3'; <i>EcoRI</i> sequence underlined
MLHI 5'(B)	5'- <u>GGTCTAGAAAATGTGTGAAAGAGCGG</u> -3'; <i>XbaI</i> sequence underlined
MLHID1	5'-AAGCGAGACAGAAGAGAACC-3'
MLHID2	5'-AGTTTGIGATGTACCCGCGCG-3'
MLHID3	5'-GGCTTCGGCATCAGTCACGG-3'
MLHID4	5'-GGCACAGCAGAGAGAGAGG-3'
MLHID5	5'-CGAACGAAGACAAACACGG-3'
MLHID6	5'-CCAGCAACACTACCATCG-3'
MLHID7	5'-GAGTGAATCTGCCGAATATCC-3'
MLHID8	5'-CCA'FCAGCACTGATTTCC-3'
MLHIU1	5'-GCATTCAGGTGGTTGTTTCAGG-3'
MLHIU2	5'-AGTGTACAGTGGTGTGTTGG-3'
MLHIU3	5'-TGGTGGTGTGGCTGCTGTGG-3'
MLHIU4	5'-GAGGATGGTGAAGAAGTGG-3'
MLHIU5	5'-GGAAATCAGTGCTGATGG-3'
MLHIU6	5'-CGCGGGTACATCACAAACTGC-3'
MLHIU7	5'-ATGCACACCAACAGACCG-3'
MSH2 3'(1)	5'- <u>CCGGATCCDA'IGCCCAIGCIARNCCRAA</u> -3'; <i>BamHI</i> sequence underlined
MSH2 5'(1)	5'- <u>CCGAATTCATHRTIACIGGICCNAAAYATG</u> -3'; <i>EcoRI</i> sequence underlined
MSH2 3' (A)	5'- <u>GGAAGCTTGATATCTGGTGGAGGTGTGAGAAGG</u> -3'; <i>HindIII</i> sequence underlined and <i>EcoRV</i> sequence double underlined
MSH2 3'(B)	5'- <u>AACTCGAGCGAGGTACTGAAGTAAGG</u> -3'; <i>XhoI</i> sequence underlined

Table 2.1 continued

Primer Name	Primer Sequence
<i>MSH2</i> 5'(A)	5'-GTGAATTCGCACAGATGTGGAATGAGGG-3'; <i>EcoRI</i> sequence underlined
<i>MSH2</i> 5'(B)	5'-CCAAGCTTGTTAACTTGCTGACTACACACCCG-3'; <i>HindIII</i> sequence underlined and <i>HpaI</i> sequence double underlined
<i>MSH2D1</i>	5'-CCGAATTCGGAGGTAAGTCAACTTTCATG-3', <i>EcoRI</i> sequence underlined
<i>MSH2D2</i>	5'-GGAGCATTGGTGTGGTGTGTCG-3'
<i>MSH2D3</i>	5'-GCTACGGATTGTACGTGGC-3'
<i>MSH2D4</i>	5'-GCTGAGCGAGCAGTATAAGG-3'
<i>MSH2D5</i>	5'-GCTGATCACGCAGGTAAGTGC-3'
<i>MSH2F1</i>	5'-GTTGTGTCAGTTATGTGGAGGG-3'
<i>MSH2F2</i>	5'-CGAGTCCTTTCTTCTGTGCC-3'
<i>MSH2F3</i>	5'-ATACGACCGACAGACGAGCG-3'
<i>MSH2F4</i>	5'-AAAGCCTTCTAGAATCCGCC-3'
<i>MSH2J1</i>	5'-CAAGAGAGAAAGGGACTGGG-3'
<i>MSH2J2</i>	5'-GTTGTGTCATTTCGTGG-3'
<i>MSH2J3</i>	5'-AAGCTGCTGTCGTGCCCGCC-3'
<i>MSH2J4</i>	5'-ATCGAATGGATCTATGGCGG-3'
<i>MSH2U1</i>	5'-CCGGATCCTCCATCGTACGTTGACGTCC-3'; <i>BamHI</i> sequence underlined
<i>MSH2U2</i>	5'-CCTTATACTGCTCGCTCAGC-3'
<i>MSH2U3</i>	5'-GCAGTACCTGCGTGATCAGC-3'
<i>MSH2U4</i>	5'-TCAATAACCACACTCCC-3'
<i>MSH2U5</i>	5'-CACAGAAGAAAGGACTCGG-3'
<i>MSH2U6</i>	5'-CAGCGTCATCAGAGGAGG-3'
<i>MSH8</i> 5'	5'-TACGAACGTCGCGAAGACCC-3'
<i>MSH8</i> 3'	5'-CGAGCTCCAGAGACTTAACC-3'
<i>MSH8D1</i>	5'-CGAGCACTCGTACTTCCG-3'

Table 2.1 continued

Primer Name	Primer Sequence
<i>MSH8D2</i>	5'-AGCTGCTGAGGTGTCAGGGG-3'
<i>MSH8D3</i>	5'-CAGTGGTGTTCGCTACTTGC-3'
<i>MSH8D4</i>	5'-CCTTCTACTCTTCTCTACCC-3'
<i>MSH8D5</i>	5'-GTGTCATCACTACAAGCCGG-3'
<i>MSH8D6</i>	5'-TACGACCAAGATGCCGC-3'
<i>MSH8D7</i>	5'-CGAATGGCAAGGAGAGCG-3'
<i>MSH8D8</i>	5'-CCATTGCGAGTTTCTGCG-3'
<i>MSH8 DN(1)</i>	5'-CACCTTCTCCATCGCCTCC-3'
<i>MSH8U1</i>	5'-GGTAGAGAAGAGTAGAAGGG-3'
<i>MSH8U2</i>	5'-GCATAGACGTAGCAAACACTGG-3'
<i>MSH8U3</i>	5'-GACACACGCTGAAGTACGGG-3'
<i>MSH8U4</i>	5'-GATGCTTCTTGTACTGCACG-3'
<i>MSH8U5</i>	5'-CGGAAGTACGAGTGCTCG-3'
<i>MSH8U6</i>	5'-CCTCCTTTAAGCTCCTCTCC-3'
<i>MSH8U7</i>	5'-GCTGATCATCATGGGATCGG-3'
<i>MSH8U8</i>	5'-ACAAAGAGCGAAAGGC-3'
<i>MSH8U9</i>	5'-GGGTGAGATCATGAAAGC-3'
<i>MSH8 UP(1)</i>	5'-TGGTGCTTTCTAAGGGTCCG-3'
Neo 5'	5'-TTGCACGCAGGTTCTCCG-3'
Neo 3'	5'-GAACTCGTCAAGAAGGCG-3'
PLCG	5'-CAACGACGTTGGAAGAGTGTGAAC-3'; labelled with 6-FAM
PLCH3	5'- CCACTGACCTTTCATTTGATCGCTTTC-3'
<i>PMS1 3'(2)</i>	5'-ATGACGAGCATTAGGTTCCG-3'
<i>PMS1 5'(2)</i>	5'-TACGACTCGATGTGGAGC-3'
<i>PMS1D1</i>	5'-GGTTGGTGGTGCTGGTTGG -3'
<i>PMS1D2</i>	5'-GCACCAGCAACTTATTGG-3'

Table 2.1 continued

Primer Name	Primer Sequence
<i>PMSID3</i>	5'-CAGCAAGGGCTCGCTGGG-3'
<i>PMSID4</i>	5'-GCTTCGCCACGAAAACCC-3'
<i>PMSID5</i>	5'-CAATGGTGCACGTGTGGG-3'
<i>PMSID6</i>	5'-TTCGACACCTTTGCCTCG-3'
<i>PMSID7</i>	5'-TAGTTGCCATTGAGTGCC-3'
<i>PMSIU1</i>	5'-CCTTCTTCCTGCATGTCAGC-3'
<i>PMSIU2</i>	5'-GTTGCTGGTGCTATCAGTGC-3'
<i>PMSIU3</i>	5'-GGTCTTTCGTTCGCAGGC-3'
<i>PMSIU4</i>	5'-GCTCACGTTGTATTTGGG-3'
<i>PMSIU5</i>	5'-GTCGCTTTTTCTACACGG-3'
<i>PMSIU6</i>	5'-GATGGACTAAAGAAGACG-3'
PolI 5'	5'-CAGGAGGATCGTTCGGCACCTTGGC-3'
PolI 3'	5'-CATGCGCCTGTGGTTCAGCATAGC-3'
ResFor	5'-TTCGTAAGTGGTGGTGG-3'
ResRev	5'-GGAGAGAGAGAAAGAGG-3'
T3	5'-AATTAACCCTCACTAAAGGG-3'
T7	5'-TAATACGACTCACTATAGGG-3'
VSG221 5'	5'- <u>CCGAATTC</u> CGCATGCCTTCCAATCAGGAGGC-3'; <i>EcoRI</i> sequence underlined
VSG221 3'	5'-CGCGGATCCGCTGTATCGGCGACAACCTGCAG-3'; <i>BamHI</i> sequence underlined

**Table 2.2** Unicellular parasite genome sequencing initiatives

For each organism the centre undertaking the genome sequencing project is shown, along with the name of the project. The web address (URL) of the project homepage is also given in each case, along with any publications pertinent to that genome sequencing initiative. This list is not exhaustive, for a number of organisms several centres are involved in independent or collaborative sequencing projects. For each organism, only the details of the sequencing databases searched during this study are shown.

<b>Organism</b>	<b>Sequencing Centre/Project</b>	<b>Web Address (URL)</b>	<b>Publications</b>
<i>Babesia bovis</i>	The Wellcome Trust Sanger Centre / <i>Babesia bovis</i> EST Sequencing Project	<a href="http://www.sanger.ac.uk/Projects/B_bovis/">http://www.sanger.ac.uk/Projects/B_bovis/</a>	
<i>Encephalitozoon cuniculi</i>	Genoscope / <i>Encephalitozoon cuniculi</i> – Complete Genome	<a href="http://www.genoscope.cns.fr/externe/English/Projets/Projet_AD/AD.html">http://www.genoscope.cns.fr/externe/English/Projets/Projet_AD/AD.html</a>	Katinka <i>et al.</i> , 2001
<i>Entamoeba histolytica</i>	The Wellcome Trust Sanger Centre / <i>Entamoeba histolytica</i> Whole Genome Shotgun	<a href="http://www.sanger.ac.uk/Projects/E_histolytica/">http://www.sanger.ac.uk/Projects/E_histolytica/</a>	
<i>Giardia lamblia</i>	Josephine Bay Paul Center for Comparative Molecular Biology and Evolution, Marine Biological Laboratory, Woods Hole / <i>Giardia lamblia</i> Genome Project	<a href="http://www.mbl.edu/Giardia/index2.html">http://www.mbl.edu/Giardia/index2.html</a>	McArthur <i>et al.</i> , 2000
<i>Leishmania major</i>	The Wellcome Trust Sanger Centre / <i>Leishmania major</i> Friedlin Genome Project	<a href="http://www.sanger.ac.uk/Projects/L_major/">http://www.sanger.ac.uk/Projects/L_major/</a>	
<i>Plasmodium falciparum</i>	The Wellcome Trust Sanger Centre / Malaria Genomic Project and PlasmoDB, The Plasmodium Genome Resource (Release 3.3)	<a href="http://www.sanger.ac.uk/Projects/P_falciparum/">http://www.sanger.ac.uk/Projects/P_falciparum/</a> and <a href="http://plasmodb.org/PlasmoDB.shtml">http://plasmodb.org/PlasmoDB.shtml</a>	Kissinger <i>et al.</i> , 2001; Bahl <i>et al.</i> , 2002
<i>Theileria annulata</i>	The Wellcome Trust Sanger Centre / <i>Theileria annulata</i> Sequence Projects	<a href="http://www.sanger.ac.uk/Projects/T_annulata/">http://www.sanger.ac.uk/Projects/T_annulata/</a>	

**Table 2.2 continued**

<i>Toxoplasma gondii</i>	The Wellcome Trust Sanger Centre / <i>Toxoplasma gondii</i> Sequencing Project	<a href="http://www.sanger.ac.uk/Projects/T_gondii/">http://www.sanger.ac.uk/Projects/T_gondii/</a>	
<i>Trypanosoma brucei</i>	The Institute for Genomic Research / The TIGR <i>Trypanosoma brucei</i> Genome Project and The Wellcome Trust Sanger Centre / <i>Trypanosoma brucei</i> Genome Project	<a href="http://www.tigr.org/tdb/mdb/tbdb/">http://www.tigr.org/tdb/mdb/tbdb/</a> and <a href="http://www.sanger.ac.uk/Projects/T_brucei/">http://www.sanger.ac.uk/Projects/T_brucei/</a>	
<i>Trypanosoma cruzi</i>	The Institute for Genomic Research / The TIGR <i>Trypanosoma cruzi</i> Genome Project and Parasite-Genome, the parasite genome databases and genome research resources at the European Bioinformatics Institute (EMBL-EBI)	<a href="http://www.tigr.org/tdb/e2k1/tea1/">http://www.tigr.org/tdb/e2k1/tea1/</a> and <a href="http://www2.ebi.ac.uk/blast2/parasites.html">http://www2.ebi.ac.uk/blast2/parasites.html</a>	

## **CHAPTER 3**

**Identification of three putative MutS  
homologues and two putative MutL  
homologues from *T. brucei***

### **3.1 Cloning and characterisation of a putative MutS Homologue 2 (MSH2) orthologue in *T. brucei***

#### **3.1.1 Introduction**

The amino acid sequences of MutS homologues from a diverse range of organisms have been shown to be highly conserved, especially over the ATP-binding domain and the helix-turn-helix motif (Eisen, 1998; Culligan *et al.*, 2000). The primary MutS homologue involved in eukaryotic MMR is MSH2, as this protein is required for all nuclear MMR reactions to proceed (Harfe and Jinks-Robertson, 2000). Orthologues of MSH2 can be distinguished from other eukaryotic mismatch-interacting MutS-related proteins by the closer proximity of their mismatch interacting domains to their N-termini (except *T. brucei* MSH8; see below). Furthermore, MSH2 orthologues evolved from a common ancestor (Eisen, 1998; Culligan *et al.*, 2000) and this is reflected in multiple alignments of MutS homologues, which reveal that there is a greater similarity between the ATP-binding domains of MSH2 proteins than there is between MSH2 orthologues and other MutS homologues (data not shown). Knowledge of the characteristics of MSH2 orthologues in other eukaryotes allowed the isolation of a gene from *T. brucei* which putatively encodes a homologue of MSH2.

#### **3.1.2 Identification of a putative *T. brucei* MSH2 orthologue by degenerate PCR**

The putative *T. brucei* MSH2 gene was initially identified by degenerate PCR using primers directed against the highly-conserved ATP-binding motif. A GCG pileup alignment of MSH2 polypeptides from a number of representative eukaryotes revealed several candidate regions which could be back-translated allowing the design of suitable degenerate primers. The 5' primer (*MSH2* 5' (1)) was designed to recognise a short motif known to be present in most *MSH2*, *MSH3* and *MSH6* homologues (Figure 3.1). However, to increase the likelihood of amplifying *MSH2* rather than any other homologue, the 3' primer (*MSH2* 3' (1)) was designed to recognise a short motif specific to *MSH2*. PCR amplification from MITat 1.2 genomic DNA using these primers at 2  $\mu$ M each (see section 2.9.1) produced a 315 bp product which was cloned and sequenced (Figure 3.2). From the sequence data it was evident that this product encoded the predicted region of a putative MSH2 homologue (Figure 3.9).

### 3.1.3 Cloning of the complete ORF of the putative *T. brucei* *MSH2* gene from a genomic lambda library

The clone of the 315 bp degenerate PCR product was then restriction-digested with *EcoRI* and *Bam*HI and the sequence encoding the ATP-binding domain used as a probe to screen an ILTat 1.2 genomic lambda library in order to recover the complete *MSH2* gene sequence. As described in Materials and Methods (see section 2.6), the library was put through four rounds of screening, and a single clonal plaque (designated Plaque 1) was recovered. DNA was prepared from this phage clone and mapped by restriction digestion. Initially, the DNA was singly digested to assess which restriction enzymes produced fragments of a suitable size for cloning. Southern analysis of the products of these digestions revealed that *EcoRI*, *SacI* and *SalI* gave fragments of between 4.0 and 8.0 kb which could therefore potentially encode the entire ORF of *MSH2* (Figure 3.3). Furthermore, *SacII* cut twice within the sequence amplified in the degenerate PCR and gave two informative bands on a Southern blot, allowing orientation of the restriction fragments as one is brighter since it is complementary to a larger region of the probe (the third fragment was too short for the probe to illuminate). The phage DNA was then double digested using *SacII* and either *EcoRI*, *SacI* or *SalI*. The products of these digestions were then Southern blotted and the results analysed (Figure 3.4). Although the *SacII* and *EcoRI* double digestion appeared to be partial, in the Southern blot the probe seemed to have hybridised only to bands of the expected size. The final restriction map of Plaque 1 was compiled using the restriction digestion data, the known sequence of the bacteriophage- $\lambda$  arms, and the short sequence identified by degenerate PCR (Figure 3.5). Analysis of these data revealed that the 4.5 kb *EcoRI* fragment should contain the entire putative *MSH2* ORF along with its processing flanks. DNA from Plaque 1 was therefore digested with *EcoRI* and 'shot-gun' cloned into pBluescript II KS. From the resulting sub-clones one (pJB100) that contained a 4.5 kb insert was chosen for sequencing. This revealed an ORF encoding a putative polypeptide 951 amino acids in length, as well the 5' and 3' processing flanks required for splicing of the mRNA. A sequence map of the 4.5 kb insert is shown in Appendix 1.

### 3.1.4 Southern analysis of the putative *T. brucei* *MSH2* locus

Both MITat 1.2 and ILTat 1.2 genomic DNA were restriction-digested with a panel of enzymes to analyse the genomic environment of the putative *T. brucei* *MSH2* ORF. The restriction digestions were separated by agarose gel electrophoresis, Southern blotted, and then probed with the same probe as was used in section 3.1.3, revealing the gene to be

present in single copy within the *T. brucei* genome (Figure 3.6). Unfortunately, the *EcoRV* restriction digestion of both types of genomic DNA appeared to be partial, and in the lane containing the ILTat 1.2 *Sall* digestion there was an air bubble in the agarose. Despite this, the restriction pattern in both the MITat 1.2 and ILTat 1.2 digestions were identical and it can be concluded that the genomic environment of this gene is the same in both strains of *T. brucei*. In each genomic digestion the probe hybridised to fragments of the same size as in the lambda digestions, thereby confirming the map of clone Plaque 1 (Figure 3.5).

### 3.1.5 Analysis of the expression of the putative *T. brucei* MSH2 gene

The expression of the putative *MSH2* gene was investigated by RT-PCR using gene-internal primers specific to the *MSH2* ORF (*MSH2D5* and *MSH2U2*; see Appendix 1 and section 2.4.2). The integrity of the cDNA was confirmed using primers specific to the large subunit of *T. brucei* RNA polymerase I (PolI 5' and PolI 3'; see Figure 3.7; Rudenko *et al.*, 1996), and it is apparent that the putative *MSH2* gene is expressed in both procyclic and bloodstream stage cells. No conclusions can be drawn about the level of expression of the gene in the different life cycle stages since, under the conditions used, RT-PCR is not quantitative.

### 3.1.6 Pairwise comparison of the putative *T. brucei* MSH2 polypeptide with other MutS homologues

A conceptual translation of the putative *T. brucei* *MSH2* ORF was compared with a range of MutS homologues from *E. coli*, *A. thaliana*, *H. sapiens*, and *S. cerevisiae* using the GCG gap program. The results of these pairwise alignments are shown in Table 3.1. A graphical representation of these data shows that the putative *T. brucei* *MSH2* polypeptide shares between 5.5 and 13.0% greater identity (or 6.0 and 12.0% greater similarity) with the *MSH2* orthologues than with the *MSH3* or *MSH6* homologues shown (Figure 3.8). While similarity does not necessarily equate to homology, this evidence suggests that the *T. brucei* gene encodes an orthologue of *MSH2*.

### 3.1.7 Multiple alignment of the putative *T. brucei* MSH2 polypeptide with eukaryotic MSH2 orthologues

A global multiple alignment of the putative *T. brucei* *MSH2* polypeptide with a number of eukaryotic *MSH2* orthologues was produced with CLUSTAL X (Thompson *et al.*, 1997) using the default settings (Figure 3.9). The alignment shows diffuse conservation

throughout the polypeptides, including the N-terminal mismatch-interacting domain (Malkov *et al.*, 1997; Culligan *et al.*, 2000; Schofield *et al.*, 2001), the middle conserved domain (Culligan *et al.*, 2000) and the helix-turn-helix motif (Adé *et al.*, 1999; Biswas *et al.*, 2001). A region of high identity corresponds to the ATP binding domain (Gorbalenya and Koonin, 1990). The putative *T. brucei* MSH2 shows conservation across the four functional domains characteristic of MSH2 orthologues, but is divergent compared with the other MSH2 orthologues shown, especially over the N-terminal region required for interaction with mismatches. From this alignment, the putative *T. brucei* MSH2 appears to contain three insertions of 8, 14 and 10 amino acids between the N-terminal mismatch-interacting domain and middle conserved domain and a 12 amino acid insertion between the middle conserved and ATP-binding domains. Since the function of the amino acid residues which lie between the conserved domains is unknown and the 3-dimensional structure of MSH2 has not been solved, the functional implications of these insertions, if any, cannot be predicted.

### 3.1.8 The genomic environment of the putative *T. brucei* MSH2 ORF in TREU 927/4

Recent searching of the Sanger *T. brucei* genome sequence database revealed a large contig, TRYP10.0.001883 (29,749 bp), composed of 324 reads from chromosome X, which encodes the putative *MSH2* gene. While this sequence is unfinished, it can be concluded that the putative *MSH2* gene is present on chromosome X. Comparison of the putative *MSH2* ORFs from ILTat 1.2 and TREU 927/4 shows the nucleotide sequences differ at 15 positions, and this is reflected in the conceptual translations of the ORFs as the two polypeptide sequences differ by 5 amino acid residues (data not shown). Although some of these changes may be sequencing artefacts, they have affected amino acid residues which are not conserved between the MSH2 orthologues shown in the alignment in Figure 3.9. If these changes do reflect real differences between the two trypanosome cell lines, they are unlikely to impact upon the function of the protein as the residues affected are unlikely to have catalytic roles. Further analysis of the contig sequence (see section 2.8.4) reveals that the putative *MSH2* gene is approximately 8.6 kb downstream of a hypothetical ORF which bears homology to a *T. brucei* flagellar antigen (Figure 3.10). The contig sequence also contains a number of other hypothetical ORFs, but BLAST searches of nucleotide and protein databases retrieve only insignificant alignments.

**Figure 3.1 Design of degenerate PCR primers to amplify an *MSH2* orthologue from *T. brucei***

A GCG pileup of a portion of the sequence of the *MSH2* orthologues from *A. thaliana* (Ath), *S. cerevisiae* (Sce), *S. pombe* (Spo), *H. sapiens* (Hsa) and *N. crassa* (Ncr) is shown. The pileup was shaded using the BOXSHADE server ([http://www.ch.embnet.org/software/BOX\\_form.html](http://www.ch.embnet.org/software/BOX_form.html)) and identical residues are boxed in black, conserved residues in grey. Only the sequences around the ATP-binding domain are shown. Residues marked with an asterisk were back-translated to generate degenerate primers to PCR amplify the intervening *MSH2* gene sequence; blue asterisks correspond to primer *MSH2* 5'(1), and red to primer *MSH2* 3'(1).

```

NcrMSH2 620 .....GRTVLT*EARHPCMEVQDDVTFITNDVTLTREDSSFLIITGPNMG *****
SceMSH2 649 .....RRTHLISSRH*PVL*EMODDISFISNDVTLESGKGDFLIITGPNMG
HsaMSH2 630 .....GRIILKASRHAC*VEVODETAFIPNDVYFEKDKQMFHIITGPNMG
SpoMSH2 666 TPNFEEIRR*TL*ENHCARLYLKQARHPCLEAQQDVKFI*PNDVNLEHGSSELLIITGPNMG
AthMSH2 626 .....AGDIVLEGSRH*PCVEAQQDW*NFIPND*CR*LMRGKSWFQIVTGP*NMG

NcrMSH2 664 GKSTYIRQIGVIALMAQIGCFVPCSSAELTTFDSILARVGASDSQLKGVSTFMAEMLETA
SceMSH2 693 GKSTYIRQIGVVISLMAQIGCFVPC*EAEIATVDAILCRVGAGDSQLKGVSTFMVEILETA
HsaMSH2 674 GKSTYIRQIGVIVLMAQIGCFVPC*EAEVSTVDCILARVGAGDSQLKGVSTFMAEMLETA
SpoMSH2 726 GKSTYIRQIGVITVMAQIGCFVPC*EVAIDIDAILARVGASDSQLKGVSTFMAEMLETA
AthMSH2 671 GKSTYIRQIGVIVLMAQIGSEFVPCDKASISIRDCIFARVGAGDCQLR*GVSTFEMOEMLETA

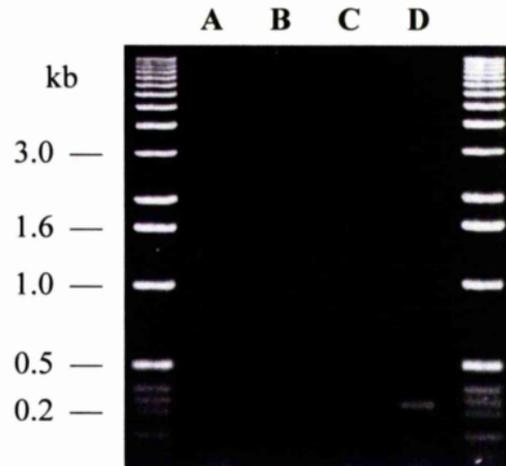
NcrMSH2 724 NILKSA*TAESLIIDE*LG*RG*TS*TYDGFGLAWAISEHTVKEIGCFALFATHFHELTALADQ *****
SceMSH2 753 SILKNASKNSLIIVDELGRG*TS*TYDGFGLAWAIAEHTASKIGCFALFATHFHELT*ELSEK
HsaMSH2 734 SILRSATKDSLIIIDE*LG*RG*TS*TYDGFGLAWAISEYATKIGAF*CMFATHFHELTALANQ
SpoMSH2 786 TILRAATPRSLIIIDE*LG*RG*TS*TYDGFGLAWAITEHTV*IQIGCFCLFATHYHEM*TKLSEE
AthMSH2 731 SILKASDKSLIIIDE*LG*RG*TS*TYDGFGLAWAICEHTVQVKR*APTLEATHFHELTALAQ.

NcrMSH2 784 YPNVKNLHVTAHISGTD*TDVITDEDEKAKKKREVTLLYKVEPCICTQSFGIHVAELVR
SceMSH2 813 LPNVKNMHVVAHIE.....KNLKEQKH*DD*EITLLYKVEPGISDQSFGIHVAEVVQ
HsaMSH2 794 IPTVNNLHVTA.....LTTE*TL*MLYQVKKGVCDQSFGIHVAELAN
SpoMSH2 846 ITTVKNLHVTA*YVGDSE*.....KDVALLYNVCEGASDRSFGIHVAKLAH
AthMSH2 790 .....ANSEVSGNTVGVANFHVSAHIDTESRKLTMLYKVEPCACTQSFGIHVAEFAN

```

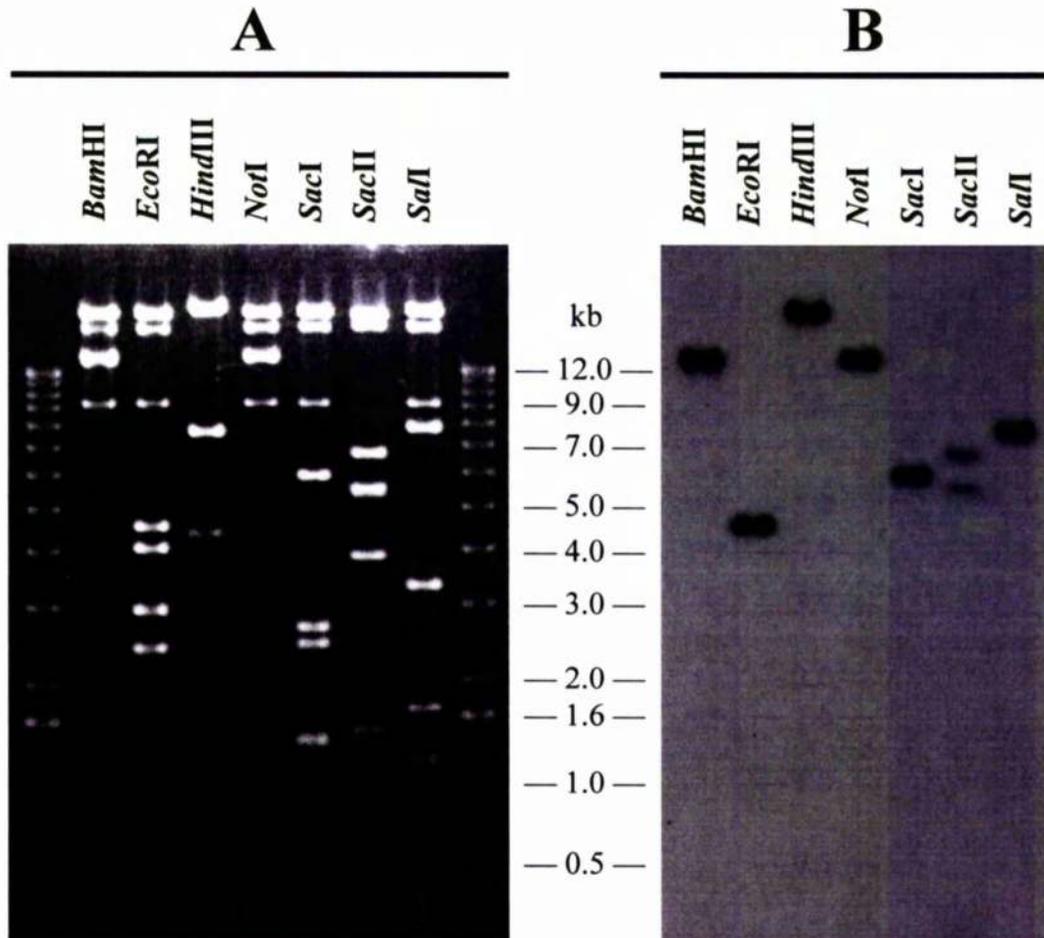
**Figure 3.2 Degenerate PCR of *MSH2***

Primers *MSH2* 5'(1) and *MSH2* 3'(1) were used with MITat 1.2 genomic DNA for PCR. Control reactions were performed with both primers in the absence of *T. brucei* DNA template (A), the *MSH2* 5'(1) primer only (B), and the *MSH2* 3'(1) primer only (C). Lane D shows the reaction containing both primers and template.



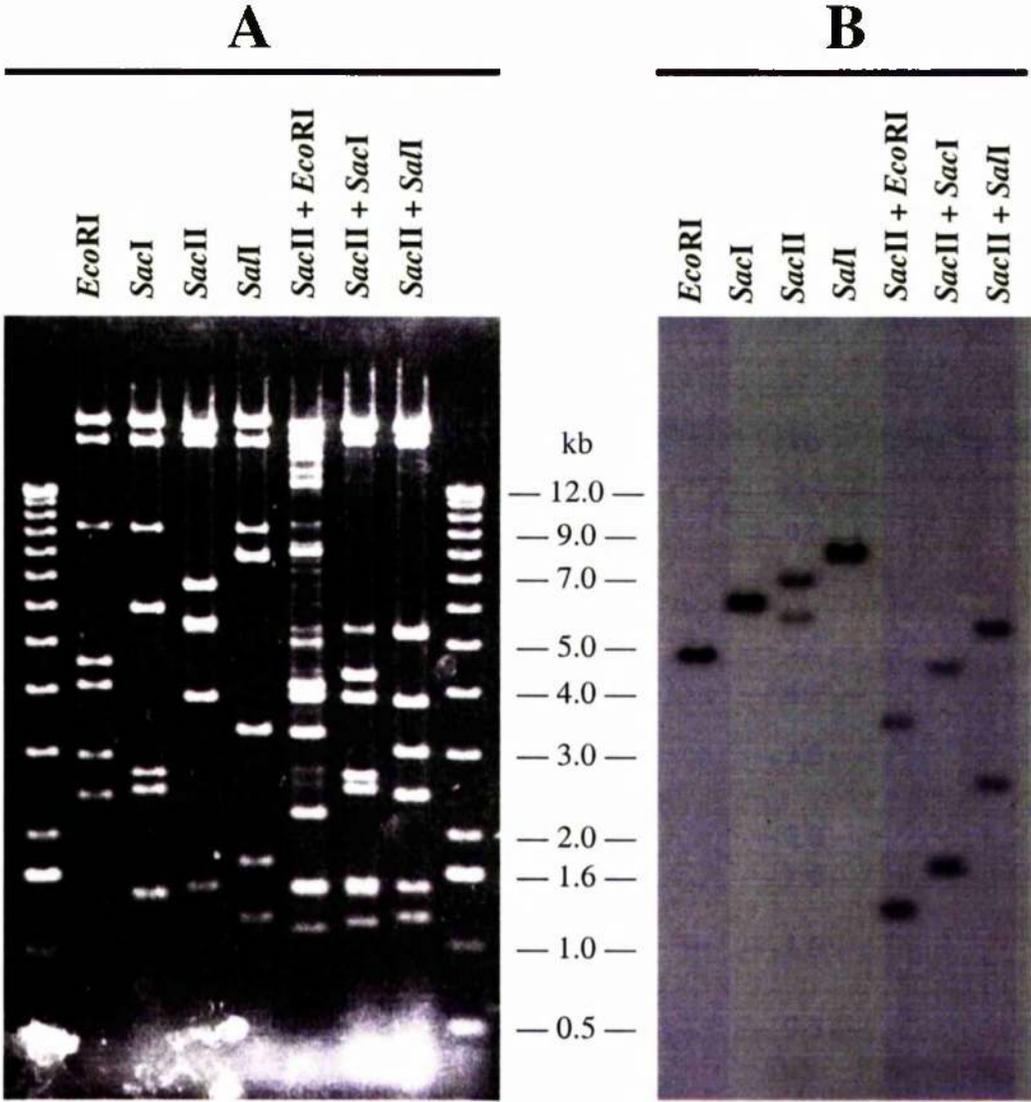
**Figure 3.3 Restriction mapping of Plaque 1 - single digestions**

(A) DNA isolated from the genomic lambda clone Plaque 1 was digested with the enzymes shown and separated on a 0.6% agarose gel. (B) A Southern blot of the restriction digestions shown in (A) probed with a 315 bp degenerate PCR product amplified from the putative *T. brucei* MSH2 gene.



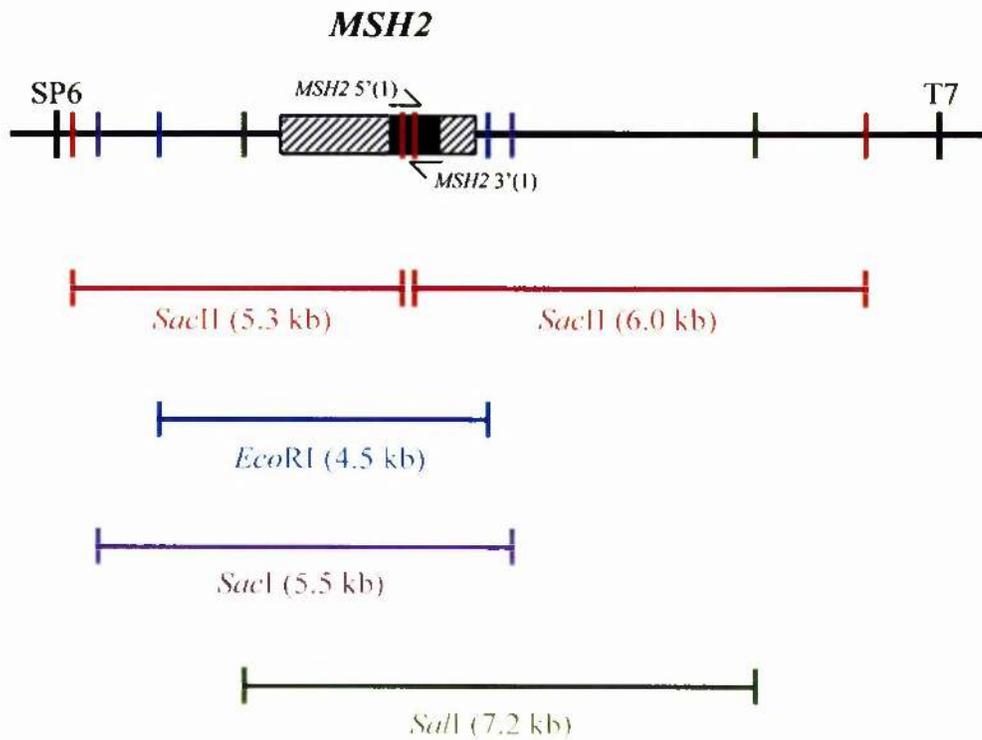
**Figure 3.4** Restriction mapping of Plaque 1 - double digestions

(A) DNA isolated from the genomic lambda clone Plaque 1 was digested with the enzymes shown and separated on a 0.6% agarose gel. (B) A Southern blot of the restriction digestions shown in (A) probed with a 315 bp degenerate PCR product amplified from the putative *T. brucei* MSH2 gene.



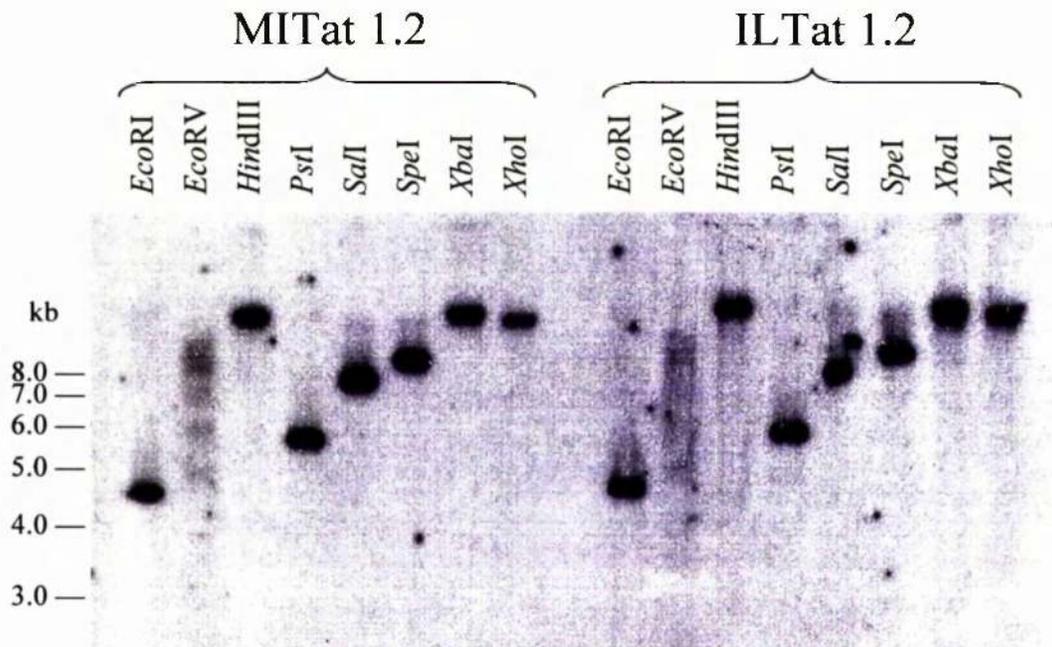
**Figure 3.5 Restriction map of the putative *T. brucei* MSH2 lambda clone Plaque 1**

A restriction enzyme map of Plaque 1, a lambda clone encoding the putative *T. brucei* MSH2 gene, which was retrieved from the ILTat 1.2 genomic lambda library (not to scale). The putative MSH2 ORF is denoted by hatched shading. The probe region is shown as a black box with the primers MSH2 5' (1) and MSH2 3'(1) (shown as black half arrows) which were used to produce the probe fragment indicated above and below. Restriction enzyme sites are shown as coloured bars, with the resulting fragments indicated below: *Sac*II (red); *Eco*RI (blue); *Sac*I (purple); *Sal*I (green).



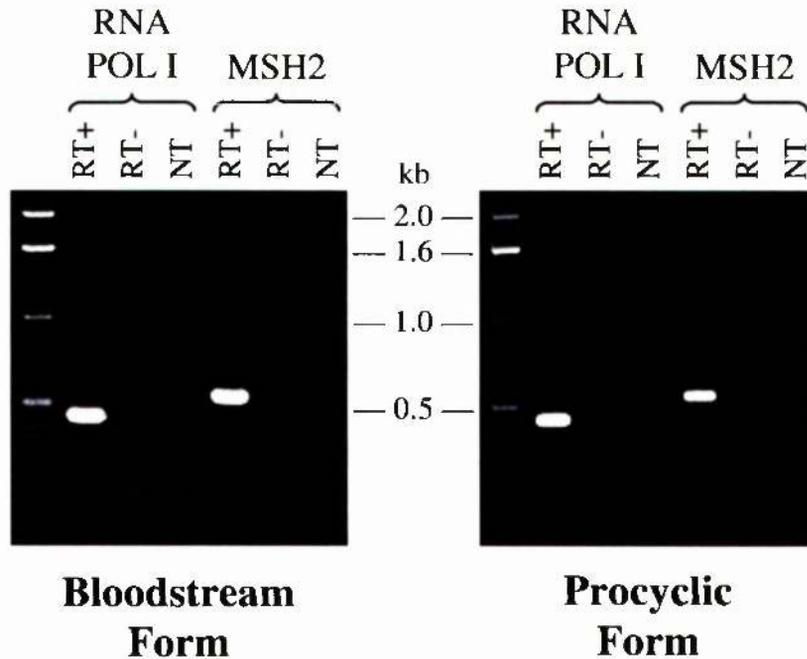
**Figure 3.6 Genomic Southern blot probed for the putative *T. brucei* MSH2 gene**

A Southern blot of MITat 1.2 and ILTat 1.2 genomic DNA which was restriction-digested with the enzymes shown, separated on a 1.0% agarose gel, probed with a 315 bp degenerate PCR product amplified from the putative *T. brucei* MSH2 gene and washed to 0.2 x SSC, 0.1% SDS at 65°C.



**Figure 3.7 Analysis of the expression of the putative *T. brucei* MSH2 gene**

RT-PCR was performed on both bloodstream and procyclic form total RNA isolated from MITat 1.2. Internal primers complementary to the putative *T. brucei* MSH2 gene were used and the integrity of the RNA was confirmed using primers directed against the large subunit of *T. brucei* RNA polymerase I (Rudenko *et al.*, 1996). For both genes RT positive (RT+), RT negative (RT-) and no template (NT) reactions were performed.



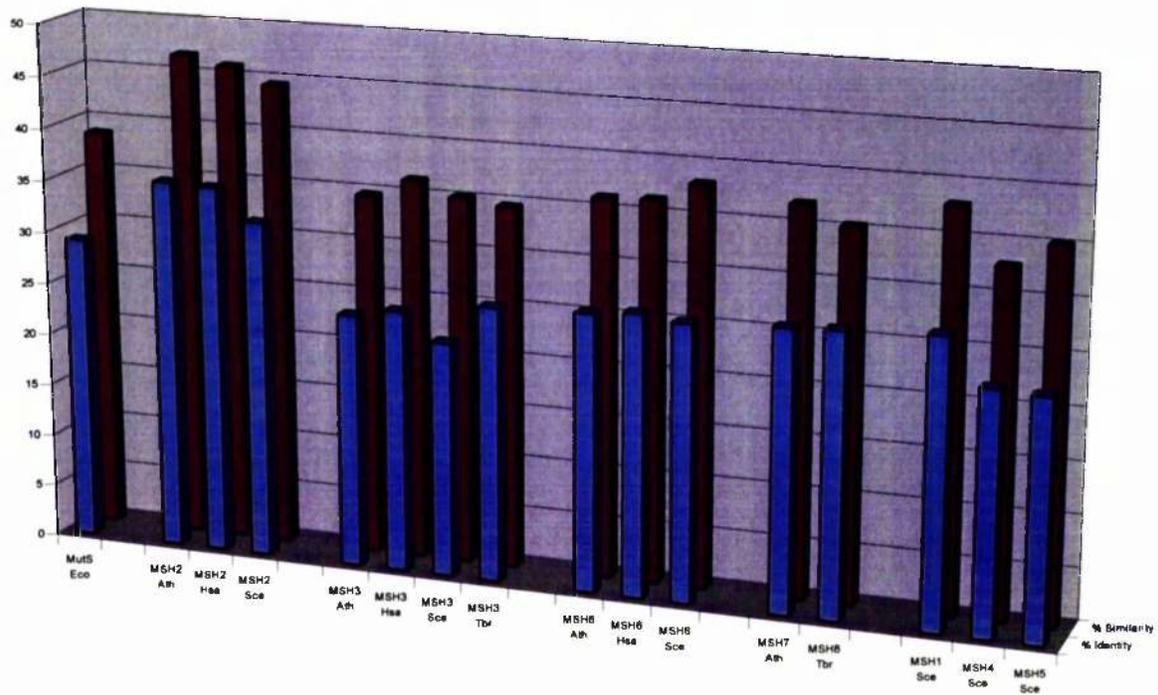
**Table 3.1** Pairwise comparison of the putative *T. brucei* MSH2 polypeptide with a range of MutS homologues

Pairwise gap alignments were performed using the GCG gap program. The full length putative *T. brucei* MSH2 polypeptide sequence was compared against the full length polypeptide sequences for a range of MutS homologues. For each alignment the percentage identity and similarity between the two proteins was calculated.

MutS homologue	<i>T. brucei</i> MSH2	
	% Identity	% Similarity
<i>E. coli</i> MutS	29.14	38.57
<i>A. thaliana</i> MSH2	35.43	46.68
<i>H. sapiens</i> MSH2	35.27	45.91
<i>S. cerevisiae</i> MSH2	32.28	44.43
<i>A. thaliana</i> MSH3	24.14	34.59
<i>H. sapiens</i> MSH3	24.94	36.36
<i>S. cerevisiae</i> MSH3	22.38	35.06
<i>T. brucei</i> MSH3	26.15	34.58
<i>A. thaliana</i> MSH6	26.42	36.17
<i>H. sapiens</i> MSH6	26.80	36.31
<i>S. cerevisiae</i> MSH6	26.33	38.11
<i>A. thaliana</i> MSH7	26.80	37.03
<i>T. brucei</i> MSH8	26.98	35.35
<i>S. cerevisiae</i> MSH1	27.38	37.98
<i>S. cerevisiae</i> MSH4	22.98	32.86
<i>S. cerevisiae</i> MSH5	22.74	35.12

**Figure 3.8** A graphical representation of the similarity between the putative *T. brucei* MSH2 polypeptide and other MutS-related proteins

Pairwise alignments were performed as described in Table 3.1 to compare the putative *T. brucei* (Tbr) MSH2 polypeptide sequence with other MutS-related polypeptide sequences from: *A. thaliana* (Ath), *E. coli* (Eco), *H. sapiens* (Hsa) and *S. cerevisiae* (Sc). Percentage identity is shown in blue and percentage similarity in red.



**Figure 3.9 Global multiple alignment of the putative *T. brucei* MSH2 polypeptide with a range of MSH2 orthologues**

Multiple sequence alignment of the putative *T. brucei* MSH2 (TbrMSH2) polypeptide (shown in red) with homologues of MSH2 from other eukaryotes: *A. thaliana* MSH2 (AthMSH2), *D. melanogaster* Spellchecker 1 (DmeMSH2), *H. sapiens* (HsaMSH2), *S. cerevisiae* MSH2 (SceMSH2), and *X. laevis* MSH2 (XlaMSH2). Sequences were aligned using CLUSTAL X (Thompson *et al.*, 1997) and shaded using the BOXSHADE server ([http://www.ch.embnet.org/software/BOX\\_form.html](http://www.ch.embnet.org/software/BOX_form.html)): identical residues are shaded in black, and conserved residues in grey. MIS is the N-terminal mismatch interacting domain (shown in purple; Malkov *et al.*, 1997; Culligan *et al.*, 2000; Schofield *et al.*, 2001), MCD is the middle conserved domain (shown in turquoise; Culligan *et al.*, 2000), the peptides identified as I, II, III and IV are the conserved nucleotide binding motifs (shown in blue; Gorbalenya and Koonin, 1990) and HTH is the conserved helix-turn-helix motif (shown in green; Adé *et al.*, 1999; Biswas *et al.*, 2001). The region identified beneath the *T. brucei* sequence corresponds to the region of the *T. brucei* MSH2 gene that was amplified by degenerate PCR and used as a probe (shown in orange).

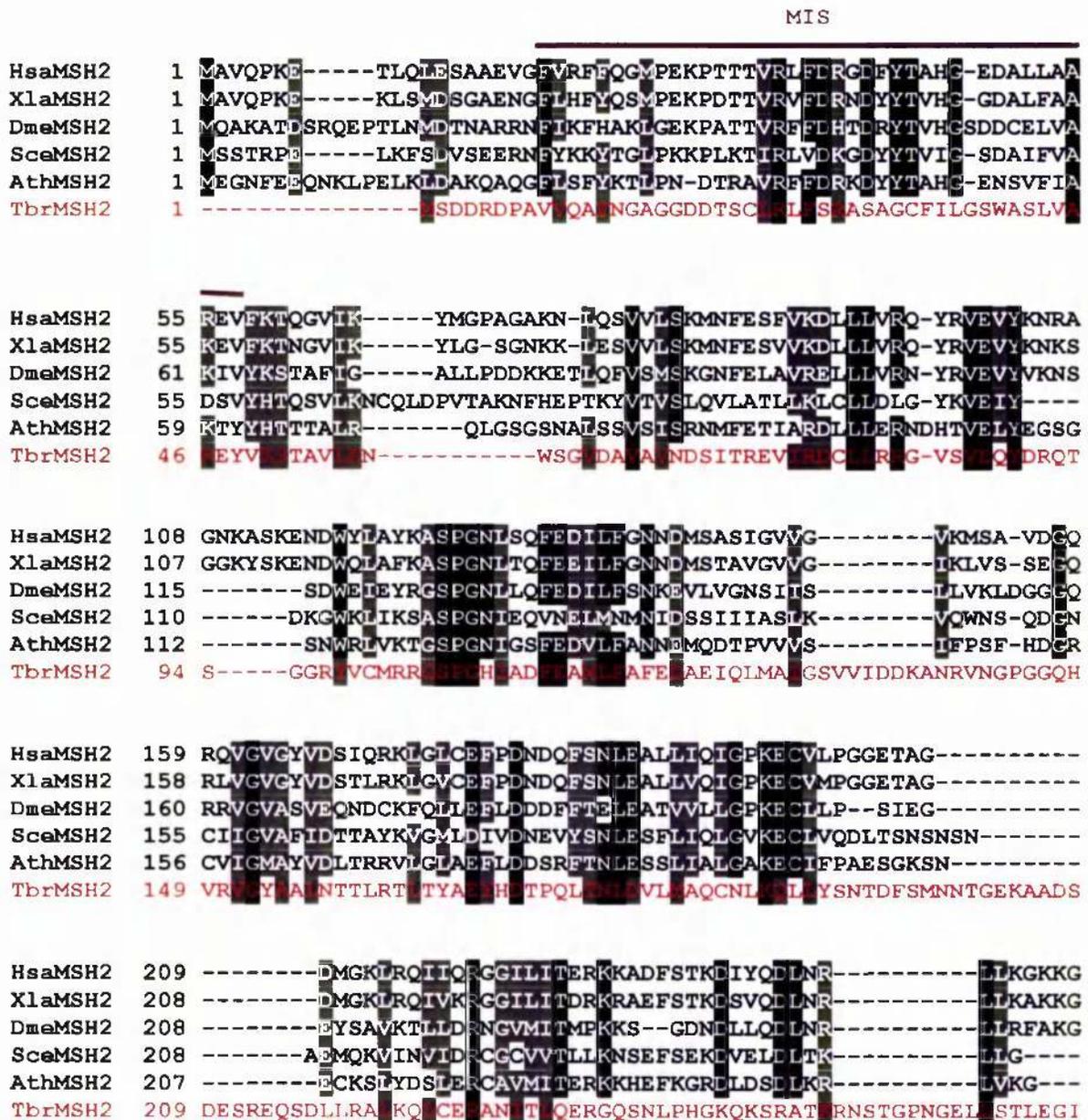


Figure 3.9 continued

HsaMSH2 251 EQMNSAVLPENENQVAVSSLSAVIKLELLSDDSNFGQFELTTDFDFOYMKLIDIAAVRAL  
 XlaMSH2 250 EQVTSAAPEMEKQVAMSALAAVMKYLELLSDESNFGQFVMTNFDLSOYMKLIDIAAVGAL  
 DmeMSH2 248 QQEDATGLKELQQLASNAKTAIKYLLDVLNDAGNLGHYEIKQLDLNRFVHLDIAAVAL  
 SceMSH2 247 -DDLALSPLQKYSKLSMGACNALICYLQLLSEQDOVQKVELVEHKLKEFMKLDASAIAKAL  
 AthMSH2 246 -NIEPVRDLVSGFDLAPALGALLSESELLSNEDNYENFTIRRYDIGGFMRLLDIAAMRAL  
 TbrMSH2 269 LRPEDRHLNSFP--SR--ES--ESAIDPF--STQHTY--KHVIPST--A--E--

MCD

HsaMSH2 311 NLFQGSVEDTTGS-----QSLAALLN-KCKTPQGGQLVNQWIKQPLMD  
 XlaMSH2 310 NLFPGSAEDTSGT-----QSLAGLLN-KCKTPQGGQLVNQWIKQPLMD  
 DmeMSH2 308 NIMPKPGTHPSMPS-----YRWQSVLGVLD-HCRTPQGHILMGQVWKQPLRS  
 SceMSH2 306 NLFPGQPQNPFGSNNLAVSGFTSAGNSGKVTSLFOLLN-HCKTNAGVRLINEWLKQPLTN  
 AthMSH2 305 NVMES-KTDANKN-----FSLFGLMNRCTAGMCKKLLHMWLKQPLVD  
 TbrMSH2 329 H--IHRKPEARSGMP-----T--Y--S--W--T--G--M--S--Q--Q--L--L--R--S--

HsaMSH2 353 KNRTEERLNLVEAFVDEAELRQTEQEDLLRRFPDLNRLAKKFORQAAANLQTCYRLYQGIN  
 XlaMSH2 352 KNRVEERLNLVEAFVMDVELRQCQEDLLRRFPDLNRLAKKFORQTANLQTCYRLYQAVN  
 DmeMSH2 354 RNILNDRHNIVQCLLESPTMETSLDYLKRIPOILMLTKKLMRRKANLQILERYQVIL  
 SceMSH2 365 IDEINKRHLDLVDYLIDQIELRMLTSEYLPMPDIRRLTKKLNKR-GNDELVLKIYQFSK  
 AthMSH2 347 LNEIKTRLDIVQCFVEEAGLRQDQR-QHLKRISQVERLLRSLERRRGLQHIKLYQSTI  
 TbrMSH2 372 IED--NQ--LS--QIM--SPI--DA--ITQV--CT--ED--N--LQ--RTVA--K--LQS--LVFAN

HsaMSH2 413 QLPNVIQALEKHEGKH-----QKLLLAVFVTPPLTDLRSD--FSKFQEMIEITLLMDQVE  
 XlaMSH2 412 HLPTVQALEKYEGTH-----QMLLAVFATPLSDLSSD--FSKFQEMIEITLLMDQVE  
 DmeMSH2 414 RTPKILKVLHIELD-----NSTTESVICAPFKSFLKD--LTGLKQMVQVWLFEEAIE  
 SceMSH2 424 RIPEIVQVFTSFLEDDSPTEPVNELVRSVWLAPLSSHVPEP--LSKFEEVETTVDDLDAYE  
 AthMSH2 406 RLFFIKTAAQQYTGEF-----ASLISERYLKKLEALSQDHLGKFDLVECSVDLDQLE  
 TbrMSH2 432 T--LA--DV--RTYHGGHGHD-----SS--LKG--VT--EDISEH--L--NLRT--NA--SDEN

HsaMSH2 465 -NHDFLVKPSFDPNLSREIMNDEKKYQSTLISAARDLGLDPGKQIKLDSQAQFCYYE  
 XlaMSH2 464 -NHDFLVKASFDPNLTDREKMDLEKKNMQGALGGAARELGLDAGKSIKLESNSQICHYE  
 DmeMSH2 463 -RGEYLVKASFDRLMELQOMMTELYSKMEELQFKCSQELNLDGKNQVKLESVAKLGHHE  
 SceMSH2 482 ENNDFMKVEFNEELGKRSKLDTRDEIHSIHLSAEDLGFDPDKKLENNHLLHWCM  
 AthMSH2 460 -NGFYMSSSYDTKLASLKDQKELLEQQIHELHKKTAIELDLQVDKALKLDDAAQFQHV  
 TbrMSH2 485 ---TVR--NPE--DD--SF--ERQRQNVKAEKENHRVLKQCGWT-E--Q--C--YHASY--YV--

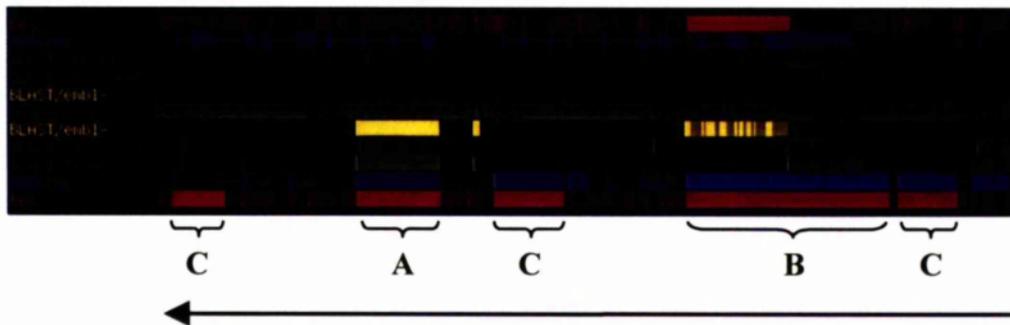
HsaMSH2 524 RVTCKEEKVLRN--NKNSTVDIQKNGVKFTNSKLTSLNEEYTKNKTEYEEA--DAIVKEI  
 XlaMSH2 523 RVTCKEEKALRN--NKNSTVDIQKNGVRFNTSKLSSLSSEYMRNREYEEA--NAIVKEI  
 DmeMSH2 522 RRTVKDDSVLRK--NKNRIVDVIKGVRFNTSKLSDKLEGYADEEASCRTRYEEQ--LSIVEEI  
 SceMSH2 542 RLTRNDAKELRK--HKYIELSTVKAGIFFSTKQLKSIANETNILEQYDKQ--SALVREI  
 AthMSH2 519 RITKKEEPKIRKLLTQELIVLETRKQGVKFTNTKLLKLGDOYQSVVDDYRSCCKELVDRV  
 TbrMSH2 541 --PR--HQ--T--S--E--I--T--S--T--A--D--V--V--S--G--Q--S--S--Q--K--G--I--T--E--K--T--R--Q--V--K--K--K--

HsaMSH2 582 VNISSGYVEPMQTLNDVLAQLDAVVSFAHVSNGAPVPPYVRPAILEK-----G  
 XlaMSH2 581 ITISAGYVDPIQTLNDVIAQLDAVVSFAHVSNSAPVPPYVRPVILEK-----G  
 DmeMSH2 580 IHVAVGYAAPLTLNNELAQLDCLVSAFAIAARSAPTPYVRPKMLEE-----G  
 SceMSH2 600 INITLTYTPVFEKLSLVLHLDVITASFAHTSSYAPIPYIRPKLHMPD-----S  
 AthMSH2 579 VETVTSESEVFEDLACLLSEMDVLLSFADLAASCPTPYCRPEITSS-----D  
 TbrMSH2 599 --DTVAT--LPVLDDAKE--A--VFA--L--LVVKDSSR--M--T--RATQSEEVKGNVDNNSD



**Figure 3.10 Analysis of the genomic environment of the putative *T. brucei* *MSH2* gene**

The nucleotide sequence of contig TRYP10.0.001883 was analysed using the Nucleotide Identification (NIX) server (<http://www.hgmp.mrc.ac.uk/Registered/webapp/nix/>) provided by the UK Human Genome Mapping Project Resource Centre. The output shows the results of selected programs which were used in the analysis. The sequence regions identified correspond to: (A) the *T. brucei* *MSH2* ORF; (B) a hypothetical ORF bearing similarity to a *T. brucei* flagellar antigen; and (C) a number of hypothetical ORFs suggested by the HMMGene gene-prediction and Fex exon-prediction programs. The predicted direction of transcription is indicated by a black arrow.



## **3.2 Identification of two further putative MutS homologues, MSH3 and MSH8, in *T. brucei***

### **3.2.1 Introduction**

During nuclear mismatch repair in yeast and humans MSH2 forms a heterodimeric complex with another MutS homologue, either MSH3 or MSH6 (Acharya *et al.*, 1996; Kolodner and Marsischky, 1999). This situation is more complicated in plants as another protein, called MSH7, which is closely related to MSH6 and which probably arose via duplication of this gene, has also been identified and forms a third heterodimer with MSH2 in these organisms (Culligan and Hays, 2000). Furthermore, in eukaryotes such as *D. melanogaster* and *C. elegans*, for whom the complete genome sequence is available, no homologues of MSH3 have been identified and these organisms are thought to rely solely on the mismatch repair capabilities of the MSH2-MSH6 heterodimer (Sekelsky *et al.*, 2000). As the sole partner for MSH2 in these organisms it is possible that MSH6 might fulfil at least some, if not all, of the roles that MSH3 performs in yeast and mammals. However, the functions of the MSH6 homologues in *D. melanogaster* and *C. elegans* have yet to be determined experimentally.

In the absence of functional data, putative homologues of MSH3, MSH6 and MSH7 can be distinguished from MSH2 orthologues as they possess N-terminal extensions upstream of the mismatch interacting domain (data not shown). It is also possible to distinguish between MSH3, MSH6 and MSH7 homologues as they vary greatly in length, and differ within the sequence of their mismatch interacting domains (see Figure 3.21). MSH3 and MSH7 proteins are generally 1000 to 1190 amino acids in length, whereas MSH6 proteins are usually between 1250 and 1370 amino acids in length (data not shown). However, the sequence of the mismatch interacting domain of MSH7 is similar to that of MSH6 (Culligan and Hays, 2000). Knowledge of the characteristics of the MSH3/MSH6 sub-family of MutS-related proteins allowed identification of two homologues from *T. brucei*, firstly a gene designated *MSH8*, and more recently a probable orthologue of MSH3.

### **3.2.2 Identification of a second putative MutS homologue by searching the *T. brucei* genome sequencing databases**

Similarity searching of the TIGR and Sanger sequencing databases using the MSH3 and MSH6 polypeptide sequences from *A. thaliana*, *H. sapiens* and *S. cerevisiae* revealed a

number of sequences bearing homology to these MutS-related proteins. Alignment of these sequences using the GCG fragment assembly programs produced two contigs which appeared to span the C-terminal half of the same gene because one of the contigs contained the M13 forward sequence of TIGR clone 30I14 and the other contained the M13 reverse sequence (Figure 3.11). From the available sequence it was possible to assign the gene as a putative member of the *MSH3/MSH6* sub-family, but as the sequence was incomplete, no further conclusions could be drawn.

### 3.2.3 Cloning and sequencing of the complete ORF of the putative *T. brucei* *MSH8* gene

PCR primers (*MSH8* 5' and *MSH8* 3') were designed from these contigs which would amplify the majority of the available gene sequence. PCR was performed on MITat 1.2 genomic DNA with high fidelity polymerase (see section 2.9.2) and generated a product approximately 2.0 kb in length (Figure 3.12). This product was cloned into pPCR-Script Amp SK (+) (the resulting plasmid was called pJB301), and sequencing revealed an incomplete ORF 631 amino acids in length.

The sequence generated above was used to search the *T. brucei* sequencing databases at regular intervals in an attempt to identify further homologous sequences recently added to the databases. After several months, a number of shotgun clone sequences became available which aligned with the sequence of the 2.0 kb PCR product (Figure 3.13). Where available, the sequence from the opposite end of these clones would therefore have to lie either upstream or downstream of the known sequence for *MSH8*. The M13 forward sequence of clone 109F12 allowed the design of a primer (*MSH8* UP(1)) which, when used with a primer complementary to the sequence of the 2.0 kb PCR product (*MSH8*U3), would amplify the 5' end of the gene. Furthermore, the M13 reverse sequence of clone 90D12 revealed that the *MSH8* gene was upstream of *T. brucei* *IISP60* (Bringaud *et al.*, 1995). A primer directed against this sequence (*MSH8* DN(1)) used in conjunction with a primer complementary to the sequence of the 2.0 kb PCR product (*MSH8*D3) would amplify the 3' end of the gene and confirm this linkage. Using these two sets of primers, two overlapping PCR products were amplified from MITat 1.2 genomic DNA (see section 2.9.2): the first, approximately 2.9 kb in length, corresponded to the 5' end of the gene (*MSH8* Upstream), and the second, approximately 2.2 kb in length, contained the sequence of the 3' end of the gene (*MSH8* Downstream; Figure 3.14). These products were cloned into pPCR-Script Amp SK(+) (generating the plasmids pJB302 and pJB303 respectively). The plasmids were sequenced and the data assembled with the sequence of the original

2.0 kb PCR product using the GCG package. A single contig resulted which upon translation revealed an ORF 997 amino acids in length, along with 5' and 3' processing flanks. The *HSP60* start codon was located approximately 1.0 kb downstream of the *MSH8* stop codon. A sequence map of the *MSH8* gene is shown in Appendix 2.

### 3.2.4 Southern analysis of the putative *T. brucei* *MSH8* locus

Both MITat 1.2 and ILTat 1.2 genomic DNA were restriction-digested with a panel of enzymes to analyse the genomic environment of the putative *T. brucei* *MSH8* ORF. The restriction digestions were separated by agarose gel electrophoresis and Southern blotted. PCR primers were designed from the sequence encoding the ATP-binding domain (*MSH8D3* and *MSH8U1*; see section 2.9.2) and used to amplify a 540 bp product from MITat 1.2 genomic DNA (Figure 3.15). This PCR product was then used as a probe for the Southern blot. This hybridisation revealed the gene to be present in single copy in the *T. brucei* genome (Figure 3.16). Furthermore, since the restriction pattern in both the MITat 1.2 and ILTat 1.2 digestions were identical it can be concluded that the genomic environment of this gene is the same in these strains of *T. brucei*.

### 3.2.5 Analysis of the expression of the putative *T. brucei* *MSH8* gene

The expression of the putative *MSH8* gene was investigated by RT-PCR using gene-internal primers specific to the *MSH8* ORF (*MSH8D3* and *MSH8U1*; see section 2.4.2). The integrity of the cDNA was confirmed as described in section 3.1.5. The results demonstrate that the putative *MSH8* gene is expressed in both bloodstream and procyclic form cells (Figure 3.17). Since non-quantitative RT-PCR was performed no conclusions about the level of expression in the different life cycle stages can be drawn.

### 3.2.6 Identification of a third putative MutS homologue by searching of the *T. brucei* genome sequencing databases

As the *T. brucei* genome sequencing databases are under continual development, BLAST searching of these resources has been performed at regular intervals to investigate whether any further mismatch repair proteins have come to light. In the last month of my Ph.D. (January 2001), a BLAST search of the Sanger database using the putative *T. brucei* *MSH8* polypeptide sequence identified a contig forming part of chromosome IX which encodes a hypothetical ORF whose conceptual translation bears strong sequence similarity to *MSH3* from *M. musculus* and *H. sapiens*. This hypothetical ORF (TRYP9.0.000864\_4) encodes a

polypeptide 935 amino acids in length which possesses the four conserved domains characteristic of a MutS homologue involved in nuclear MMR (see Figures 1.9 and 3.21). Although the chromosome IX sequence released by Sanger is unfinished and therefore may contain errors, this contig, TRYP9.0.000864, is 12,066 bp in length and composed of 127 reads, suggesting extensive overlaps between the constituent sequence fragments. Furthermore, investigation of the conceptual translation of the hypothetical ORF suggests that the sequence is in frame throughout its length and that the putative MSH3 polypeptide sequence encoded is likely to be correct. Therefore, for the purposes of this analysis the sequence of the hypothetical *MSH3* ORF will be considered correct. A sequence map of this hypothetical ORF is shown in Appendix 3.

### **3.2.7 Pairwise comparison of the putative *T. brucei* MSH3 and MSH8 polypeptides with other MutS homologues**

Using the GCG gap program conceptual translations of the putative *T. brucei* *MSH8* and *MSH3* ORFs were compared with a range of MutS homologues from *E. coli*, *A. thaliana*, *H. sapiens*, and *S. cerevisiae*. The results of these pairwise alignments are shown in Table 3.2. The data are represented in the form of a graph in Figure 3.18 for MSH8 and in Figure 3.19 for MSH3. Analysis of these data reveals that the putative *T. brucei* MSH8 polypeptide shares between 3.5 and 6.0% greater identity (or 1.5 and 8.0% greater similarity) with the MSH6 and MSH7 homologues shown than it does with either the MSH2 or MSH3 homologues used in the analysis. Before the discovery of the putative *T. brucei* *MSH3*, the overall levels of similarity suggested that the *MSH8* gene encoded a member of the MSH3/MSH6 sub-family of MutS homologues, but the data was inconclusive as to whether it was an orthologue of MSH6 or a composite protein encompassing the functions of both MSH6 and MSH3. Since the putative *T. brucei* MSH3 polypeptide shares between 1.0 and 6.0% greater identity (or 4.0 and 8.5% greater similarity) with MSH3 than with either MSH2, MSH6 or MSH7 homologues from other eukaryotes (Figure 3.19), it can be concluded that this gene probably encodes the orthologue of MSH3 in *T. brucei*. This new data from *T. brucei* and extrapolation from the MMR systems of *S. cerevisiae* and *H. sapiens* suggests that MSH8 is an orthologue of MSH6 as this would fulfil the requirements for MMR in these organisms. However, MSH8 is truncated in comparison with MSH6 orthologues from other eukaryotes (see sections 3.2.1, 3.2.8 and 3.2.9), and it may not therefore perform the same functions. Furthermore, plants possess a third member of the MSH3/MSH6 subfamily, MSH7, and since the genome sequence for *T. brucei* is not yet complete it is possible that another

MutS-related gene bearing a greater similarity to *MSH6* will come to light in the future. It is for these reasons we have chosen the nomenclature *MSH8*, rather than *MSH6* or *MSH7*.

### 3.2.8 Investigation of the 5' splicing of the putative *T. brucei* *MSH8* mRNA

As discussed in section 3.2.7, the predicted *T. brucei* *MSH8* polypeptide is truncated in comparison with homologues of *MSH6* at only 997 amino acids in length. Therefore, in order to confirm the conceptual translation of the putative *MSH8* gene, RT-PCR was performed (see section 2.4.2) using a primer directed against the conserved splice acceptor sequence trans-spliced onto all mature mRNAs in *T. brucei* (*EcoRISL(1)*; Ullu *et al.*, 1996) and primers complementary to the 5' end of the *MSH8* gene (*MSH8U7* and *MSH8U8*; Figure 3.20). The integrity of the cDNA was confirmed using primers complementary to the *MSH8* ORF (*MSH8D6* and *MSH8U6*). Amplification from cDNA generated from MITat 1.2 bloodstream form total RNA using the *MSH8D6* and *MSH8U6* gene-internal primers produced two products: a larger specific product (approximately 600 bp) confirming transcription of the *MSH8* gene; and a smaller non-specific product (approximately 350 bp). The primers *EcoRISL(1)* and *MSH8U7* gave a single specific band of the expected size (approximately 600 bp), while amplification using *EcoRISL(1)* and *MSH8U8* gave three bands. The product generated by RT-PCR using the primers *EcoRISL(1)* and *MSH8U7* was cloned, and sequencing confirmed that the hypothesised splice acceptor site was correct, lying 72 bp upstream of the predicted start codon. *MSH8* likely forms part of a multi-gene transcription unit, and so the products generated in the *EcoRISL(1)* and *MSH8U8* RT-PCR probably arose from splicing events further upstream of the *MSH8* splice acceptor site (see Appendix 2). These results suggest that the proposed start codon is correct and that the *MSH8* polypeptide is indeed truncated and no longer than 997 amino acids.

### 3.2.9 Multiple alignment of the putative *T. brucei* *MSH3* and *MSH8* polypeptides with eukaryotic *MSH3*, *MSH6* and *MSH7* homologues

Global multiple alignment of the *MSH3*, *MSH6* and *MSH7* orthologues from *A. thaliana*, *H. sapiens* and *S. cerevisiae* together with the putative *T. brucei* *MSH3* and *MSH8* polypeptides was performed using the CLUSTAL X program (Figure 3.21; Thompson *et al.*, 1997). The multiple alignment parameters used were as follows: gap opening penalty (12.00), gap extension penalty (2.0), delay divergent sequences (40%) and the protein weight matrix used was the gonnet series. In comparison with the putative *T. brucei* *MSH3* and *MSH8* polypeptides, the *MSH3*, *MSH6* and *MSH7* proteins from other

eukaryotes all possess N-terminal extensions, the longest of which belong to the MSH6 orthologues, and which show virtually no sequence conservation. The first region of conservation between these MutS homologues corresponds to the N-terminal mismatch recognition domain (Malkov *et al.*, 1997; Culligan *et al.*, 2000), where the putative *T. brucei* MSII3 polypeptide shows greater similarity to the MSH3 orthologues, whereas the putative *T. brucei* MSH8 polypeptide is more similar to the MSH6 and MSH7 homologues shown. Within the N-terminal mismatch interacting domain a conserved Phe-X-Glu DNA binding motif (Schofield *et al.*, 2001) can be identified in the MSH6 and MSH7 homologues shown. This motif is also present in the *T. brucei* MSH8 polypeptide but not in *T. brucei* MSH3. The middle-conserved domain which is thought to have a role in protein-protein interactions shows a similar pattern of conservation (Culligan *et al.*, 2000), although both MSH3 from *A. thaliana* and the putative *T. brucei* MSH8 polypeptides possess insertions of approximately 25 amino acids in this region. The C-terminal half of the alignment shows a diffuse conservation between the polypeptides, except across the ATP-binding domain where the levels of similarity are very high (Gorbalenya and Koonin, 1990). From this alignment, it can be deduced that the putative *T. brucei* MSH3 polypeptide is more similar to the MSH3 orthologues from other eukaryotes. However, while the putative *T. brucei* MSH8 polypeptide would appear to belong to the MSH6 sub-family of MutS homologues it is impossible to designate it as an orthologue of either MSH6 or MSH7, or as yet another more divergent member of this family. From this alignment it is obvious that the *T. brucei* MSH3 and MSH8 polypeptide sequences are divergent compared with the sequences of the other members of the MSH3/MSH6 subfamily used in the analysis.

Recently, a conserved PCNA binding motif has been identified in MSH3, MSH6 and MSH7 proteins from *A. thaliana*, *H. sapiens* and *S. cerevisiae* (Culligan and Hays, 2000; Warbrick, 2000). This represents a fifth region of conservation in these MutS homologues, and it is normally found near the extreme N-terminus of these proteins. The consensus of this amino acid motif is **Qxxhxxaa**, where **h** indicates hydrophobic residues (commonly isoleucine or leucine) and **a** represents aromatic residues (usually phenylalanine; Flores-Rozas *et al.*, 2000). Figure 3.22 shows the best candidate motif sequences identified by manual local multiple alignment of a number of MSH3/MSH6 family members. Interestingly, the full consensus sequence appears to be absent from the *T. brucei* MSH3 and MSH8 polypeptides, as well as the *D. melanogaster* MSH6 and *Z. mays* MSH7 proteins. It is possible to assign possible partial PCNA binding motifs in all these polypeptides, but whether these would be able to interact with PCNA is unknown. Furthermore, the position of the partial motifs in the *T. brucei* MSH3 and MSH8

polypeptides is incorrect in comparison with the motifs found in other organisms. The MSH8 partial motif is located towards the N-terminus, but the amino acids residues involved form part of the conserved N-terminal mismatch interacting domain and are unlikely to be available for interactions with PCNA. The partial motif identified in *T. brucei* MSH3, on the other hand, is located between the N-terminal mismatch interacting domain and the middle-conserved domain, thereby disrupting the highly conserved domain organisation of the MSH3/MSH6 sub-family, again making it unlikely to be a functional PCNA binding domain.

### **3.2.10 The genomic environments of the putative *T. brucei* MSH8 and MSH3 genes in TREU 927/4**

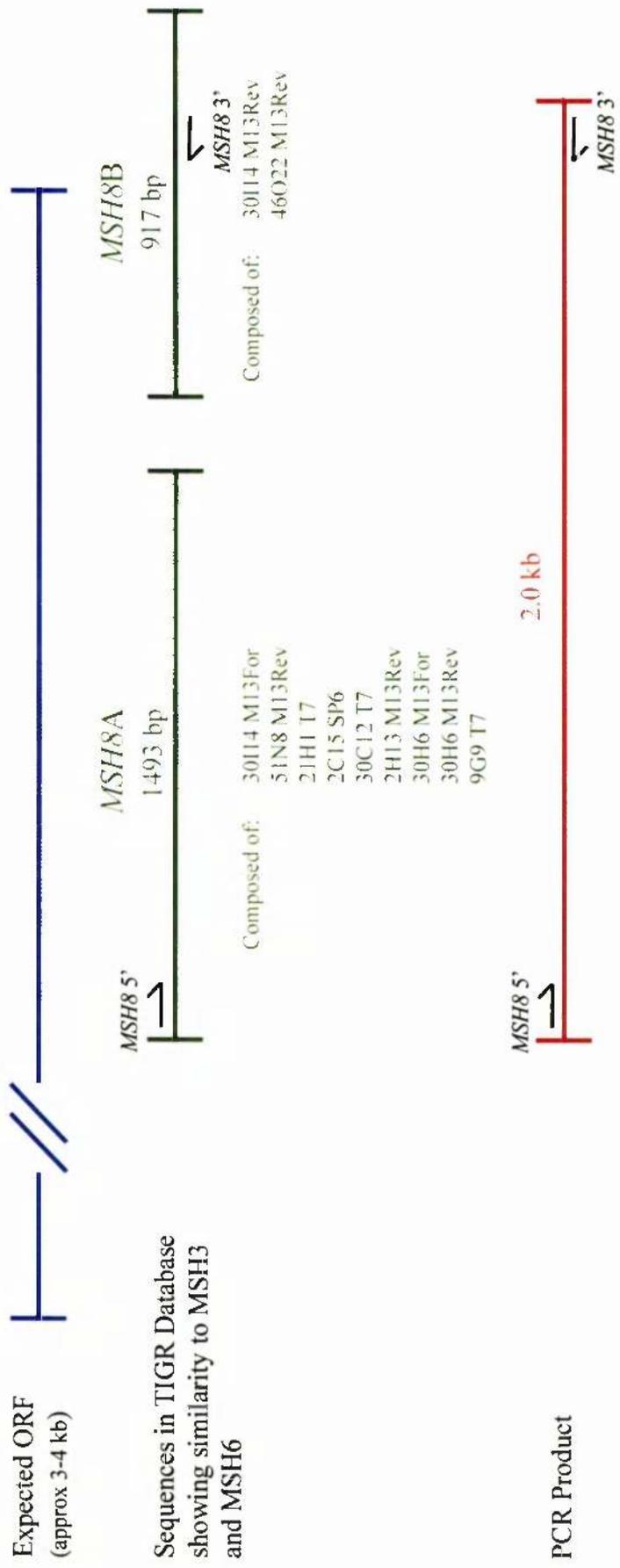
A recent BLAST search of the Sanger *T. brucei* genome sequencing database using the MSH8 ORF retrieved a large contig, TRYP10.0.001896 (37,475 bp), composed of 461 reads from chromosome X. Although the TREU 927/4 genome sequence is unfinished and preliminary it can be concluded that the putative MSH8 gene is present on chromosome X. A comparison of the conceptual translations of the putative MSH8 ORF from MITat 1.2 and TREU 927/4 shows the sequences to differ at only 3 amino acid residues (data not shown). Similarly, the nucleotide sequences only differ at seven positions. Although these changes could be sequencing artefacts, if they are real differences between the two trypanosome stocks they are unlikely to be significant to the overall functioning of the protein because the amino acids residues affected are not conserved between members of the MSH3/MSH6 sub-family of proteins and so are unlikely to have a catalytic role. Further analysis of the contig sequence (see section 2.8.4) confirmed that the putative MSH8 ORF is approximately 1.0 kb upstream of the HSP60 ORF (see section 3.2.3; Bringaud *et al.*, 1995), and reveals that another previously described *T. brucei* gene, encoding a T lymphocyte triggering factor (Vaidya *et al.*, 1997), is approximately 10.0 kb downstream from the gene (Figure 3.23). Upstream of the putative MSH8 ORF, but transcribed from the opposite strand and in the opposite direction, is a gene which has been identified as a CDC2-related protein kinase in *T. brucei* (Mottram and Smith, 1995). There are also a number of hypothetical ORFs whose functions cannot be predicted from the results of BLAST searches of nucleotide and protein databases.

During the BLAST search of the Sanger *T. brucei* genome sequencing database described above another contig was retrieved encoding a 12,066 bp region of chromosome IX (TRYP9.0.000864). Analysis of this contig revealed that it encoded a putative orthologue of MSH3, as described in section 3.2.6. Further analysis of the sequence encoded by this

contig (see section 2.8.4) reveals that the putative *MSH3* gene lies approximately 2.0 kb downstream of a hypothetical ORF encoding a probable glutaminyl-tRNA synthetase, and 4.6 kb downstream of another hypothetical ORF encoding a putative PCNA homologue in *T. brucei* (Figure 3.24). Also encoded within this contig sequence are a number of other hypothetical ORFs, but BLAST searches of the nucleotide and protein databases do not reveal any significant similarities to known sequences.

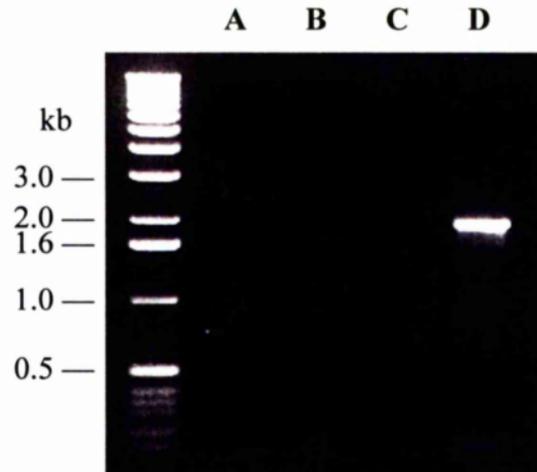
**Figure 3.11 Strategy for PCR amplification of a fragment of the putative *T. brucei* MSH8 ORF**

Comparison of eukaryotic MSH3 and MSH6 homologues indicates that the ORFs encoding these polypeptides are usually 3.0-4.0 kb in length. The expected *T. brucei* MSH8 ORF is shown as a blue bar. Similarity searches of the *T. brucei* genome sequencing databases revealed a number of clones encoding fragments of a putative MutS homologue (shown in green). The nucleotide sequences of these clones were aligned into two fragments: MSH8A and MSH8B (indicated by green bars), which are aligned approximately against the expected ORF. PCR primers (MSH8 5' and MSH8 3'; shown as black half-arrows) were designed from the *T. brucei* MSH8 fragments, and used to amplify a PCR product 2.0 kb in length (shown as a red bar).



**Figure 3.12 PCR of a 2.0 kb region of the putative *T. brucei* *MSH8* gene**

Primers *MSH8* 5' and *MSH8* 3' were used with 20 ng of MITat 1.2 genomic DNA for PCR. Control reactions were performed with both primers in the absence of *T. brucei* DNA template (A), the *MSH8* 5' primer only (B), and the *MSH8* 3' primer only (C). Lane D shows the reaction containing both primers and template.



**Figure 3.13 Strategy for PCR amplification of the putative *T. brucei* MSH8 ORF**

A 2.0 kb fragment encoding the C-terminal end of MSH8 which was cloned and sequenced is shown in red. The expected *T. brucei* MSH8 ORF is shown in blue. Similarity searching of the trypanosome sequencing databases revealed further clones overlapping and extending the original 2.0 kb fragment and showed that *MSH8* lies upstream of the *HSP60* gene (Bringaud *et al.*, 1995); these sequences and the contigs they comprise are shown in purple for upstream sequence and green for downstream sequence. PCR primers (*MSH8* UP(1), *MSH8*U3, *MSH8* DN(1) and *MSH8*D3; shown as black half-arrows) were designed from these sequences and used to amplify two overlapping PCR products: *MSH8* Upstream (shown in orange) and *MSH8* Downstream (shown in black).

over page

Figure 3.13 continued

## MSH8

## HSP60

Expected ORF  
(approx 3-4 kb)



Sequences in TIGR  
Database showing  
similarity to MSH3  
and MSH6

MSH8 UP(1)



Composed of:  
109F12 M13For  
109E12 M13 For



Composed of:  
109F12 M13Rev  
51N8 M13For  
6C11 T7

MSH8D3



MSH8U3



Composed of:  
90D12 M13For



MSH8 DN(1)

Composed of:  
90D12 M13Rev  
21L17 M13Rev

PCR Products

MSH8 UP(1)



MSH8 Upstream

2.9 kb

MSH8U3

MSH8D3



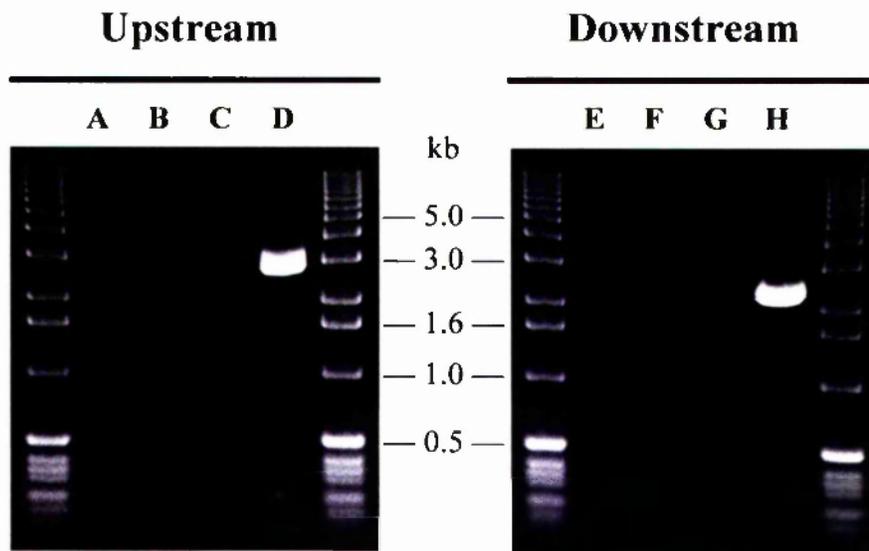
MSH8 Downstream

2.2 kb

MSH8 DN(1)

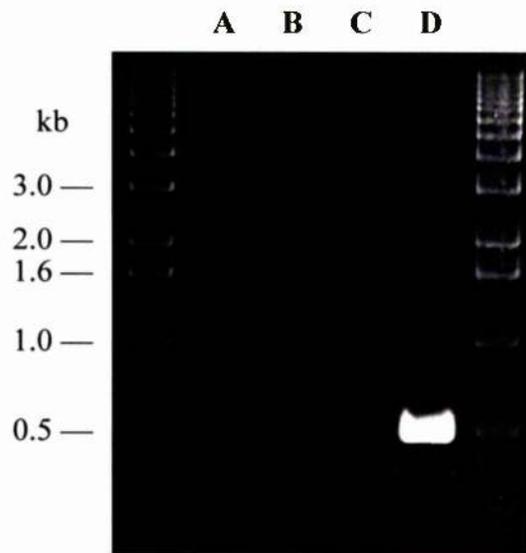
**Figure 3.14 PCR of the upstream and downstream regions of the putative *T. brucei* *MSH8* gene**

To amplify the sequence encoding the 5' end of the *MSH8* gene the primers *MSH8* UP(1) and *MSH8*U3 were used with MITat 1.2 genomic DNA for PCR. Control reactions were performed with both primers in the absence of *T. brucei* DNA template (A), the *MSH8* UP(1) primer only (B), and the *MSH8*U3 primer only (C). Lane D shows the reaction containing both primers and template. Similarly, the primers *MSH8*D3 and *MSH8* DN(1) were used to amplify the 3' end of the gene. Lanes E, F and G show the control reactions performed with both primers in the absence of *T. brucei* DNA template, the *MSH8*D3 primer only, and the *MSH8* DN(1) primer only respectively. Lane H shows the reaction containing both template and primers.



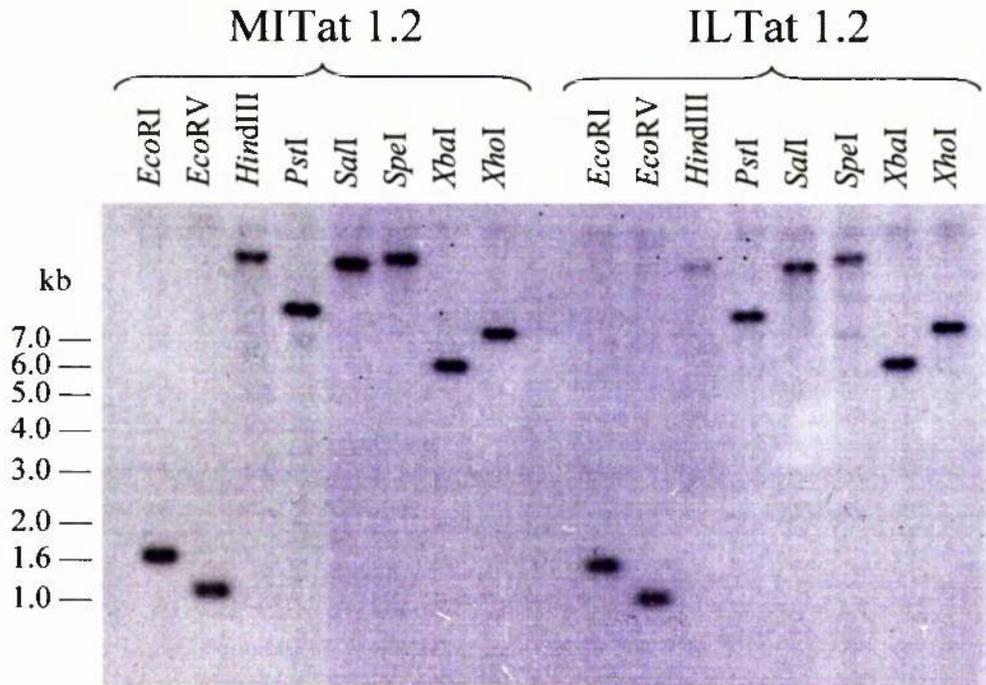
**Figure 3.15 PCR of *T. brucei* MSH8 probe sequence**

Primers *MSH8D3* and *MSH8U1* were used with MITat 1.2 genomic DNA for PCR. Control reactions were performed with both primers in the absence of *T. brucei* DNA template (A), the *MSH8D3* primer only (B), and the *MSH8U1* primer only (C). Lane D shows the reaction containing both primers and template.



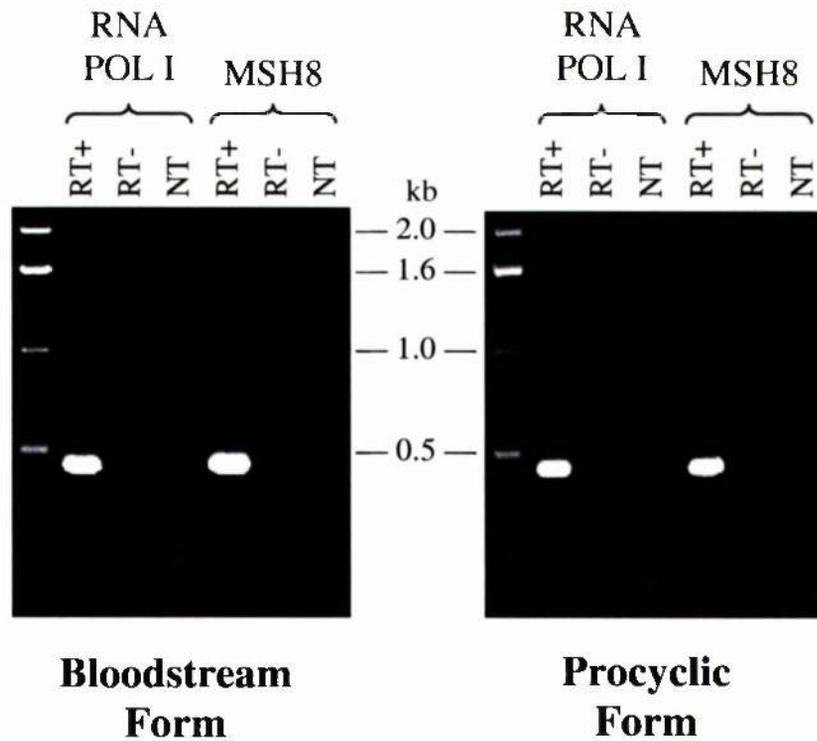
**Figure 3.16** Genomic Southern blot probed for the putative *T. brucei* *MSH8* gene

A Southern blot of MITat 1.2 and ILTat 1.2 genomic DNA which was restriction-digested with the enzymes shown, separated on a 0.6% agarose gel, probed with a 540 bp PCR product amplified from the *T. brucei* *MSH8* gene and washed to 0.2 x SSC, 0.1% SDS at 65°C.



**Figure 3.17 Analysis of the expression of the putative *T. brucei* MSH8 gene**

RT-PCR was performed on both bloodstream and procyclic form total RNA isolated from MITat 1.2. Internal primers complementary to the putative *T. brucei* MSH8 gene were used and the integrity of the RNA was confirmed using primers directed against the large subunit of *T. brucei* RNA polymerase I (Rudenko *et al.*, 1996). For both genes RT positive (RT+), RT negative (RT-) and no template (NT) reactions were run.



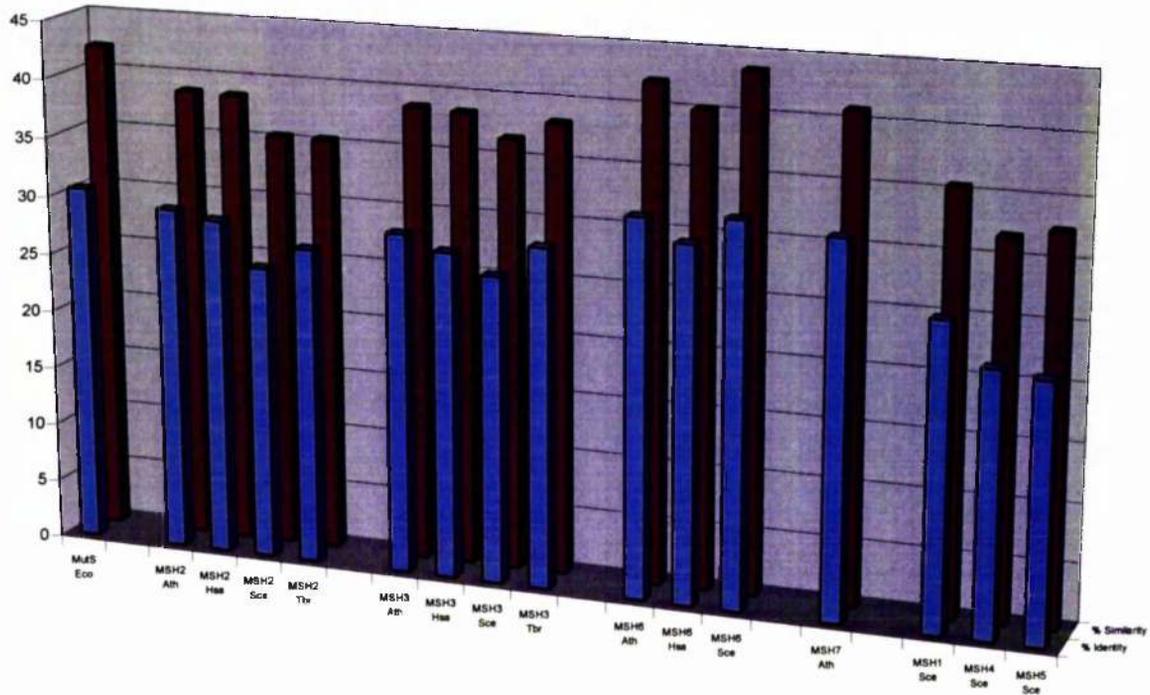
**Table 3.2** Pairwise comparison of the putative *T. brucei* MSH8 and MSH3 polypeptides with a range of MutS homologues

Using the GCG gap program, pairwise alignments between the full length putative MSH8 polypeptide and a range of full length MutS-related protein sequences were performed. Similar alignments were also completed using the full length putative *T. brucei* MSH3 polypeptide sequence. For each alignment the percentage identity and similarity between the two polypeptides was calculated.

MutS homologue	<i>T. brucei</i> MSH8		<i>T. brucei</i> MSH3	
	% Identity	% Similarity	% Identity	% Similarity
<i>E. coli</i> MutS	30.55	42.02	26.84	36.58
<i>A. thaliana</i> MSH2	29.36	38.60	25.42	34.26
<i>H. sapiens</i> MSH2	28.74	38.52	27.74	36.98
<i>S. cerevisiae</i> MSH2	25.00	35.34	24.97	35.34
<i>T. brucei</i> MSH2	26.98	35.35	26.15	34.58
<i>A. thaliana</i> MSH3	28.99	38.69	30.94	42.83
<i>H. sapiens</i> MSH3	27.77	38.52	30.72	42.49
<i>S. cerevisiae</i> MSH3	26.10	36.61	30.43	42.16
<i>T. brucei</i> MSH3	28.85	38.20	100.00	100.00
<i>A. thaliana</i> MSH6	32.07	42.17	28.35	38.36
<i>H. sapiens</i> MSH6	30.19	40.15	29.70	37.80
<i>S. cerevisiae</i> MSH6	32.51	43.61	25.81	37.10
<i>A. thaliana</i> MSH7	31.74	40.96	29.79	38.57
<i>T. brucei</i> MSH8	100.00	100.00	28.85	38.20
<i>S. cerevisiae</i> MSH1	25.84	35.51	27.08	36.93
<i>S. cerevisiae</i> MSH4	22.24	31.84	26.49	36.43
<i>S. cerevisiae</i> MSH5	21.71	32.57	24.42	36.89

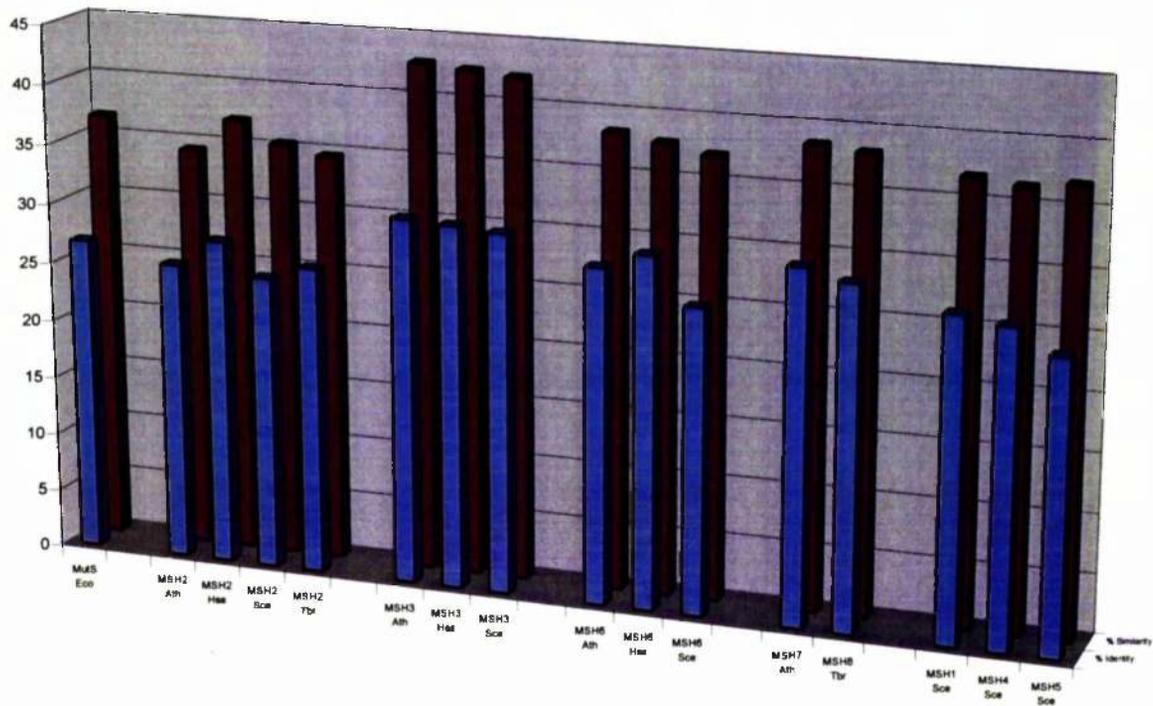
**Figure 3.18** A graphical representation of the similarity between the putative *T. brucei* MSH8 polypeptide and other MutS-related proteins

Pairwise alignments were performed as described in Table 3.2 to compare the putative *T. brucei* (Tbr) MSH8 polypeptide sequence with other MutS-related polypeptide sequences from: *A. thaliana* (Ath), *E. coli* (Eco), *H. sapiens* (Hsa) and *S. cerevisiae* (Sce). Percentage identity is shown in blue and percentage similarity in red.



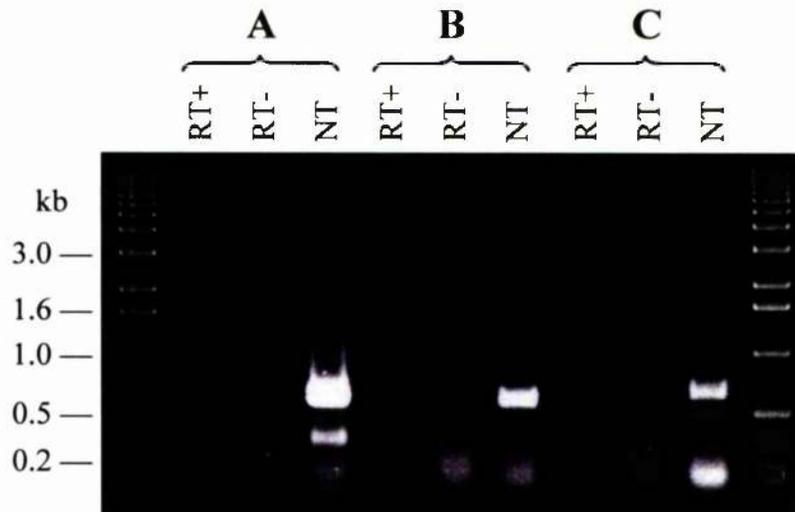
**Figure 3.19** A graphical representation of the similarity between the putative *T. brucei* MSH3 polypeptide and other MutS-related proteins

Pairwise alignments were performed as described in Table 3.2 to compare the putative *T. brucei* (Tbr) MSH3 polypeptide sequence with other MutS-related polypeptide sequences from: *A. thaliana* (Ath), *E. coli* (Eco), *H. sapiens* (Hsa) and *S. cerevisiae* (Sce). Percentage identity is shown in blue and percentage similarity in red.



**Figure 3.20 Analysis of the 5' splicing of the putative *T. brucei* *MSH8* gene**

RT-PCR was performed on bloodstream form total RNA isolated from MITat 1.2. Internal primers complementary to the putative *T. brucei* *MSH8* gene (*MSH8D6* and *MSH8U6*) were used to confirm the integrity of the cDNA (shown in panel A). A primer complementary to the conserved *T. brucei* splice acceptor sequence (*EcoRISL(1)*; Ullu *et al.*, 1996) was used in conjunction with internal primers specific to the *MSH8* ORF: *MSH8U7* (shown in panel B) or *MSH8U8* (shown in panel C). For each set of primers RT positive (RT+), RT negative (RT-) and no template (NT) reactions were run.



**Figure 3.21 Global multiple alignment of the putative *T. brucei* MSH3 and MSH8 polypeptides with a range of MSH3, MSH6 and MSH7 orthologues**

Multiple sequence alignment of the putative *T. brucei* (Tbr) MSH3 (shown in blue) and MSH8 (shown in red) polypeptides with homologues of MSH3, MSH6 and MSH7 from other eukaryotes: *A. thaliana* (Ath), *H. sapiens* (Hsa), and *S. cerevisiae* (Sce). Sequences were aligned using CLUSTAL X (Thompson *et al.*, 1997), and shaded using the BOXSHADE server ([http://www.ch.embnet.org/software/BOX\\_form.html](http://www.ch.embnet.org/software/BOX_form.html)): identical residues are shaded in black, and conserved residues in grey. MIS is the N-terminal mismatch interacting domain (shown in purple; Malkov *et al.*, 1997; Culligan *et al.*, 2000; Schofield *et al.*, 2001), MCD is the middle conserved domain (shown in turquoise; Culligan *et al.*, 2000), the peptides identified as I, II, III and IV are the conserved nucleotide binding motifs (shown in green; Gorbalenya and Koonin, 1990) and HTH is the conserved helix-turn-helix motif (shown in pink; Adé *et al.*, 1999; Biswas *et al.*, 2001). The region identified beneath the *T. brucei* MSH8 sequence corresponds to the region of the *T. brucei* MSH8 gene that was amplified and used as a probe (shown in orange).

```

HsaMSH3      1 -----
AthMSH3      1 -----
SceMSH3      1 -----
TbrMSH3      1 -----
AthMSH6      1 -----MAPSRRQISGRSPLVNQQRQITSFFGKSASSSS
HsaMSH6      1 MSRQSTLYSFFPKSPALSDANKASARASREGGRAAAAPGASPSPPGGDAAWSEAGPGRPL
SceMSH6      1 -----
AthMSH7      1 -----
TbrMSH8      1 -----

HsaMSH3      1 -----
AthMSH3      1 -----
SceMSH3      1 -----
TbrMSH3      1 -----
AthMSH6      34 SPSPPSPSLSNKKTPKSNPNPKSPSPSPSPPKKTPKLNPNPSSNLPARSPSPGPDTPS
HsaMSH6      61 ARSASPPKAKNLNGGLRRSVAPAAPTSCDFSPGDLVWAKMEGYFWWPCLVYNHPFDGTFI
SceMSH6      1 -----MAPATPKTSKTAHFENGSTSSQKKM
AthMSH7      1 -----
TbrMSH8      1 -----

HsaMSH3      1 -----
AthMSH3      1 -----
SceMSH3      1 -----
TbrMSH3      1 -----
AthMSH6      94 FVQSKFKKPLLVIQQTSPSPQSVVITYGDEVVGKQVRVYWPLDKKWDGSVTFYDKGEGK
HsaMSH6      121 REKGKSVRVHVQFFDDSPTRGWVSKRLLKPYTGSKSKEAQKGGHFYSAPKEILRAMQRAD
SceMSH6      26 KQSSLLSFFSKQVPSGTPSKKVQKPTPATLNTATDKITKNPQGGKTGKLFVDVDEDNDL
AthMSH7      1 -----MQRQRSILSFFQKPTAATTKGLVSGDAASGGGGSGGPR
TbrMSH8      1 -----

HsaMSH3      1 --MSRRKPASGGLAASSAPARQAVLSRFFQSTGSLKSTSSSTGAADQVDPGAAAAAAPP
AthMSH3      1 -----
SceMSH3      1 -----
TbrMSH3      1 -----
AthMSH6      154 HVVEYEDGEEESLDLGKEKTEWVVGKSGDRFNRLKRGASALRKVVTDSDDDVEMGNVEE
HsaMSH6      181 EALNKDKIKRLELAVCDEPSEPEEEEEMEVGTTYVTDKSEEDNEIESEEEVQPKTQGSRR
SceMSH6      86 TIAEETVSTVRSDIMHSQEPQSDTMLNSNTTEPKSTTTDEDLSSSQSRRNHKRRVNYAES
AthMSH7      39 FNVKEGDAKGDASVRFVAVSKSVDEVVGTDTTPPEKVP RRVLPSGFKPAESAGDASSLFSNI
TbrMSH8      1 -----

```

Figure 3.21 continued

HsaMSH3 59 APAFPPQLPPHVATEIDRRKKRPLENDGPVKKKVKVQOKEGGSDLGMSGNSEPKKCLRT  
 AthMSH3 1 -----MG  
 SceMSH3 1 -MVIGNEPKLVLLRAKSSANRFILLNLLTIMAQPTISRFFKAVKSELTHKQEQEVAVG  
 TbrMSH3 1 -----  
 AthMSH6 214 DKSDGDDSSDEDWGKNGVKEVCESEEDDVELVDENEMDEEELVEEKDEETS KVN RVSKTD  
 HsaMSH6 241 SSRQIKKRRVISDSESDIGGSDVEFKPDTKEEGSSDEISSGVGDSESEGLNSPVKVARKR  
 SceMSH6 146 DDDSDTTF TAKRKKGVVDSEDEDEYLPDKNDGDEDDIADDKEDIKGE LAEDSGDDD  
 AthMSH7 99 MHKFVKVDDRDCSGERSREDVVPLNDSSLCKM KANDVIPOFRSNNGKTQERNHAFSFSGRA  
 TbrMSH8 1 -----

HsaMSH3 119 RNVKSLEKLKEFCCDSALPQSRVQTESLQERFAVLPKCTDFDDISLLHAKNAVSSSED--  
 AthMSH3 3 KQKQQTISRFFAPKPKSPHEPNPVAESSTPPPKISATVSFSPSKRKLSDHLAAASP--  
 SceMSH3 60 NGAGSEICLDTDEEDNLSSVASTTVTND SFPLKGSVSSKNSKNSEKTSGTSTTFNDI--  
 TbrMSH3 1 -----  
 AthMSH6 274 SRKRKTSEVTKSGGEKSKTDTGTILKGFKASVVEPAKKIGQADRVVKGLEDNVL DGD--  
 HsaMSH6 301 KRMVTGNGSLKRKSSRKEIPSATKQATSISSETKNTLRAFSAPQNSESOAHVSGGGDD--  
 SceMSH6 206 DLISLAETTSKKKFSYNTSHSSSPFTRNISRDNSSKKS RPNQAPRSRSYNPSHSQPSAT--  
 AthMSH7 159 ELRSVEDIGVDGDPGPEIPGMRPRASRLKRVLEDEMTFKEDKVPVLD SNKRLKMLQDPV  
 TbrMSH8 1 -----

HsaMSH3 177 -SKRQINQKDTTLFDLSQFGSSNTSH---ENLQKTASKSANKRSKSIYTPLELOYIEMK  
 AthMSH3 61 -KKPKLSPHTQNPVDPNLHQRF LQR---FLEPSPEEYVPE TSSSRKYTPLEQQVVELK  
 SceMSH3 118 -DFAKLDRI MKRRSDENVAEEDDEE---EGEEDFVKKKARKSPTAKLTPLDKQVKDLK  
 TbrMSH3 1 -----MSKRRRGDDES IFRKAI---LLCHRRATTEELSVAT VVC  
 AthMSH6 332 -ALARFGARDSEKFRFLGVDRRAKRRPTDENYIPRTLYLPPDFVKKLTGGOROWWEIK  
 HsaMSH6 359 -SSRPTVWYHETLEWLKEEKRRDEHRRRPDHPDFDASTLYVPEDFLNSCTPGMRKWWQIK  
 SceMSH6 264 -SKSSKFNKQNEERYQWLVDERDAQRRPKSDPEYDPTLYIPSSAWNKETPEEKQYWEIK  
 AthMSH7 219 CGEKKEVNEGTKFEWLESSRIDANRRRPDPLYDRKTLHIIPDVFKMSASOKOYWSVK  
 TbrMSH8 1 -----MPEVSLECEDICP NVSYPFRLGI PKRPPS T PRDLEAA

## MIS

HsaMSH3 232 QQHK-DAVLCVECGYKYRFFGEDAEIAARELN IYCHLDHN-----FMTASII  
 AthMSH3 116 SKYP-DVVL MVEVGYRYRFFGEDAEIAARVLGIYAHMDHN-----FMTASV  
 SceMSH3 173 MHR-DKVLVIRVGYKYKCFEADAVTVSRILHIKLV PCKLTIDE SNPQDCNHRQFAYCSF  
 TbrMSH3 49 ESLPP MVAC YRVKF GR R VS RFG MC Q TP-----FEYSS  
 AthMSH6 391 AKHM-KVVF FKMGFYELFEMDAHVCAKELDIQYMKGEQ-----PHCCF  
 HsaMSH6 418 EQNF-DLVICYKVGKFYELYHMDALIGVSELGLVFMKCNW-----AHSCE  
 SceMSH6 323 SKMW-DCIVFFKKGKFFELYEKDALLANALFDLKIAGGGRAN-----MQLAGI  
 AthMSH7 279 SEYM-DIVLFFKVGKFYELYELDAELGHKELDWKMTMSGVG-----KCRQVGI  
 TbrMSH8 55 HY- F K DQ A H FG KL VDTTNRG-----KMRLA

HsaMSH3 277 PTHRLFVHVRRLVAKGYKVG VVKQTETAALKAIGD-----NRSSIFSRKLTALYTKST  
 AthMSH3 161 PTFRLNFHVRRLVNAGYKIGVVKQTETA AAIKSHGA-----NRTGPFERGLSALYTKAT  
 SceMSH3 232 PDVRLNVHLERLVHNLKVAVVECAETS AIIKHDPG-----ASKSVFERKISNVETKAT  
 TbrMSH3 95 YTGNIYR AM FA ESASIRSTSGN-----SKGFS GR Y  
 AthMSH6 435 PEKNFSVNIKLVKGYRVLVVEQTE TPDQLEQRRKET---GSKDKVVKREVCVVTKGT  
 HsaMSH6 462 PEIAFGRYSDSLVQKGYKVARVEQTE TPEMMEARCRKMAHISKYDRVVREICRIITKGT  
 SceMSH6 370 PEMSFYWAAQFLQMGYKVAQVDQRESMLAKEMREG-----SKGIVKRELOCILTSGT  
 AthMSH7 326 SESGIDEAVQKLLARGYKVGRIEQLETSDQAKARG-----ANTIIPRKLQVLTTPST  
 TbrMSH8 104 QTFSEWARLF F R MKEEGESSKNAR-----PKIPVE P

Figure 3.21 continued

HsaMSH3 330 IIGEDVNPILIKLDDAVNVDEIMTDTST-SYLLCISENKENVRDKKKGNIFIGIVGVQPAT  
 AthMSH3 214 IEAARDISGGCGGEEFGFSQSNFLVCV-VDERVKSETLG-CGIEMSFDRVGVVGVVEIST  
 SceMSH3 287 FGVNSTFVLRGKRILGDTNSIHWALSRLD-VHQQKVA-----KYSLSISVNLNN  
 TbrMSH3 147 ILPDVAVTAGTPQEGSATGQGDEGAP-EGGDPVGVVEELLPLKEGSSELFICFVWPSTG  
 AthMSH6 492 ILDGEMLLTNPDAASYLMALTEGGESELT-NPTAEHNFVCLVDVATQKIILGQFKDDQDCS  
 HsaMSH6 522 QIYSVLEGDPSENYSKYLSSVKEKEED-SSGHTRAYGVCVVDTSLGKFFICQFSDDRHCSS  
 SceMSH6 423 LTDGDMHS---DLATFCLAIREEPGN-FYNETQLDSSTIVQKLNTKIFGAAFIDTATGE  
 AthMSH7 378 ASEGNTGPD-----AVHLLAIKEIKME-LQKCVTYGFAFVDCALRFWVGSISDDASCA  
 TbrMSH8 156 IADPMISGYGAVFVLALYPIGSGSVDGMAVDLSRRVVFHCPCGTNGKESVAGFVEEVLN

HsaMSH3 389 CEVVFDSFQDS-ASRSELETRMSSLOPVEILLPSALSEQTEALIHRRATSVSVQDDRIIRVE  
 AthMSH3 272 CEVVYEEFNDN-FYRSGLEAVILSLSPAELLGQPLSQQTEKFLVAHAGPTSN---VVRVE  
 SceMSH3 332 CEVVYDEFEEENLADEKIQIRIKYLQPIEVLVNTDDLPLHVAKFFKDISCPLIH-----  
 TbrMSH3 206 IGSSEVTLLESFTYS SRHHLSSGGEIDILNRYNIVE LFYDEDGARELDR-HQNA  
 AthMSH6 551 ALSCLISEMRPVEIKKPAKVSATERTIVRQTRNPLVNNLVPLSEFDWSEKTIYEVGII  
 HsaMSH6 581 RFRTLVAHYPEVQVLFKGNLSKETKTIKSSLSLCSLQEGLIIPGSQFWDASKTLRITLEE  
 SceMSH6 479 LQVLEFEDDSECTKDDTMSQVRPMEVVMERNNLSTLANKIVKFNAPNAIFNEVKAGEE  
 AthMSH7 432 ALGALLMOVSEKEVLYDSKGLSREAQKALRKYTLTGSTAVQVAPVPOVMGDTDAAGVRNI  
 TbrMSH8 216 EVSALQQIRREIIPRGAIDAPGEEPKGSFGRRLFEWVEGEGFQVELVEEVGTSRKL

HsaMSH3 448 RMDNIYFEYSHAFQAVTEFYAKDITVDIKGSQ-----IISGIVNLEKPVIC  
 AthMSH3 328 RASLDCFSNGNAVDEVISLCEKISAGNLEDDKEMKLEAAEKMSCLTVHTIMNMPHLTVQ  
 SceMSH3 386 --KQDYDLEDHVVAIAKVMNEKIQLSP-----SLIR  
 TbrMSH3 265 AATDGTLPFRLCGLPEAFYTPLNTILSLHHGPTVNGEEDNNSVTVCVTSPIGSIIDSA  
 AthMSH6 611 YKRINCQPSSAYSSEKILGDGSSFLPKMLSELATEDK-----NGSLALSALGGATY  
 HsaMSH6 641 EYFREKLSDGIGVMLPQVLKGMTSESDSIGLTPGEKSELALSALGGCVFYLLKCLIDQEL  
 SceMSH6 539 FYDCDKTYAEIISSEYFSTEEDWPEVLKSYDYTG-----KKGVSFAF  
 AthMSH7 492 IESNGYFKGSSESWNCAVDG-----LNECDVALS  
 TbrMSH8 276 PLEERSLKEAGRFLAQYFRSLKLNVDSE-----ILLEARPYNFH

HsaMSH3 493 SLAAAIKYLKEFNLEKMLSKPENFKQLSSKMEFMTINGTTLRNLEILQNOTDMKTGKSLIL  
 AthMSH3 388 ALALTECHLKQGFERILYQASFRSLSSNTEMTLS-ANTLQOLEVVKNNSDGSESGSLE  
 SceMSH3 415 LVSKLYSHMVEYNNEQVMLIPSIYSPFASKIHMLLD-PNSLQSLDIFTHDGG---KGSLE  
 TbrMSH3 325 EYLKPSRFDTFRKMCSEKPPLLARSTAGGVAEIMPGTTMSADIFHSSIGLIL  
 AthMSH6 663 YTRQAFGLDESLLRFAKFEISLPYCDFSNVNEKQHMVLDAAALENLEIFENSRRNGGYSGLY  
 HsaMSH6 701 LSMANEEYIPLDSDTVSTR-SGAIFTKAYQRMVLDVATLNNLEIFLNGTNGSTEGTLL  
 SceMSH6 581 GGLLYLKWKLKDKNLISMKNIEYDFVKSQHSMLVDGITLQNLIFSNSEFDGSDKGTLE  
 AthMSH7 521 ALGELINHLRSLKLEDVVKHGDIFPYQVYRG-CLRIDGQTMVNLEIFNNSCDGVLQGLN  
 TbrMSH8 315 LKQQVTSQVPSNDRSRCIDSTLLWYERREDPGVLAASVAVGVLRLDSE

MCD

HsaMSH3 553 WLDHTKTSFGRRLKQWVTPPLLKLREINARLDAVSEVLIHSESS-----  
 AthMSH3 447 HNNHTLTVYGSRLLRHWVTHPLCDRNLSARLDAVSEISACMGSHSSSQLSSELVEEGS  
 SceMSH3 471 WLLDHTRTSFLRMLREWILKPLVDVHCIEERLDAIECITSEINN-----  
 TbrMSH3 385 A S V P L R S A A CDLRA S R AF RGEAGD-----  
 AthMSH6 723 AOLNOCITASGKRLKLTWLRPLYNTELIKERQDAVAIDRGENLP  
 HsaMSH6 760 ERVDTCHTTPFGKRLKQWLCAPLCNHYAINDRLDAIEDIVVVPDK  
 SceMSH6 641 KLFNRAITPMGKRMMKWLMLPLLRKNDIESRLDSVDSLQDITLR-----  
 AthMSH7 580 KYLENCVSPTGKRLLRNWHCHPLKDVESINKRLDVVEEFTANSES-----  
 TbrMSH8 375 NENCCNGSFSLLSASPRVLAQVRFENNLDLWAKTEESADVTPP

Figure 3.21 continued

HsaMSH3 598 -----VFGQIENHLRKLDPDIERGLCSIYHKKCSST-----QEEFLIVKTTYHLK  
 AthMSH3 507 ERAIVSPEFYLVLSVLTAMSRSSDIQRGITRIFHRTAKA-----TEEIAVMEAILLAG  
 SceMSH3 516 -----SIFFESLNQMLNHTPDLLRRLNRI MYGTTSR-----KEVYFYLKQITTSFV  
 TbrMSH3 431 -----SVVG LREFA FG AT RAQRC V-----T LR RA KVT  
 AthMSH6 768 -----YSLEFRKSLSRLPDMERLIARMFSSIEASG-----RNGDKVLYEDTAK  
 HsaMSH6 805 -----ISEVVELLKKLPDLERLLSKIHNVGSPLKSNHPDSRAIMYEETYSK  
 SceMSH6 687 -----EQEITFSKLPDLERMLARIHSRTIKVK-----DEEKVITAFETII  
 AthMSH7 625 -----MQITGQYLHKLPLDERLLGRIKSSVRS-----  
 TbrMSH8 435 ICTPNSSTRTSEGPTQEFQASGTQCGSKR TTNTFE RFTNLFAT ERN SR ADL

HsaMSH3 641 SEFOATTPAVNSHIQSDLLRTV-----IIEIPELLSPV-----EHYDKI  
 AthMSH3 561 KQIQRLGIKQDSEMRSMQSATVR-STLLRKLISVISSPVVDNA-----GKLLSA  
 SceMSH3 561 DHFKMHQSYISEHFKSSDGRIGKQSPLLFRLFSELNELLSTTQL-----PHEFTM  
 TbrMSH3 474 KLALD SLCGEGMCDQIRDA V-----A TSENV-----ELF QS  
 AthMSH6 812 KQVQEFISTURGCE TMAEACSSLRAILKHDTSRRLLHLITPGQS-----LPNISS  
 HsaMSH6 853 KKI IDFI SALEGFKVMCKI IGIMEEVADGFKSKILKQVLSLQTKN-----PEGRFPDITV  
 SceMSH6 728 ELQDSLKNNDLKGDVSKYISSFP-----EGIVE  
 AthMSH7 653 ---ASVLPALG-----  
 TbrMSH8 495 GDSQQ AFVDPLVQYKKHLQL ISTVVAFEEMLDWSNN QKECAPPLLQELWGTMGALAP

HsaMSH3 680 LNEQAAKVGDKTELFKDLSD-----FPLIKKRKDEIQGVIDEIRMHLOBIRK-ILKNPS  
 AthMSH3 610 LNKEAAVRGDLLDILITSSDQ-----FPPELAEARQAVLVIREKLDSSIASERK-KLAIRN  
 SceMSH3 611 INVSVMKNSDKQVMDFFNLNNYDCSEGHIKIQRESESVRSQKKEELAEIRK-YLKRPY  
 TbrMSH3 510 CKCELNLDA SPQEYAAALGS ---LPDL QTHAKERDE LRA DVE EC ---T LPA  
 AthMSH6 862 SIKYFKDAFDWVEAHNSGRVIPHGAEDEYDCACKTVEEFESSDKKHDKQQRK-LLGDAS  
 HsaMSH6 908 ELNRWDTAFDHEKARKTGLITPKAGFDSYDQALADIRENEQSDLEYDEKORN-RIGCRT  
 SceMSH6 756 AVKSWTNAFEROKAINENIIVPQRFDFIEFDKSMDRLOELEDLMEIDMTYRK-QFKCSN  
 AthMSH7 662 -----KKVVKQRVKAFGQIVKGFRCGIDLILA-LQESN  
 TbrMSH8 555 AVASIKACF RKAAEVSGVIV SQGACPAYDEATEC D I EKK DEV RE DNI FNAA

HsaMSH3 733 AOYVTVSGQEFMIEIKN-SAVSCIPTDWKVGSTKAVSRFHSPEIVENYRHLNQLREQLV  
 AthMSH3 664 DEFLQVSGITHLIELP---VDSKVPNNVWKNSTKKTIRYHPPEIVAGLDELALATEHLA  
 SceMSH3 670 LNER--DEVLYLIEVKN-SQIKDLPDDWIKVNNTKMVSRETTPRTOKLTKOLEYKDLLI  
 TbrMSH3 566 E RT GT N-VRANDA K V TR THV H R V N TV CS RLA  
 AthMSH6 921 INYVTVGKDEYLLEVPE-SLSGSPHDIYELCSSLKGVSRYPPTIKKILKELSQAKSEK  
 HsaMSH6 967 IVYWGIGRNRYOLEIPENFTTRNLPEEYELKSTKKGCKRYWTKTIEKKLANLINAEERRD  
 SceMSH6 815 IQYKDSCKEITYTIEIPI-SATKNVPSNWWQMAANKTYKRYYSDEVRALARSMAEAKEIHK  
 AthMSH7 695 MMSLLYLKCKLPILVKG-----SGLELELSQFEAAIDSDFPNYQNQDV  
 TbrMSH8 615 T SH REN LMEAPKRC PG ERSR S CV TVAG EF VE HKR TKKA

HsaMSH3 792 LDCSAEWLDFLEKFS---EHYHSLCKAVHHLATVDCIFSLAKVAKQGD--YCRPTVQBER  
 AthMSH3 721 IVNRSWDSFLKSF---RYITDEKAAVQALALDCLHSLSTLSRKN--YVRPEFVDC  
 SceMSH3 727 RESELQYKEFLNKIT---AEYTELRKITLNLQAQYDCILSLAATSCNVN--YVRPTFVNGQ  
 TbrMSH3 625 IAANEAWLAKQAELEGSVDTEI KSV NS V A HC VA SAPG--YTA S SD E  
 AthMSH6 980 SALKSISORLIIGRFC---EHQEKWRQLVSATAELDVLISLAFASDSYECVRCRPVSGST  
 HsaMSH6 1027 VSLKDCMRRIFYNED---KNYKDWQSAVECIAVLDVLLCLANYSRGGDCPMCRPVILLPE  
 SceMSH6 874 TLEEDLNRLCOKFD--AHYNTIEMPTLOAISNIDCLLAI TRTSEYLGAPSO PAVVDE  
 AthMSH7 738 TDENAETLTHLIELF--IERATONSEVIHTISCLDVLRSFAIASLSACSMAREVIFPES  
 TbrMSH8 675 DALLLVVRN AS IFN---Y PVLYEATAA CYF SLHTSAVTCYPVVOECLA

Figure 3.21 continued

I

HsaMSH3	847	K-----IVIKNGRHPVIVDVLGEQDQYVPNNIDLS-----EDSERVMIITGPN
AthMSH3	776	EP-----VEINIQSRHPVLETHLQDN--FVPNDTILH-----AAGEYCOIITGPN
SceMSH3	782	Q-----AIIAKNARNPIIESLDVHY--VPNDIMMS-----PENGKINIITGPN
TbrMSH3	683	Q-----S-----DG-----ES MRGG--GC S V-----KHG--AW
AthMSH6	1037	SDGV----PHISATGLGHPVLRGDSLGRGSEFVPNNVKIGG-----A EKAS FILLTGPNN
HsaMSH6	1084	DTP-----PFDELKGSRHPCTTKTFFGDD-FIPNDILIGCEEEEQENGKAYCVLVTGPN
SceMSH6	932	DSKTNTQLNGFDKFKSLRHPCFNLGATTAKDFIPNDIELG-----KIQPRLGLLTGAN
AthMSH7	796	EATDQNKTKGPIIKIQGLWHPFAVAADGQLPVPNDILIGEARSSSGSIHPRSILLTGPNN
TbrMSH8	732	G-----AYSAEEL-----FKSD-----S-----TND-----ATHG I

I

HsaMSH3	890	MGGKSSYIKOVALITIMAQIGSYVPAEEATIGIVDGI FTRMGAADNIYKGRSTFMEELTD
AthMSH3	820	MGGKSCYIROVALISIMAQVGSFVPASFAKLHVLDGVFTRMGASDSIQHGRSTFLEELSE
SceMSH3	823	MGGKSSYIROVALITIMAQIGSYVPAEEIRLSIFENVLIRIGAHDDIINGDSTFKVEMLD
TbrMSH3	723	A Y TF C K A Q P F G A C S S E S K E
AthMSH6	1086	MGGKSTLLRQVCLAVILAQIGADVPAETFEVSPVDKICVRMGAKDHIMAGOSTFLTELSE
HsaMSH6	1137	MGGKSTLMRQAGLLAVMAQMGCVPAEVCRLTPIDRVFTRLGASDRIMSGESTFFVELSE
SceMSH6	985	AAGKSTILRMACTAVIMAQMGCVPCESAVLTPIDRIMTRLGANDNIMQKSTFFVELAE
AthMSH7	856	MGGKSTLLRATCLAVIFAQLGCVPCESCEISLVDTI FTRLGASDRIMTGSTFFLVECTE
TbrMSH8	771	A P S T E T A N N P A N G P F T M R A P T R R A T H K Q L Y

III

HsaMSH3	950	TAEIIR-KATQSLVILDELGRGTSTHDGIAIAYATLEYEIRDVKS--LTLFVTHYPPVC
AthMSH3	880	ASHIIR-TCSSRLVILDELGRGTSTHDGVAIAYATLQHLAEKRC--LVLFVTHYPEIA
SceMSH3	883	ILHLIK-NCNKRSLLLLEDEVGRGTCTHDGIAISYALIKYFSELSDCP-LILFTTHEPMLG
TbrMSH3	783	R SEIVSS Y I A T E Y R G A T ---T F V S Q C
AthMSH6	1146	TAVMLT-SATRNSLVVDELGRGTATSDGCAIAESVLEHELEKVOC--RGFFSTHYHRLS
HsaMSH6	1197	TASILM-HATAHSLVLDDELGRGTATEDGTAIANAVVKEIAETIKC--RTLFFSTHYHSLV
SceMSH6	1045	TKKILD-MATNRSLLVDELGRGCSSSDGFAIAESVLHHVATHIQS--LGFFATHYGTLA
AthMSH7	916	TASVLQ-NATQDLSVILDELGRGTSTEDGMAIAMSVERHLVEKVOC--RMLFATHYHPLT
TbrMSH8	831	E E V E - F G P W C L V S D L D H Y T A H A A A K K H P V P P L S H A A

HsaMSH3	1007	E--LEKNYSHQVGNVHMCFLVSEDESKLDP-GAAEQVPDFVTFLYQITRIGIAARSYGLNV
AthMSH3	937	E--ISNGFPGSVGTYHVSYLTLQDK-----GSYDHDVTYLYKLVRLGCSRSFSGFKV
SceMSH3	941	E--IKS---PLIRNYHMDYVEEQKTG-----EDWMSVIFLYKLLKGLTYNSYGMNV
TbrMSH3	840	P V N S S N N G L S C Y Y F H E E K I V S R -----E G E G E V K V T P T L T P S A R
AthMSH6	1203	VDYQTN---PKVSLCHMACQIGEGIG-----GVEEVTFLYRLTPGACPKSYGVNV
HsaMSH6	1254	EDYSON---VAVRLGHMACMVENECED-----PSQETITFLYKFIKACPKSYGENA
SceMSH6	1102	SSPKHH---PQVRPLKMSILVDEATRNN-----VTFLYKMLEGQSEGSFSGMHV
AthMSH7	973	KEEASH---PRVTSKHMACAFKRSRDYQ-----PRGCDQDLVFLYRLTEGACPESYGLQV
TbrMSH8	890	E E H K S M Q K S T S S A A S E T G G Q L G Y M D F A V S A A S D S N I P T V P A A E

## HTH

HsaMSH3	1064	AKLADVPEIILKKAHKSKELEGLINTKQKRLKYFAKLTWTHNAQDLQKWTEEFNMEETQ
AthMSH3	988	AQLAQIPPSCIIRRAISMAAKLEAEVRARERNTRYGEPEGHEEPRGAEEESISALGDLFADL
SceMSH3	987	AKLARLDKDIINRAFSISEELRKES-INEDALKQFSSIKRILKSDNITATDKLAKLLSLD
TbrMSH3	894	S A T E Q L I E A T H D W C L S L L E R R F K N N S -----
AthMSH6	1250	ARLAGLPDYVLQRAVIKSEFEALYGNHRKTDHKLAMIKQIISVASDSDYSASKDSL
HsaMSH6	1303	ARLANLPPEEVIQKCHRKAREFEKMQSLRFLREYCLASERSTVDAEAVHKLLTLIKEL--
SceMSH6	1146	ASVCGISKEIIDNAQIADNLEHTSRLVKERDLAANNNGEVVSVPGGLQSDFVRIAYGD
AthMSH7	1025	ALMAGIENOVVETASCAAOAKRSIGENFKSSEDRSEFSSSLHEDWLKSLVGISRVAHNNA
TbrMSH8	950	L A S P G A N T R V K L L A W Y E R Q D L G T R G F T P S G T Q F S H R -----

**Figure 3.21 continued**

```
HsaMSH3 1124 TSL LH-----  
AthMSH3 1048 KFALSEEDPWKAFEF LKHAWKIAGKIR LKPTCSF---  
SceMSH3 1046 IH-----  
TbrMSH3 -----  
AthMSH6 1310 CELHSMANTFLRLTN-----  
HsaMSH6 -----  
SceMSH6 1206 GLKNTKLGSGEGVLNYDWN IKRNV LKSLFSI IDDLQS  
AthMSH7 1085 PIGEDDYDTLFC L WHEIKSSYCVPK-----  
TbrMSH8 -----
```

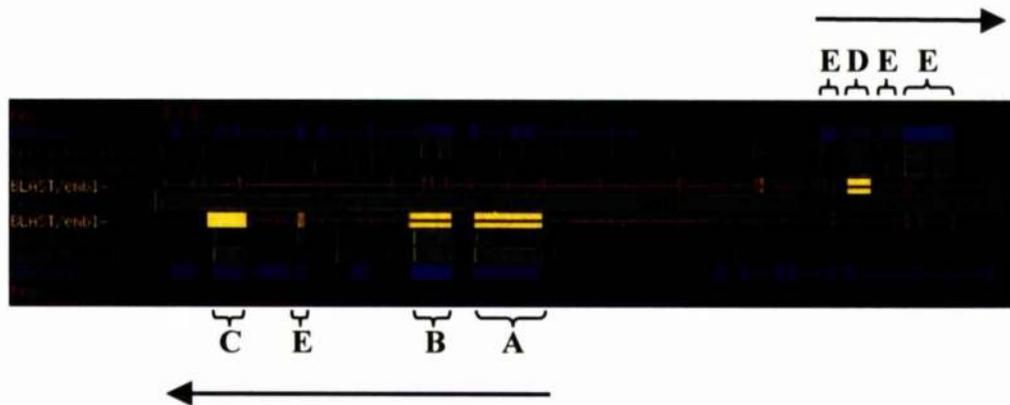
**Figure 3.22 Alignment of the conserved PCNA binding motifs present in the MSH3/MSH6 sub-family of MutS homologues**

A manual local alignment of the putative *T. brucei* (Tbr) MSH3 and MSH8 polypeptides with homologues of MSH3, MSH6, and MSH7 from other eukaryotes: *A. thaliana* (Ath), *D. melanogaster* (Dme), *H. sapiens* (Hsa), *S. cerevisiae* (Sce), *S. pombe* (Spo) and *Z. mays* (Zma). Only sequences showing similarity to the consensus sequence of the conserved PCNA binding motif are shown (Flores-Rozas *et al.*, 2000). In the consensus line **h** indicates hydrophobic residues and **a** represents aromatic residues. Conserved residues are shown in bold.

Consensus	<b>Qxxhxxaa</b>
AthMSH3	<b>QQTISRFF</b>
HsaMSH3	<b>QAVLSRFF</b>
SceMSH3	<b>QPTISRFF</b>
TbrMSH3	<b>IVEVVLFF</b>
AthMSH6	<b>QRQITSFF</b>
DmeMSH6	<b>TNTLLNYF</b>
HsaMSH6	<b>QSTLYSFF</b>
SceMSH6	<b>QSSLLSFF</b>
SpoMSH6	<b>QKTLFGFF</b>
AthMSH7	<b>QRSILSFF</b>
ZmaMSH7	<b>YLSLAPYF</b>
TbrMSH8	<b>HYDVVIFP</b>

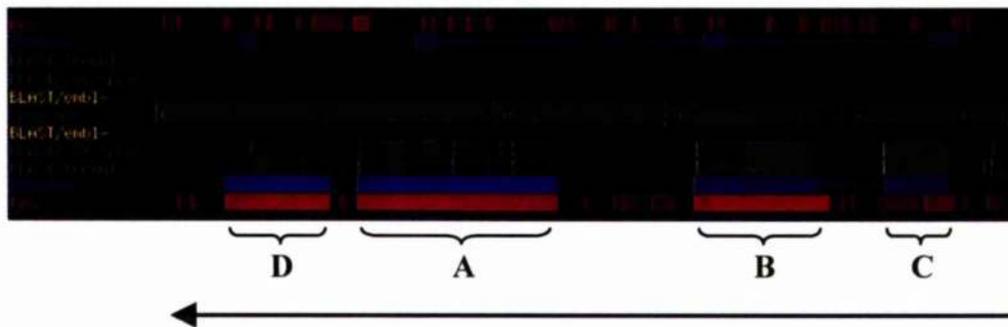
**Figure 3.23 Analysis of the genomic environment of the putative *T. brucei* MSH8 gene**

The nucleotide sequence of contig TRYP10.0.001896 was analysed using the Nucleotide Identification (NIX) server (<http://www.hgmp.mrc.ac.uk/Registered/webapp/nix/>) provided by the UK Human Genome Mapping Project Resource Centre. The output shows the results of selected programs which were used in the analysis. The sequence regions identified correspond to: (A) the *T. brucei* MSH8 ORF; (B); the *T. brucei* HSP60 ORF (Bringaud *et al.*, 1995); (C) a *T. brucei* gene encoding a T lymphocyte triggering factor (Vaidya *et al.*, 1997); (D) a *T. brucei* gene encoding a CDC2-related protein kinase (Mottram and Smith, 1995); and (E) a number of hypothetical ORFs suggested by the HMMGene gene-prediction program. The predicted direction of transcription is indicated by black arrows.



**Figure 3.24 Analysis of the genomic environment of the putative *T. brucei* *MSH3* gene**

The nucleotide sequence of contig TRYP9.0.000864 was analysed using the Nucleotide Identification (NIX) server (<http://www.hgmp.mrc.ac.uk/Registered/webapp/nix/>) provided by the UK Human Genome Mapping Project Resource Centre. The output shows the results of selected programs which were used in the analysis. The sequence regions identified correspond to: (A) the putative *T. brucei* *MSH3* ORF; (B); a hypothetical ORF bearing similarity glutaminyl-tRNA synthetase; (C) a hypothetical ORF bearing similarity to proliferating cell nuclear antigen (PCNA); and (D) a hypothetical ORF suggested by the HMMGene gene-prediction and Fex exon-prediction programs. The predicted direction of transcription is indicated by a black arrow.



### **3.3 Cloning and characterisation of a putative MutL Homologue 1 (MLH1) orthologue in *T. brucei***

#### **3.3.1 Introduction**

During MMR, the MutS-related proteins (discussed above) function in conjunction with MutL homologues (Kolodner and Marsischky, 1999). These proteins are also highly conserved at the amino acid level, especially across the ATP-binding domain and, to a lesser extent, over the region involved in dimer formation (Schär *et al.*, 1997). The primary MutL homologue in eukaryotes is MLH1, since this protein is required in the formation of all MutL-related heterodimers and is therefore required for all nuclear MMR reactions to proceed (Wang *et al.*, 1999). In addition to the conserved ATP-binding and dimer formation domains, orthologues of MLH1 possess a characteristic carboxy-terminal homology (CTH) domain which can be used to distinguish these proteins from other eukaryotic MutL homologues (Pang *et al.*, 1997). Knowledge of the characteristics of MutL-related proteins allows workers to identify putative MLH1 orthologues and distinguish them from other MutL homologues. This approach allowed the isolation of a gene from *T. brucei* which putatively encodes an orthologue of MLH1.

#### **3.3.2 Identification of the putative *T. brucei* MLH1 gene by searching the *T. brucei* genome sequencing databases**

BLAST searching of the *T. brucei* genome sequencing databases using the MLH1 polypeptide sequences from *A. thaliana*, *H. sapiens* and *S. cerevisiae* retrieved a number of clones bearing similarity to these proteins. These sequences were aligned using the GCG fragment assembly programs revealing three contigs: *MLH1A* and *MLH1B* comprised the N-terminal region of the gene and *MLH1C* appeared to encode the C-terminus (Figure 3.25). *MLH1A* and *MLH1B* could be recognised as part of the same gene as they both contained sequence from the TIGR clone 2D6. A PCR was performed using MITat 1.2 genomic DNA with primers complementary to regions just upstream and downstream of the predicted start and stop codons respectively (*MLH1* 5'(1) and *MLH1* 3'(1); see section 2.10.1) which gave a product 2.9 kb in length, confirming the linkage of these contigs (Figure 3.26).

### 3.3.3 Cloning a complete ORF from a genomic lambda library putatively encoding the *T. brucei* *MLH1* gene

PCR primers were designed from the sequence encoding part of the ATP binding domain of the putative *MLH1* gene (*MLH1U1* and *MLH1D6*; see section 2.10.1), and used to amplify a 472 bp product from MITat 1.2 genomic DNA (Figure 3.27). This product was then used as a probe to screen an ILTat 1.2 genomic library in order to recover the complete *MLH1* gene sequence, including upstream and downstream ORF-flanking regions. As described in Materials and Methods (section 2.6), the library was put through four rounds of screening, and a single clonal plaque (designated Plaque 6) was recovered. DNA was prepared from this phage clone for use in sub-cloning experiments.

Examination of the restriction sites within the available sequence for the putative *MLH1* gene (see Appendix 4) revealed that digestions with several restriction enzymes might potentially release fragments containing the entire ORF along with processing flanks which could be sub-cloned into pBluescript II KS, thereby avoiding the need for complex and time-consuming restriction mapping of the phage clone. MITat 1.2 genomic DNA was restriction-digested with a panel of enzymes and the products were Southern blotted to investigate the size of the fragments which hybridised to the probe for each digestion (Figure 3.28). A number of enzymes gave fragments between 4.0 and 8.0 kb (*Bam*HI, *Hind*III, *Kpn*I, *Pvu*II and *Sac*II) which could potentially encode the entire *MLH1* ORF. Unfortunately, the *Bam*HI restriction digestion appeared to be partial, so this enzyme was ignored. *Sal*I also gave a fragment within the desired size range; however, this enzyme gave two bands because it recognises a site within the region of the *MLH1* ORF to which the probe hybridised. From the Southern blot it was concluded that the 4.0 kb fragment released in the *Kpn*I digestion would be suitable for sub-cloning since the available sequence data for the putative *MLH1* gene showed that a *Kpn*I site was located approximately 480 bp upstream of the proposed start codon for the gene. Since *MLH1* ORFs are usually between 2.8 and 3.5 kb in length this fragment was expected to contain the entire ORF along with the 5' and 3' processing flanks.

DNA from Plaque 6 was digested with *Acc*65I (an isoschizomer of *Kpn*I which cleaves to give 5' overhangs), and shot-gun cloned into pBluescript II KS. The resulting sub-clones were analysed and a clone containing the 4.0 kb fragment was chosen for sequencing. This plasmid (called pJB200) contained an ORF encoding a putative polypeptide 887 amino acids in length flanked by the 5' and 3' untranslated regions of the gene. A sequence map of the 4.0 kb insert is shown in Appendix 4.

### 3.3.4 Southern analysis of the putative *T. brucei* *MLH1* locus

In order to investigate the genomic environment of the putative *T. brucei* *MLH1* ORF both MITat 1.2 and ILTat 1.2 genomic DNA were restriction-digested with a range of enzymes. The restriction digestions were separated by agarose gel electrophoresis, Southern blotted, and probed with the PCR product described in section 3.3.3. The pattern of hybridisation revealed the gene to be present in single copy in the *T. brucei* genome (Figure 3.29). Unfortunately, the *SpeI* restriction digestion appeared to be partial and *SalI* digestion produced two bands as the enzyme recognises a site within the sequence used as the probe. Furthermore, it can be concluded that the genomic environment of this gene is the same in MITat 1.2 and ILTat 1.2 since the restriction pattern in both sets of digestions were identical.

### 3.3.5 Analysis of the expression of the putative *T. brucei* *MLH1* gene

The expression of the putative *MLH1* gene was investigated by RT-PCR using gene-internal primers specific to the *MLH1* ORF (*MLHIU1* and *MLHID6*; see section 2.4.2 and Appendix 4) while the integrity of the cDNA used in the experiment was confirmed as described in section 3.1.5. The putative *MLH1* gene is expressed in both bloodstream and procyclic stage cells (Figure 3.30). No conclusions can be drawn about the levels of expression in the different life cycle stages since non-quantitative RT-PCR was used.

### 3.3.6 Pairwise comparison of the putative *T. brucei* *MLH1* polypeptide with other MutL homologues

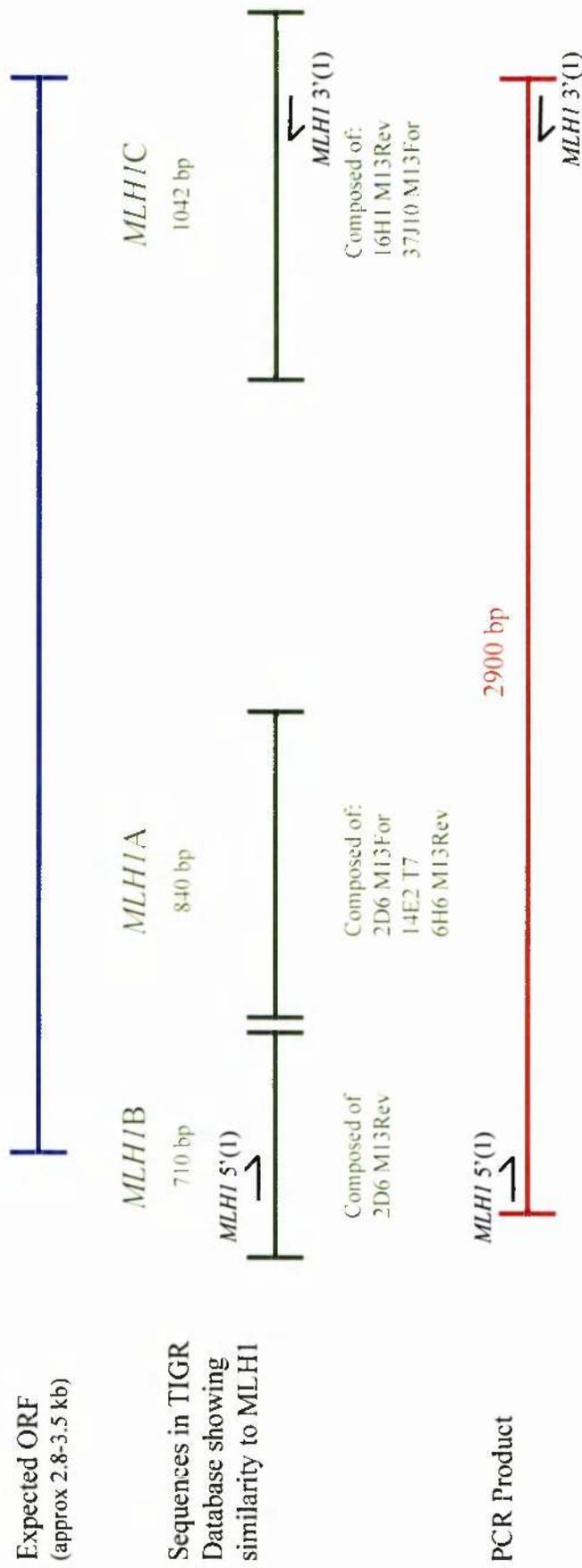
Pairwise gap alignments were performed using the GCG gap program to compare a conceptual translation of the putative *T. brucei* *MLH1* ORF with a broad spectrum of MutL-related peptides from *E. coli*, *A. thaliana*, *H. sapiens*, and *S. cerevisiae*. Table 3.3 and Figure 3.31 show the results of these pairwise alignments. These data indicate that the polypeptide is indeed a MutL homologue, bearing between 4.0 and 15.0% greater sequence identity (or 6.0 and 15.0% greater similarity) to the *MLH1* orthologues than to the PMS1 orthologues shown. This evidence suggests that the gene encodes an orthologue of *MLH1*, although it must be noted that similarity does not necessarily equate to homology.

### 3.3.7 Multiple alignment of the putative *T. brucei* MLH1 polypeptide with eukaryotic MLH1 orthologues

A global multiple alignment of the putative *T. brucei* MLH1 polypeptide with MLH1 orthologues from a range of eukaryotes was generated using CLUSTAL X (Thompson *et al.*, 1997) using the default parameters (Figure 3.32). The alignment shows a region of high identity corresponding to the ATP binding domain (Bergerat *et al.*, 1997; Mushegian *et al.*, 1997; Pang *et al.*, 1997; Ban and Yang, 1998; Dutta and Inouye, 2000). The diffuse conservation throughout the C-terminal half of the multiple alignment identifies the PMS1-interacting domain (Pang *et al.*, 1997; Kondo *et al.*, 2001). At the C-terminus a short motif which is also required for MMR activity shows high identity (Pang *et al.*, 1997). The *T. brucei* MLH1 polypeptide is conserved across the three domains characteristic of MLH1 orthologues from other eukaryotes. However, the polypeptide is divergent compared with the other proteins shown, especially over the region required for dimerisation. It should also be noted that this alignment suggests the *T. brucei* MLH1 polypeptide possesses an 85 amino acid insertion in comparison with the other MLH1 orthologues shown. However, this insertion is located within the central region of the polypeptide which has not been associated with any properties necessary for the functioning of MLH1.

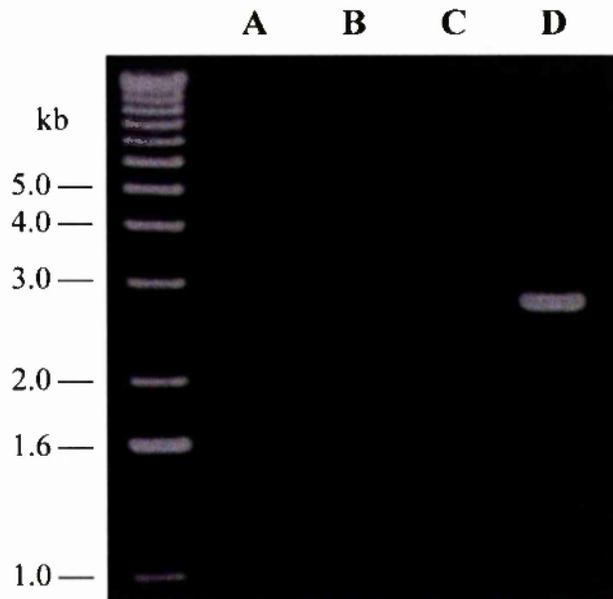
**Figure 3.25 Strategy for PCR amplification of the putative *T. brucei* *MLHI* ORF**

Comparison of eukaryotic *MLH1* homologues reveals that the ORFs encoding these proteins are between 2.8 to 3.5 kb in length. The expected *T. brucei* *MLHI* ORF is shown as a blue bar. Similarity searches of the *T. brucei* genome sequencing databases revealed a number of clones encoding fragments of a putative *MLH1* homologue in *T. brucei* (shown in green). The nucleotide sequence of these clones were aligned into three contigs: *MLH1A*, *MLH1B* and *MLH1C* (indicated by green bars), which are aligned approximately against the expected ORF. PCR primers (*MLH1* 5'(1) and *MLH1* 3'(1); shown as black half-arrows) were designed from the putative *T. brucei* *MLH1* fragments, and used to amplify a PCR product approximately 2.9 kb in length (shown as a red bar).



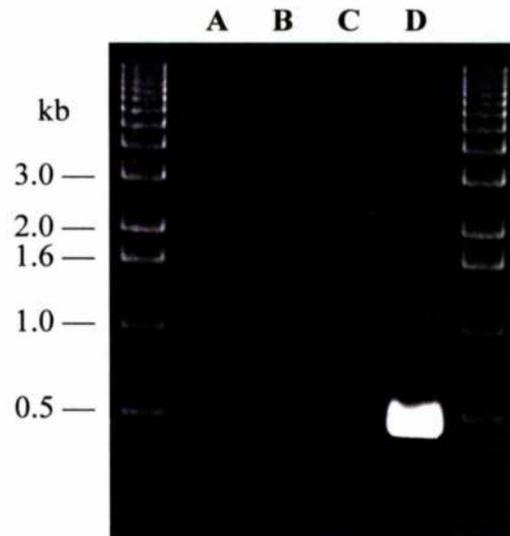
**Figure 3.26 PCR of a 2.9 kb region of the putative *T. brucei* *MLH1* gene**

Primers *MLH1* 5'(1) and *MLH1* 3'(1) were used with 20 ng of MITat 1.2 genomic DNA for PCR. Control reactions were performed with both primers in the absence of *T. brucei* DNA template (A), the *MLH1* 5'(1) primer only (B), and the *MLH1* 3'(1) primer only (C). Lane D shows the reaction containing both primers and template.



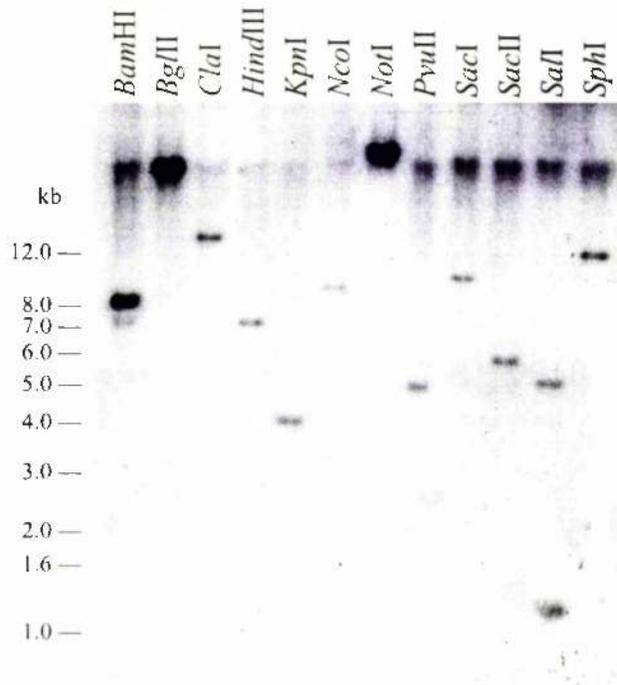
**Figure 3.27 PCR of the putative *T. brucei* *MLH1* probe sequence**

Primers *MLHIU1* and *MLHID6* were used with MITat 1.2 genomic DNA for PCR. Control reactions were performed with both primers in the absence of *T. brucei* DNA template (A), the *MLHIU1* primer only (B), and the *MLHID6* primer only (C). Lane D shows the reaction containing both primers and template.



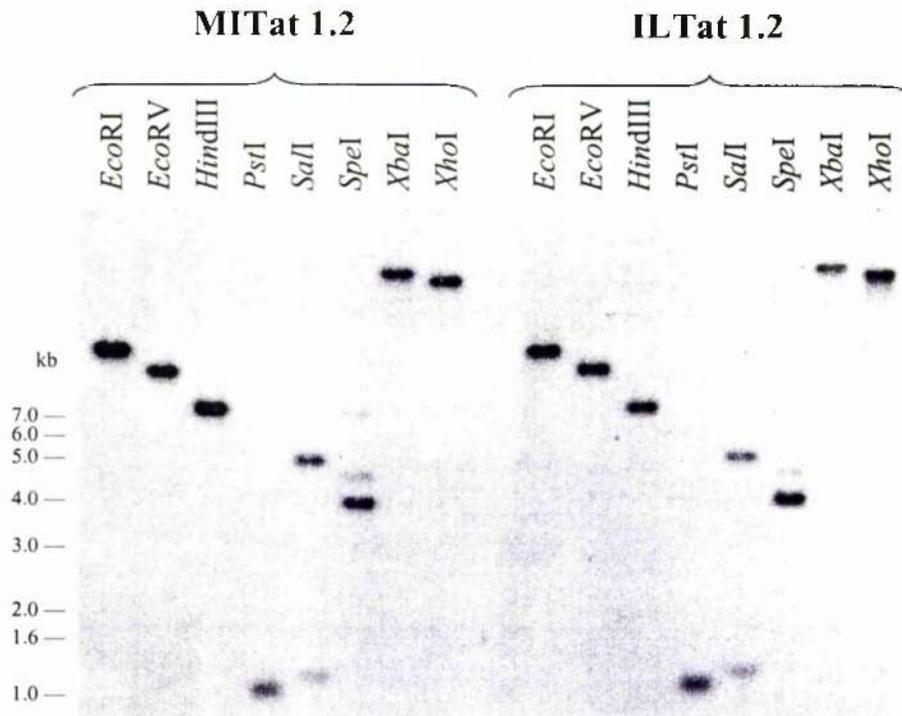
**Figure 3.28 Genomic Southern blot probed for the putative *T. brucei* *MLH1* gene**

A Southern blot of MITat 1.2 genomic DNA which was restriction-digested with the enzymes shown, separated on a 0.6% agarose gel, probed with a 472 bp PCR product amplified from the *T. brucei* *MLH1* gene and washed to 0.2 x SSC, 0.1% SDS at 65°C.



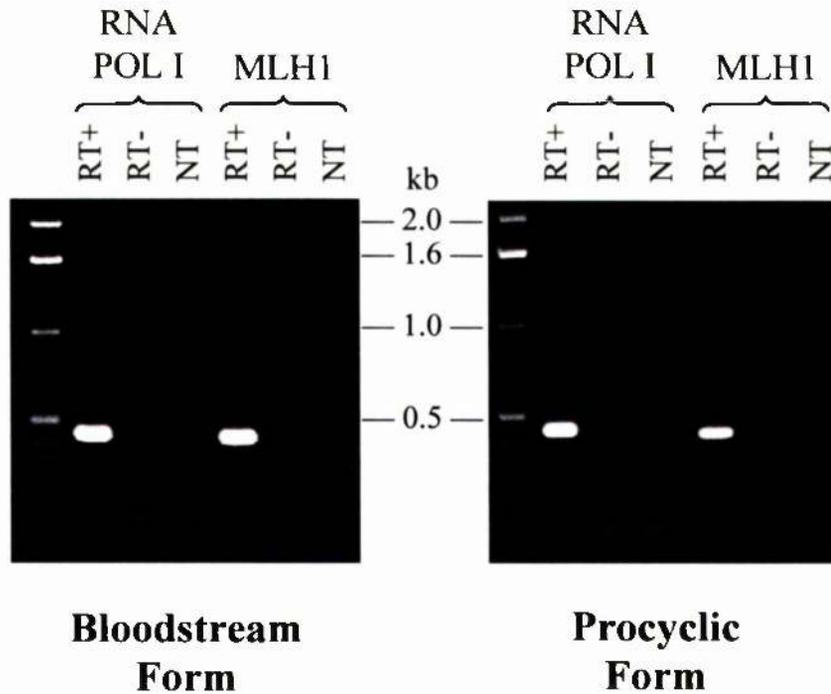
**Figure 3.29** Genomic Southern blot probed for the putative *T. brucei* *MLH1* gene

A Southern blot of MITat 1.2 and ILTat 1.2 genomic DNA which was restriction-digested with the enzymes shown, separated on a 0.6% agarose gel, probed with a 472 bp PCR product amplified from the *T. brucei* *MLH1* gene and washed to 0.2 x SSC, 0.1% SDS at 65°C.



**Figure 3.30 Analysis of the expression of the putative *T. brucei* MLH1 gene**

RT-PCR was performed on both bloodstream and procyclic form total RNA isolated from MITat 1.2. Internal primers complementary to the putative *T. brucei* MLH1 gene were used and the integrity of the RNA was confirmed using primers directed against the large subunit of *T. brucei* RNA polymerase I (Rudenko *et al.*, 1996). For both genes RT positive (RT+), RT negative (RT-) and no template (NT) reactions were run.



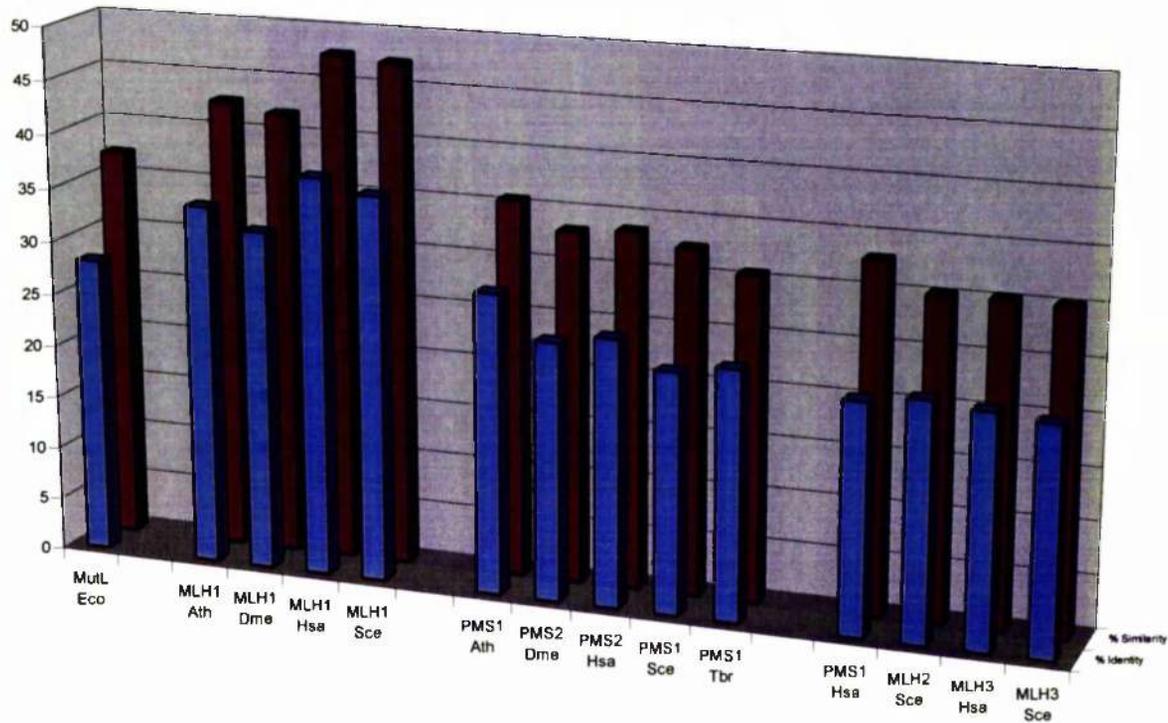
**Table 3.3** Pairwise comparison of the putative *T. brucei* MLH1 polypeptide with a range of MutL homologues

Pairwise alignments between the putative *T. brucei* MLH1 polypeptide and a number of MutS homologues from other organisms were performed using the GCG gap program. In each case the full length polypeptide sequences were compared, and the percentage identity and similarity between the two sequences was calculated.

MutL homologue	<i>T. brucei</i> MLH1	
	% Identity	% Similarity
<i>E. coli</i> MutL	28.12	37.11
<i>A. thaliana</i> MLH1	34.07	42.67
<i>D. melanogaster</i> MLH1	32.19	41.84
<i>H. sapiens</i> MLH1	37.62	47.66
<i>S. cerevisiae</i> MLH1	36.38	47.25
<i>A. thaliana</i> PMS1	28.38	35.63
<i>D. melanogaster</i> PMS2	24.33	33.21
<i>H. sapiens</i> PMS2	25.19	33.68
<i>S. cerevisiae</i> PMS1	22.68	32.58
<i>T. brucei</i> PMS1	23.47	30.67
<i>H. sapiens</i> PMS1	21.68	33.01
<i>S. cerevisiae</i> MLH2	22.33	30.23
<i>H. sapiens</i> MLH3	21.83	30.33
<i>S. cerevisiae</i> MLH3	21.31	30.22

**Figure 3.31** A graphical representation of the similarity between the putative *T. brucei* MLH1 polypeptide and other MutL-related proteins

Pairwise alignments were performed as described in Table 3.3 to compare the putative *T. brucei* (Tbr) MLH1 polypeptide sequence with other MutL-related polypeptide sequences from: *A. thaliana* (Ath), *D. melanogaster* (Dme), *E. coli* (Eco), *H. sapiens* (Hsa) and *S. cerevisiae* (Sce). Percentage identity is shown in blue and percentage similarity in red.



**Figure 3.32 Global multiple alignment of the putative *T. brucei* MLH1 polypeptide with a range of MLH1 orthologues**

Multiple sequence alignment of the putative *T. brucei* MLH1 (Tbr) polypeptide (shown in red) with homologues of MLH1 from other eukaryotes: *A. thaliana* (Ath), *D. melanogaster* (Dme), *H. sapiens* (Hsa), and *S. cerevisiae* (Sce). Sequences were aligned using CLUSTAL X (Thompson *et al.*, 1997), and shaded using the BOXSHADE server ([http://www.ch.embnet.org/software/BOX\\_form.html](http://www.ch.embnet.org/software/BOX_form.html)): identical residues are shaded in black, and conserved residues in grey. The sequences identified as I, II, III and IV (shown in blue) are the four characteristic nucleotide binding motifs of the GHKL ATPase/kinase superfamily (Bergerat *et al.*, 1997; Mushegian *et al.*, 1997; Ban and Yang, 1998; Dutta and Inouye, 2000); the carboxy-terminal homology region identified as CTH (shown in green) is also required for MMR activity (Pang *et al.*, 1997), and the peptide labelled DIMER (shown in pink) contains the PMS1-interacting domain of *S. cerevisiae* MLH1 (Pang *et al.*, 1997; Kondo *et al.*, 2001). The region identified beneath the *T. brucei* MLH1 gene corresponds to the region of the *T. brucei* MLH1 gene that was amplified by PCR and used as a probe (shown in orange).

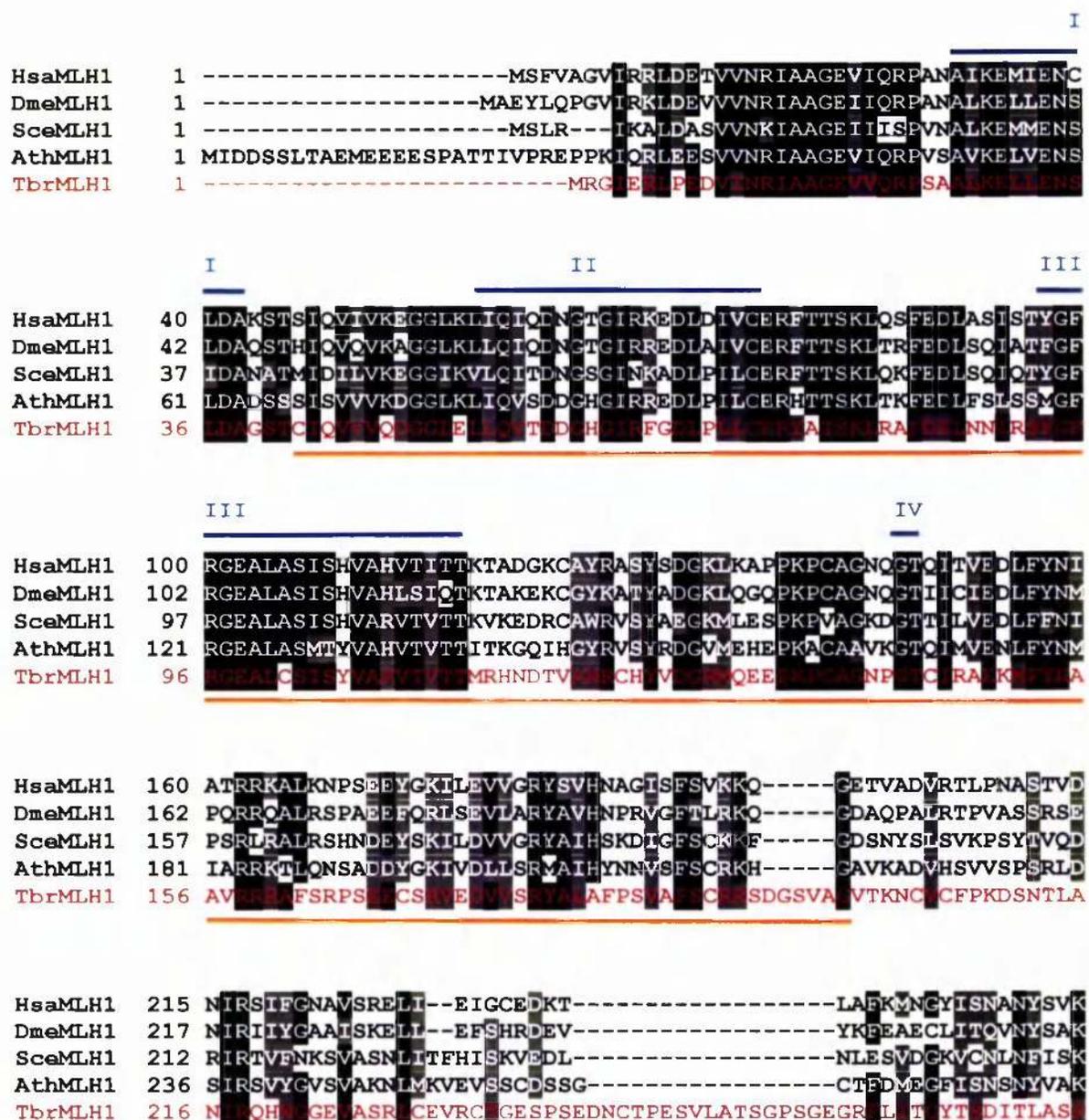


Figure 3.32 continued

HsaMLH1 255 KCIF-LLFINHRLVVESTSLRKAETVYAAYLPKNTHPFYLYLSLEISPNQVDVNVHPTKHE  
 DmeMLH1 257 KCOM-LLFINQRLVVESTALRTSVDSIYATYLPGRHHPFVYMSLTLPQNLQVDVNVHPTKHE  
 SceMLH1 254 KSISPIEFINNRLVTCDLLRRALNSVYSNYLPKGNRPFIYLGIVIDPAAVDVNVHPTKRE  
 AthMLH1 279 KTIIL-VLFINDRLVECSALKRAETVYAAATLPKASKPFVYMSINLPREHVDINIHPKKE  
 TbrMLH1 276 **SYL-CVFN** **NSVLSAAR** **LA** **SGV** **V** **NS** **TV** **F** **T** **P** **DR**

HsaMLH1 314 VHFLHEESILERVQOHTESKLLCSNSRMYFTQTLLPGLAG-----PSGEMVKSTTSIT  
 DmeMLH1 316 VHFLYQEEIVDSIKQOVEARLLCSNATRTTEYKQLRLPGAP-----DID  
 SceMLH1 314 VRFLSQDEIIEKIANQLHAELSAIDTSRTTEKASSISTNKPESLIPFNDTIESDRNRKSR  
 AthMLH1 338 VSLLNQEIIEEMIQSEVEVKLRNANDTRTEQEQKVEYIQS-----TLTSQKSDSPYS  
 TbrMLH1 335 **CL** **DE** **I** **SQ** **SEC** **QGA** **Q** **SAAR** **QMD** **IRQ** **HSKAV** **MLG** **---** **DRESQRS** **NQPMQ**

HsaMLH1 368 SSSTSGSS-----DKVYAHQVVRTDSRE--OKLDAFLQ-ELSKPLSSQPQAIUTE  
 DmeMLH1 359 ETQLADKT-----QRTYPKEMVRTDSTE--OKLDKFLA-ELVK-----  
 SceMLH1 374 QAQVVENSYTTANSQLRKAKRQENKLVRIDASQ--AKITSFLS-SSQQFNEFGSSTKRQL  
 AthMLH1 390 QKPSGQKT-----QKVPVNKMVRTDSSDPAGRLHAFLO-EPKQ-----  
 TbrMLH1 391 HSSTSPFNPLPTGARGGVAA**APCS** **AV** **PQR** **---** **GA** **DA** **RR** **KPTA** **EGNGDAPLRSE**

HsaMLH1 415 DKTDISSGRARQDDEEMLELPA--PAEVAAKNQSLEGDTTKGTSEMSEKRGPTSSNPRKR  
 DmeMLH1 394 SDSGVSS-----SSQEQASR  
 SceMLH1 431 SEPKVTVNSHSQAEKLTLINESEQPRDANTINDNDLKDQPKKKQKLGDKYKVPSTADDEKN  
 AthMLH1 427 SLPDKVS-----SLSVVRSS  
 TbrMLH1 449 **IEERAGGGVDAQ** **QRSAGT** **GSS** **IKS** **QDTAT** **GGNFQ** **VARTADDG** **KAPSG** **LTC** **AA** **S** **SPAVT**

HsaMLH1 473 HRESDVEMVEDDS-----  
 DmeMLH1 409 LPEES-----  
 SceMLH1 491 ALPIS-----  
 AthMLH1 442 VRQR-----  
 TbrMLH1 509 **TVTDAEASSGSGARGWSENQ** **STG** **TLSMT** **PVLL** **DDTT** **DE** **DGEE** **VEY** **TME** **HFK** **KHR** **KEV** **QDV**

HsaMLH1 487 -----RKEMTAACTPRRRINLTSVL  
 DmeMLH1 414 -----FRVTAAKKSREVRSSVL  
 SceMLH1 496 -----KDGYIRVPKERVNVLTSIK  
 AthMLH1 447 -----NPKETADLSSVQ  
 TbrMLH1 569 **VSSVIDTAGVGVGRGASDY** **IAAED** **NKAA** **ASTGDA** **AVRM** **VEGAD** **SAGS** **QEEAG** **F** **L** **SS**

HsaMLH1 508 SIQEEHNEQGHEVIREMLHNHSFVGCYNPQW--ALACHQTKLYDLNNTKLSSELFYQILI  
 DmeMLH1 432 DMRKVERQCSVQURSTLKNLVYVGCVDERR--ALFCHETRLYCNTRSFSELFYORMI  
 SceMLH1 516 KUREKVDDSIHREITDIFANLNYGVVDEERRLAICHDLKLEFLIDYGSVCYELFYQIGI  
 AthMLH1 459 EIIAGVDSCHPGYLETVRNCTYVGMADDVF--ALVCYNTHLYLANVNLKELAYQOTL  
 TbrMLH1 629 **T** **VSN** **RAG** **TSQ** **TAQ** **S** **L** **FQ** **LA** **V** **V** **K** **G** **H** **L** **---** **FFA** **SG** **T** **V** **D** **S** **L** **R** **V** **R** **H** **V** **R** **I** **F**

## DIMER

HsaMLH1 566 YDFAN-----FGVLRISPPAPLFDLAMLALDSP---ESGWTEELGPKEGLAEYIVEFLKK  
 DmeMLH1 490 YEFQN-----CSEITISPPPLPKELLILSLESE---AAGWTFEDCDKAELADGAADILLK  
 SceMLH1 576 TDFAN-----FGKINLQSTNVSDDIVLYNLLSE---FDELN-DLASK---EKIISKIWD  
 AthMLH1 517 RRFAN-----FNATQLSDPAPLSELILLALKEED-LDPGNDTKDOLKERTAEAMNTELLKE  
 TbrMLH1 687 **L** **R** **V** **T** **P** **S** **L** **S** **T** **V** **P** **Q** **S** **F** **E** **E** **I** **H** **S** **S** **F** **A** **Q** **N** **V** **Q** **L** **P** **P** **S** **Q** **K** **R** **A** **G** **P** **G** **S** **L** **S** **R** **L** **G** **R** **R** **C** **N**

Figure 3.32 continued

---

```

HsaMLH1 618 KAEMLADYFSLEIDEEN-----LIGLPLLIDN-YVPELEGLEIFILRLATEVNWDE
DmeMLH1 542 KAPIMREYFGLRISEDM-----LESLPSSLHQ-HRPCVAHLPEVYLLRLATEVDWEQ
SceMLH1 623 MSSMLNEYYSTIEVNDGLDNDLKSVKLKS LPLLLKG-YIPSLVKLPEFTYRLGKEVDWED
AthMLH1 571 KAEMLEEYFSVHIDSSAN-----LSRLPVILDQ-YTPDMDRVPEFLLCLGNDVEWED
TbrMLH1 747 WRYLQKSAE SAH-----IA SGTSPPARALWIA PNA

```

---

```

HsaMLH1 669 -EKECFESISKECAMFY SIRQYISEES-----TLSGQQ
DmeMLH1 593 -ETRCFETFCRETARY-----AQL
SceMLH1 682 -EQECLDGLREIALLYIPDMVPKVDTS-----DASLSE
AthMLH1 622 -EKSCFQGVSAATGNFYAMHPPLLNPSPGDGIQFYSKRGESSQEKSDLEGNVDMEDNLDQ
TbrMLH1 799 GIEATAAHIETLYGVQLHSSWLP-----NVIK

```

CTH

---

```

HsaMLH1 702 SEVPGSIPNSWKWT--VEHVYKALRSHILPPKHFTEDGNLQIANLPDLYKVFERC
DmeMLH1 612 LWREGATAGFSRWT--MEHVLFPAPFKKYLPPRIKD--QIYEITNLPTLYKVFERC
SceMLH1 715 DEKAQFINRKEHISLLEHVLFPCKRRFLAPRHILK--DVEITANLPDLYKVFERC
AthMLH1 681 DLLSDAENAWAQREWSIQHVLFPMSRLFLKPPASMASNGTFVKVASLEKLYKIFERC
TbrMLH1 831 GIRQDDVPPFCDAIRFGLPCATNSTFFVCDALVDGTQAIVSDE

```

### 3.4 Cloning and characterisation of a *T. brucei* gene encoding a putative orthologue of Post-Meiotic Segregation 1 (PMS1) from *S. cerevisiae*

#### 3.4.1 Introduction

Post-meiotic segregation 1 (PMS1) in *S. cerevisiae* (and its mammalian homologue PMS2) interact with MLH1 to form the primary MutL-related heterodimer involved in nuclear MMR (Prolla *et al.*, 1994b; Kolodner and Marsischky, 1999). The MLH1-PMS1 heterodimer forms ternary complexes with both MSH2-MSH3 and MSH2-MSH6 thereby facilitating the repair of both base-base mismatches and insertion-deletion loops (IDLs) (Habraken *et al.*, 1997; Habraken *et al.*, 1998). PMS1 orthologues are also highly conserved with the regions required for ATP-binding and dimerisation with MLH1 showing the greatest levels of conservation. Examination of a multiple sequence alignment of eukaryotic MutL homologues performed by Schär *et al.* (1997) reveals that the dimerisation domain of *S. pombe* PMS1 exhibits higher levels of identity with the dimerisation domains of other orthologues of *S. cerevisiae* PMS1 than with other MutL-homologues. The sequence PWNCPHGRPTMRH (residues 775-787 in *S. pombe* PMS1) was identified in this alignment as a highly conserved peptide motif which could be considered characteristic of PMS1, and thus is useful in the identification of possible PMS1 orthologues (Schär *et al.*, 1997). Knowledge of characteristics of the dimerisation domain of PMS1 allowed the isolation of a gene from *T. brucei* which putatively encodes an orthologue of PMS1 from *S. cerevisiae*.

#### 3.4.2 Identification of a second putative MutL homologue by searching of the *T. brucei* genome sequencing databases

Similarity searching of the TIGR and Sanger *T. brucei* genome sequencing databases using homologues of the *S. cerevisiae* PMS1 protein from *A. thaliana* and *H. sapiens* revealed a number of clones showing high levels of sequence conservation with these proteins. These sequences were aligned into three contigs using the GCG fragment assembly programs: *PMS1A* appeared to encode the C-terminus of the gene, *PMS1B* spanned the N-terminal region of the gene, and *PMS1C* seemed to encode the central region of the gene (Figure 3.33). *PMS1B* and *PMS1C* obviously encoded regions of the same gene because they both contained sequence from the TIGR clone 42G23. In order to confirm the linkage of these contigs a PCR was performed (Figure 3.34) using MITat 1.2 genomic DNA and primers complementary to regions just upstream and downstream of the predicted start and stop codons respectively (*PMS1* 5'(2) and *PMS1* 3'(2); see section 2.10.2).

### **3.4.3 Cloning of the complete ORF of the putative *T. brucei* *PMS1* gene from a genomic lambda library**

Using primers complementary to the sequence encoding the C-terminal dimerisation domain of the putative PMS1 (*PMSIU2* and *PMSI 3'(2)*; see section 2.10.2), a 460 bp PCR product was generated from MITat 1.2 genomic DNA (Figure 3.35). This product was then used as a probe for an ILTat 1.2 genomic library from which several overlapping clones were retrieved and restriction mapped. A 5.5kb *EcoRI* fragment was cloned into pBluescript II KS (creating the plasmid pJB400). The library screening, restriction mapping and sub-cloning were generously performed by T. I. Harvey in a manner similar to that described in section 3.1.3. Sequencing of this clone revealed a putative ORF 788 amino acids in length flanked by the 5' and 3' processing signals for the gene. A sequence map of the 5.5 kb insert is shown in Appendix 5.

### **3.4.4 Southern analysis of the putative *T. brucei* *PMS1* locus**

Both MITat 1.2 and ILTat 1.2 genomic DNA were restriction-digested with a panel of enzymes to analyse the genomic environment of the putative *T. brucei* *PMS1* ORF. The restriction digestions were separated by agarose gel electrophoresis, Southern blotted, and probed with the 460 bp PCR product described in section 3.4.3 revealing the gene to be present in single copy in the *T. brucei* genome (Figure 3.36). The genomic Southern blot reveals a 5.5 kb *EcoRI* fragment equivalent to the one sub-cloned in section 3.4.3, confirming that the lambda clone reflected the genomic context of the putative *PMS1* gene. Furthermore, it can be concluded that the genomic environment of this gene is the same in the MITat 1.2 and ILTat 1.2 strains since the restriction pattern in both sets of digestions were identical.

### **3.4.5 Analysis of the expression of the putative *T. brucei* *PMS1* gene**

In order to investigate the expression of the putative *PMS1* gene RT-PCR was performed using gene-internal primers specific to the *PMS1* ORF (*PMSIU1* and *PMSID1*; see section 2.4.2 and Appendix 5). The integrity of the cDNA used during this experiment was confirmed as described in section 3.1.5. The results show that *PMS1* is expressed in both procyclic and bloodstream form cells (Figure 3.37). However, since non-quantitative RT-PCR conditions were employed no conclusions can be drawn about the levels of expression in the different life cycle stages.

### 3.4.6 Pairwise comparison of the putative *T. brucei* PMS1 polypeptide with other MutL homologues

Using the GCG gap program, pairwise gap alignments were performed to compare a range of MutL homologues from *E. coli*, *A. thaliana*, *H. sapiens*, and *S. cerevisiae* with a conceptual translation of the putative *T. brucei* PMS1 ORF. The data generated in these alignments is shown in Table 3.4 and as a graphical representation in Figure 3.38. The *T. brucei* PMS1 polypeptide does indeed appear to be related to MutL, and shows between 6.5 and 11.5% greater sequence identity (or 5.0 and 12.0% greater similarity) with the eukaryotic PMS1 homologues than it shows with homologues of MLH1. Although similarity does not necessarily equate to homology, these data suggest that the gene is an orthologue of PMS1 from *S. cerevisiae*.

### 3.4.7 Multiple alignment of the putative *T. brucei* PMS1 polypeptide with orthologues of PMS1 from *S. cerevisiae*

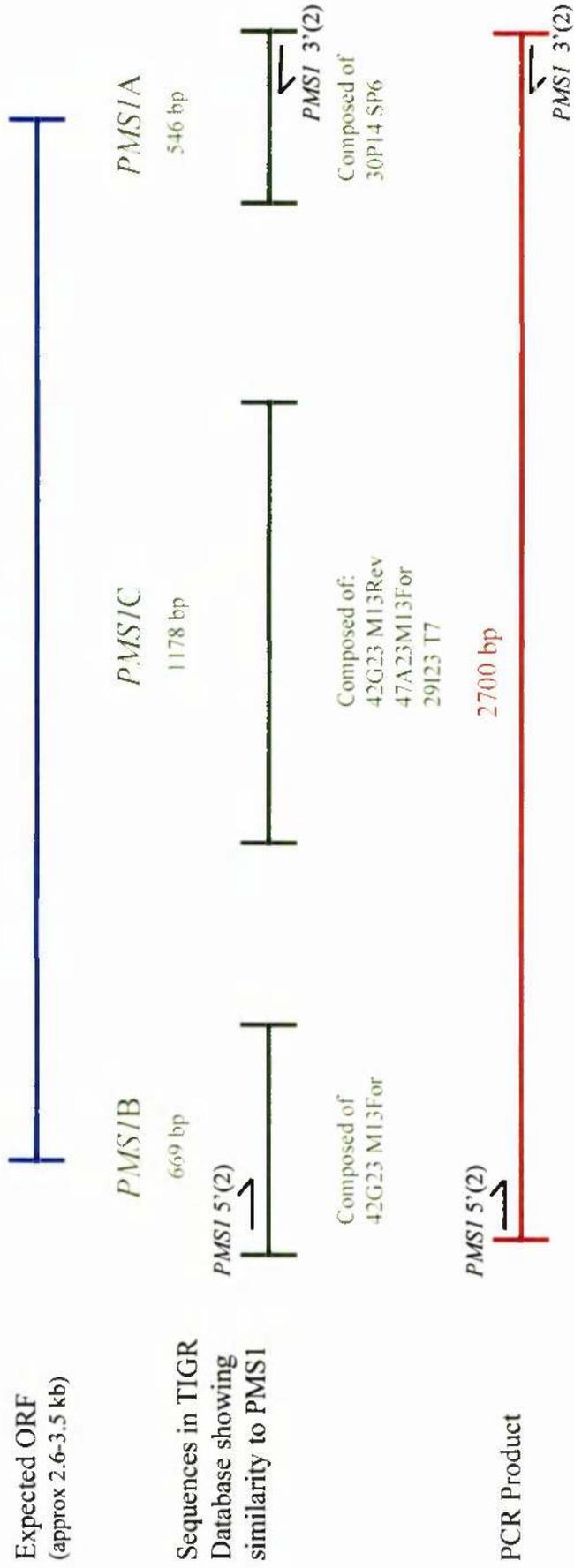
A global multiple alignment of the putative *T. brucei* PMS1 polypeptide with a number of homologues of *S. cerevisiae* PMS1 was performed using CLUSTAL X (Thompson *et al.*, 1997) using the default settings (Figure 3.39). The alignment reveals two regions of high identity corresponding to the ATP-binding domain and the MLH1-interacting domain (Bergerat *et al.*, 1997; Mushegian *et al.*, 1997; Pang *et al.*, 1997; Schär *et al.*, 1997; Ban and Yang, 1998; Dutta and Inouye, 2000; Kondo *et al.*, 2001). *T. brucei* PMS1 shows conservation over these domains although the polypeptide is divergent compared with the other orthologues of *S. cerevisiae* PMS1 shown. It is clear from this alignment that the putative *T. brucei* PMS1 polypeptide contains a sequence which differs at only one residue from the PWNCPHGRPTMRH sequence which is characteristic of homologues of PMS1 from *S. cerevisiae* (residues 759-771 in *T. brucei* PMS1; Schär *et al.*, 1997). A central region, approximately 300 amino acids in length, shows little or no sequence conservation, and the *T. brucei* PMS1 polypeptide appears to have two large deletions in this area in comparison with the other proteins in the alignment. Notably, *T. brucei* PMS1 appears to have three insertions within the region encompassing the ATP-binding motif: the first is 13 amino acids in length and interrupts the consensus of the second nucleotide binding motif; the second and third are two and three amino acids in length respectively. The functional implications of these insertions, if any, are unknown and would require 3-D structure determination or experimental analysis to assess.

### 3.4.8 The genomic environment of the putative *T. brucei* *PMS1* gene in TREU 927/4

Recent searching of the Sanger *T. brucei* genome sequencing database using the predicted PMS1 polypeptide sequence retrieved a contig, TRYP9.0.000851 (10,249 bp), which encodes the putative *PMS1* gene. This contig is composed of sequence from 121 reads from chromosome IX. This suggests the putative *PMS1* gene is present on chromosome IX. Furthermore, comparison of the putative PMS1 sequences from ILTat 1.2 and TREU 927/4 shows the sequences to be identical at the polypeptide level, and 99.96% similar at the nucleotide level. Further analysis of this contig (see section 2.8.4) reveals that the putative *PMS1* gene lies upstream of three hypothetical ORFs (Figure 3.40), whose functions cannot be predicted by BLAST searches of the nucleotide and protein databases as no significant similarities to known sequences are found.

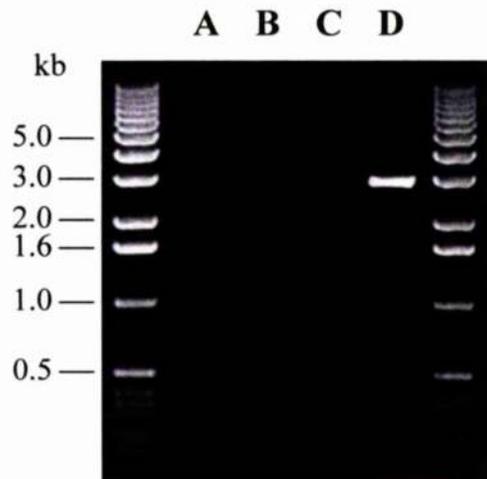
**Figure 3.33 Strategy for PCR amplification of the putative *T. brucei* PMS1 ORF**

Comparison of eukaryotic homologues of *Saccharomyces cerevisiae* PMS1 indicates that the ORFs encoding these polypeptides are usually between 2.6 to 3.5 kb in length. The expected *T. brucei* PMS1 ORF is shown as a blue bar. Similarity searches of the *T. brucei* genome sequencing databases revealed a number of clones encoding fragments of a putative PMS1 homologue in *T. brucei* (shown in green). The nucleotide sequences of these clones were aligned into three fragments: PMS1A, PMS1B and PMS1C (indicated by green bars), which are aligned approximately against the expected ORF. PCR primers (PMS1 5'(2) and PMS1 3'(2); shown as black half-arrows) were designed from the putative *T. brucei* PMS1 fragments, and used to amplify three overlapping PCR products (shown as red bars).



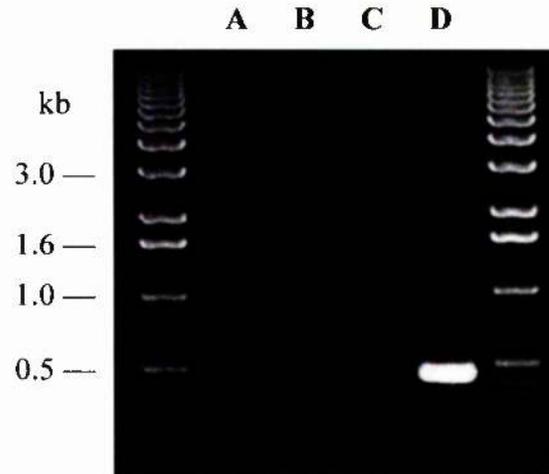
**Figure 3.34 PCR of a 2.7 kb region of the putative *T. brucei* *PMS1* gene**

Primers *PMS1* 5'(2) and *PMS1* 3'(2) were used with 20 ng of MITat 1.2 genomic DNA for PCR. Control reactions were performed with both primers in the absence of *T. brucei* DNA template (A), the *PMS1* 5'(2) primer only (B), and the *PMS1* 3'(2) primer only (C). Lane D shows the reaction containing both primers and template.



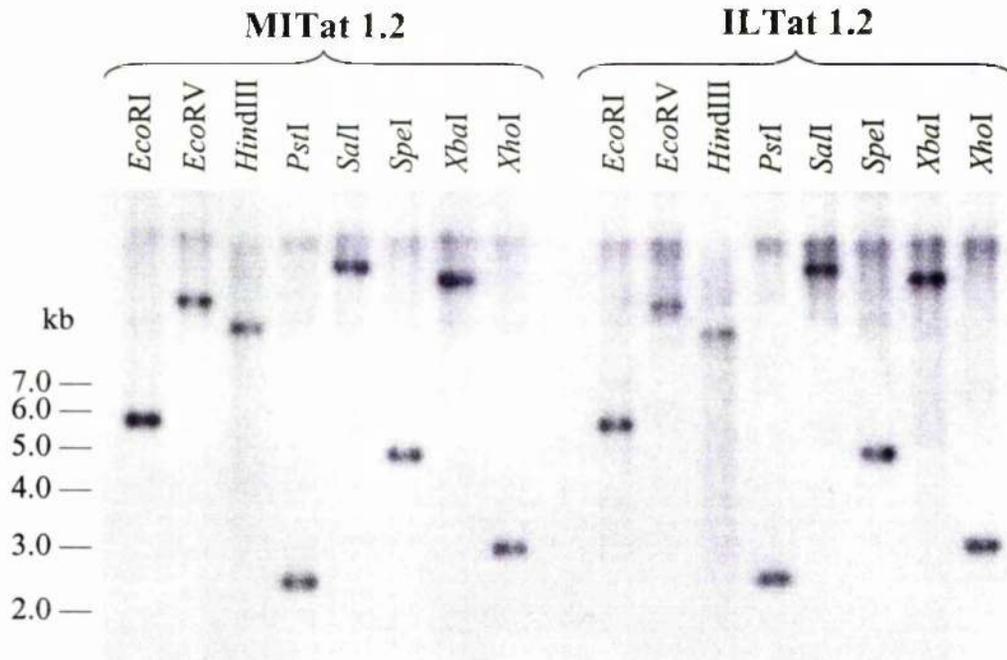
**Figure 3.35 PCR of *T. brucei* PMSI probe sequence**

Primers *PMSIU2* and *PMSI 3'(2)* were used with MITat 1.2 genomic DNA for PCR. Control reactions were performed with both primers in the absence of *T. brucei* DNA template (A), the *PMSIU2* primer only (B), and the *PMSI 3'(2)* primer only (C). Lane D shows the reaction containing both primers and template.



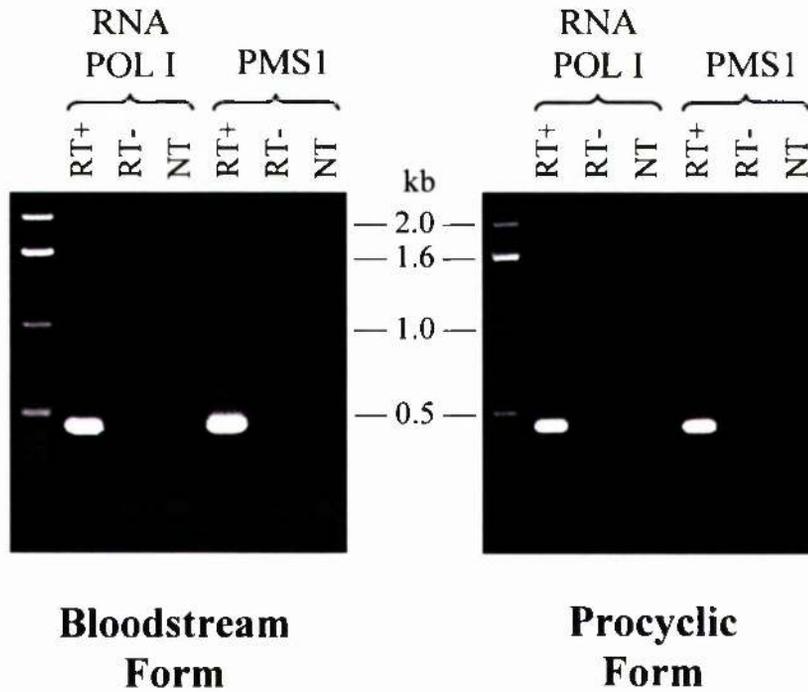
**Figure 3.36** Genomic Southern blot probed for the putative *T. brucei* *PMS1* gene

A Southern blot of MITat 1.2 and ILTat 1.2 genomic DNA which was restriction-digested with the enzymes shown, separated on a 0.6% agarose gel, probed with a 460 bp PCR product amplified from the *T. brucei* *PMS1* gene and washed to 0.2 x SSC, 0.1% SDS at 65°C.



**Figure 3.37 Analysis of the expression of the putative *T. brucei* PMS1 gene**

RT-PCR was performed on both bloodstream and procyclic form total RNA isolated from MITat 1.2. Internal primers complementary to the putative *T. brucei* PMS1 gene were used and the integrity of the RNA was confirmed using primers directed against the large subunit of *T. brucei* RNA polymerase I (Rudenko *et al.*, 1996). For both genes RT positive (RT+), RT negative (RT-) and no template (NT) reactions were run.



**Table 3.4** Pairwise comparison of the putative *T. brucei* PMS1 polypeptide with a range of MutL homologues

The percentage identity and similarity between the putative *T. brucei* PMS1 polypeptide sequence and a number of MutS homologues was calculated by performing pairwise alignments using the GCG gap program. In each case the full length polypeptide sequences were compared.

MutL homologue	<i>T. brucei</i> PMS1	
	% Identity	% Similarity
<i>E. coli</i> MutL	28.03	36.16
<i>A. thaliana</i> MLH1	24.49	34.75
<i>D. melanogaster</i> MLH1	22.62	31.39
<i>H. sapiens</i> MLH1	24.55	33.05
<i>S. cerevisiae</i> MLH1	22.13	30.77
<i>T. brucei</i> MLH1	23.47	30.67
<i>A. thaliana</i> PMS1	33.64	42.22
<i>D. melanogaster</i> PMS2	32.94	42.03
<i>H. sapiens</i> PMS2	31.66	42.91
<i>S. cerevisiae</i> PMS1	31.23	39.79
<i>H. sapiens</i> PMS1	24.16	32.12
<i>S. cerevisiae</i> MLH2	24.02	32.78
<i>H. sapiens</i> MLH3	23.67	31.33
<i>S. cerevisiae</i> MLH3	25.26	36.63

**Figure 3.38** A graphical representation of the similarity between the putative *T. brucei* PMS1 polypeptide and other MutL-related proteins

Pairwise alignments were performed as described in Table 3.4 to compare the putative *T. brucei* (Tbr) PMS1 polypeptide sequence with other MutL-related polypeptide sequences from: *A. thaliana* (Ath), *D. melanogaster* (Dme), *E. coli* (Eco), *H. sapiens* (Hsa) and *S. cerevisiae* (Sce). Percentage identity is shown in blue and percentage similarity in red.

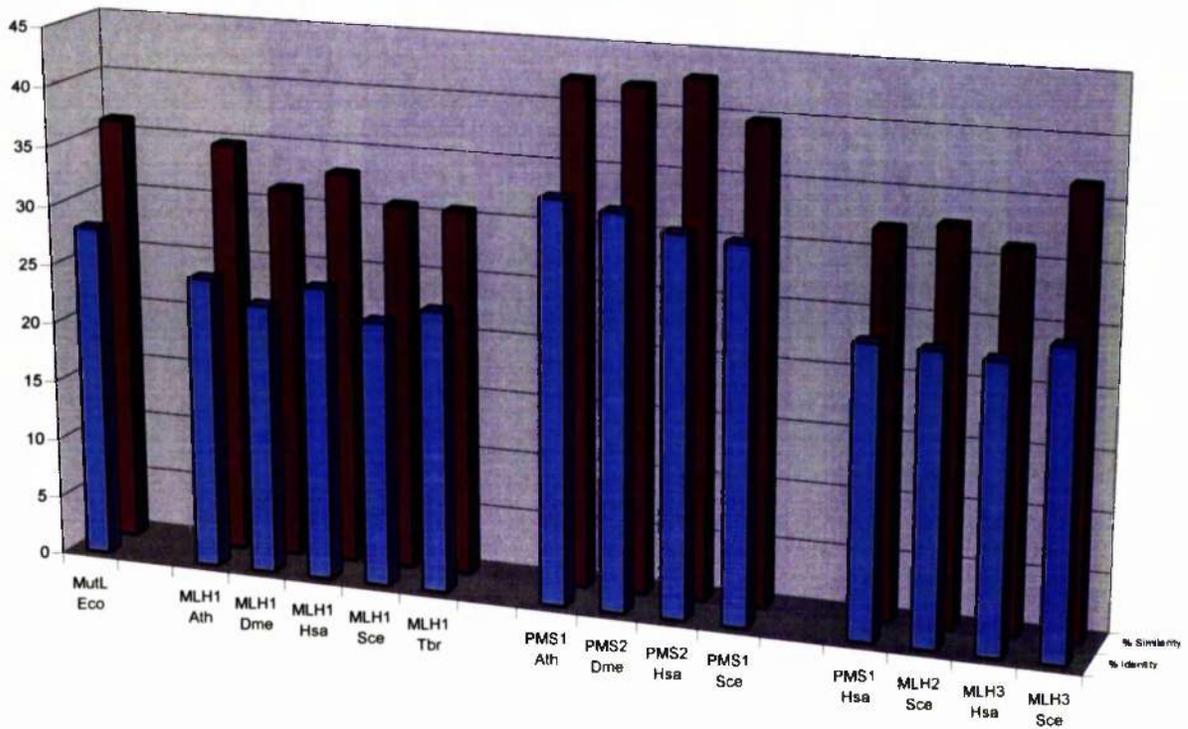




Figure 3.39 continued

HsaPMS2 256 G-----LSCSDALHN---LFYISGFISSQCTHGVRSSSTDRQFFFINRRPQDPAK  
 DmePMS2 274 SGADLADTTSPQISTEDVERLNQADFQLEGFISSCRHGAGRSSRDQFFFNNSRPQDPAK  
 ScePMS1 276 G-----KYTDDPDFLDLDYKIRVKGYSQNSFGGRNSKDRQFIYVNRKPVEYST  
 AthPMS1 246 -----ICVSEDCRVEGFLSKPGQETGRNLADROFFFINGRPVDMPK  
 TbrPMS1 250 -----WELTFGTSTVYKGN--AQLSMLVALLGLV LPM

HsaPMS2 302 VCRLVNEVYHMYNR---HOYPFVVLNLSVDSEC--VDINVTPDKROILLQEEKLLLAVLK  
 DmePMS2 334 IAKVMNEVYHRYNV---QQQPFYIENLITARSQ--VDVNLTPDKROLLINNERILLALK  
 ScePMS1 326 LLKCCNEVYKTFNN---VQFPVAVLNLLELPMSL--IDVNVTPDKRVILLHNERAVIDIEK  
 AthPMS1 287 VSKLVNELYKOTSS---RKYPVTILDFIVPGGA--CDLNVTPDKRKVFFSDETSVIGSLR  
 TbrPMS1 288 SAALAAAE.SLPNAAQRTTAAFFHSSGESLPYVAVKQKASDERHAGEE

HsaPMS2 357 TSLIGMDSQSDVNKLNYSQQPLLDVEGNLIKMHAAADLEKPMVEK-----  
 DmePMS2 389 KSLDTEGQTPATFOYQNTTIVSMLEP--KTNSGKTKFPKES-----  
 ScePMS1 381 TTSDDYNRQELALPKRMCSQSEQQAQKRLKTEVFDDRSTTHE-----  
 AthPMS1 342 EGVNEIYSSSNASYIYNRFEENSEQPKAGVSSFQKSNLLSEGIVLDVSSKTRLGEAIE  
 TbrPMS1 348 TCGLRTQASTDGIDLPVRNEG-----

HsaPMS2 400 -----QDQSPSLRTGEEKKQVSIISRLREAFS--LRHTTENKPHSPKTPPPRSPPL  
 DmePMS2 430 -----KDDQHEEVSEEDVPTTSTORFMDVLTQWRRRTGDTKGTAPSVPVKRCSE  
 ScePMS1 424 -----SDNENYHTARSQSNQSNHAFNSTTGVIDKSNGETELTSVMDGNYTNVTD  
 AthPMS1 402 KENPSLREVEIDNSSPMEKFKFKIKACGTKKGEGLSVHVDVTHLDKTPSKGLPQLNVTEK  
 TbrPMS1 371 -----WRHVPERRNTQETMPTQPLSAISIAQFIYQRRETSQGDNL

HsaPMS2 448 GQKRGMLSSSTSGAISDKGVLRPQKEAVSSSHGSPSOPTDRAEVE--KDSGHGSTVSDSEG  
 DmePMS2 479 SEE-LVTRILKMKIHTFLSQESPNEQSSKCEAESDAASDVDA--KKD-KTQVSLDNSF  
 ScePMS1 473 VIGSECEVSDSSVLDENSSPTTKKLSIKTDSQNLSDLNINNFNPEFQNTISPDKA  
 AthPMS1 462 VTDASKDLSRSSFQASTLNTFVTMGKRKHENISILSETPVLRNQTSSYRVEKSKFEVR  
 TbrPMS1 412 IDNAAVAQVQPVVCSQLLSGSSPIGRSTSSPDATASPTSTTNRAPERSAGSVEVLEYP

HsaPMS2 506 FSIPTDGTSHCSSEYAASSPGDRGSQEHVDSQEKAP--ETDDSFSDVDCHSNQEDTGCKFRV  
 DmePMS2 535 LNLKELAKESEAYDLLTKPQN---TPRIDCKVLTPIKSRRSIFEFKTDAMALDK----QE  
 ScePMS1 533 RSLEKVVEE PVYFDIDGEKFQEKAVLSQADGLVFDVNECHEHTNDCCHQERRGSTDEQD  
 AthPMS1 522 ALASRCLVEGDQLDDMVISKEDMTPSERDSELGNRISPGTQADNVERHEREHEKPIRFEF  
 TbrPMS1 472 LTFEPTQKRQ-----RLES

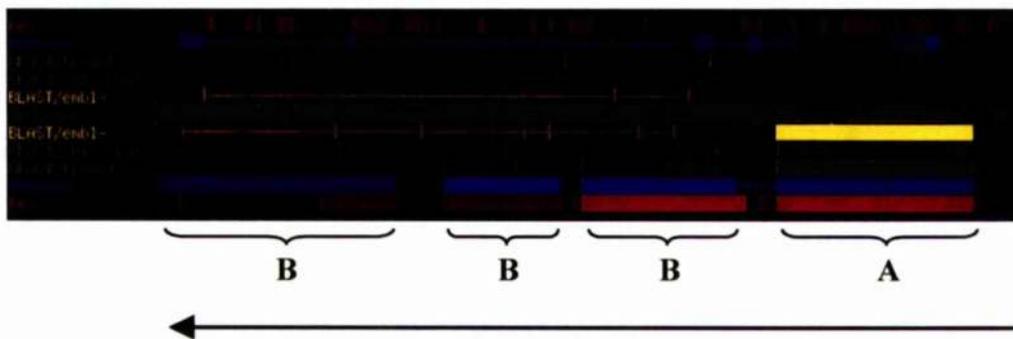
HsaPMS2 565 LPQPTNLATPNTKRFKKEILSSSDICQKLVNTQDMSASQVDVAVKINKKVVPTDFSS-M  
 DmePMS2 588 SPAELNSQPPSFTQLSEETDIASDEDDHIELPTRIEFDEEMEEGLPSNFSSGE-ITTS-L  
 ScePMS1 593 DEADSIYAEIEPVEINVRTPLKNSRKSISKDNYRSLSDGLTHRKFEDILEYN-ITSTKNF  
 AthPMS1 582 PTDNTLTKGDVERVSEDNPRCSQPLRSVATVLDSPAQSTGPKMFSTLEFSFQNLRTRRLL  
 TbrPMS1 486 SAEIGNTGDTDGSSWGEDPRELGPPEMMGTDDAVVNYVTDDTNQQRPGPPRSSRFPPF

HsaPMS2 623 SSIKAKRIKQLHHEAQQSEGEQNYR--KERAKICPGENQ-----AAEDELKKEISKTM  
 DmePMS2 646 EETASSLKAHEQQQRDRRTRTKLQRLRFKTEINPNQNT-----SAEAELOREIDKED  
 ScePMS1 652 KEISKNGKQMSIISKRKSEAQENIKKQDELEDFEQ-----GEKYLTLTVSKND  
 AthPMS1 642 ERISRLQSTGYVSKCMNTPQPKKFAAATLELSQPDDDEERKARALAAATSELERLFRKED  
 TbrPMS1 546 SVIAEMPLVHSLGEWAAPSQPSDGGGVRFSRLQKQT-----EITLYGES



**Figure 3.40 Analysis of the genomic environment of the putative *T. brucei* *PMS1* gene**

The nucleotide sequence of contig TRYP9.0.000851 was analysed using the Nucleotide Identification (NIX) server (<http://www.hgmp.mrc.ac.uk/Registered/webapp/nix/>) provided by the UK Human Genome Mapping Project Resource Centre. The output shows the results of selected programs which were used in the analysis. The sequence regions identified correspond to: (A) the putative *T. brucei* *PMS1* ORF; and (B) a number of hypothetical ORFs suggested by the HMMGene gene-prediction and Fex exon-prediction programs. The predicted direction of transcription is indicated by a black arrow.



### 3.5 Discussion

A total of seven genes have been identified in *T. brucei* on the basis of their similarity to *mutS*- and *mutL*-related genes from other organisms. Two putative *mutS* homologues, *MSH2* and *MSH8*, have been cloned and characterised, while the sequence of a third, a putative orthologue of *MSH3*, has been analysed in this chapter. I propose that these three genes encode polypeptides active in MMR (see below). During BLAST searches of the *T. brucei* genome sequencing databases, two further putative MutS homologues were encountered (data not shown). These genes probably encode the MutS homologues MSH4 and MSH5 which form a heterodimer active during meiotic crossing-over (Ross-Macdonald and Roeder, 1994; Hollingsworth *et al.*, 1995; Pochart *et al.*, 1997). Analysis of the sequence and functions of these genes is beyond the scope of this work. Furthermore, two *mutL*-related genes have been isolated from *T. brucei*, the first encoding a putative orthologue of MLH1, and the second encoding a putative orthologue of PMS1 from *S. cerevisiae*. No genes bearing similarity to any of the other *mutL* homologues identified in eukaryotes (for instance *MLH2* and *MLH3* in *S. cerevisiae*) were encountered during this study. However, the *T. brucei* genome sequencing project is still in progress, and it is possible that more *mutL*- and *mutS*-related genes may come to light.

The three putative MutS homologues identified in *T. brucei* which are thought to be active in MMR, *MSH2*, *MSH3* and *MSH8*, share sequence characteristics with all MSH family members. Towards the C-terminus, two highly conserved motifs comprising an ATP-binding domain and a helix-turn-helix motif are present in each polypeptide (Gorbalenya and Koonin, 1990; Adé *et al.*, 1999; Biswas *et al.*, 2001). A middle-conserved domain, which is thought to be involved in protein-protein interactions and which is found in all MutS homologues is also present in the three polypeptides (Culligan *et al.*, 2000). A further motif has only been observed in bacterial MutS and the eukaryotic *MSH1*, *MSH2*, *MSH3*, *MSH6*, and *MSH7* proteins, and can therefore be considered characteristic of the MutS homologues involved in MMR. This motif is a conserved N-terminal domain which is involved in mismatch interaction (Malkov *et al.*, 1997; Culligan *et al.*, 2000; Schofield *et al.*, 2001), and it is present in the three *T. brucei* polypeptides under investigation, suggesting they are indeed involved in MMR.

The three *T. brucei* MutS-related polypeptides have been classified on the basis of sequence comparisons with MutS homologues from other eukaryotes. Sequence comparison of the first *T. brucei* MutS homologue identified that this gene encodes an orthologue of *MSH2*. Clear evidence for this is provided by the fact that *T. brucei* *MSH2*

shows greater homology to its *A. thaliana*, *H. sapiens* and *S. cerevisiae* counterparts (47%, 46% and 44% similarity respectively) than it does to either *T. brucei* MSH3 (35% similarity) or MSH8 (35% similarity). The predicted *T. brucei* MSH2 polypeptide is also similar in length to other MSH2 orthologues. Sequence comparisons using the second *T. brucei* MutS homologue to be identified indicated that this gene encodes a member of the MSH3/MSH6 subfamily. However, although it bore marginally greater similarity to MSH6 and MSH7 homologues (between 1.5 and 8.0%) than to MSH3 orthologues, it could not be classified beyond doubt. The third *T. brucei* MutS homologue was identified very recently, and sequence comparisons using this polypeptide suggested that this gene also encodes a member of the MSH3/MSH6 sub-family, although in this case the polypeptide showed a marginally greater similarity (between 1.0 and 6.0%) to the MSH3 orthologues used in the analysis. Global multiple alignment of these polypeptides with MSH3, MSH6 and MSH7 homologues allowed the identification of short sequence motifs in the *T. brucei* sequences characteristic of either MSH3 or MSH6/MSH7 family members. On the basis of sequence similarity and the data gleaned from the global multiple alignment the third *T. brucei* MutS homologue was designated MSH3, but the second *T. brucei* MutS-related gene was assigned the name *MSH8* because while it can be identified as a member of the MSH6/MSH7 sub-family it lacks the N-terminal extension characteristic of these proteins.

*T. brucei* MSH3 is similar in length to other MSH3 orthologues, a fact which strengthens its classification within this group. However, the putative MSH8 polypeptide is also similar in length to MSH3 orthologues. This observation was confirmed by RT-PCR which identified the 5' splice site of the gene, which lies 72 bp upstream of the proposed start codon, indicating that the predicted ORF is correct. This means that the MSH8 polypeptide is N-terminally truncated in comparison with other members of the MSH6/MSH7 sub-family. There are a number of possible explanations for this. It may be that *MSH8* simply encodes an orthologue of MSH6 but because of differences in the *T. brucei* MMR system in comparison with that of yeast and mammals, the N-terminal extension has been lost, although the function of the protein remains similar. Since the *T. brucei* genome sequencing project is not yet complete, an alternative possibility is that another member of the MSH6/MSH7 sub-family will come to light which encodes the true MSH6 orthologue in *T. brucei*, and that MSH8 recognises a different set of mismatches to either MSH3 or MSH6. This may suggest it performs a similar role to that of MSH7 in plants, or maybe a function unique to trypanosomes. The biochemical consequences of a third MutS-related binding partner for MSH2 in plants remains to be determined. From the global multiple alignment it is apparent that *T. brucei* MSH8 is divergent in comparison with the MSH6 and MSH7 homologues, although this is also true of *T. brucei* MSH2 and

MSH3 with respect to the MSH2 and MSH3 orthologues, making it difficult to ascertain if this reflects sequence differences due to trypanosomes being early-diverging eukaryotes (Sogin *et al.*, 1986; Sogin, 1989) or because of altered functional properties of the MSH8 polypeptide. A final explanation may be that trypanosomes diverged from the main eukaryote line after the duplication of the progenitor of MSH3 and MSH6, but before MSH6 acquired its N-terminal extension. Without performing mutational analysis or other functional characterisations of *T. brucei* MSH8 it is impossible to determine the exact function of this gene.

Future analysis of the conserved N-terminal mismatch interacting domain would provide a definitive indication as to the function of the *T. brucei* MutS homologues. MutS-related polypeptides which possess this domain probably function in MMR, as the sequence is absent from the type II MutS homologues (for instance, MSH4 and MSH5) which do not interact with mismatches (Ross-Macdonald and Roeder, 1994; Rossolillo and Albertini, 2001). Proteins which bind base-base mismatches, including *E. coli* MutS and *Thermus aquaticus* MutS possess a conserved Phe-X-Glu DNA binding motif, where the glutamic acid residue hydrogen bonds with either an unpaired thymidine or the thymidine of a G-T mismatch (Schofield *et al.*, 2001). This motif forms part of the conserved N-terminal mismatch interacting domain and is present in MSH6 and MSH7 homologues from *A. thaliana*, *H. sapiens* and *S. cerevisiae*, although MSH6 proteins are also required for the repair of small IDLs. Examination of the predicted *T. brucei* MSH8 polypeptide reveals the presence of this conserved DNA binding domain suggesting this protein may interact with base-base mismatches. The MSH3 proteins, which are involved in the repair of larger IDLs, do not have the conserved Phe-X-Glu motif but do show sequence conservation within this region. Across these residues the predicted *T. brucei* MSH3 polypeptide shows greater identity with the MSH3 orthologues studied than with the MSH6 and MSH7 homologues from other eukaryotes. The mismatch interacting domain of MSH2 orthologues also does not contain the Phe-X-Glu motif, probably because MSH2 is not responsible for mismatch specificity, although the equivalent residues in MSH2 orthologues do show conservation. Surprisingly, the predicted *T. brucei* MSH2 polypeptide possesses dissimilar amino acid residues in this region and is divergent in comparison with the other MSH2 proteins shown over the length of the mismatch interacting domain. The functional implications of this, as well as the mismatch recognition predictions discussed above for the putative *T. brucei* MSH8 and MSH3 polypeptides, could be investigated by DNA binding studies *in vitro* combined with site-directed mutagenesis.

Another conserved motif often present at the extreme N-terminus of members of the MSH3/MSH6 sub-family, but not in MSH2 orthologues, is the PCNA binding domain (Clark *et al.*, 2000; Flores-Rozas *et al.*, 2000). This motif is found in the MSH3, MSH6 and MSH7 homologues from *A. thaliana*, *H. sapiens* and *S. cerevisiae*. Analyses of the putative *T. brucei* MSH3 and MSH8 amino acid sequences did not reveal convincing PCNA binding motifs anywhere along the length of these polypeptides, nor was a consensus motif apparent in *T. brucei* MSH2, MLH1 or PMS1. Although partial PCNA binding sites could be identified in *T. brucei* MSH3 and MSH8, the location of the amino acids residues in question, in terms of domain organisation, would argue against their serious consideration. PCNA binding is required not only at an early stage in mismatch repair preceding mismatch excision (Umar *et al.*, 1996), but also during DNA resynthesis following the excision of the mismatch (Gu *et al.*, 1998). Lack of a PCNA binding motif is not unprecedented, however, as *S. pombe* SWI4 (an orthologue of MSH3) also lacks the consensus motif although it is present in *S. pombe* MSH6 (Kleczkowska *et al.*, 2001), and analyses of MSH6 from *D. melanogaster* and MSH7 from *Zea mays* retrieve only partial consensus motifs where both proteins lack the conserved glutamine residue. Moreover, interaction with PCNA is not limited to proteins possessing the conserved binding motif. For instance, Gadd45, which may be involved in regulating DNA repair, and CAF1, which links chromatin assembly to DNA replication, both interact with PCNA using different amino acid residues (Kearsey *et al.*, 1995; Shibahara and Stillman, 1999). Furthermore, two-hybrid experiments in yeast indicate that PCNA interacts with MLH1 (Umar *et al.*, 1996), suggesting mismatch repair proteins can bind at sites other than the interdomain connector loop of PCNA as MLH1 does not contain the PCNA consensus binding motif. Whether the likely interactions of MSH8 or MSH3 and PCNA are mediated by a standard PCNA binding motif that is simply highly diverged in sequence, or whether they employ non-canonical binding interactions, would require further analysis.

The two putative *T. brucei* MutL-related genes were classified on the basis of sequence homology with MLH family members from other eukaryotes. The first MutL homologue identified in *T. brucei* has been designated MLH1 because it is evident from sequence comparisons that this polypeptide bears greater similarity to MLH1 orthologues from *A. thaliana*, *D. melanogaster*, *H. sapiens*, and *S. cerevisiae* (43%, 42%, 48% and 47% similarity respectively) than to any of the other MutL-related proteins used in the analysis. The *T. brucei* MLH1 polypeptide is also within the expected size range for MLH1 orthologues. The second *T. brucei* MLH family member has been designated PMS1 since it shows greater homology to orthologues of PMS1 from *A. thaliana* and *S. cerevisiae* (42% and 40% similarity respectively), and their closest homologues in animals, PMS2

from *D. melanogaster* and *H. sapiens* (42% and 43% similarity respectively) than to the other MutL-related proteins analysed. The predicted *T. brucei* PMS1 polypeptide is also similar in length to orthologues of PMS1 from *S. cerevisiae*.

MLH family members are characterised by two sequence motifs: a highly conserved N-terminal ATP-binding domain, and a C-terminal dimerisation domain which shows more diffuse sequence conservation (Bergerat *et al.*, 1997; Mushegian *et al.*, 1997; Pang *et al.*, 1997; Ban and Yang, 1998; Dutta and Inouye, 2000; Kondo *et al.*, 2001). The N-terminal ATP-binding domain of MutL homologues has been associated with a weak ATPase activity that is required for MMR (Ban and Yang, 1998; Spampinato and Modrich, 2000; Hall *et al.*, 2002). Both the putative MutL homologues isolated from *T. brucei*, MLH1 and PMS1, possess regions bearing high identity to this domain suggesting the ATPase activity is conserved in *T. brucei*. However, the predicted *T. brucei* PMS1 polypeptide contains several insertions within this domain and without further study it is impossible to determine what the functional implications of these are. The conserved C-terminal dimerisation domain is required, as implied in its name, for interactions between two MutL homologues. The regions required for interaction between *S. cerevisiae* MLH1 and PMS1 have been well characterised and equivalent regions are present in the two putative *T. brucei* MutL homologues, indicating they are also likely to interact. A further motif, which has only been observed in orthologues of MLH1 and which is therefore considered characteristic of these proteins, is the highly conserved C-terminal homology domain (CTH). The putative *T. brucei* MLH1 orthologue shows a high level of identity with MLH1 orthologues from other eukaryotes within this region. Deletion and mutation analyses of the CTH of yeast MLH1 indicated that this sequence was essential for MMR activity, but was not required for interaction with PMS1 (Pang *et al.*, 1997). Given the high level of conservation of this region in the putative *T. brucei* MLH1 polypeptide, this is probably also the case in trypanosomes.

At the genomic level, the *T. brucei* *MSH2*, *MSH8*, *MLH1* and *PMS1* genes are single copy, as evidenced by their detection in Southern analyses as single bands following restriction digestion of genomic DNA with enzymes which do not recognise any sites in the probe sequence. Furthermore, since the restriction pattern is identical in digestions of both MITat 1.2 and ILTat 1.2 genomic DNA it can be concluded that the genomic environment of each gene is the same in these two strains of *T. brucei*. This is important because the *MSH2*, *MLH1* and *PMS1* gene sequences were derived from ILTat 1.2 genomic DNA (cloned into the LambdaGEM-12 vector to form an ILTat 1.2 genomic library), whilst most of the downstream analyses of these genes, for instance analysis of the expression of

the genes, were performed using DNA or RNA isolated from MITat 1.2 cells. Moreover, this rules out the possibility that the difference in *VSG* switching rates between the pleomorphic ILTat 1.2 strain and the monomorphic MITat 1.2 strain are due to loss, or other gross structural alterations, in any of these genes.

RT-PCR using total RNA isolated from both bloodstream and procyclic form MITat 1.2 trypanosomes shows that *MSH2*, *MSH8*, *MLH1* and *PMS1* are all expressed in both life cycle stages. The expression of *T. brucei MSH3* was not analysed. Since expression levels in *T. brucei* are predominantly regulated by post-translational mechanisms, it was deemed sufficient to demonstrate expression of the target ORFs by RT-PCR, as the technique is simpler and more sensitive than Northern analysis. In order to determine the expression levels of a protein in *T. brucei* it would be necessary to perform Western analysis using antibodies specific to that protein. At present such antibodies are not available for any of the mismatch repair proteins in *T. brucei*, although this would clearly be of interest in terms of understanding whether or not their levels of expression are modulated by any environmental or cellular signals.

Despite the obvious similarities between the *T. brucei* MMR genes and their counterparts in other eukaryotes, definite proof that their gene products are active in MMR in *T. brucei* can only be obtained through analysis of mutant cell lines. Since *MSH2* is the primary MutS homologue involved in MMR, and *MLH1* is the most important *mutL*-related gene, these two genes were chosen for mutant analysis in *T. brucei*. The gene knockout strategies and functional analyses of the putative *MSH2* and *MLH1* genes are described in chapter 4.

# CHAPTER 4

**Generation and analysis of homozygous  
*MSH2* and *MLH1* knockouts in *T. brucei***

## 4.1 Introduction

In order to determine the roles of MMR in *T. brucei* it was decided that functional studies of both the putative *MSH2* and *MLH1* genes were warranted as their gene products are believed to be required to form all complexes active in nuclear MMR (Harfe and Jinks-Robertson, 2000). Studies of *MSH2* and *MLH1* null mutants in this organism will reveal whether *T. brucei* possesses a functional MMR system, and if this system acts in a manner similar to that described in other eukaryotes. Since the primary reason for studying MMR in *T. brucei* is its possible role in regulating DNA recombination, and thereby antigenic variation, it was deemed necessary to study both *MSH2* and *MLH1* because, while they act in concert during MMR, *MSH2* has been associated with activities independent of *MLH1* during a number of recombination processes. The *MSH2-MSH3* complex in *S. cerevisiae* has been shown to impact upon SSA and gene conversion, where it is required to stabilise the recombination intermediates and allow the *RAD1-RAD10* endonuclease to process non-homologous 3' tails (Sugawara *et al.*, 1997; Studamire *et al.*, 1999). Also, the *MSH2-MSH6* heterodimer has been found to bind Holliday junctions and enhance their cleavage by phage resolution enzymes *in vitro* (Marsischky *et al.*, 1999). Furthermore, in *S. cerevisiae* loss of *MSH2* resulted in the greatest increase in the rate of homeologous recombination relative to the wild-type (Nicholson *et al.*, 2000). Similar experiments using *MLH1* mutants revealed smaller increases in recombination between divergent sequences, suggesting that *MSH2* acts to block recombination by a pathway independent of *MLH1*. If *MSH2* can also act independently of *MLH1* during some processes in *T. brucei*, then these differences may be revealed by comparison between the phenotypes of null mutants of both these genes in this organism.

*T. brucei* is diploid in all life cycle stages that have been described, and for this reason any strategy to create null mutants must disrupt both alleles. Recently, strains enabling researchers to disrupt gene expression using RNA interference (RNAi) have been developed in *T. brucei* (Wirtz *et al.*, 1999; Bastin *et al.*, 2000; LaCount *et al.*, 2000; Shi *et al.*, 2000; Wang *et al.*, 2000), but the use of this strategy was inappropriate for this work for two reasons. Firstly, given that neither *MSH2* or *MLH1* are essential genes in either prokaryotes or unicellular eukaryotes, it is unlikely that stable knockouts would be lethal in *T. brucei*. Secondly, the main aim of this research was to determine whether MMR is involved in the regulation of *VSG* switching in this organism. In order to study the effect of *MSH2* and *MLH1* gene disruption upon antigenic variation, the null mutants had to be created in the 3174 transgenic strain of *T. brucei* described in McCulloch *et al.* (1997).

This strain underpins the only assay available which allows researchers to assay both the frequency of *VSG* switching and determine the relative contribution of the *in situ* and gene conversion switching mechanisms (see section 4.2.8). During this assay cells which have switched their *VSG* coat are selected in mice, and it is not known whether it is possible to sustain RNAi induction in trypanosomes grown in a host organism. This study has used the well-established method of gene disruption by performing two rounds of transformation with linear resistance cassettes possessing terminal sequence homology to the gene of interest (Ray and Hines, 1995; Li *et al.*, 1996; McCulloch and Barry, 1999). This strategy creates stable integrants in which the gene of interest has been inactivated, enabling long-term functional studies. The 3174 strain was derived from MIT at 1.2a S427 bloodstream forms, and contains antibiotic resistance markers for hygromycin B and G418 within its active expression site. Since the 3174 strain is unaltered from its parent background at all other loci, other functional studies of the *MSH2* and *MLH1* mutants reported in this chapter were also performed in this strain.

## 4.2 Generation and analysis of homozygous *MSH2* knockout trypanosomes

### 4.2.1 Introduction

*MSH2* underpins the entire MMR system in eukaryotes. This protein is required for the formation of all the MutS-related mismatch-interacting complexes, *MSH2-MSH3*, *MSH2-MSH6* and *MSH2-MSH7*, and so is necessary for the recognition of all mismatches repaired by MMR (Culligan and Ilyas, 2000; Harfe and Jinks-Robertson, 2000). Loss of *MSH2* therefore leads to the complete elimination of nuclear MMR activity (Harfe and Jinks-Robertson, 2000). Several phenotypic changes are associated with the loss of *MSH2*, and a number of experimental approaches are available to characterise these in *T. brucei*. Primarily, cells develop a mutator phenotype, and although the frequency of point mutations is elevated, there is a greater effect on the stability of simple, repetitive sequences called microsatellites (Levinson and Gutman, 1987; Strand *et al.*, 1993; de Wind *et al.*, 1995). For this reason, the effect of *MSH2* mutation on a number of microsatellite loci in *T. brucei* will be determined in this study (see section 4.2.7). Furthermore, inactivation of *MSH2* has been shown to lead to an increase in resistance to a number of cytotoxic chemicals, including alkylating agents and the anti-cancer drug cisplatin (Karran and Marinus, 1982; Aebi *et al.*, 1997; Marra and Schär, 1999). The survival of *MSH2* mutant trypanosomes in the presence of the alkylating agent *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine (MNNG) will therefore be assessed during this work (see section 4.2.6). Loss of *MSH2* has also been shown to lead to an increase in recombination between non-identical or homeologous DNA sequences, because mismatches formed during strand invasion are no longer recognised, and the recombination reaction, which would be prevented in MMR competent cells, is allowed to proceed (Alani *et al.*, 1994; de Wind *et al.*, 1995; Selva *et al.*, 1995; Datta *et al.*, 1996). As the frequency of *VSG* switching is greatly decreased in trypanosomes lacking RAD51, a primary component of the homologous recombination machinery, it is hypothesised that other proteins involved in the catalysis or regulation of homologous recombination may also affect *VSG* switching (McCulloch and Barry, 1999). While a system to study the effect of *MSH2* mutation on recombination between divergent sequences is described in Chapter 5, the possibility that *MSH2* is involved in regulating recombination between DNA sequences during *VSG* switching was investigated by examining the role of *MSH2* in trypanosome antigenic variation (see section 4.2.8).

#### 4.2.2 Design and generation of *MSH2* knockout constructs

In order to examine the function of *MSH2* in *T. brucei*, a strategy to create transgenic trypanosomes from which both copies of the *MSH2* ORF were deleted was developed. Alignment of the protein sequence of *T. brucei* *MSH2* (see section 3.1.7) with *MSH2* proteins from other eukaryotes revealed that the functional domains are spaced throughout the length of the polypeptide, and it has been reported that mutations in one domain do not necessarily inactivate all the activities of the *MSH2* protein (Drotschmann *et al.*, 2001). For this reason it was determined that both copies of the entire *MSH2* ORF should be deleted from the genome, in order that no residual *MSH2* activity could remain.

*T. brucei* *MSH2* knockout constructs were designed to contain the sequences immediately 5' and 3' of the *MSH2* ORF to act as targeting sequences for replacement of the entire *MSH2* ORF by homologous recombination (Figure 4.1). These targeting flanks were separated by one of two antibiotic resistance cassettes, containing either the puromycin-N-acetyltransferase ORF (*PUR*) or the blasticidin-S-deaminase ORF (*BSD*), flanked by 5' and 3' processing signals from the tubulin locus. After integration, the resistance genes should be expressed by endogenous *MSH2* transcription, since the knockout constructs contain no promoter elements.

PCR primers were designed to amplify 430 bp of 5' flank (*MSH2* 5'(A) and *MSH2* 5'(B)) and 370 bp of 3' flank (*MSH2* 3'(A) and *MSH2* 3'(B)) from the *MSH2* locus. PCR from MITat 1.2 genomic DNA using these primers generated products of the expected sizes (Figure 4.2; see section 2.11.1). The *MSH2* targeting flanks were then cloned simultaneously into pBluescript II KS, before the *BSD*- or *PUR*-containing resistance cassettes were cloned between the *MSH2* flanks to generate the  $\Delta$ *MSH2*::*BSD* and  $\Delta$ *MSH2*::*PUR* constructs respectively (see section 2.11.1).

#### 4.2.3 Design and generation of *MSH2* re-expression construct

In order to determine whether any effects on MMR or antigenic variation were due to the loss of *MSH2* and not incidental to changes which arose during electroporation, selection or cloning of the transformants, it was necessary to generate a construct capable of re-expressing the *MSH2* gene in homozygous mutants (Figure 4.3). The  $\Delta$ *MSH2*::*MSH2-BLE* construct was designed to contain the *MSH2* ORF flanked by the same 5' and 3' *MSH2* processing flanks used as targeting flanks in the knockout constructs (see section 4.2.2). This construct also incorporated a bleomycin resistance protein ORF

(*BLE*); to allow the selection of transformants this was flanked by processing signals derived from the actin locus at the 5' end, and the  $\alpha$ - $\beta$  tubulin intergenic region at the 3' end, positioned downstream of the 3' end of the *MSH2* 3' flank. This construct was designed to allow the re-expression of *MSH2* from its original genomic location, and since it does not contain a promoter element both *MSH2* and the resistance gene will be expressed by endogenous *MSH2* transcription, circumventing the potential problem of altered *MSH2* expression levels which could occur if an ectopic locus was used. The *MSH2* 5' flank and the  $\alpha$ - $\beta$  tubulin intergenic region (also used as the 3' processing flank for the *BSD* and *PUR* ORFs in the *MSH2* knockout constructs) form the targeting flanks allowing integration of the construct into either the *MSH2::BSD* or *MSH2::PUR* alleles in the *MSH2*<sup>-/-</sup> mutants.

The  $\Delta$ *MSH2::MSH2-BLE* construct was derived from the *MSH2* lambda sub-clone which contains the *MSH2* ORF and its processing flanks (pJB100) (see section 3.1.3). The *BLE* resistance cassette was cloned into the *EcoRV* site of the multiple cloning site downstream of the *MSH2* 3' flank (see section 2.12).

#### 4.2.4 Generation of *MSH2* mutant cell lines in *T. brucei*

Three independent *MSH2* bloodstream stage mutants were generated by direct transformation of 3174 bloodstream cells with the linearised knockout constructs (see section 2.11.1). Initially, the  $\Delta$ *MSH2::PUR* construct was electroporated into a number of independent cell populations and transformants selected on semi-solid agarose plates containing 1.0  $\mu\text{g}\cdot\text{ml}^{-1}$  puromycin dihydrochloride to generate putative heterozygous (*MSH2*<sup>+/-</sup>) mutants (see section 2.1.3). Three independent first round transformants were then electroporated with the  $\Delta$ *MSH2::BSD* construct and transformants selected on semi-solid agarose plates containing 2.5  $\mu\text{g}\cdot\text{ml}^{-1}$  blasticidin S hydrochloride to generate putative homozygous (*MSH2*<sup>-/-</sup>) mutants. Finally, a single second round transformant was electroporated with the linearised re-expression cassette  $\Delta$ *MSH2::MSH2-BLE* and transformants selected in HMI-9 containing 2.5  $\mu\text{g}\cdot\text{ml}^{-1}$  phleomycin in 24-well plates to generate a putative re-expressor (*MSH2*<sup>-/+</sup>) mutant (see sections 2.1.3 and 2.12).

Correct integration of the constructs into the *MSH2* locus was determined by Southern analysis. Genomic DNA from each transformant was digested with *SacI*, separated by agarose gel electrophoresis, Southern blotted, and probed with the *MSH2* 5' flank described in section 4.2.2. Figure 4.4 shows the results of this analysis for untransformed

3174 trypanosomes, the three independent heterozygous *MSH2* knockouts, the three homozygous *MSH2* knockouts derived from these heterozygotes, and the single *MSH2* re-expressor generated from one of the homozygotes. In the untransformed cells, a 5.8 kb restriction fragment was detected, corresponding to the intact wild-type *MSH2* locus. This fragment was also detected in the three *MSH2*<sup>+/+</sup> mutants, but was absent from the three *MSH2*<sup>-/-</sup> cell lines. The three first round transformants also possessed a 2.85 kb fragment corresponding to the *MSH2* allele replaced by the  $\Delta$ *MSH2*::*PUR* construct, and this allele was retained in the three second round knockout cell lines. The digestions of the three homozygous mutants also revealed a 2.65 kb fragment derived by replacement of the second *MSH2* allele by the  $\Delta$ *MSH2*::*BSD* construct. The two fragments present in the *MSH2*<sup>-/-</sup> mutants designated 1.1 and 2.1 merged into one another in both digestions because the gel was overloaded with DNA resulting in an over-intense hybridisation signal. Finally, two bands were detected in the digestion of the *MSH2*<sup>-/+</sup> cell line. The 2.85 kb fragment corresponded to the *MSH2*::*PUR* allele, while the *MSH2*::*BSD* allele was replaced by the  $\Delta$ *MSH2*::*MSH2-BLE* construct, thus restoring a copy of the *MSH2* gene to the genome and creating a 6.2 kb fragment. For each cell line, Southern analysis indicated that the knockout and re-expression constructs integrated into the *MSH2* locus as expected to generate the desired *MSH2* mutant trypanosomes.

PCR using primers internal to the *MSH2* ORF (*MSH2D5* and *MSH2U2*; Figure 4.1) was performed using genomic DNA isolated from the putative 3174 *MSH2* mutants in order to show that no copies of the *MSH2* ORF existed within the *MSH2*<sup>-/-</sup> mutant cell lines (Figure 4.5; see section 2.11.1). The integrity of the DNA template was confirmed using primers specific to the large subunit of *T. brucei* RNA polymerase I (Poll 5' and Poll 3'; Rudenko *et al.*, 1996). PCR amplification using the *MSH2* internal primers revealed that the untransformed cells, the three heterozygous mutants and the *MSH2* re-expressor all gave a product of the expected size. No product was visible in the three homozygous mutants, however, indicating that no further copy of *MSH2* had been created upon disruption of the two wild-type alleles.

Finally, in order to confirm that no wild-type *MSH2* mRNA was transcribed in the homozygous mutants, total RNA was isolated from the transformants and cDNA generated by reverse-transcription (see section 2.4). The cDNA was used as a template for PCR amplification using the *MSH2* internal primers used above (Figure 4.6), and the integrity of the cDNA was confirmed as above. Reverse transcription PCR was used in this analysis as antibodies against the *T. brucei* *MSH2* polypeptide are not currently available. A PCR

product of the expected size was amplified from the untransformed cells, the three *MSH2*<sup>+/+</sup> mutants and the *MSH2*<sup>-/-+</sup> cell line, but not from the *MSH2*<sup>-/-</sup> mutants. It can therefore be concluded that it is possible to remove all intact transcripts of the *MSH2* gene, and thus all expression of the MSH2 protein, and trypanosomes remain viable. These results again confirm that the knockout and re-expression constructs integrated into the cells as expected, and that the desired *MSH2* mutant trypanosomes were generated.

#### 4.2.5 Growth of *MSH2* knockout trypanosomes *in vitro*

The *in vitro* growth rate of the *MSH2* mutants was determined by diluting mid-log phase trypanosomes to a concentration of  $1 \times 10^5$  cells.ml<sup>-1</sup> in HMI-9 and measuring the trypanosome concentrations every 24 h thereafter (Figure 4.7). It is clear that *MSH2*<sup>-/-</sup> cells are not only viable but show no change in growth compared to the wild-type cells. The average population doubling time during exponential growth was 7.4 h for the *MSH2* wild-type cells, 6.5 ( $\pm$  0.5) h for the *MSH2*<sup>+/+</sup> cell lines, 6.5 ( $\pm$  0.7) h for the *MSH2*<sup>-/-</sup> mutants and 6.69 h for the *MSH2*<sup>-/-+</sup> cell line. All the cultures reached similar maximal densities (data not shown).

#### 4.2.6 Survival of *MSH2* mutant trypanosomes in the presence of the alkylating agent, *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine (MNNG)

Alkylating agents, including MNNG, are an important group of mutagenic and cytotoxic chemicals whose toxicity is affected by MMR. Primarily, these drugs act by generating O<sup>6</sup>-methylguanine (O<sup>6</sup>-meG) lesions by methylating the O<sup>6</sup> position of guanine. During replication, thymine is misincorporated opposite these lesions, forming mismatches that are recognised but not repaired by the MMR system (Griffin *et al.*, 1994; Duckett *et al.*, 1996). Inactivation of this repair pathway confers resistance to the cytotoxicity of alkylating agents, and results in cellular hypermutability, a phenomenon known as methylation tolerance (reviewed in Karran and Bignami, 1994; Aquilina and Bignami, 2001). Given that O<sup>6</sup>-meG is a highly mutagenic and genotoxic modified base which can arise during normal cellular processes, a highly conserved protein has evolved, O<sup>6</sup>-methylguanine DNA methyltransferase (MGMT), which repairs these lesions in a single-step, error-free reaction. However, in many cell types this protein is expressed at low levels, and treatment with MNNG may overwhelm the normal repair pathway, leading to tumorigenesis or cell death in MMR-proficient cells (reviewed in Bignami *et al.*, 2000).

Several models have been advanced to explain methylation tolerance in MMR mutants. The most widely accepted is termed the 'futile cycle model' (Goldmacher *et al.*, 1986; Karran and Bignami, 1992; Vaisman *et al.*, 1998). It has been suggested that attempts by the MMR system to correct O<sup>6</sup>-meG-containing mispairs are abortive since correction will be directed to the newly synthesised strand while the methylated base is retained in the parental strand. Since resynthesis will again result in an O<sup>6</sup>-meG-containing mismatch, this process leads to repetitive cycles of excision and resynthesis and thus single-stranded regions at the site of O<sup>6</sup>-meG residues. Replication across these regions results in the formation of double-strand breaks, and hence chromosomal aberrations or cell death (D'Atri *et al.*, 1998; Toft *et al.*, 1999). An alternative model proposes that binding of the mispair by the MutS- and MutL-related heterodimers shields the lesion from other DNA repair proteins, allowing such adducts to persist in the genome (Mello *et al.*, 1996). Further models suggest that assembly of the MMR machinery at O<sup>6</sup>-meG-containing mismatches directly activates a damage-signalling cascade which leads to cell cycle arrest or cell death (Carethers *et al.*, 1996; Moreland *et al.*, 1999; Berardini *et al.*, 2000). These processes are avoided in MMR-deficient cells, and loss of MSH2 has been shown to result in methylation tolerance or resistance to alkylating agents in mammalian cells (de Wind *et al.*, 1995; Humbert *et al.*, 1999).

In order to determine if loss of *MSH2* in *T. brucei* resulted in increased resistance to MNNG, the survival of *MSH2* wild-type cells, *MSH2*<sup>+/-</sup> mutants, *MSH2*<sup>-/-</sup> mutants and *MSH2*<sup>-/+</sup> cells was assayed in the presence of increasing concentrations of MNNG. Cells were plated in HMI-9 containing 0, 0.25, 0.5, 0.75 or 1.0 µg.ml<sup>-1</sup> MNNG at a density of 2 cells per well in 96-well plates in triplicate, and after 14 days the number of wells containing growth were counted. The results of this assay are shown in tabular form in Table 4.1, and as a graph in Figure 4.8. It is clear from these data that loss of *MSH2* results in increased resistance to MNNG in *T. brucei*. Exposure to 0.25 µg.ml<sup>-1</sup> MNNG results in a reduction in survival of 27% in *MSH2* wild-type cells, but of only 6% and 12% in the *MSH2*<sup>+/-</sup> mutants. At this concentration the *MSH2*<sup>-/-</sup> mutants and *MSH2*<sup>-/+</sup> cells were barely affected with reductions in survival of only 3% and 1% respectively. At 0.5 µg.ml<sup>-1</sup>, the survival of *MSH2* wild-type cells was reduced by 85%, while in the *MSH2*<sup>+/-</sup> cell lines a reduction in survival of 37% and 33% was seen. The survival of *MSH2*<sup>-/-</sup> mutants at this concentration remained at 97%, but was reduced to 84% in *MSH2*<sup>-/+</sup> cells. At a concentration of 0.75 µg.ml<sup>-1</sup>, the survival of *MSH2* wild-type cells was severely affected with only 2% of wells showing any growth, and similarly reduced survival was seen in the *MSH2*<sup>+/-</sup> cells at 19% and 12%, and the *MSH2*<sup>-/+</sup> cells at 19%.

The increased resistance of the *MSH2*<sup>-/-</sup> mutants became obvious at this concentration as 91% and 87% of wells showed growth, equating to a 4.5-fold increase in survival over cells expressing *MSH2*. At the final concentration, 1.0 µg.ml<sup>-1</sup>, no growth was observed for the *MSH2* wild-type cells, or one of the *MSH2* heterozygous cell lines. The second *MSH2*<sup>+/-</sup> cell line showed only 3% survival and survival in the *MSH2*<sup>-/-+</sup> cell line was reduced to 10%. At this concentration, the survival of the *MSH2*<sup>-/-</sup> mutants was also affected but 52% and 43% of wells still showed growth, which represents a minimum increase in survival of 33% over cells expressing *MSH2*. These results are consistent with the loss of MMR activity in the *MSH2*<sup>-/-</sup> cells as a result of the deletion of *MSH2* from these cell lines.

#### 4.2.7 Microsatellite instability in *MSH2* knockout cell lines

Microsatellites are an extremely abundant class of repetitive elements which are composed of tandem arrays of short repetitive sequences, generally 1 to 6 bp in length (Beckmann and Weber, 1992). Many of these sequences show extensive variation in the number of repeat units per tract between alleles, and these polymorphisms have proved invaluable to studies of linkage analysis and population genetics in many organisms, including trypanosomes (Oliveira *et al.*, 1998; Hope *et al.*, 1999; Tait *et al.*, 2002). Owing to their repetitive nature, microsatellites are especially prone to slippage of DNA polymerase during replication which leads to the formation of insertion-deletion loops (Figure 4.9; reviewed in Sia *et al.*, 1997a; Strauss, 1999; Ellegren, 2000b). If uncorrected, these loops result in base additions or deletions which are collectively termed frameshift mutations. The MMR system is responsible for correcting these misalignments, and inactivation of this repair pathway induces high levels of mutation at such repetitive loci (termed microsatellite instability or MSI; Buermeier *et al.*, 1999). In bacteria, inactivation of either the *mutS* or *mutL* genes leads to greater than 13-fold increases in the rate of tract instability (Levinson and Gutman, 1987). Destabilisation of microsatellites has also been reported for MMR mutants in *S. cerevisiae* (Strand *et al.*, 1993; Strand *et al.*, 1995; Johnson *et al.*, 1996b; Sia *et al.*, 1997b; Sia *et al.*, 2001). Mutations in the yeast *MSH2* and *MSH3* genes decrease the stability of repetitive tracts composed of repeat units 1 to 13 bp in length (Strand *et al.*, 1993; Sia *et al.*, 1997b), while inactivation of *MSH6* decreases the stability of microsatellites with repeat units of 1 or 2 bp (Sia *et al.*, 1997b). Given the requirement for the MLH1-PMS1 heterodimer in both MSH2-MSH3 and MSH2-MSH6 mediated mismatch repair it is expected that these proteins are also required to correct replication slippage events in microsatellites of all sizes (Kolodner and Marsischky, 1999; Harfe and Jinks-Robertson, 2000). In support of this argument is the finding that

inactivation of *MLH1* or *PMS1* led to similar increases in tract instability as *MSH2* mutation for a poly(GT) microsatellite (Strand *et al.*, 1993). Also, mutations in *MLH3* have been reported to increase the rate of 1 bp insertions and deletions in naturally occurring homopolymeric sequences (Flores-Rozas and Kolodner, 1998). MSI is also an important consequence of MMR deficiency in mammalian cells (de Wind *et al.*, 1995), and is known to occur in tumours from patients with HNPCC (Aaltonen *et al.*, 1993). Currently, mutations in five human MMR genes have been described in patients suffering from HNPCC; the majority of mutations occur within *MSH2* and *MLH1*, while mutations in *MSH6*, *PMS1* and *PMS2* are relatively rare (Liu *et al.*, 1996; Viel *et al.*, 1998). Furthermore, the stability of a repetitive tract is also dependant upon the number of repeat units within the tract in that the frequency of mutation increases with the number of repeat units (Wierdl *et al.*, 1997). Tract stability is also affected by whether the tract contains perfect or imperfect repeats (Petes *et al.*, 1997). Perfect repeats are uninterrupted by base insertions or deletions, while imperfect repeats are punctuated by bases which are different to those comprising the repeat units. The rate of replication slippage in imperfect repeats is likely to be reduced because misalignments will be disrupted by the formation of mismatches between the imperfect repeat units.

To investigate whether *MSH2* is involved the repair of slipped replication intermediates in *T. brucei*, the stability of five naturally occurring microsatellite loci was compared in *MSH2* wild-type cells, *MSH2*<sup>+/+</sup> cells, *MSH2*<sup>-/-</sup> mutants and *MSH2*<sup>-/+</sup> cells. To this end, each cell line was cloned (see section 2.1.3), ten clones from each were grown for approximately 25 generations (data not shown), and genomic DNA isolated from the cultures (see section 2.1.6). The five microsatellite loci used in this analysis were chosen to reflect different types of repeat unit and different genomic locations (see below).

The first microsatellite chosen for this analysis has been mapped to Chromosome IV, and is composed of GT-dinucleotide repeats (JS-2; Figure 4.10; Hope *et al.*, 1999). Since polymorphisms at this microsatellite locus have been characterised by electrophoresis on high percentage agarose gels previously, it was decided that a similar approach might yield data about the *MSH2* mutants. The JS-2 locus was amplified by PCR using the primers JS-2A and JS-2B from all the clonal genomic DNA samples, and the resulting PCR products separated by electrophoresis on 3% agarose gels (Figure 4.11). It is evident that the JS-2 locus is stable in all ten *MSH2* wild-type clones, yielding a PCR product approximately 320 to 340 bp in length. This is significantly longer than the larger JS-2 allele in TREU 927/4, which is only 177 bp in length (containing 47 GT-dinucleotide repeat units: Sasse, 1998). Since JS-2 alleles vary greatly in length between different

trypanosome stocks (Hope *et al.*, 1999) this is not surprising, and probably represents an increase in the number of repeat units contained within the microsatellite. This is also suggested by the fact that it proved impossible to clone the JS-2 PCR product amplified from 3174 genomic DNA, probably owing to the inability of bacterial DNA polymerase to replicate such long repetitive sequences. Analysis of the clones derived from the two *MSH2*<sup>+/+</sup> cell lines indicates that the JS-2 locus is stable in this background also. However, in the *MSH2*<sup>-/-</sup> background a number of clones show changes in the size of the PCR products amplified, suggesting that the JS-2 locus has mutated in these cells (*MSH2*<sup>-/-</sup> 1.1 clones C, G and H; *MSH2*<sup>-/-</sup> 1.1 clone A). Finally, the products generated from the *MSH2*<sup>-/+</sup> clones reveal that the JS-2 locus is different from that amplified from the *MSH2* wild-type strain, in that two PCR products, presumably representing the two JS-2 alleles, are apparent in the re-expressor but only one size is distinguished in the *MSH2* wild-type cells. Since this difference is found in every *MSH2*<sup>-/+</sup> clone it probably arose prior to transformation with the  $\Delta$ *MSH2*::*MSH2*-*BLE* construct when the cells were *MSH2*<sup>-/-</sup> mutants. For this reason the JS-2 locus was considered to be stable in the *MSH2*<sup>-/+</sup> cell line. These results are summarised in Table 4.2 and shown as a graph in Figure 4.12. These data indicate that *MSH2* homozygous mutants display a mutator phenotype in *T. brucei*. Indeed, small changes may have occurred in clones derived from every cell line but it is probable that agarose gel electrophoresis cannot distinguish insertions or deletions under a threshold size. It is clear, however, that loss of *MSH2* leads to greater changes than are observed in cells expressing *MSH2*. To determine the size of microsatellite alleles more accurately, and determine what types of mutation event occur to alter the lengths of repetitive tracts in *MSH2* mutants in *T. brucei*, a more sensitive method of microsatellite analysis was required.

A more precise method for determining the sizes of microsatellites is fluorescence-based PCR (Figure 4.13; see section 2.16). In this system, often called GeneScan analysis after the software required to analyse the data (GeneScan software; Applied Biosystems), a microsatellite locus is amplified using one unmodified primer and a primer labelled with a fluorescent dye. The resulting PCR products are separated by acrylamide-urea gel electrophoresis alongside GeneScan internal size standards which are included in the same well. The size standards are also fluorescently labelled (with a different dye), and the automated DNA sequencer interprets the fluorescent bands as peaks. Since the sizes of the size standards are known, the length of the microsatellite PCR products can be determined automatically by comparison. This system is able to distinguish PCR products that differ in length by only 1 bp, and while a cluster of fragments is observed for each allele owing

to modifications caused by *Taq* polymerase during amplification, called 'stutter' or 'shadow' bands, the highest peak is interpreted as the primary product, and can be compared with equivalent peaks from other samples (Oda *et al.*, 1997; Paulson *et al.*, 1999). In this study, PCR product sizes are given in base-pairs to one decimal point, owing to slight differences in migration. In most cases, PCR products amplified from multiple clones derived from the 3174 wild-type cell line, and repetitions of the same clone, only vary in size by 0.5 bp by comparison with the internal GeneScan size markers (see below). However, large alleles (over 200 bp in length) showed greater variation of up to 1.5 bp between samples which probably share the same sequence.

Two microsatellite loci identified by searching the *T. brucei* sequencing databases are located on Chromosome I, and are termed ChrI-7 and ChrI-15 (A. MacLeod, A. Tweedie and A. Tait, personal communication). These were chosen for this analysis, because while they both consist of GT-dinucleotide repeats (Figure 4.14), they are located in distinct regions of the chromosome, approximately 290 kb apart (see The Wellcome Trust Sanger Institute *T. brucei* genome sequencing database; [http://www.sanger.ac.uk/Projects/T\\_brucei/](http://www.sanger.ac.uk/Projects/T_brucei/)). Two loci identified on Chromosome II were also studied (Figure 4.15). The first contains a run of TA-dinucleotide repeats as well as a T-monomucleotide and an A-monomucleotide run (PLC; A. MacLeod, S. McLellan, C. M. R. Turner and A. Tait, personal communication). The second, ChrII-6, was again identified by searching the *T. brucei* databases and consisted of a CT-dinucleotide repeat (A. MacLeod, A. Tweedie and A. Tait, personal communication). These loci are separated by approximately 420 kb of sequence (see The Wellcome Trust Sanger Institute *T. brucei* genome sequencing database). Where possible, the PCR products generated from these loci were cloned and sequenced from wild-type cells in order to determine the composition of the microsatellite alleles in the 3174 trypanosome strain (see section 2.16).

The microsatellite locus ChrI-7 was amplified using the primers ChrI-7A and ChrI-7B (see section 2.16), and the resulting PCR products analysed using the GeneScan system. The results of this analysis are shown in Table 4.3, and the data summarised in Table 4.4. A graph of the data shown in Table 4.4 is depicted in Figure 4.16. At the ChrI-7 locus the two alleles differ considerably in size (Figure 4.13 and 4.14). In clones which appear not to have mutated at this locus, termed stable clones, GeneScan analysis gave sizes of between 102.2 and 102.6 bp for the smaller allele, and 218.2 to 218.9 bp for the larger allele. In all *MSH2* wild-type clones, both sets of *MSH2*<sup>+/+</sup> clones and the *MSH2*<sup>-/-</sup> clones both alleles appear to be stable. However, comparison of the sizes of the PCR products amplified from this locus in the *MSH2*<sup>-/-</sup> clones reveals a number of clones which appear to

have mutated at one or both alleles (Figure 4.17). Out of the 19 clones analysed, 11 appear to contain mutated versions of the larger allele, while only one clone has a mutated smaller allele. This probably reflects the increased likelihood of replication slippage across longer tracts of repeats (Wierdl *et al.*, 1997). Most of the mutation events appear to be losses of a single repeat unit (-2 bp), including the only mutation in the smaller allele, although several larger deletions can be seen within the larger allele, probably of 2, 3 or 6 repeat units (-4, -6 and -12 bp). Only 3 of the events increase the length of the repeat tracts; two of these cases appear to have inserted a single repeat (+2 bp), and the other probably inserted 2 repeats (+4 bp). Without sequencing the mutated alleles, it is impossible to say whether these events truly represent changes in the number of repeats or whether other mutational events have led to these increases or decreases in microsatellite length. However, given that the mechanism of microsatellite instability, and its basis in replication slippage over repeat units, has been well characterised it seems highly probable that the mutations documented above arose in this way. These data provide further evidence that loss of *MSH2* in *T. brucei* leads to the phenotype of microsatellite instability.

The second microsatellite locus analysed on Chromosome I was ChrI-15. This was amplified by PCR using the primers ChrI-15A and ChrI-15B, and analysed using the GeneScan system. The data generated in this analysis is shown in Table 4.5, and summarised in Table 4.6. A graph showing the summary of this data is shown in Figure 4.18. Genescan analysis of this locus revealed that the two alleles present in 3174 wild-type trypanosomes differed in size by only 8 bp (Figure 4.13), and in stable clones (defined as above) the size of the smaller allele ranged between 166.8 and 167.3 bp, while the size range of the larger allele was 174.9 to 175.4 bp. For all clones analysed from the *MSH2* wild-type, *MSH2*<sup>+/-</sup> and *MSH2*<sup>+/+</sup> backgrounds, both alleles appear to be stable as no changes in length could be discerned. Once again, however, a number of *MSH2*<sup>-/-</sup> clones appeared to contain mutated alleles (Figure 4.19). Out of 20 *MSH2*<sup>-/-</sup> clones analysed, 50% were mutated at one or both alleles. In this sample set, the smaller allele was mutated in 7 clones, while the larger allele was mutated in 4 cases. For both alleles, the most common change was either addition or deletion of 2 bp, equivalent to one repeat unit. These events appeared to occur with equal frequency in the smaller allele, as three 2 bp additions and three 2 bp deletions were observed. Mutation of the larger allele appeared to favour a decrease in length, as three 2 bp deletions were apparent while only one 2 bp addition was observed. For both alleles, only one mutation event leading to a change in length greater than 2 bp was seen. In this case the smaller allele decreased in length by 4 bp, probably due to a loss of 2 repeat units. Again these data indicate that loss of *MSH2* in *T. brucei* leads to the phenotype of MSI.

The first microsatellite locus analysed on Chromosome II was PLC, using amplification with the primers PLCG and PLCH3. Table 4.7 gives the results of this analysis, while Table 4.8 shows a summary of the data analysed. A graphical representation of these results is shown in Figure 4.20. GeneScan analysis shows that for all *MSH2* wild-type, *MSH2*<sup>+/-</sup>, and *MSH2*<sup>-/-</sup> clones the PCR products generated were 158 bp in length (a size range of 157.8 to 158.2 bp after electrophoretic separation), indicating that the length of this microsatellite was stable in these backgrounds. Once again, alleles of a different length were observed in the *MSH2* homozygous mutant clones (Figure 4.21). Since only one PCR product size was observed, PCR using the PLCG and PLCH3 primers appeared to have amplified only one allele, possibly due to variation in the primer sequences between the two alleles. During the initial work characterising this microsatellite locus numerous primers were tested before a primer pair which would amplify PLC alleles from several trypanosomic stocks was identified (A. MacLeod, personal communication). The alternative explanation, that both alleles mutated to an identical size is highly unlikely. This analysis revealed that out of 20 clones 3 contained mutated versions of the PLC allele, which was higher than the mutation frequency observed for the smaller allele of the ChrI-7 microsatellite locus, even though the number of dinucleotide repeats in PLC is fewer than in the ChrI-7 smaller allele. This perhaps suggests that the mutation rate of TA-dinucleotide repeats is higher than for CA-repeats in this organism, although extensive analyses at a number of microsatellite loci would be required to confirm this. All of the mutation events in PLC resulted in deletions of sequence, 2 led to 2 bp deletions and the other to a 1 bp deletion. Sequencing of the wild-type PLC allele revealed that it contained a T<sub>9</sub>-mononucleotide run and an A<sub>11</sub>-mononucleotide run in addition to the TA-dinucleotide repeat units. Therefore, the 1 bp deletion probably arose by a mutation within one of the mononucleotide runs as these are more likely to undergo to replication slippage and mutation than the intervening, non-repetitive sequence in the PLC allele. Once again, these results indicated that inactivation of *MSH2* in trypanosomes results in MSI and a mutator phenotype.

The final microsatellite locus studied was ChrII-6. It was amplified using the primers ChrII-6A and ChrII-6B, and the results of this analysis are documented in Table 4.9, and summarised in Table 4.10. Genescan analysis of the PCR products amplified from this locus gave sizes of between 96.2 and 96.7 bp. Every clone analysed from the *MSH2* wild-type, *MSH2*<sup>+/-</sup>, *MSH2*<sup>-/-</sup> and *MSH2*<sup>-/-</sup> backgrounds fell within this size range, indicating that all clones were stable at this locus. This was surprising as this locus was chosen as it contained 32 perfect CT-dinucleotide repeats in the TREU 927/4 strain (Figure 4.15). However, sequencing of the ChrII-6 locus in 3174 trypanosomes revealed that the

microsatellite had degenerated in this strain and contained only 12 imperfect repeats. It is predicted that this microsatellite is unlikely to mutate owing to its small size and imperfect sequence. This locus therefore provided a control for the other microsatellite analyses performed above. Since no mutated alleles were observed it is impossible to say whether the ChrII-6A and ChrII-6B primers only amplified one allele from this locus as occurred for PLC, or if the ChrII-6 locus is homozygous in 3174 trypanosomes.

The mutational spectrum of all the microsatellite loci studied by GeneScan analysis using clones derived from *MSH2* wild-type cells, *MSH2* heterozygous mutants, *MSH2* homozygous mutants and *MSH2* re-expressor cells is shown in Table 4.11. For each locus the number of alleles which increased or decreased in length by a given number of base-pairs was assessed and the total number of mutations of each type was calculated for all the loci. The most common events involved the loss or gain of 2 bp, equating to a single repeat unit in the dinucleotide microsatellites. These events accounted for 77% of the observed mutations, and approximately two thirds of these resulted in a decrease in microsatellite length. Larger insertions or deletions were less common, accounting for only 23% of mutations, and again deletions were more frequently observed than increases in length. Indeed, only one increase greater than 2 bp was seen, and this probably reflected an increase of only two repeat units. The larger deletions were found only in the larger allele of ChrI-7, which probably contained the greatest number of repeat units of any of the microsatellites studied. The sequence of this allele is unknown as it could not be cloned and sequenced, probably owing to the inability of our *E. coli* strain to maintain long microsatellite clones. Overall, this study revealed that 69% of the observed mutations resulted in a decrease in microsatellite length, indicating that loss of repeat units was more common than gain. The observation that inactivation of *MSH2* in *T. brucei* results in a mutator phenotype and therefore MSI suggests that the protein is required for the repair of slipped replication intermediates, as is the case in both yeast and mammalian cells. This confirms that *T. brucei* *MSH2* is required for nuclear MMR.

#### 4.2.8 Antigenic variation in *MSH2* knockout trypanosomes

In order to determine the effect of the *MSH2*<sup>-/-</sup> mutation on antigenic variation in *T. brucei*, it was necessary to generate the mutant cell lines in the transgenic 3174 background (McCulloch *et al.*, 1997; McCulloch and Barry, 1999). The active expression site within this trypanosome strain has been modified to allow its use in experiments to determine the frequency of VSG switching, and to assess the relative contribution of the *in situ* transcriptional switching and gene conversion switching mechanisms. The exact

conditions used during this assay are described in section 2.15, but a general description is given below. The modified expression site in the 3174 cell line contains antibiotic resistance markers for hygromycin and G418 (Figure 4.22). Growth of cells on both hygromycin and G418 allows the selection of cells expressing only the modified expression site and therefore the VSG coat encoded by the *VSG221* gene. These cells were used to immunise mice against this VSG coat. Later, a measured number of trypanosomes which had been allowed to grow for a fixed number of generations without antibiotic selection, and which were therefore able to undergo *VSG* switching, were injected into these immunised mice. 24 h after injection, surviving trypanosomes, those which had switched their VSG coat and so were not killed by immune lysis, were recovered from the mice and plated over 96-well plates in HMI-9. After 10 days the number of wells which showed growth were counted, and the frequency of VSG switching was estimated. Furthermore, the drug sensitivities of the cells from the wells showing growth was assessed, and genomic DNA was isolated from the switched trypanosomes to allow PCR amplification of marker genes from the modified expression site. Using this system, three types of VSG switching event can be distinguished which result in the expression of a novel *VSG* gene. Cells which are sensitive to both hygromycin (Hyg<sup>S</sup>) and G418 (G418<sup>S</sup>) may have undergone either an *in situ* transcriptional switch or a long-range expression site gene conversion. PCR amplification using primers directed against the hygromycin phosphotransferase (Hyg), neomycin phosphotransferase (Neo) and *VSG221* (221) ORFs differentiates between these possibilities. During an *in situ* switch, the active expression site is transcriptionally silenced, while another expression site is activated, leaving all markers intact and allowing amplification of the Hyg, Neo and 221 PCR products (Hyg+, Neo+, 221+). Note that in principle these events could also be reciprocal exchanges upstream of the hygromycin marker, but in practise transcriptional switching is most common in these MITat 1.2a S427-derived cells (Rudenko *et al.*, 1996; McCulloch *et al.*, 1997; McCulloch and Barry, 1999). In an expression site gene conversion, however, all the markers are replaced by sequence from another expression site so they cannot be amplified by PCR (Hyg-, Neo-, 221-). Cells which have undergone the third type of switching event, *VSG* gene conversion, are resistant to hygromycin (Hyg<sup>R</sup>), but are G418<sup>S</sup>. *VSG* gene conversion involves the replacement of the sequence from the 70-bp repeats to the 3' end of *VSG221* with sequence from another expression site, removing the Neo and 221 markers but leaving the Hyg marker intact (Hyg+, Neo-, 221-).

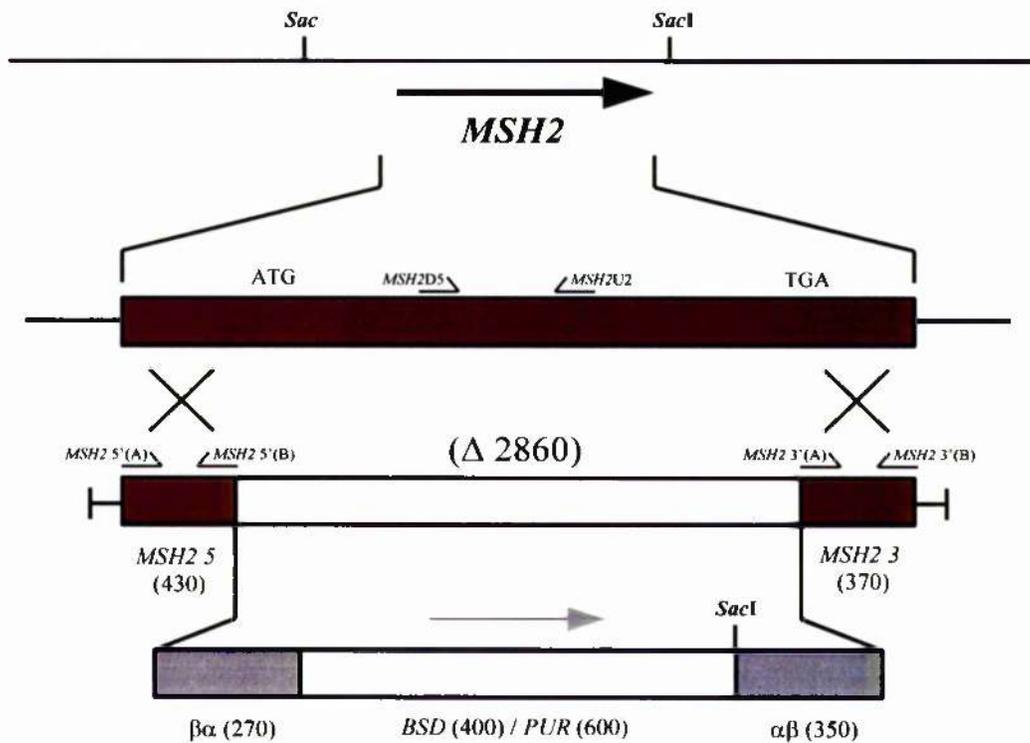
The approach described above was used to isolate switched variants arising after immune selection for *MSH2* wild-type cells, *MSH2*<sup>+/-</sup> mutants, *MSH2*<sup>-/-</sup> mutants and *MSH2*<sup>+/+</sup> trypanosomes. For each cell line a number of repetitions were performed from entirely

independently grown populations, allowing an average VSG switching frequency to be calculated. The data retrieved during these experiments are shown in Table 4.12, and are represented graphically in Figure 4.23. For each cell line, the estimated VSG switching frequency varied greatly between repetitions. This is probably due to a founder effect. During growth of the trypanosome population in the absence of antibiotic prior to injection into an immunised mouse, switched variants can arise at any cell division. Thus, if a switching event occurred early in the growth of the population, a greater number of cells expressing that novel VSG will be present than if the event had occurred at a later division. On average, *MSH2* wild-type cells switched their VSG coat at a frequency of  $0.55 \times 10^{-6}$  switches/cell/generation. Similar average switching frequencies were observed for each cell line tested, including three independent *MSH2* heterozygous mutants, three independent *MSH2* homozygous mutants and a cell line re-expressing *MSH2*. These results indicate that inactivation of *MSH2* has no substantial effect on the frequency of VSG switching in 3174 trypanosomes.

Where possible, ten switched variants were analysed from each repetition for *MSH2* wild-type cells, the three *MSH2*<sup>+/+</sup> cell lines, the three *MSH2*<sup>-/-</sup> mutants and the *MSH2* re-expressor cell line to determine the switching mechanisms used by each cell line. The results are shown in Table 4.13, and as a graph in Figure 4.24. Inactivation of *MSH2* appeared to make no substantial alteration to the relative use of the different switching mechanisms, although this is complicated by variation in their frequency from experiment to experiments, again because of founder effects. For most cell lines, except *MSH2*<sup>+/+</sup> 3.1, expression site gene conversion was the most commonly used switching mechanism, being used in 36% to 86% of the switched variants analysed. Switched variants that had arisen by *in situ* switching were also observed in every cell line, and this was the most common event seen in the *MSH2*<sup>+/+</sup> 3.1 cell line. *In situ* transcriptional switching accounted for between 5% and 50% of the switched variants observed. Finally, VSG gene conversion was the least frequently observed switching mechanism, and was not used to generate any switched variants in the *MSH2*<sup>+/+</sup> 2.1, *MSH2*<sup>-/-</sup> 2.1 or *MSH2*<sup>-/-</sup> 3.1 cell lines. However, it accounted for between 3% and 37% of the switched variants in the remaining cell lines, and could be detected in *MSH2* wild-type cells, and *MSH2* heterozygous and homozygous mutants. Taken together, these results indicate that loss of *MSH2* has little or no effect on the frequency of antigenic variation in trypanosomes, or upon the types of VSG switching mechanisms used.

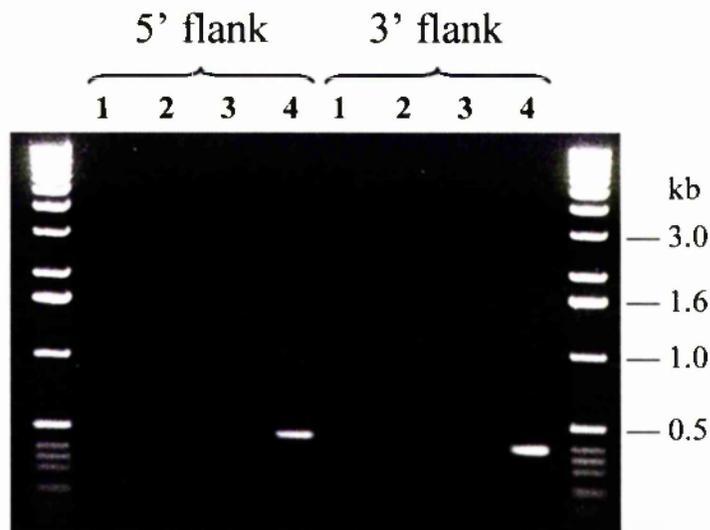
**Figure 4.1** A strategy for the generation of *T. brucei* *MSH2* knockout mutants

The *MSH2* locus is indicated as a red box, and the constructs used to disrupt its expression are shown below. *MSH2* 5' and *MSH2* 3' are sequences flanking the *MSH2* ORF (ATG to TGA) which allow for integration of the knockout constructs, as indicated by crosses. Two constructs were used, one containing the 400 bp blasticidin S deaminase (*BSD*) gene and the other the 600 bp puromycin acetyltransferase gene (*PUR*); both are flanked by sequences derived from the tubulin locus ( $\beta\alpha$  and  $\alpha\beta$ ; grey boxes) that allow transcripts of the resistance markers to be trans-spliced and polyadenylated into mature mRNAs. In each construct, the 2860 bp *MSH2* ORF is precisely replaced by the resistance cassettes. Arrows represent the direction of transcription of the genes, numbers in brackets denote lengths in base-pairs, and half-arrows indicate the approximate positions of primers used in the analysis. *SacI* restriction sites are shown as vertical bars.



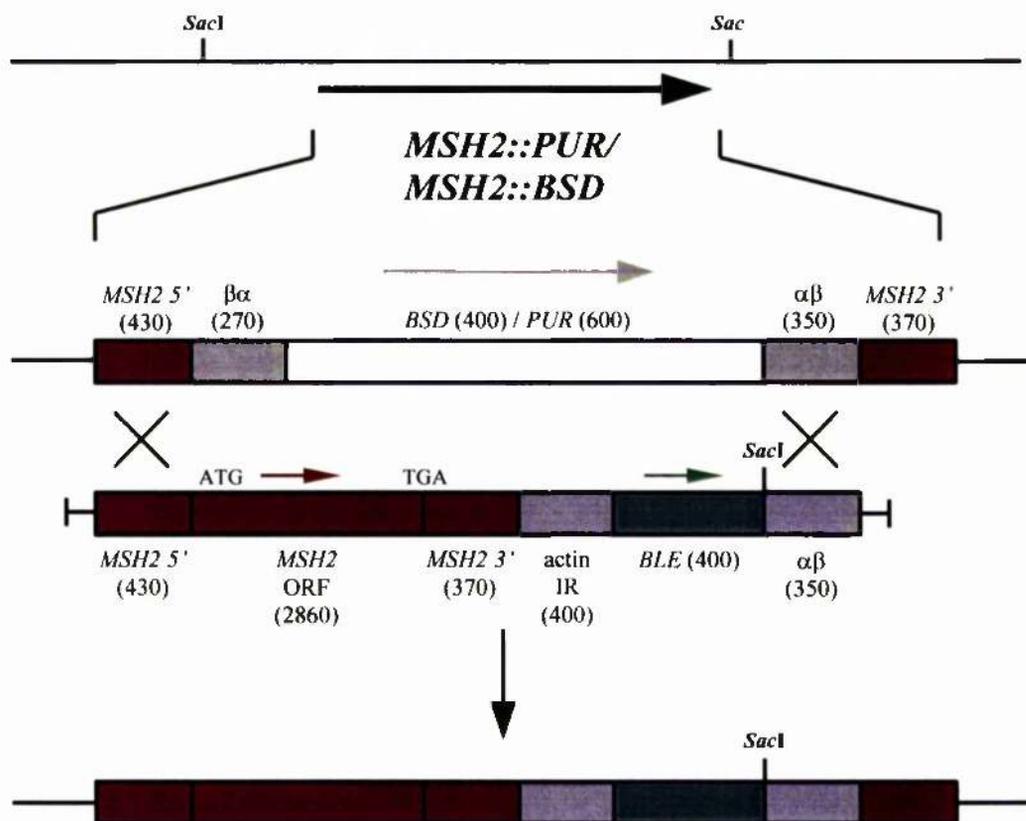
**Figure 4.2 PCR amplification of *MSH2* knockout construct flanks**

The *MSH2* integration flanks were amplified from MITat 1.2 genomic DNA using the primers *MSH2* 5'(A) and *MSH2* 5'(B) (5' flank), and *MSH2* 3'(A) and *MSH2* 3'(B) (3' flank). For each primer pair control reactions were performed with both primers in the absence of *T. brucei* DNA template (1), the (A) primer only (2) and the (B) primer only (3). Lane 4 shows the reaction containing both primers and template for each primer pair.



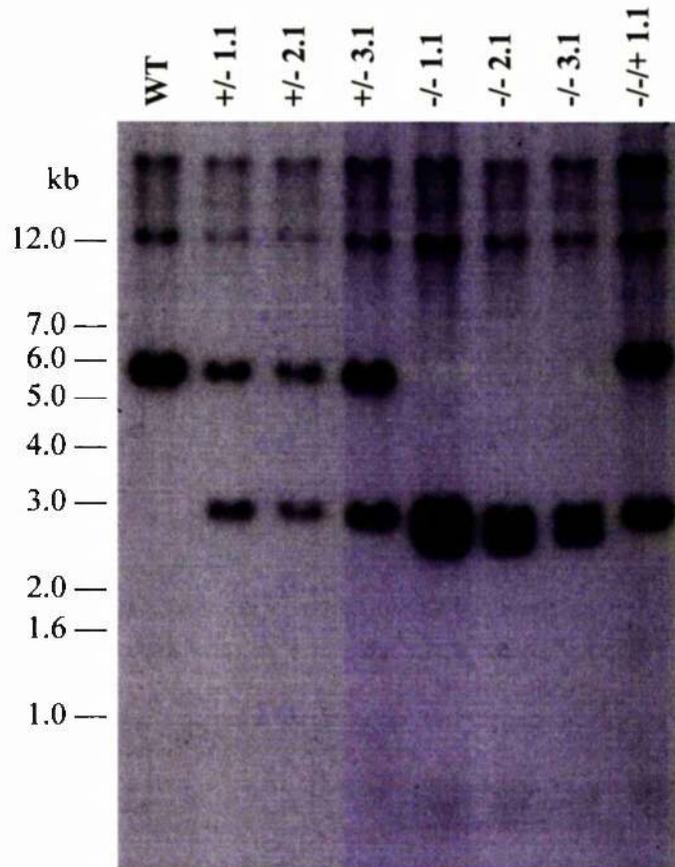
**Figure 4.3** *MSH2* re-expression construct

The *MSH2* locus, following inactivation by replacing the *MSH2* ORF with either *BSD* or *PUR* resistance cassettes, is shown and (below) the construct  $\Delta$ *MSH2*::*MSH2*-*BLE* for re-expression of *MSH2* in the mutants is depicted. The construct integrates an intact *MSH2* ORF into the altered loci: *MSH2* 5' and  $\alpha\beta$  are the sequences that allow integration of the construct to replace either the *MSH2*::*PUR* or *MSH2*::*BSD* alleles (indicated by crosses). Correct integration of the re-expression construct will result in expression of *MSH2* from its original genomic location and resistance to phleomycin. The *MSH2* ORF is flanked by its original processing sequences (*MSH2* 5' and *MSH2* 3') while the bleomycin phosphotransferase ORF (*BLE*) is flanked at the 5' end by sequence derived from the actin locus (actin IR) and at the 3' end by sequence from the tubulin locus ( $\alpha\beta$ ) that allow the transcripts to be processed into mature mRNAs. ATG and TGA indicate the approximate positions of the start and stop codons of *MSH2* respectively. Arrows represent the direction of transcription of the genes and numbers in brackets represent lengths in base-pairs. *Sac*I restriction sites are shown as vertical bars.



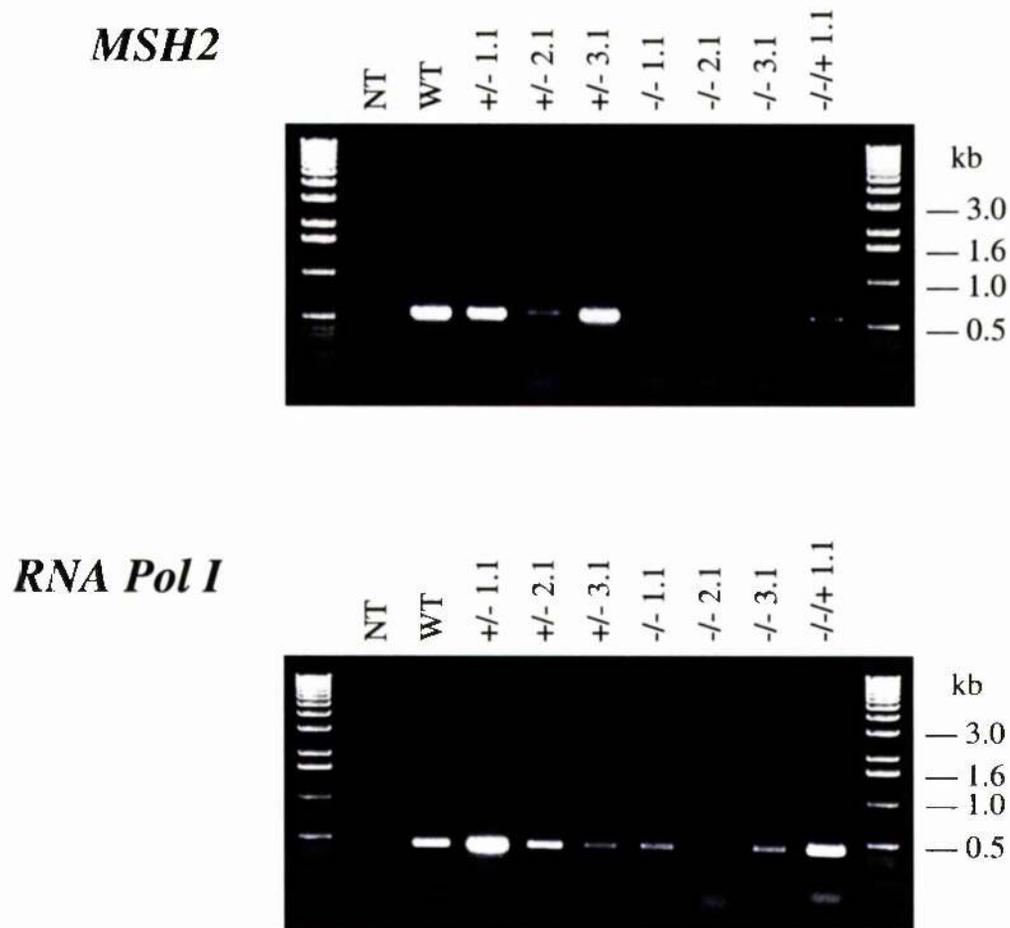
#### Figure 4.4 Southern analysis of the putative 3174 *MSH2* mutant cell lines

Genomic DNA isolated from the putative 3174 *MSH2* mutant cell lines was digested with *SacI*, separated on a 0.6% agarose gel, Southern blotted and probed with the *MSH2* 5' flank used as the 5' integration flank in the *MSH2* knockout construct. The following samples were analysed: (WT) untransformed 3174 cells; (+/-) 3174 cells transformed with  $\Delta$ *MSH2*::*PUR*; (-/-) 3174 cells transformed with  $\Delta$ *MSH2*::*PUR* and  $\Delta$ *MSH2*::*BSD*; (-/-/+) 3174 cells transformed with  $\Delta$ *MSH2*::*PUR*,  $\Delta$ *MSH2*::*BSD* and  $\Delta$ *MSH2*::*MSH2-BLE*.



### Figure 4.5 PCR analysis of the putative 3174 *MSH2* mutant cell lines

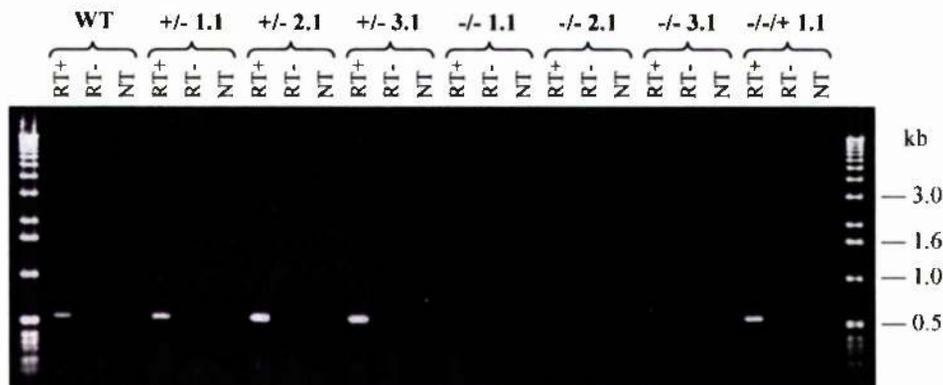
PCR was performed using genomic DNA isolated from the 3174 cell lines transformed with the *MSH2* knockout constructs. Internal primers complementary to the *T. brucei MSH2* gene were used (*MSH2*) and the integrity of the DNA was confirmed using primers directed against the large subunit of *T. brucei* RNA polymerase I (*RNA Pol I*; Rudenko *et al.*, 1996). For both sets of primers the following reactions were run: (NT) no template DNA; (WT) untransformed 3174 cells; (+/-) 3174 cells transformed with  $\Delta$ *MSH2*::*PUR*; (-/-) 3174 cells transformed with both  $\Delta$ *MSH2*::*PUR* and  $\Delta$ *MSH2*::*BSD*; (-/-+) 3174 cells transformed with  $\Delta$ *MSH2*::*PUR*,  $\Delta$ *MSH2*::*BSD* and  $\Delta$ *MSH2*::*MSH2-BLE*.



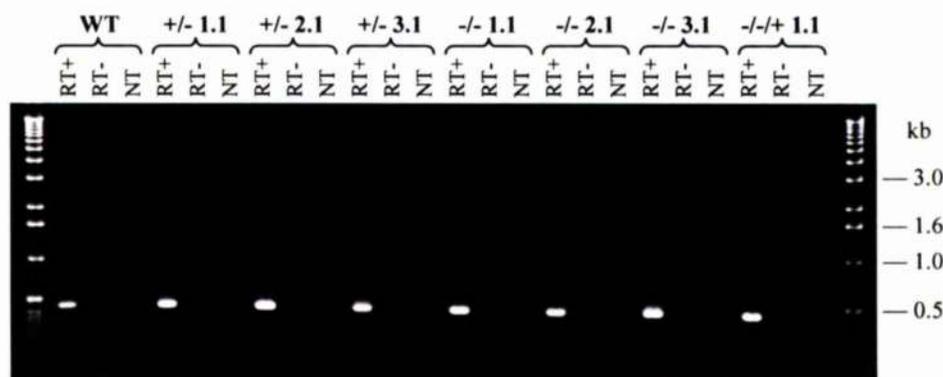
**Figure 4.6** Analysis of the expression of *MSH2* in the putative 3174 *MSH2* mutant cell lines

RT-PCR was performed using total RNA isolated from the 3174 cell lines transformed with the *MSH2* knockout constructs. Internal primers complementary to the *T. brucei MSH2* gene were used (*MSH2*) and the integrity of the DNA was confirmed using primers directed against the large subunit of *T. brucei* RNA polymerase I (*RNA Pol I*; Rudenko *et al.*, 1996). For both sets of primers RT positive (RT+), RT negative (RT-), and no template (NT) reactions were performed using total RNA isolated from the following cell lines: (WT) untransformed 3174 cells; (+/-) 3174 cells transformed with  $\Delta$ *MSH2*::*PUR*; (-/-) 3174 cells transformed with both  $\Delta$ *MSH2*::*PUR* and  $\Delta$ *MSH2*::*BSD*; (-/-+) 3174 cells transformed with  $\Delta$ *MSH2*::*PUR*,  $\Delta$ *MSH2*::*BSD* and  $\Delta$ *MSH2*::*MSH2-BLE*.

### *MSH2*



### *RNA Pol I*





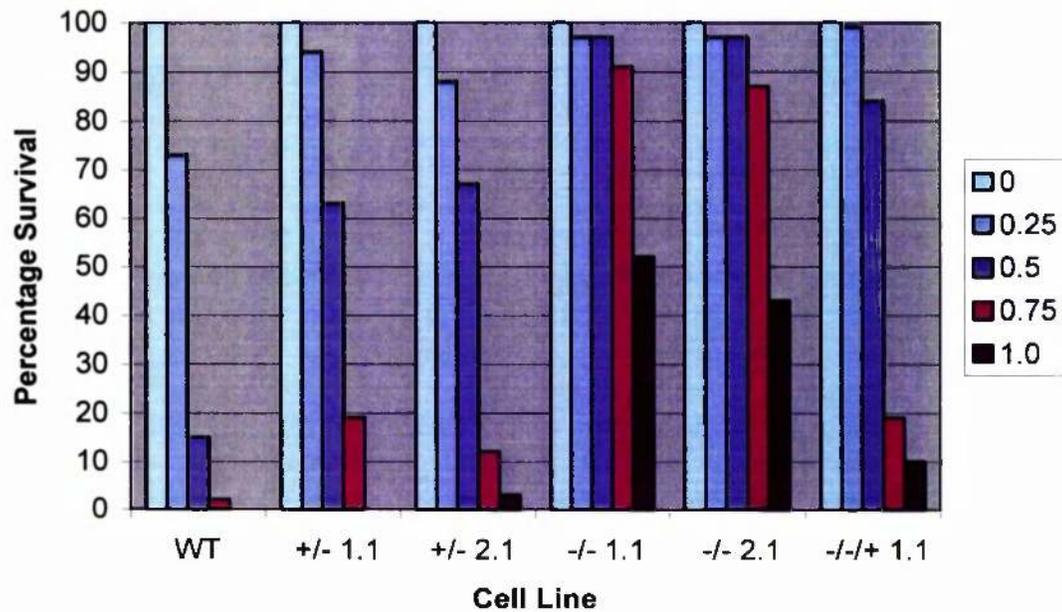
**Table 4.1 Survival of *MSH2* knockout trypanosomes in the presence of MNNG**

Bloodstream trypanosomes that had been unaltered at the *MSH2* locus (WT); that had one *MSH2* allele disrupted (+/-); that had both *MSH2* alleles disrupted (-/-); or that had both alleles disrupted but the *MSH2* gene restored to the genome (-/-+) were plated in HMI-9 containing increasing concentrations of MNNG (as shown) at a density of 2 cells per well in 96-well plates. For each cell line, at each concentration, 3 plates were prepared, and the number of wells containing growth were counted after 14 days. Percentage survival is presented relative to the number of wells growing in the absence of MNNG, which is given as 100%.

Cell Line	MNNG Conc. ( $\mu\text{g.ml}^{-1}$ )	No. of wells growing				Percentage Survival
		Plate 1	Plate 2	Plate 3	Average	
WT	0	72/96	70/96	75/96	72/96	100
	0.25	49/96	53/96	57/96	53/96	73
	0.50	13/96	12/96	7/96	11/96	15
	0.75	0/96	1/96	3/96	1/96	2
	1.00	0/96	0/96	0/96	0/96	0
+/- 1.1	0	89/96	91/96	86/96	89/96	100
	0.25	86/96	78/96	86/96	83/96	94
	0.50	58/96	58/96	51/96	56/96	63
	0.75	11/96	17/96	22/96	17/96	19
	1.00	0/96	1/96	0/96	0/96	0
+/- 2.1	0	84/96	73/96	76/96	78/96	100
	0.25	65/96	74/96	66/96	68/96	88
	0.50	47/96	52/96	56/96	52/96	67
	0.75	7/96	10/96	10/96	9/96	12
	1.00	3/96	1/96	3/96	2/96	3
-/- 1.1	0	76/96	72/96	76/96	75/96	100
	0.25	72/96	76/96	69/96	72/96	97
	0.50	74/96	71/96	72/96	72/96	97
	0.75	62/96	67/96	74/96	68/96	91
	1.00	38/96	38/96	37/96	38/96	52
-/- 2.1	0	73/96	81/96	81/96	78/96	100
	0.25	78/96	80/96	69/96	76/96	97
	0.50	81/96	68/96	80/96	76/96	97
	0.75	71/96	65/96	68/96	68/96	87
	1.00	30/96	37/96	30/96	32/96	43
-/-+ 1.1	0	76/96	72/96	78/96	75/96	100
	0.25	73/96	77/96	73/96	74/96	99
	0.50	72/96	63/96	54/96	63/96	84
	0.75	13/96	17/96	14/96	15/96	19
	1.00	8/96	8/96	7/96	8/96	10

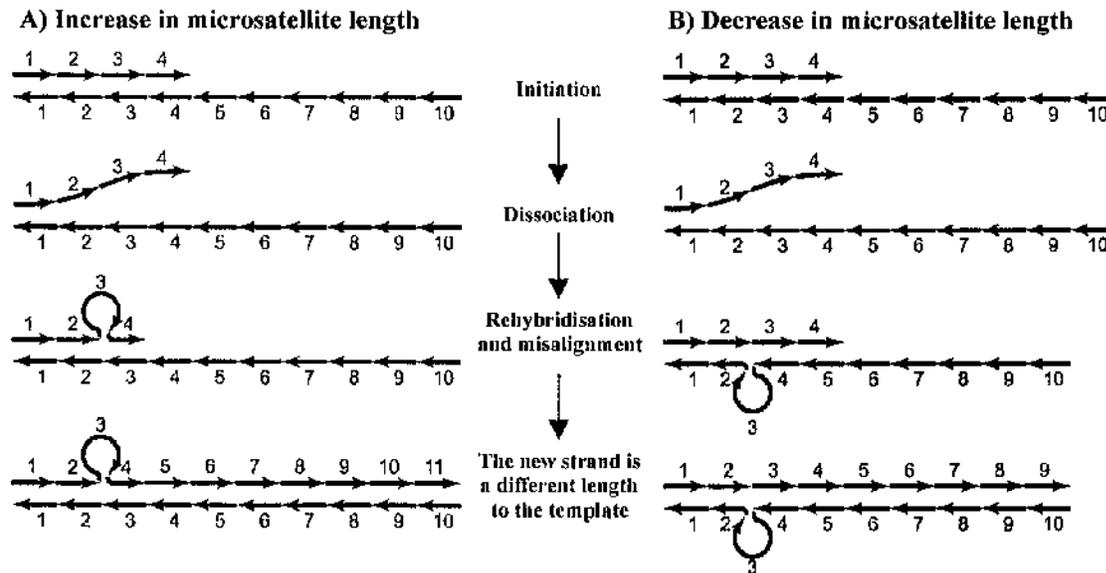
**Figure 4.8 Survival of *MSH2* knockout trypanosomes in the presence of MNNG**

The survival of bloodstream trypanosomes in increasing concentrations of MNNG (shown in  $\mu\text{g}\cdot\text{ml}^{-1}$ ) that had been unaltered at the *MSH2* locus (WT); that had one *MSH2* allele disrupted (+/-); that had both *MSH2* alleles disrupted (-/-); or that had both alleles disrupted but the *MSH2* gene restored to the genome (-/-+) was determined as described in Table 4.1.



**Figure 4.9 Model of microsatellite mutation by replication slippage**

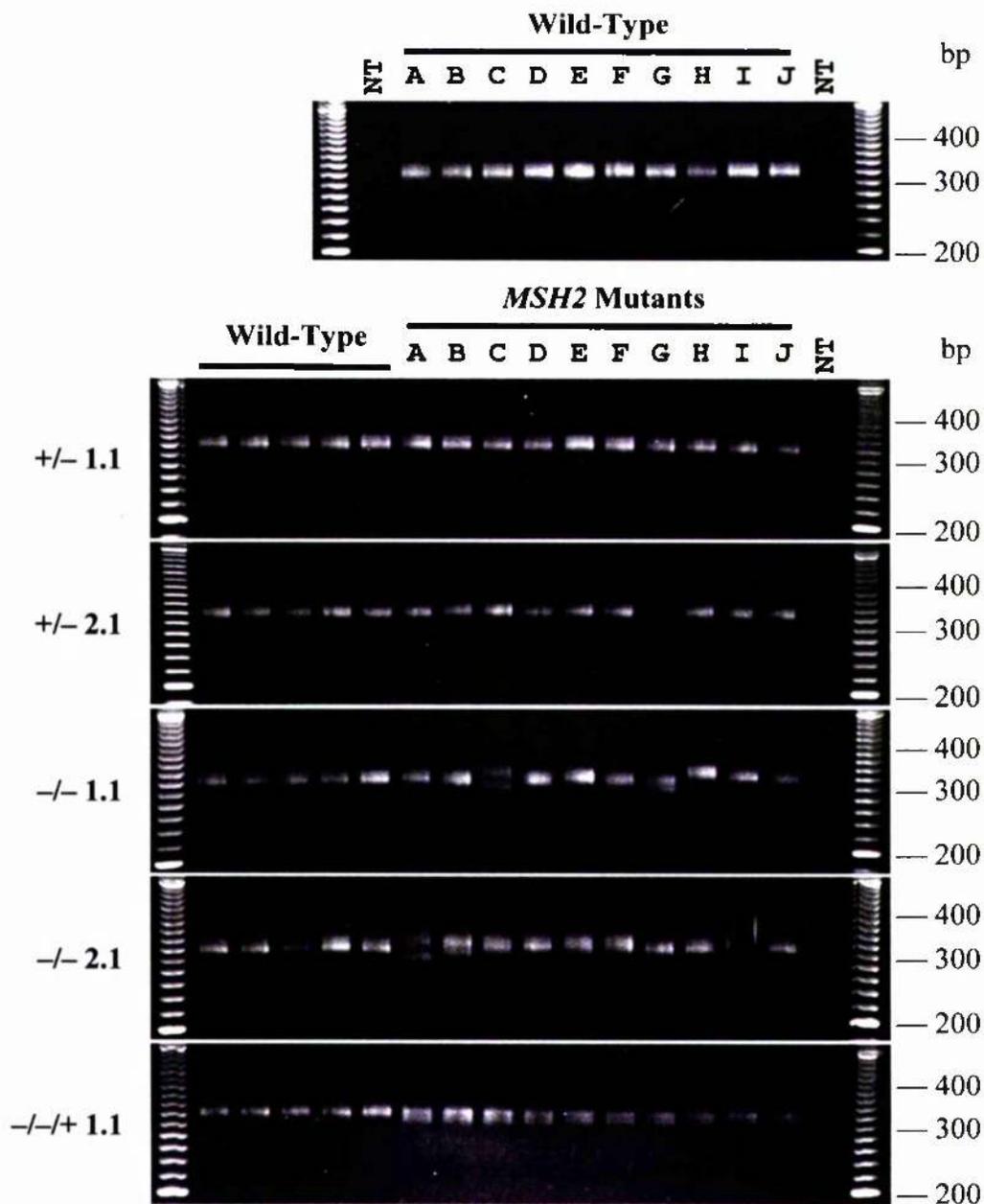
In this scheme arrows denote repeat units and numbers indicate the repeat unit number within each DNA strand. During replication the two strands dissociate, and can be misaligned upon reassociation, resulting in an out-of-register alignment of the two strands. (A) If the most 3' repeat unit of the nascent strand rehybridises with a downstream repeat unit on the template strand, a loop is formed in the new strand. After elongation the new sequence will be correspondingly longer than the template. (B) Alternatively, if the most 3' repeat unit of the new strand reassociates with an upstream repeat unit on the template strand, after synthesis the new strand will be shorter than the template. (Adapted from Ellegren, 2000b)





**Figure 4.11 PCR amplification of microsatellite JS-2 in 3174 *MSH2* mutant cell lines**

PCR amplification of the JS-2 locus, using the primers JS-2A and JS-2B (see section 2.16), was performed using genomic DNA isolated from 10 clones of each 3174 cell line (labelled A-J in each case): *MSH2* wild-type cells (Wild-Type); *MSH2* heterozygous mutants (+/-); *MSH2* homozygous mutants (-/-); and *MSH2* re-expressor cells (-/-/+). Control reactions with both primers but lacking template DNA were performed for each set of clones (NT). The PCR products were separated by electrophoresis on 3% agarose gels. In each case, the PCR products derived from five of the *MSH2* wild-type clones were run along side the products amplified from the *MSH2* mutant clones for ease of comparison.



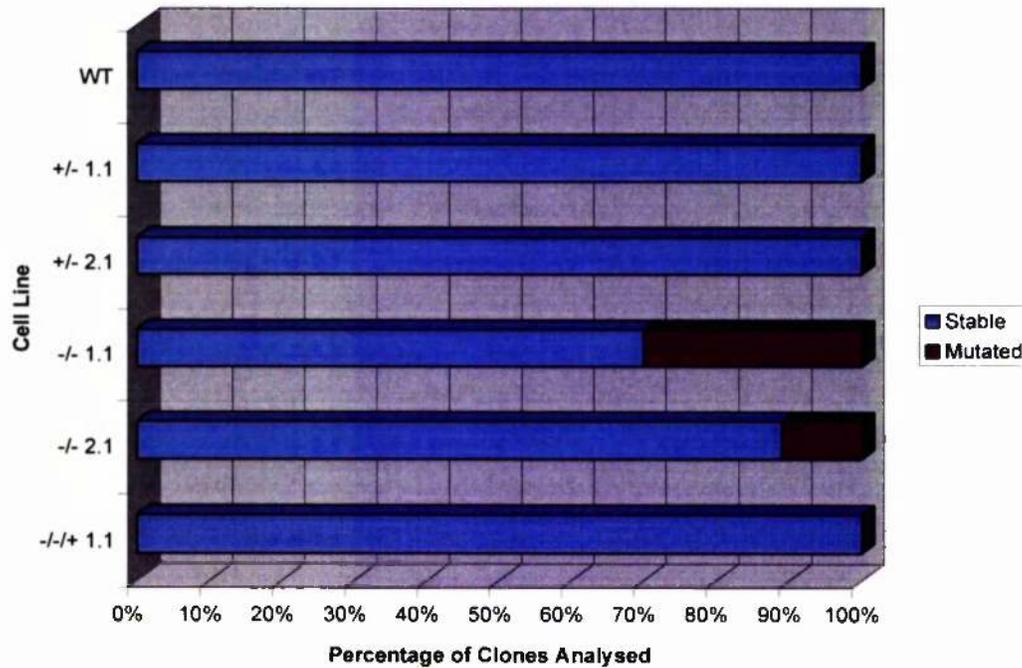
**Table 4.2 Synopsis of results for microsatellite JS-2 in 3174 *MSH2* knockout trypanosomes**

The JS-2 microsatellite locus was analysed as shown in Figure 4.11. Ten clones each, derived from the 3174 *MSH2* wild-type cell line (WT); two heterozygous *MSH2* cell lines (+/-); two *MSH2* homozygous mutants (-/-); and the *MSH2* re-expressor cell line (-/-+) were analysed. The amplified JS-2 alleles were compared between the 10 clones derived from the same *MSH2* mutant cell line, and with alleles amplified from 5 *MSH2* wild-type clones. Alleles which appeared identical between the majority of clones derived from the same cell line and with the products amplified from the *MSH2* wild-type clones were considered stable. If all the alleles amplified from all the clones derived from the same cell line appeared identical then they were also considered stable. Alleles which appeared different from the majority of alleles derived from the same cell line were considered mutant versions of the microsatellite. For each cell line, the number and percentage of clones in which the JS-2 locus was shown to be stable or mutated is given.

Cell line	No. of clones analysed	Stable clones	Percentage	Mutated clones	Percentage
WT	10	10	100	0	0
+/- 1.1	10	10	100	0	0
+/- 2.1	9	9	100	0	0
-/- 1.1	10	7	70	3	30
-/- 2.1	9	8	89	1	11
-/-+ 1.1	10	10	100	0	0

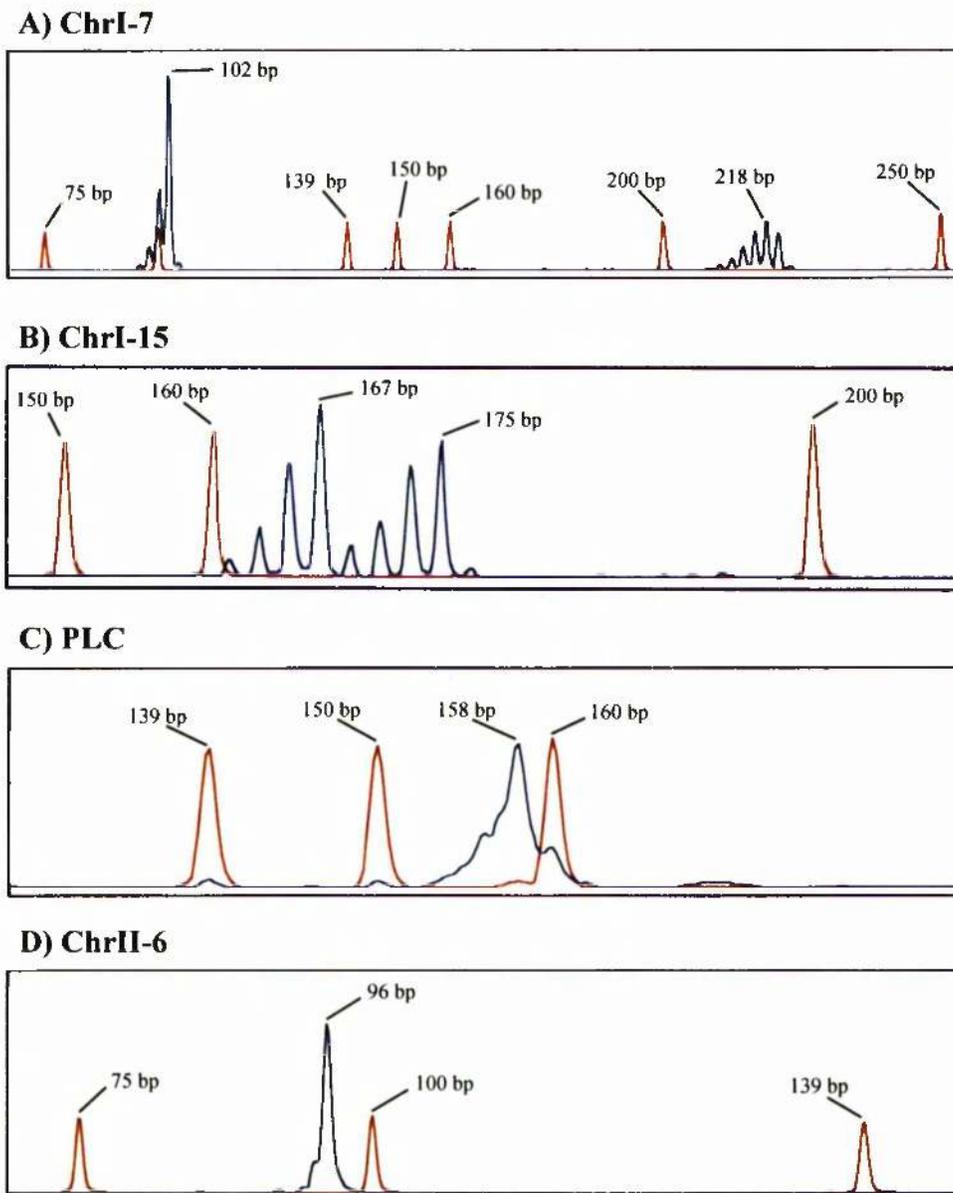
**Figure 4.12** Analysis of microsatellite JS-2 in *MSH2* knockout trypanosomes

The percentage number of clones which were stable (shown in blue) and mutated (shown in red) at the JS-2 microsatellite locus was determined as described in Table 4.2. (WT) 3174 *MSH2* wild-type trypanosomes; (+/-) 3174 *MSH2* heterozygous mutants; (-/-) 3174 *MSH2* homozygous mutants; (-/-+) 3174 *MSH2* re-expressor cells.



**Figure 4.13 Genescan analysis of fluorescently-labelled PCR products derived from four independent microsatellite loci from 3174 transgenic trypanosomes**

PCR amplifications were performed with a 6-FAM-labelled primer, and the PCR products for each microsatellite locus were run in an acrylamide-urea gel (see section 2.16). Bands produced fluorescent peaks (alleles; shown in blue), and their molecular sizes were automatically determined by comparison to the TAMRA-labelled GeneScan-500 internal size standards loaded in each well (shown in red; the numbers refer to the sizes in base-pairs). Where a cluster of peaks are identified for each allele the size is determined by considering the fluorescent peak with the maximum height, as the other peaks arose as a consequence of modifications introduced by amplification with *Taq* polymerase (Oda *et al.*, 1997; Paulson *et al.*, 1999). The products shown in each panel below were amplified from genomic DNA isolated from 3174 trypanosomes. Panel A shows the two alleles amplified from the ChrI-7 microsatellite locus using the ChrI-7A and ChrI-7B primers. Panel B shows the two alleles amplified from the ChrI-15 microsatellite locus using the primers ChrI-15A and ChrI-15B. Panel C shows the single allele amplified from the PLC microsatellite locus using the primers PLCG and PLCH3. Panel D shows the allele(s) generated from the ChrII-6 locus using the ChrII-6A and ChrII-6B primers.







**Table 4.3 Genescan analysis of microsatellite ChrI-7 in 3174 *MSH2* mutant cell lines**

PCR amplification of the ChrI-7 locus was performed using the primers ChrI-7A (labelled with 6-FAM) and ChrI-7B (see section 2.16) on genomic DNA isolated from 10 clones of each 3174 cell line (designated A-J in each case): (WT) *MSH2* wild-type trypanosomes; (+/-) *MSH2* heterozygous mutants; (-/-) *MSH2* homozygous mutants; (-/+ ) *MSH2* re-expressor cells. The PCR products were separated by acrylamide-urea gel electrophoresis and the fluorescent bands detected by an automated sequencer. The sizes (shown in bp) of the PCR products were determined using the GeneScan software as described in Figure 4.13. The size of stable alleles, taking migration error into account to give a size range, was determined by comparing the size of alleles in every clone analysed, and assessing which sizes occurred most frequently. The most frequently observed size was considered the stable, or wild-type, allele length. Where a PCR product differed from this size by 1 bp or more it was considered to have mutated, and the change in length from the stable allele length was determined.

Cell Line	Clone	Size of Smaller Allele	Stable/Mutated (change in bp)	Size of Larger Allele	Stable/Mutated (change in bp)
WT	A	120.5	Stable	218.8	Stable
	B	102.5	Stable	218.8	Stable
	C	102.4	Stable	218.8	Stable
	D	102.6	Stable	218.7	Stable
	E	102.4	Stable	218.7	Stable
	F	102.4	Stable	218.7	Stable
	G	102.3	Stable	218.6	Stable
	H	102.3	Stable	218.7	Stable
	I	102.4	Stable	218.6	Stable
	J	102.3	Stable	218.6	Stable
+/- 1.1	A	102.5	Stable	218.8	Stable
	B	102.5	Stable	218.8	Stable
	C	102.4	Stable	218.7	Stable
	D	102.4	Stable	218.7	Stable
	E	102.4	Stable	218.8	Stable
	F	102.3	Stable	218.6	Stable
	G	102.3	Stable	218.7	Stable
	H	102.3	Stable	218.6	Stable
	I	102.3	Stable	218.7	Stable
	J	102.4	Stable	218.6	Stable
+/- 2.1	A	102.5	Stable	218.8	Stable
	B	102.5	Stable	218.8	Stable
	C	102.4	Stable	218.8	Stable
	D	102.4	Stable	218.7	Stable
	E	102.4	Stable	218.7	Stable
	F	102.3	Stable	218.7	Stable
	G	102.3	Stable	218.6	Stable
	H	102.3	Stable	218.6	Stable
	I	102.4	Stable	218.6	Stable
	J	102.4	Stable	218.6	Stable

Table 4.3 continued

Cell Line	Clone	Size of Smaller Allele	Stable/ Mutated (change in bp)	Size of Larger Allele	Stable/ Mutated (change in bp)
-/- 1.1	A	102.5	Stable	218.7	Stable
	B	102.5	Stable	<b>216.7</b>	<b>Mutated (-2)</b>
	C	102.6	Stable	<b>216.6</b>	<b>Mutated (-2)</b>
	D	102.4	Stable	<b>221.0</b>	<b>Mutated (+2)</b>
	E	ND	ND	ND	ND
	F	102.3	Stable	218.6	Stable
	G	102.3	Stable	<b>212.3</b>	<b>Mutated (-6)</b>
	H	102.4	Stable	<b>216.6</b>	<b>Mutated (-2)</b>
	I	102.4	Stable	<b>206.0</b>	<b>Mutated (-12)</b>
	J	102.3	Stable	218.7	Stable
-/- 2.1	A	102.5	Stable	<b>216.6</b>	<b>Mutated (-2)</b>
	B	<b>100.6</b>	<b>Mutated (-2)</b>	<b>220.9</b>	<b>Mutated (+2)</b>
	C	102.4	Stable	218.9	Stable
	D	102.5	Stable	218.9	Stable
	E	102.2	Stable	<b>223.1</b>	<b>Mutated (+4)</b>
	F	102.4	Stable	<b>214.4</b>	<b>Mutated (-4)</b>
	G	102.3	Stable	<b>216.5</b>	<b>Mutated (-2)</b>
	H	102.3	Stable	218.7	Stable
	I	102.4	Stable	218.7	Stable
	J	102.4	Stable	218.7	Stable
-/+ 1.1	A	102.6	Stable	218.8	Stable
	B	102.6	Stable	218.8	Stable
	C	102.6	Stable	218.2	Stable
	D	102.5	Stable	218.4	Stable
	E	102.4	Stable	218.8	Stable
	F	102.3	Stable	218.6	Stable
	G	102.3	Stable	218.6	Stable
	H	102.3	Stable	218.6	Stable
	I	102.5	Stable	218.4	Stable
	J	ND	ND	ND	ND
Size range of allele in stable clones (bp)		102.2 – 102.6		218.2 – 218.9	

**Table 4.4** Synopsis of results for microsatellite ChrI-7 in 3174 *MSH2* knockout trypanosomes

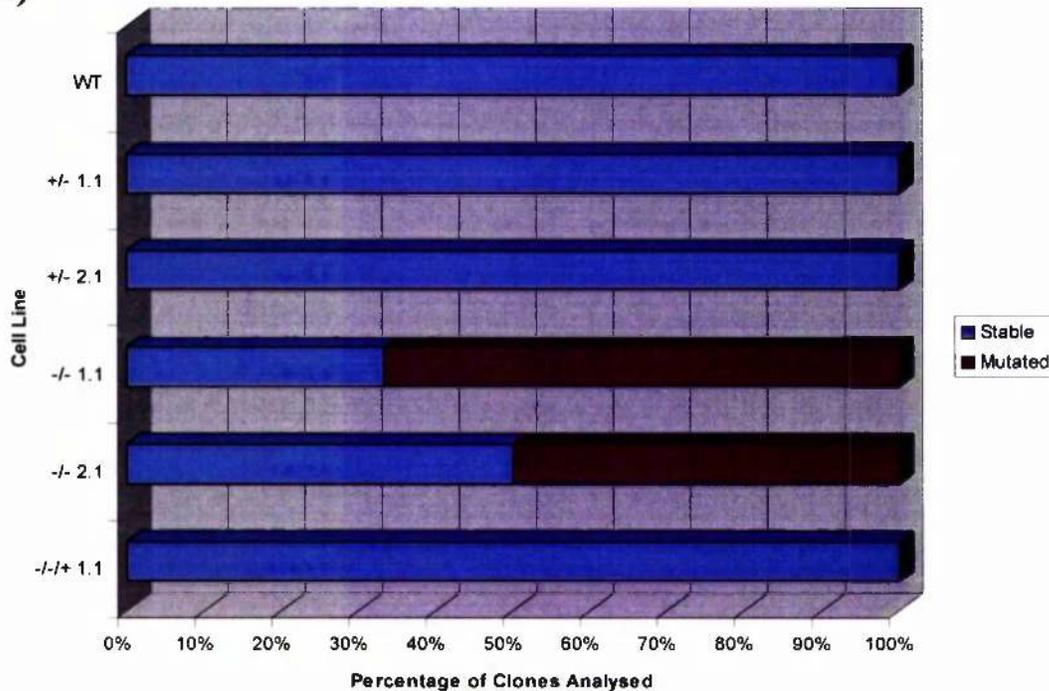
The ChrI-7 microsatellite locus was analysed as described in Table 4.3. Ten clones (where possible) were analysed from each of the following 3174-derived cell lines: *MSH2* wild-type (WT); two heterozygous *MSH2* mutants (+/-); two *MSH2* homozygous mutants (-/-); and the *MSH2* re-expressor cell line (-/-/+). For each cell line, the number and percentage of clones and alleles which were stable or mutated at the ChrI-7 locus is given.

Cell line	Smaller allele			Larger allele						
	No. analysed	Stable	Percentage	Mutated	Percentage	No. analysed	Stable	Percentage	Mutated	Percentage
WT	10	10	100	0	0	10	10	100	0	0
+/- 1.1	10	10	100	0	0	10	10	100	0	0
+/- 2.1	10	10	100	0	0	10	10	100	0	0
-/- 1.1	9	9	100	0	0	9	3	33	6	67
-/- 2.1	10	9	90	1	10	10	5	50	5	50
-/-/+ 1.1	9	9	100	0	0	9	9	100	0	0
<b>Total alleles</b>										
Cell line	No. analysed	Stable	Percentage	Mutated	Percentage	No. analysed	Stable	Percentage	Mutated	Percentage
WT	20	20	100	0	0	10	10	100	0	0
+/- 1.1	20	20	100	0	0	10	10	100	0	0
+/- 2.1	20	20	100	0	0	10	10	100	0	0
-/- 1.1	18	12	67	6	33	9	3	33	6	66
-/- 2.1	20	12	60	6	40	10	5	50	5	50
-/-/+ 1.1	18	18	100	0	0	9	9	100	0	0

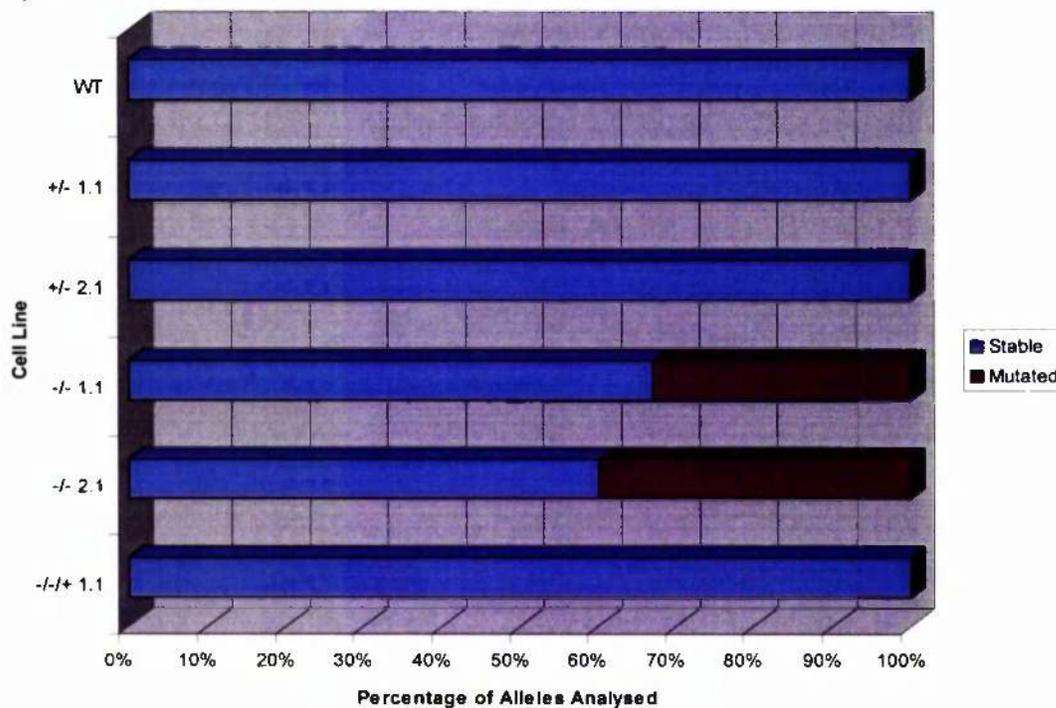
**Figure 4.16** Analysis of microsatellite ChrI-7 in *MSH2* knockout trypanosomes

The percentage number of clones (A) and alleles (B) which were stable (shown and blue) and mutated (shown in red) at the ChrI-7 locus was determined as described in Table 4.4. (WT) 3174 *MSH2* wild-type trypanosomes; (+/-) 3174 *MSH2* heterozygous mutants; (-/-) 3174 *MSH2* homozygous mutants; (-/-/+) 3174 *MSH2* re-expressor cells.

**A)**

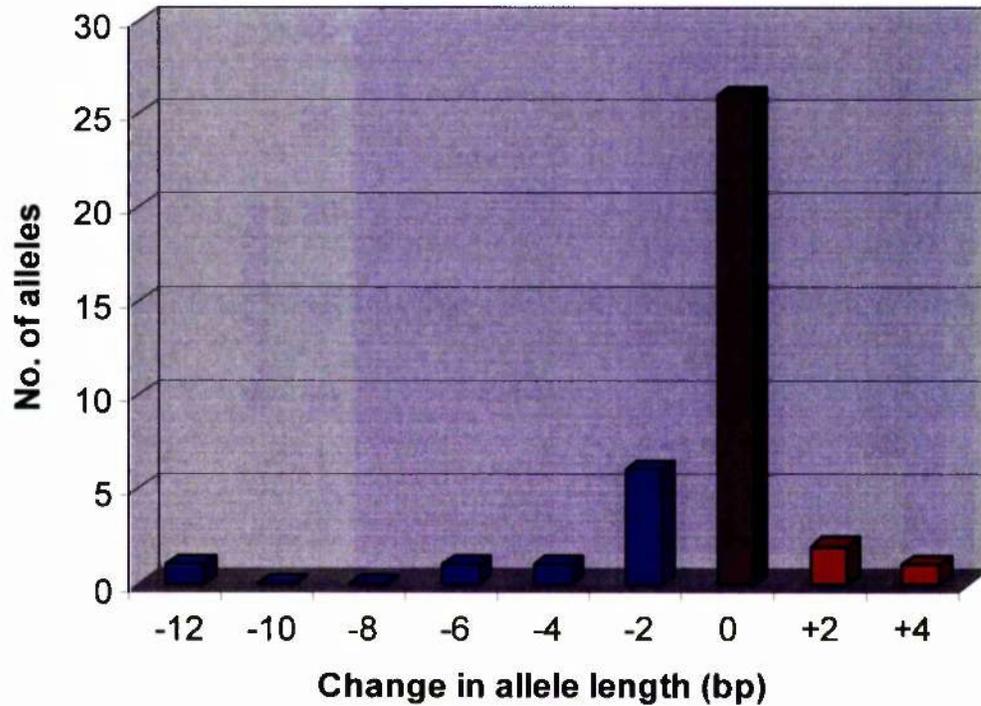


**B)**



**Figure 4.17** Spectrum of mutations observed at microsatellite ChrI-7 in *MSH2* homozygous mutants

The ChrI-7 microsatellite locus was analysed as described in Table 4.3. The change in allele length from the stable allele length, indicated as either no change (shown in purple), addition (+; shown in red) or deletion (-; shown in blue), was calculated for each allele from 9 clones derived from the *MSH2*<sup>-/-</sup> 1.1 cell line, and 10 clones derived from the *MSH2*<sup>-/-</sup> 2.1 cell line.



**Table 4.5 Genescan analysis of microsatellite ChrI-15 in 3174 *MSH2* mutant cell lines**

PCR amplification of the ChrI-15 locus was performed using the primers ChrI-15A (labelled with 6-FAM) and ChrI-15B (see section 2.16) on genomic DNA isolated from 10 clones of each 3174 cell line (designated A-J in each case): (WT) *MSH2* wild-type trypanosomes; (+/-) *MSH2* heterozygous mutants; (-/-) *MSH2* homozygous mutants; (-/-/+) *MSH2* re-expressor cells. The PCR products were separated by acrylamide-urea gel electrophoresis and the fluorescent bands detected by an automated sequencer. The sizes (shown in bp) of the PCR products were determined using the GeneScan software as described in Figure 4.13. The size of stable alleles, taking migration error into account to give a size range, was determined by comparing the size of alleles in every clone analysed, and assessing which sizes occurred most frequently. The most frequently observed size was considered the stable, or wild-type, allele length. Where a PCR product differed from this size by 1 bp or more it was considered to have mutated, and the change in length from the stable allele length was determined.

Cell Line	Clone	Size of Smaller Allele	Stable/Mutated	Size of Larger Allele	Stable/Mutated
WT	A	167.3	Stable	175.4	Stable
	B	167.2	Stable	175.3	Stable
	C	167.2	Stable	175.3	Stable
	D	167.2	Stable	175.3	Stable
	E	167.3	Stable	175.4	Stable
	F	166.9	Stable	175.0	Stable
	G	166.9	Stable	175.1	Stable
	H	166.9	Stable	175.1	Stable
	I	166.8	Stable	175.0	Stable
	J	166.9	Stable	175.1	Stable
+/- 1.1	A	167.2	Stable	175.3	Stable
	B	167.2	Stable	175.3	Stable
	C	167.2	Stable	175.3	Stable
	D	167.2	Stable	175.3	Stable
	E	167.3	Stable	175.4	Stable
	F	166.8	Stable	175.0	Stable
	G	166.9	Stable	175.2	Stable
	H	166.9	Stable	175.2	Stable
	I	166.8	Stable	175.0	Stable
	J	166.8	Stable	175.0	Stable
+/- 2.1	A	167.1	Stable	175.2	Stable
	B	167.1	Stable	175.4	Stable
	C	167.1	Stable	175.4	Stable
	D	167.1	Stable	175.4	Stable
	E	167.3	Stable	175.4	Stable
	F	166.8	Stable	175.1	Stable
	G	166.8	Stable	175.1	Stable
	H	166.8	Stable	174.9	Stable
	I	166.8	Stable	175.1	Stable
	J	166.8	Stable	175.1	Stable

Table 4.5 continued

Cell Line	Clone	Size of Smaller Allele	Stable/Mutated	Size of Larger Allele	Stable/Mutated
-/- 1.1	A	167.2	Stable	175.4	Stable
	B	167.1	Stable	175.4	Stable
	C	167.2	Stable	175.4	Stable
	D	<b>165.1</b>	<b>Mutated (-2)</b>	175.2	Stable
	E	<b>169.3</b>	<b>Mutated (+2)</b>	175.3	Stable
	F	<b>164.8</b>	<b>Mutated (-2)</b>	<b>173.0</b>	<b>Mutated (-2)</b>
	G	<b>162.7</b>	<b>Mutated (-4)</b>	175.0	Stable
	H	<b>164.9</b>	<b>Mutated (-2)</b>	175.1	Stable
	I	166.9	Stable	175.0	Stable
	J	166.9	Stable	175.0	Stable
-/- 2.1	A	167.2	Stable	175.3	Stable
	B	<b>169.2</b>	<b>Mutated (+2)</b>	175.3	Stable
	C	167.1	Stable	<b>173.2</b>	<b>Mutated (-2)</b>
	D	<b>169.3</b>	<b>Mutated (+2)</b>	175.3	Stable
	E	167.1	Stable	175.3	Stable
	F	166.9	Stable	<b>173.0</b>	<b>Mutated (-2)</b>
	G	166.9	Stable	175.0	Stable
	H	167.1	Stable	175.4	Stable
	I	166.9	Stable	<b>177.2</b>	<b>Mutated (+2)</b>
	J	167.2	Stable	175.4	Stable
-/+ 1.1	A	167.2	Stable	175.3	Stable
	B	167.2	Stable	175.3	Stable
	C	167.1	Stable	175.3	Stable
	D	167.2	Stable	175.3	Stable
	E	167.2	Stable	175.4	Stable
	F	167.3	Stable	175.4	Stable
	G	166.9	Stable	175.1	Stable
	H	166.8	Stable	175.0	Stable
	I	166.9	Stable	175.1	Stable
	J	166.9	Stable	175.1	Stable
Size Range of Allele in Stable Clones (bp)		166.8 – 167.3		174.9 – 175.4	

**Table 4.6 Synopsis of results for microsatellite ChrI-15 in 3174 *MSH2* knockout trypanosomes**

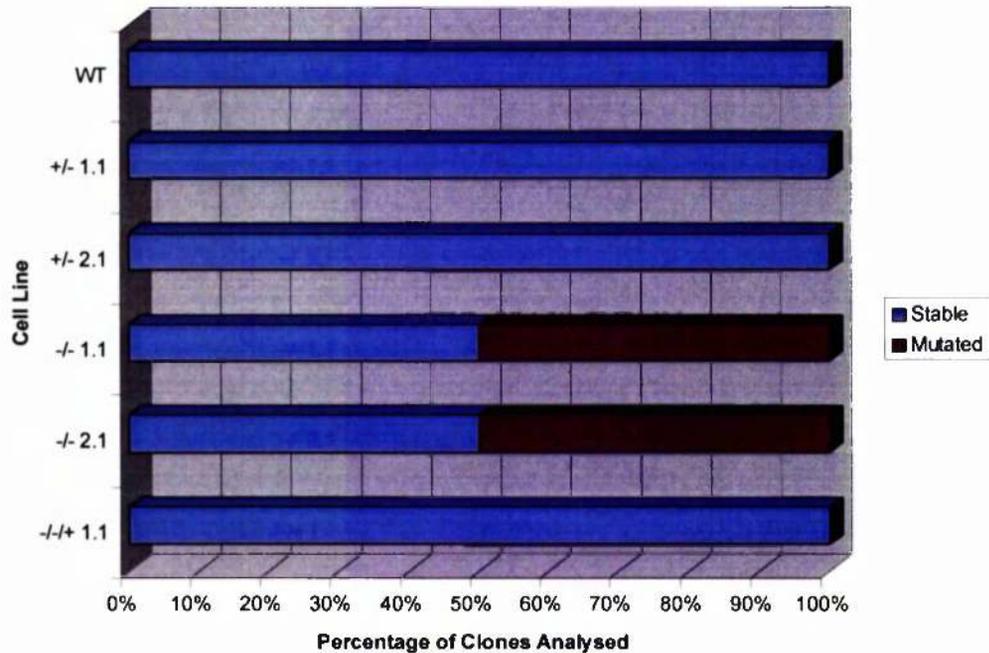
The ChrI-15 microsatellite locus was analysed as described in Table 4.5. Ten clones were analysed from each of the following 3174-derived cell lines: *MSH2* wild-type (WT); two heterozygous *MSH2* mutants (+/-); two *MSH2* homozygous mutants (-/-); and the *MSH2* re-expressor cell line (-/-/+). For each cell line, the number and percentage of clones and alleles which were stable or mutated at this locus is given.

Cell line	Smaller allele						Larger allele					
	No. analysed	Stable	Percentage	Mutated	Percentage	No. analysed	Stable	Percentage	Mutated	Percentage		
WT	10	10	100	0	0	10	10	100	0	0		
+/- 1.1	10	10	100	0	0	10	10	100	0	0		
+/- 2.1	10	10	100	0	0	10	10	100	0	0		
-/- 1.1	10	5	50	5	50	10	9	90	1	10		
-/- 2.1	10	8	80	2	20	10	7	70	3	30		
-/-/+ 1.1	10	10	100	0	0	10	10	100	0	0		
<b>Total alleles</b>												
Cell line	No. analysed	Stable	Percentage	Mutated	Percentage	No. analysed	Stable	Percentage	Mutated	Percentage		
WT	20	20	100	0	0	10	10	100	0	0		
+/- 1.1	20	20	100	0	0	10	10	100	0	0		
+/- 2.1	20	20	100	0	0	10	10	100	0	0		
-/- 1.1	20	14	70	6	30	10	5	50	5	50		
-/- 2.1	20	15	75	5	25	10	5	50	5	50		
-/-/+ 1.1	20	20	100	0	0	10	10	100	0	0		

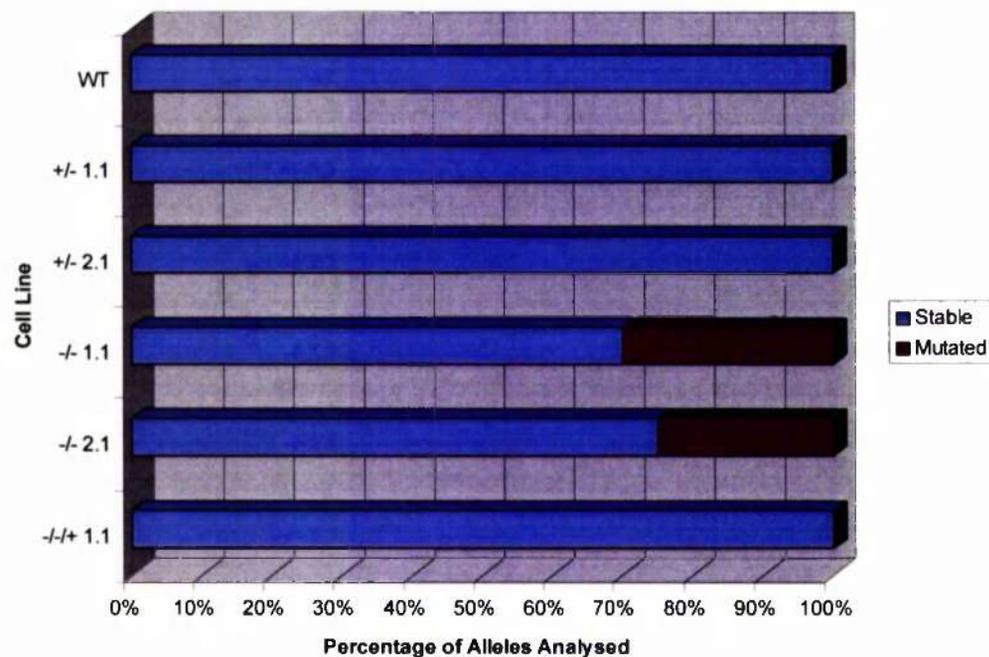
**Figure 4.18** Analysis of microsatellite ChrI-15 in *MSH2* knockout trypanosomes

The percentage number of clones (A) and alleles (B) which were stable (shown in blue) and mutated (shown in red) at the ChrI-15 locus was determined as described in Table 4.6. (WT) 3174 *MSH2* wild-type trypanosomes; (+/-) 3174 *MSH2* heterozygous mutants; (-/-) 3174 *MSH2* homozygous mutants; (-/-+) 3174 *MSH2* re-expressor cells.

**A)**

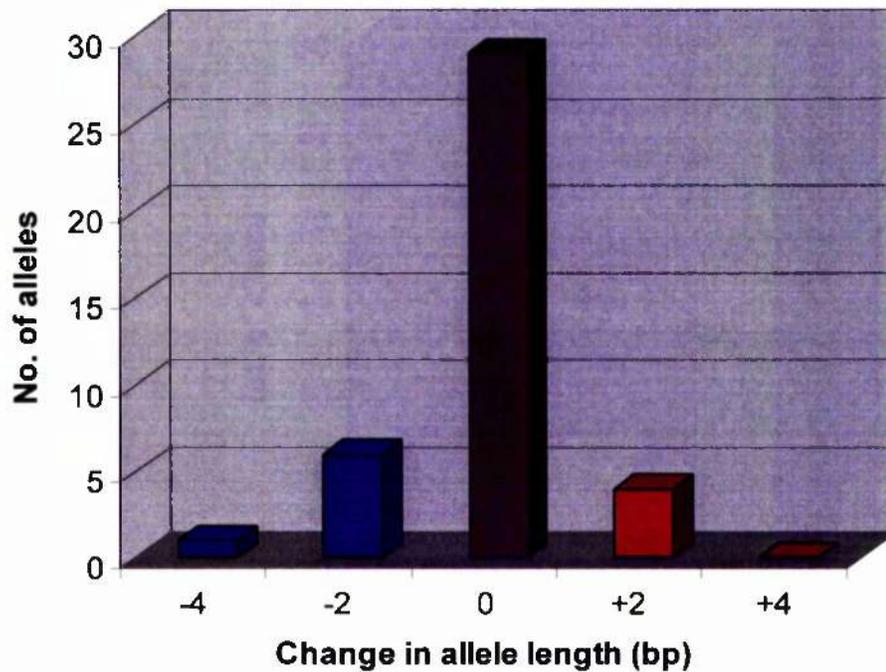


**B)**



**Figure 4.19** Spectrum of mutations observed at microsatellite ChrI-15 in *MSH2* homozygous mutants

The ChrI-15 microsatellite locus was analysed as described in Table 4.5. The change in allele length from the stable allele length, indicated as either no change (shown in purple), addition (+; shown in red) or deletion (-; shown in blue), was calculated for each allele from 10 clones each derived from the *MSH2*<sup>-/-</sup> 1.1 and 2.1 cell lines.



**Table 4.7 Genescan analysis of microsatellite PLC in 3174 *MSH2* mutant cell lines**

PCR amplification of the PLC locus was performed using the primers PLCG (labelled with 6-FAM) and PLCH3 (see section 2.16) on genomic DNA isolated from 10 clones of each 3174 cell line (designated A-J in each case): (WT) *MSH2* wild-type trypanosomes; (+/-) *MSH2* heterozygous mutants; (-/-) *MSH2* homozygous mutants; (-/-/+) *MSH2* re-expressor cells. The PCR products were separated by acrylamide-urea gel electrophoresis and the fluorescent bands detected by an automated sequencer. The sizes (shown in bp) of the PCR products were determined using the GeneScan software as described in Figure 4.13. The size of stable alleles, taking migration error into account to give a size range, was determined by comparing the size of alleles in every clone analysed, and assessing which sizes occurred most frequently. The most frequently observed size was considered the stable, or wild-type, allele length. Where a PCR product differed from this size by 1 bp or more it was considered to have mutated, and the change in length from the stable allele length was determined.

Cell Line	Clone	Size	Stable/Mutated
WT	A	158.1	Stable
	B	157.9	Stable
	C	157.9	Stable
	D	157.9	Stable
	E	158.1	Stable
	F	158.1	Stable
	G	158.1	Stable
	H	157.9	Stable
	I	157.9	Stable
	J	157.9	Stable
+/- 1.1	A	158.1	Stable
	B	158.2	Stable
	C	158.2	Stable
	D	157.9	Stable
	E	157.9	Stable
	F	158.1	Stable
	G	157.9	Stable
	H	157.9	Stable
	I	157.9	Stable
	J	157.9	Stable
+/- 2.1	A	158.1	Stable
	B	158.2	Stable
	C	157.9	Stable
	D	158.1	Stable
	E	157.9	Stable
	F	158.1	Stable
	G	157.9	Stable
	H	ND	ND
	I	157.9	Stable
	J	157.9	Stable

Table 4.7 continued

Cell Line	Clone	Size	Stable/Mutated
-/- 1.1	A	157.9	Stable
	B	158.1	Stable
	C	157.9	Stable
	D	<b>156.2</b>	<b>Mutated (-2)</b>
	E	157.9	Stable
	F	157.9	Stable
	G	158.1	Stable
	H	158.1	Stable
	I	157.9	Stable
	J	<b>155.9</b>	<b>Mutated (-2)</b>
-/- 2.1	A	157.9	Stable
	B	157.9	Stable
	C	158.0	Stable
	D	157.8	Stable
	E	<b>157.3</b>	<b>Mutated (-1)</b>
	F	158.0	Stable
	G	157.8	Stable
	H	157.9	Stable
	I	158.1	Stable
	J	157.9	Stable
-/+ 1.1	A	158.1	Stable
	B	157.8	Stable
	C	158.1	Stable
	D	158.0	Stable
	E	157.9	Stable
	F	157.8	Stable
	G	157.8	Stable
	H	157.9	Stable
	I	157.8	Stable
	J	157.9	Stable
Size Range of Allele in Stable Clones (bp)		157.8 – 158.2	

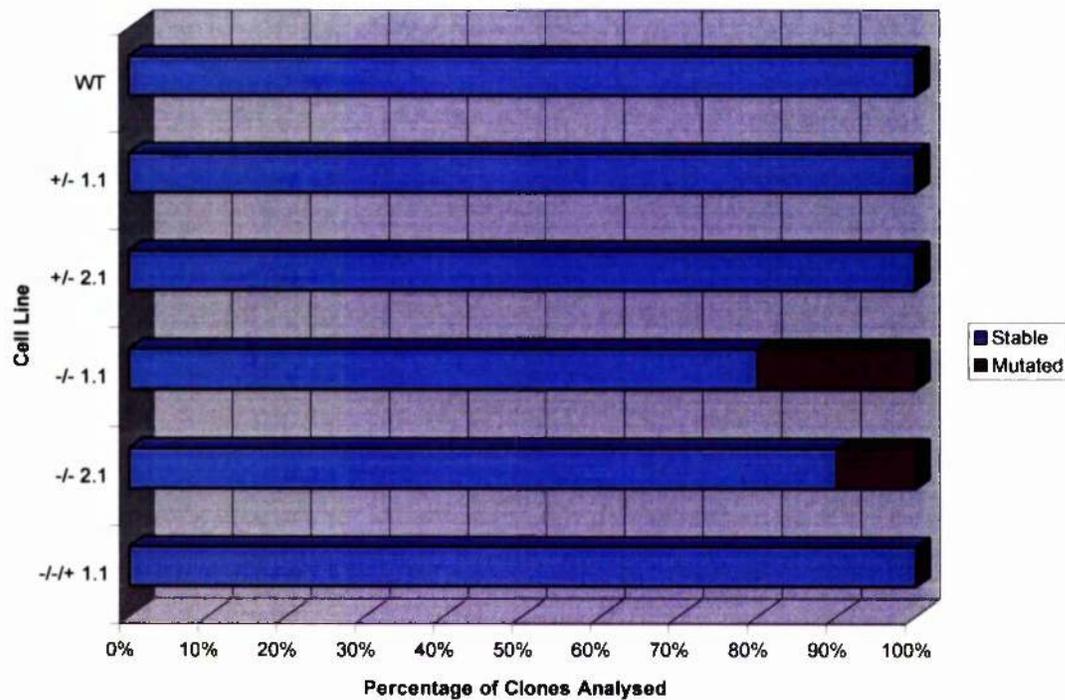
**Table 4.8 Synopsis of results for microsatellite PLC in 3174 *MSH2* knockout trypanosomes**

The PLC microsatellite locus was analysed as described in Table 4.7. Ten clones (where possible) were analysed from each of the following 3174-derived cell lines: *MSH2* wild-type (WT); two heterozygous *MSH2* mutants (+/-); two *MSH2* homozygous mutants (-/-); and the *MSH2* re-expressor cell line (-/-/+). For each cell line, the number and percentage of clones which were stable or mutated at this locus is given.

Cell line	No of clones analysed	Stable clones	Percentage	Mutated clones	Percentage
WT	10	10	100	0	0
+/- 1.1	10	10	100	0	0
+/- 2.1	9	9	100	0	0
-/- 1.1	10	8	80	2	20
-/- 2.1	10	9	90	1	10
-/-/+ 1.1	10	10	100	0	0

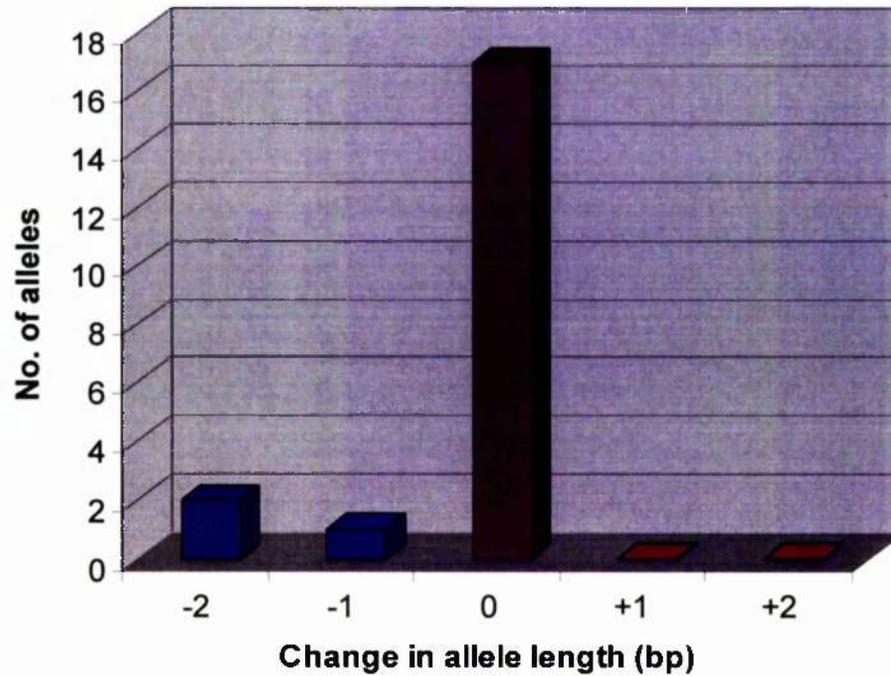
**Figure 4.20 Analysis of microsatellite PLC in *MSH2* knockout trypanosomes**

The percentage number of clones which were stable (shown in blue) and mutated (shown in red) at the PLC locus was determined as described in Table 4.8. (WT) 3174 *MSH2* wild-type trypanosomes; (+/-) 3174 *MSH2* heterozygous mutants; (-/-) 3174 *MSH2* homozygous mutants; (-/-+) 3174 *MSH2* re-expressor cells.



**Figure 4.21** Spectrum of mutations observed at microsatellite PLC in *MSH2* homozygous mutants

The PLC microsatellite locus was analysed as described in Table 4.7. The change in allele length from the stable allele length, indicated as either no change (shown in purple), addition (+; shown in red) or deletion (-; shown in blue), was calculated for each allele from 10 clones each derived from the *MSH2*<sup>-/-</sup> 1.1 and 2.1 cell lines.



**Table 4.9 Genescan analysis of microsatellite ChrII-6 in 3174 *MSH2* mutant cell lines**

PCR amplification of the ChrII-6 locus was performed using the primers ChrII-6A (labelled with 6-FAM) and ChrII-6B (see section 2.16) on genomic DNA isolated from 10 clones of each 3174 cell line (designated A-J in each case); (WT) *MSH2* wild-type trypanosomes; (+/-) *MSH2* heterozygous mutants; (-/-) *MSH2* homozygous mutants; (-/-/+) *MSH2* re-expressor cells. The PCR products were separated by acrylamide-urea gel electrophoresis and the fluorescent bands detected by an automated sequencer. The sizes (shown in bp) of the PCR products were determined using the GeneScan software as described in Figure 4.13. The size of stable alleles, taking migration error into account to give a size range, was determined by comparing the size of alleles in every clone analysed, and assessing which sizes occurred most frequently. The most frequently observed size was considered the stable, or wild-type, allele length. Where a PCR product differed from this size by 1 bp or more it was considered to have mutated, and the change in length from the stable allele length was determined.

Cell Line	Clone	Size	Stable/Mutated
WT	A	96.5	Stable
	B	96.6	Stable
	C	96.7	Stable
	D	96.4	Stable
	E	96.4	Stable
	F	96.4	Stable
	G	96.6	Stable
	H	96.7	Stable
	I	96.4	Stable
	J	96.2	Stable
+/- 1.1	A	96.4	Stable
	B	96.4	Stable
	C	96.4	Stable
	D	96.4	Stable
	E	96.4	Stable
	F	96.4	Stable
	G	96.4	Stable
	H	96.2	Stable
	I	96.4	Stable
	J	96.2	Stable
+/- 2.1	A	96.4	Stable
	B	96.4	Stable
	C	96.4	Stable
	D	96.7	Stable
	E	96.4	Stable
	F	96.4	Stable
	G	96.3	Stable
	H	96.2	Stable
	I	96.2	Stable
	J	96.4	Stable

Table 4.9 continued

Cell Line	Clone	Size	Stable/Mutated
-/- 1.1	A	96.4	Stable
	B	96.4	Stable
	C	96.4	Stable
	D	96.4	Stable
	E	96.4	Stable
	F	96.7	Stable
	G	96.3	Stable
	H	96.2	Stable
	I	96.2	Stable
	J	96.2	Stable
-/- 2.1	A	96.4	Stable
	B	96.4	Stable
	C	96.4	Stable
	D	96.4	Stable
	E	96.4	Stable
	F	96.4	Stable
	G	96.3	Stable
	H	96.2	Stable
	I	96.4	Stable
	J	96.2	Stable
-/+ 1.1	A	96.4	Stable
	B	96.4	Stable
	C	96.4	Stable
	D	96.4	Stable
	E	96.4	Stable
	F	96.4	Stable
	G	96.2	Stable
	H	96.5	Stable
	I	96.2	Stable
	J	96.4	Stable
Size Range of Allele in Stable Clones (bp)		96.2 – 96.7	

**Table 4.10** Synopsis of results for microsatellite ChrII-6 in 3174 *MSH2* knockout trypanosomes

The ChrII-6 microsatellite locus was analysed as described in Table 4.9. Ten clones were analysed from each of the following 3174-derived cell lines: *MSH2* wild-type (WT); two heterozygous *MSH2* mutants (+/-); two *MSH2* homozygous mutants (-/-); and the *MSH2* re-expressor cell line (-/-/+). For each cell line, the number and percentage of clones which were stable or mutated at this locus is given.

Cell line	No of clones analysed	Stable clones	Percentage	Mutated clones	Percentage
WT	10	10	100	0	0
+/- 1.1	10	10	100	0	0
+/- 2.1	10	10	100	0	0
-/- 1.1	10	10	100	0	0
-/- 2.1	10	10	100	0	0
-/-/+ 1.1	10	10	100	0	0

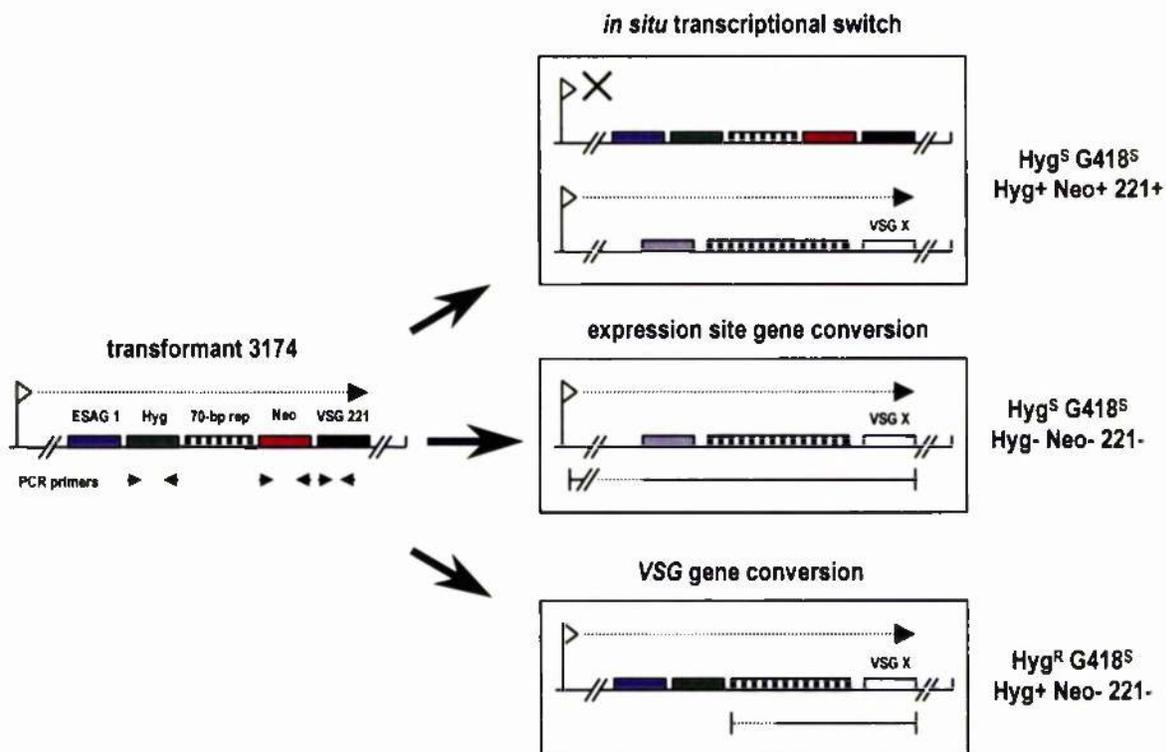
**Table 4.11 Mutational spectrum of *MSH2*<sup>-/-</sup> mutants analysed by GeneScan in *T. brucei***

The microsatellite loci were amplified by PCR, using one fluorescently-labelled primer and an unmodified primer, and the PCR products separated by acrylamide-urea gel electrophoresis. The resulting fluorescent bands were detected using an automated sequencer and analysed using the GeneScan system. After comparing the size of alleles at a given locus in every sample, the most frequently observed size was determined to be the size of the wild-type or stable allele. Where a PCR product differed from this size by 1 bp or more it was considered to have mutated, and the change in length from the size of the stable allele was calculated. For each microsatellite locus the number of clones and alleles analysed is given, and the number of alleles which increased (+) or decreased (-) in length is shown. The total number of mutations of a given size at all loci has been calculated and expressed as a percentage of all the mutations observed. For the ChrII-6 locus the number of alleles analysed could not be determined (ND).

Locus	Type of Repeat	No. of clones (alleles) analysed	Number of tracts with additions (+) or deletions (-) of base-pairs						
			-12	-6	-4	-2	-1	+2	+4
ChrI-7	GT	19 (38)	1	1	1	6	0	2	1
ChrI-15	GT	20 (40)	0	0	1	6	0	4	0
PLC	TA + T + A	20 (20)	0	0	0	1	1	1	0
ChrII-6	None	20 (ND)	0	0	0	0	0	0	0
<b>Number of Mutations (total number analysed)</b>			1/26	1/26	2/26	13/26	1/26	7/26	1/26
<b>Percentage of Mutations</b>			4	4	8	50	4	27	4

**Figure 4.22 Use of the 3174 transgenic trypanosomes strain to characterise *VSG* switching events**

This scheme depicts the telomeric region of the actively transcribed expression site in 3174 bloodstream form cells. This expression site was modified to include the resistance markers for hygromycin (Hyg; shown in green), between the most proximal *ESAG* (*ESAG* 1; shown in purple) and the 70-bp repeats (70-bp rep; shown as a vertically striped box), and G418 (Neo; shown in red), between the 70-bp repeats and the *VSG221* gene (*VSG* 221; shown as a black box). A dashed arrow indicates transcription of the site while X depicts cessation of transcription. Small arrows represent the positions of primers used to assay for the presence of genes within the expression site. Three different switching events can be distinguished in this trypanosome strain. During an *in situ* transcriptional switch, an inactive expression site containing the unknown *VSGX* (*VSG* X; shown as a white box) is transcriptionally activated and the *VSG221* site is inactivated. Cells which have undergone this type of switching event are sensitive to hygromycin (Hyg<sup>S</sup>) and G418 (G418<sup>S</sup>), and each marker gene can be amplified by PCR (Hyg<sup>+</sup>, Neo<sup>+</sup>, 221<sup>+</sup>). Cells which are Hyg<sup>S</sup> and G418<sup>S</sup> from which the markers cannot be amplified by PCR (Hyg<sup>-</sup>, Neo<sup>-</sup>, 221<sup>-</sup>), have undergone a long range expression site gene conversion event in which all the marker genes have been replaced by sequence from another expression site which introduces the *VSGX* gene. In a *VSG* gene conversion event the sequence from the 70-bp repeats to the 3' end of *VSG221* is replaced by sequence containing *VSGX*, deleting the Neo and *VSG221* markers; cells which are resistant to hygromycin (Hyg<sup>R</sup>) but G418<sup>S</sup>, and from which only the Hyg marker can be amplified by PCR (Hyg<sup>+</sup>, Neo<sup>-</sup>, 221<sup>-</sup>) have undergone this type of switching event. The integrity of the genomic DNA template was confirmed using primers directed against the large subunit of RNA polymerase I (Rudenko *et al.*, 1996). (After McCulloch and Barry, 1999)



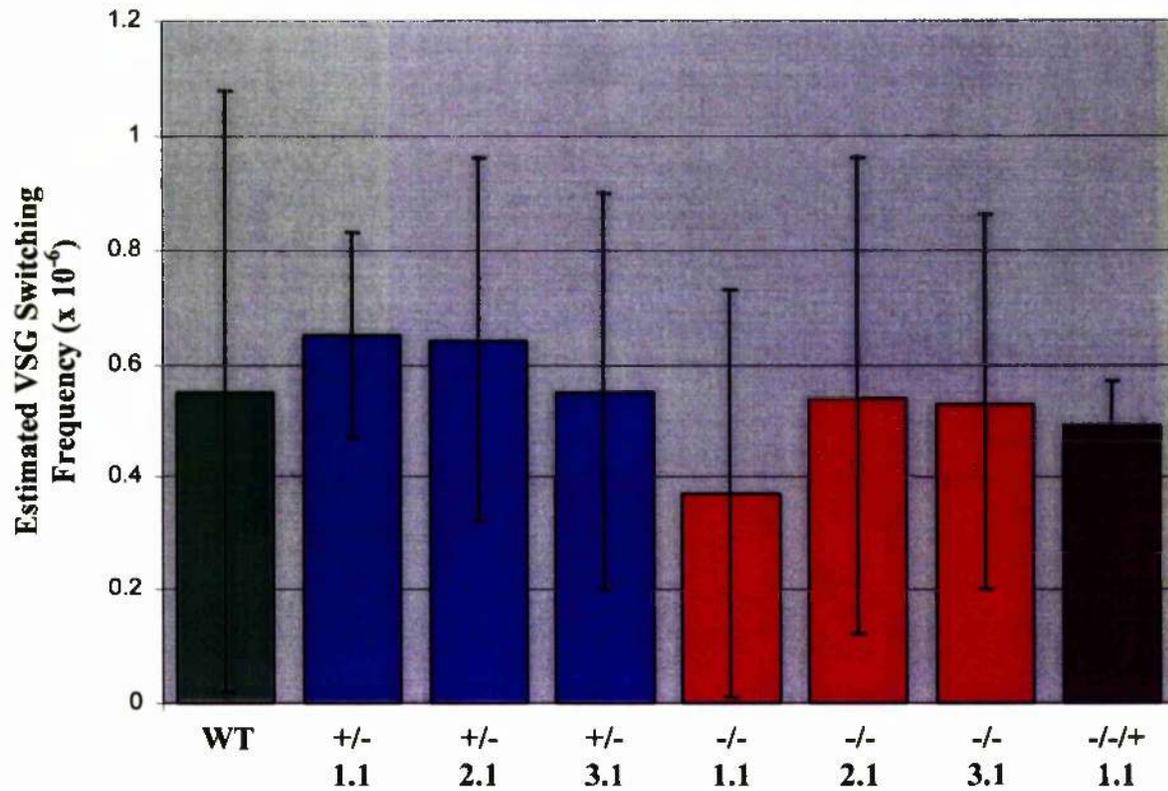
**Table 4.12 Effect of *MSH2* mutation on the frequency of VSG switching in the 3174 trypanosome strain**

The frequency of VSG switching was determined as described in Materials and Methods (see section 2.15) using 3174 bloodstream trypanosomes that had been unaltered at the *MSH2* locus (WT); that had one *MSH2* allele disrupted (+/-); that had both *MSH2* alleles disrupted (-/-); or that had both alleles disrupted but the *MSH2* gene restored to the genome (-/-/+). In each independent experiment, switched trypanosomes were plated over 96-well plates in HMI-9 and the number of wells containing growth were counted after 10 days. The number of VSG switching events/cell/generation (estimated VSG switching frequency ( $\times 10^{-6}$ )) was determined as described in McCulloch *et al.* (1997).

Cell Line	No. of wells growing	Estimated VSG switching frequency ( $\times 10^{-6}$ )
WT	64/96	1.00
	9/96	0.14
	3/96	0.05
	65/96	1.02
	Average	0.55
+/- 1.1	49/96	0.77
	33/96	0.52
	Average	0.65
+/- 2.1	57/96	0.89
	18/96	0.28
	48/96	0.75
	Average	0.64
+/- 3.1	61/96	0.95
	24/96	0.38
	20/96	0.31
	Average	0.55
-/- 1.1	5/96	0.08
	49/96	0.77
	16/96	0.25
	Average	0.37
-/- 2.1	53/96	0.83
	46/96	0.72
	4/96	0.06
	Average	0.54
-/- 3.1	26/96	0.41
	58/96	0.91
	18/96	0.28
	Average	0.53
-/-/+ 1.1	37/96	0.58
	28/96	0.44
	29/96	0.45
	Average	0.49

**Figure 4.23** Effect of *MSH2* mutation on the frequency of VSG switching in the 3174 trypanosome strain

The average frequency of VSG switching was determined as described in Table 4.12 using 3174 bloodstream trypanosomes that had been unaltered at the *MSH2* locus (WT; shown in green); that had one *MSH2* allele disrupted (+/-; shown in blue); that had both *MSH2* alleles disrupted (-/-; shown in red); or that had both alleles disrupted but the *MSH2* gene restored to the genome (-/-/+; shown in purple). Error bars depict standard deviation.



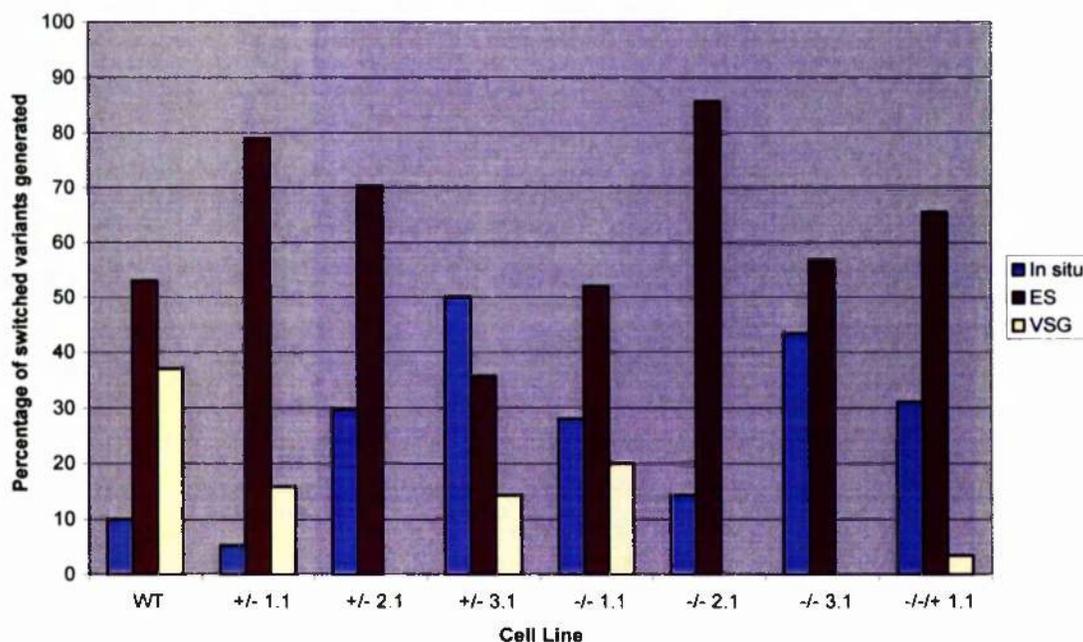
**Table 4.13** *VSG switching mechanisms used by MSH2 wild-type and MSH2 mutant trypanosomes*

The types of switching event used by untransformed 3174 bloodstream forms (WT); 3174 cells transformed with  $\Delta MSB2::PUR$  (+/-); 3174 cells transformed with both  $\Delta MSB2::PUR$  and  $\Delta MSB2::BSD$  (-/-); and 3174 cell transformed with  $\Delta MSB2::PUR$ ,  $\Delta MSB2::BSD$  and  $\Delta MSB2::MSH2-BLE$  (-/-+) was determined as described in Figure 4.22, Materials and Methods (see section 2.15) and McCulloch *et al.* (1997).

<b>Strain (no. of switched variants analysed)</b>	<b><i>In situ</i> switch</b>	<b>Expression site gene conversion</b>	<b><i>VSG</i> gene conversion</b>
<i>MSH2</i> wild-type (30)	3/30 10%	16/30 53%	11/30 37%
<i>MSH2</i> +/- 1.1 (19)	1/19 5%	15/19 79%	3/19 16%
<i>MSH2</i> +/- 2.1 (27)	8/27 30%	19/27 70%	0/27 0%
<i>MSH2</i> +/- 3.1 (28)	14/28 50%	10/28 36%	4/28 14%
<i>MSH2</i> -/- 1.1 (25)	7/25 28%	13/25 52%	5/25 20%
<i>MSH2</i> -/- 2.1 (21)	3/21 14%	18/21 86%	0/21 0%
<i>MSH2</i> -/- 3.1 (30)	13/30 43%	17/30 57%	0/30 0%
<i>MSH2</i> -/-+ 1.1 (29)	9/29 31%	19/29 66%	1/29 3%

**Figure 4.24** *VSG* switching mechanisms used by *MSH2* knockout trypanosomes

The types of switching event used were determined as described in Table 4.13 for untransformed 3174 bloodstream forms (WT); 3174 cells transformed with  $\Delta MS H2::PUR$  (+/-); 3174 cells transformed with both  $\Delta MS H2::PUR$  and  $\Delta MS H2::BSD$  (-/-); and 3174 cells transformed with  $\Delta MS H2::PUR$ ,  $\Delta MS H2::BSD$  and  $\Delta MS H2::MS H2-BLE$  (-/-+). For each cell line the percentage of switched variants generated by *in situ* transcriptional switching (shown in blue), expression site gene conversion (shown in red), and *VSG* gene conversion (shown in yellow) are presented.



### 4.3 Generation and analysis of homozygous *MLH1* knockout trypanosomes

#### 4.3.1 Introduction

MLH1 is the primary MutL-related protein in eukaryotes (Harfe and Jinks-Robertson, 2000). It is a component of every MutL-related heterodimer described in eukaryotes so far, and is believed to be required for all nuclear MMR reactions to proceed (Harfe and Jinks-Robertson, 2000). Inactivation of *MLH1* therefore leads to the complete elimination of MMR activity (Harfe and Jinks-Robertson, 2000). The roles of *MLH1* in *T. brucei* can be investigated using the same assays applied to study the functions of *T. brucei* *MSH2*, as loss of *MLH1* has also been associated with methylation tolerance in mammals (Koi *et al.*, 1994; Hawn *et al.*, 1995), and has been shown to result in the mutator phenotype, MSI (Strand *et al.*, 1993; Bakcr *et al.*, 1996). For these reasons the survival of *MLH1*-deficient trypanosomes in the presence of the alkylating agent, MNNG, will be assessed during this study (see section 4.3.5), and a number of microsatellite loci will be studied in *MLH1*-proficient and *MLH1*-deficient trypanosomes to determine the effect of *MLH1* mutation on microsatellite stability in *T. brucei* (see section 4.3.6). It is hypothesised that homeologous recombination reactions are inhibited in MMR competent cells due to their ability to recognise mismatches which form in heteroduplex intermediates (Datta *et al.*, 1996; Datta *et al.*, 1997), and it is hypothesised that since trypanosome antigenic variation involves recombination (McCulloch and Barry, 1999), MMR may also regulate *VSG* switching. As for *MSH2*, inactivation of *MLH1* has been reported to cause an increase in the frequency of recombination between non-identical sequences in yeast and mammals, although the elevation in recombination frequency seen in *MLH1* mutants is consistently less than that seen in *MSH2* mutants (Nicholson *et al.*, 2000). It is therefore believed that *MSH2* possesses some anti-recombination activity that is independent of *MLH1* (Evans and Alani, 2000). *MSH2* has also been associated with activities which facilitate the resolution of recombination intermediates, including non-homologous tail removal (Sugawara *et al.*, 1997) and Holliday junction cleavage (Marsischky *et al.*, 1999), and may therefore function both to promote recombination as well as suppress it. *MLH1* is not required to process non-homologous tails (Sugawara *et al.*, 1997), and its role in Holliday junction resolution has not been investigated. Deletion of *MSH2* did not result in a detectable alteration in *VSG* switching (section 4.2.8), but it is possible that *MSH2* acts both to promote and inhibit recombination in *T. brucei*, and these conflicting activities cancel each other out at the phenotypic level. Since *MLH1* has not been associated with activities which promote recombination, the same *VSG* switching assay used to study the

role of *MSH2* will be employed to determine if inactivation of *MLH1* affects antigenic variation in this organism.

#### 4.3.2 Design and generation of *MLH1* knockout constructs

The same strategy used to generate *MSH2*<sup>-/-</sup> trypanosomes was adopted to delete both copies of the *MLH1* ORF in order to investigate the function of MLH1 in *T. brucei*. It is clear from an alignment of the *T. brucei* MLH1 polypeptide with MLH1 protein sequences from other eukaryotes (see section 3.3.7) that the primary functional domains are located in the N- and C-terminal regions. Since mutations in one domain do not necessarily inactivate all the activities of the MLH1 protein (Pang *et al.*, 1997), it was again deemed necessary to delete both copies of the entire *MLH1* ORF from the genome, in order that no residual MLH1 activity could remain.

The design of the *T. brucei* *MLH1* knockout constructs mirrored the *MSH2* constructs: sequences immediately 5' and 3' of the *MLH1* ORF were used as targeting sequences for the replacement of the entire *MLH1* ORF by homologous recombination (Figure 4.25) and were separated by either the *BSD* or *PUR* antibiotic resistance cassettes (see section 4.2.2). The resistance genes should be expressed by endogenous *MLH1* transcription after integration, as there are no promoter elements within the knockout constructs.

PCR primers were designed to amplify 413 bp of 5' flank (*MLH1* 5'(A) and *MLH1* 5'(B)) and 520 bp of 3' flank (*MLH1* 3'(A) and *MLH1* 3'(B)) from the *MLH1* locus. PCR using these primers with MITat 1.2 genomic DNA generated products of the expected sizes (Figure 4.26; see section 2.11.2). The *MLH1* targeting flanks were then cloned simultaneously into pBluescript II KS. Finally, the *BSD*- or *PUR*-containing resistance cassettes were cloned between the *MLH1* flanks to generate the  $\Delta MLH1::BSD$  and  $\Delta MLH1::PUR$  constructs respectively (see section 2.11.2).

#### 4.3.3 Generation of *MLH1* mutant cell lines in *T. brucei*

Direct transformation of 3174 bloodstream cells with the linearised knockout constructs allowed the generation of two independent *MLH1* mutants (see section 2.11.2). In order to generate putative heterozygous (*MLH1*<sup>+/-</sup>) mutants, a number of independent cell populations were electroporated with the  $\Delta MLH1::PUR$  construct, and transformants selected in HMI-9 containing 1.0 µg.ml<sup>-1</sup> puromycin dihydrochloride in 24-well plates (see section 2.1.3). Two independent first round transformants were then electroporated with

the  $\Delta MLH1::BSD$  construct, and transformants selected in HMI-9 containing  $2.5 \mu\text{g}\cdot\text{ml}^{-1}$  blasticidin S hydrochloride in 24-well plates to generate putative homozygous ( $MLH1^{-/-}$ ) mutants.

Southern analysis was performed to determine whether or not the constructs had integrated into the  $MLH1$  locus correctly. Genomic DNA was isolated from each transformant, digested with *KpnI*, separated by agarose gel electrophoresis, Southern blotted, and probed with the  $MLH1$  5' flank described in section 4.3.2. The results of this analysis for  $MLH1$  wild-type trypanosomes, the two independent  $MLH1^{+/+}$  cell lines and the two independent  $MLH1^{-/-}$  mutants are shown in Figure 4.27. A 4.0 kb fragment, corresponding to the intact  $MLH1$  locus, was detected in both the  $MLH1$  wild-type cells and the two  $MLH1^{+/+}$  mutants. This fragment was absent from the two  $MLH1^{-/-}$  mutants. A second fragment which was 1.3 kb in length, was detected in the two first round transformants and the two second round transformants. This fragment was generated by replacement of one of the  $MLH1$  alleles by the  $\Delta MLH1::PUR$  construct. Finally, the digestions of the two  $MLH1^{-/-}$  mutants revealed a 1.1 kb fragment, corresponding to the  $MLH1::BSD$  allele, derived by replacement of the second  $MLH1$  allele by the  $\Delta MLH1::BSD$  construct. Taken together, these results indicate that the knockout constructs integrated into the  $MLH1$  locus as expected to generate the desired  $MLH1$  mutant trypanosomes.

In order to show that no copies of the  $MLH1$  ORF existed within genomic DNA isolated from the putative  $MLH1^{-/-}$  mutant cell lines, PCR using primers internal to the  $MLH1$  ORF ( $MLH1U1$  and  $MLH1D6$ ) was performed (Figure 4.28; see section 2.11.2). The integrity of the DNA template was confirmed using primers specific to the large subunit of RNA polymerase I in *T. brucei* (PolI 5' and PolI 3'; Rudenko *et al.*, 1996). PCR amplification using the  $MLH1$  internal primers showed that  $MLH1$  wild-type cells and the two  $MLH1^{+/+}$  mutants gave products of the expected size. No product was visible in the two homozygous mutants, however, indicating that no further copy of  $MLH1$  had been created upon disruption of the two wild-type alleles.

Finally, as antibodies against the *T. brucei*  $MLH1$  polypeptide are not currently available, reverse transcription PCR was used to confirm that no wild-type  $MLH1$  mRNA was being transcribed in the homozygous mutants. Total RNA was isolated from the transformants, and cDNA was generated by reverse-transcription (see section 2.4) for use as a template for PCR amplification using the  $MLH1$  internal primers as above (Figure 4.29). The integrity of the cDNA was confirmed as described above. A PCR product of the expected size was amplified from the  $MLH1$  wild-type cells and the two  $MLH1^{+/+}$  mutants, but not

from the two *MLH1*<sup>-/-</sup> mutants. This indicates that it is possible to remove all intact transcripts of the *MLH1* gene, and therefore all expression of the MLH1 protein, and trypanosomes remain viable. These results also confirm that the knockout constructs integrated into the cells as expected to generate the desired *MLH1* mutant trypanosomes.

#### 4.3.4 Growth of *MLH1* knockout trypanosomes *in vitro*

The *in vitro* growth rate of the *MLH1* mutant trypanosomes was assessed by diluting mid-log phase cultures to a concentration of  $1 \times 10^5$  cells.ml<sup>-1</sup> in HMI-9 and measuring the trypanosome concentrations every 24 h thereafter. Figure 4.30 shows that the *MLH1*<sup>-/-</sup> cells are not only viable but show no alteration in growth compared with the wild-type cells. The average population doubling time during exponential growth was 7.88 h for the *MLH1* wild-type cells, 7.96 ( $\pm 0.9$ ) h for the *MLH1*<sup>+/-</sup> mutants, and 8.73 ( $\pm 1.4$ ) h for the *MLH1*<sup>-/-</sup> cell lines. All the cultures reached similar maximal densities (data not shown). Surprisingly, the 3174 wild-type trypanosomes grew more slowly during this experiment than they did during the same experiment for the *MSH2* mutants (see section 4.2.5). Since all the cell lines used were re-cloned prior to each experiment to ensure vigorous growth, it is unlikely that the wild-type cells used during the *MLH1* experiment were unhealthy. It is more probable that the differences in growth rate were due to differences between the batches of HMI-9 used as the culture medium, or result from the use of a different incubator.

#### 4.3.5 Survival of *MLH1* mutant trypanosomes in the presence of the alkylating agent, MNNG

As described in section 4.2.6, loss of mismatch repair activity in bacteria and mammalian cells results in increased resistance to the cytotoxic effects of alkylating agents, such as MNNG (Karran and Marinus, 1982; de Wind *et al.*, 1995; Humbert *et al.*, 1999). In order to determine if inactivation of *MLH1* resulted in methylation tolerance in *T. brucei*, the survival of *MLH1* wild-type, *MLH1*<sup>+/-</sup> and *MLH1*<sup>-/-</sup> mutant cell lines was assessed by a clonal growth assay in the presence of increasing concentrations of MNNG, as described for *MSH2* mutants (section 4.2.6). The results of this assay are shown in tabular form in Table 4.14, and as a graph in Figure 4.31. It is evident from these data that inactivation of *MLH1* resulted in increased tolerance to the cytotoxic effects of MNNG in *T. brucei*. The survival of *MLH1* wild-type cells was severely reduced after exposure to 0.25  $\mu\text{g.ml}^{-1}$  MNNG as only 2% of wells showed growth. Similarly, the survival of *MLH1* heterozygous mutants was reduced by 99% and 91% under the same conditions. At this

concentration, however, the survival of the *MLH1* homozygous mutants was only reduced to 73% and 87%, equating to a 8.1-fold increase in survival over cells expressing *MLH1*. At 0.5  $\mu\text{g.ml}^{-1}$  MNNG, no growth was observed for either the *MLH1* wild-type cells or the *MLH1*<sup>+/-</sup> mutants, but 16% and 29% of wells showed growth for the *MLH1*<sup>-/-</sup> mutants. Again, at a concentration of 0.75  $\mu\text{g.ml}^{-1}$ , no growth was observed for *MLH1* wild-type cells, either of the *MLH1*<sup>+/-</sup> mutants, or the *MLH1*<sup>-/-</sup> mutant 1.1. For the second *MLH1*<sup>-/-</sup> mutant, 2.1, 2% of wells showed growth at this concentration. At the final concentration, 1.0  $\mu\text{g.ml}^{-1}$ , no growth was observed for any cell line. These results indicate that inactivation of *MLH1* results in MMR deficiency in *T. brucei*. Comparison of the data generated during the survival assays performed using the *MSH2* and *MLH1* mutants indicates that the wild-type cells used during the assay using the *MLH1* mutants were inherently more sensitive to MNNG than those used during the assay using the *MSH2* mutants (see section 4.2.6). For instance, Table 4.1 shows that for *MSH2* wild-type cells, 73% of wells showed growth at a concentration of 0.25  $\mu\text{g.ml}^{-1}$  MNNG. In contrast, Table 4.14 shows that for *MLH1* wild-type cells at the same drug concentration only 2% of wells showed growth. Since the *MSH2* and *MLH1* wild-type cell lines used during these assays were, in fact, the same 3174 transgenic trypanosome cell line it seems unlikely that this dramatic difference in drug sensitivity is genuine. A mundane explanation for this discrepancy could be that the stock solution of MNNG used for the assay performed on the *MLH1* mutants was more concentrated than that used for the assay using the *MSH2* mutants, owing to changes in the sensitivity of the balance used to weigh the drug. Alternatively, the discrepancy could have arisen as a consequence of the difference in growth rate between the two sets of mutants (discussed above).

#### 4.3.6 Microsatellite instability in *MLH1* knockout cell lines

An overview of the role of the bacterial and eukaryotic MMR systems in microsatellite instability (MSI) is given in section 4.2.7, along with a description of the experimental approaches used to assay for this phenotype during this study. To investigate whether *MLH1* is required to facilitate the repair of misaligned replication intermediates in *T. brucei*, the stability of the three naturally occurring microsatellite loci examined previously was compared in *MLH1* wild-type cells, *MLH1*<sup>+/-</sup> cells and *MLH1*<sup>-/-</sup> mutants. These microsatellite loci were chosen because they had the highest mutation frequencies of the five loci studied in *MSH2*<sup>-/-</sup> cells, and so were expected to give the most easily observable phenotypes in *MLH1* mutant trypanosomes. Each of these cell lines was cloned (see section 2.1.3), ten clones from each were grown for approximately 25 generations

(data not shown), and genomic DNA was isolated from each of the cultures (see section 2.16).

The first microsatellite locus analysed in the *MLHI* mutants was JS-2 (Figure 4.10; Hope *et al.*, 1999). The primers JS-2A and JS-2B were used to amplify the locus from all the clonal genomic DNA samples, and the resulting PCR products were separated by electrophoresis on 3% agarose gels (Figure 4.32). It is clear that the JS-2 locus is stable in each of the ten *MLHI* wild-type clones, and that the PCR products are approximately 320 to 340 bp in length. Analysis of the *MLHI*<sup>+/-</sup> clones also indicates that this locus is stable in these cell lines. In contrast, two of the clones derived from the *MLHI*<sup>-/-</sup> 1.1 cell line (A and E) appear to show changes in the sizes of the PCR products amplified, suggesting that the JS-2 locus has mutated in these cells. The PCR products amplified from the clones derived from the second *MLHI*<sup>-/-</sup> mutant do not show any obvious size changes at this locus. These results are summarised in Table 4.15 and are shown as a graph in Figure 4.33. In Figure 4.32, the PCR products are resolved into two bands in some of the agarose gels, while in others only a single band is observed. These differences are probably due to the quality of the agarose gels used to separate the samples, as the five wild-type samples shown in the *MLHI*<sup>+/-</sup> 2.1 panel were taken from exactly the same PCR reactions as samples F-J shown in the wild-type panel, and the same electrophoresis conditions were used in each case. These observations suggest that agarose gel electrophoresis is limited as a technique for the analysis of changes in microsatellite length and that it cannot be used to distinguish insertions or deletions under a threshold size. While it appears that *MLHI*-deficient trypanosomes display greater changes in microsatellite length than cells expressing the gene, it is possible that small changes have occurred in clones derived from every cell line. The more sensitive technique of GeneScan analysis was therefore employed (see section 4.2.7 and Figure 4.13), for the two remaining microsatellite loci, ChrI-7 and ChrI-15 (Figure 4.14; see section 2.16).

The first of these loci, ChrI-7, was amplified from each of the clonal genomic DNA samples using the primers ChrI-7A and ChrI-7B, and the resulting PCR products were analysed using the GeneScan system (see section 2.16). The results of this analysis are shown in Table 4.16, and the data summarised in Table 4.17. A graph of the data shown in Table 4.17 is shown in Figure 4.34. Two alleles, which differed greatly in size, were amplified in each case. In stable clones, clones which did not appear to have mutated at this locus, GeneScan analysis gave sizes of between 102.4 and 102.8 bp for the smaller allele, and 217.1 to 218.5 bp for the larger allele. For all clones analysed from the *MLHI* wild-type and *MLHI*<sup>+/-</sup> backgrounds, both alleles appear to be stable as no changes in

length could be discerned. In contrast, a number of *MLH1*<sup>-/-</sup> clones appeared to contain mutated alleles (Figure 4.35). Out of the 20 *MLH1*<sup>-/-</sup> clones analysed, 50% had mutated at one allele. The larger allele mutated more frequently, as 8 out of the 20 clones showed mutations in this allele, accounting for 80% of the mutation events observed at this locus. The smaller allele was only mutated in two clones. This is probably a reflection of the greater number of repeat units present in the larger allele, which increases the likelihood of slippage during replication of this sequence (Figure 4.14). The most frequent mutation events decreased the length of the microsatellite, in most cases probably reflecting the loss of a single repeat unit, as 6 out of the 10 mutations observed resulted in a loss of 2 bp. A single event decreased the size of the larger allele by 6 bp, probably equating to the loss of 3 repeat units. Increases in allele length were also seen, albeit less frequently. Two mutations of the larger allele probably led to the gain of a single repeat unit (+2 bp), while in one case, a single base-pair was introduced into the smaller allele. Since there is no mononucleotide-run in this allele (Figure 4.14), this 1 bp insertion may have occurred by a mechanism other than replication slippage. These data confirm that inactivation of *MLH1* in *T. brucei* leads to the mutator phenotype MSI.

The second locus analysed using the GeneScan system, ChrI-15, was amplified from all of the clonal genomic DNA samples using the primers ChrI-15A and ChrI-15B (see section 2.16). Table 4.18 gives the results of this analysis, while Table 4.19 shows a summary of the data analysed. A graphical representation of these results is shown in Figure 4.36. The length of the two alleles amplified from the ChrI-15 locus differed by only 8 bp in stable clones (defined as above). GeneScan analysis revealed the smaller allele to be between 167.0 and 167.3 bp in length, while the larger allele was 175.3 to 175.7 bp in length. In every *MLH1* wild-type clone, and both sets of the *MLH1*<sup>+/-</sup> clones, both alleles were shown to be stable. However, comparison of the sizes of the PCR products amplified from this locus in the *MLH1*<sup>-/-</sup> clones revealed a number of clones which appear to have mutated at one or both alleles (Figure 4.37). Again, the larger allele appears to have mutated more frequently, accounting for 7 out of the 10 mutation events observed, while only 3 out of the 19 smaller alleles analysed showed evidence of mutation. For both alleles, the most frequent change resulted in the loss or addition of 2 bp, probably equating to a single repeat unit. Such changes accounted for 80% of the observed mutations, with 5 out of the 8 events leading to a decrease in microsatellite length. Two events which changed the length of the alleles by 4 bp were also seen, one probably resulting in the loss of 2 repeat units and the other in an equivalent gain. Again these data indicate that inactivation of *MLH1* in *T. brucei* leads to a mutator phenotype.

The mutational spectra of the two microsatellite loci analysed using the GeneScan system are shown in Table 4.20. Since mutations were only observed in *MLH1*-deficient clones, only the data from the two sets of clones derived from the *MLH1*<sup>-/-</sup> mutant cell lines is shown. For both loci, the number of alleles which increased or decreased in length by a given number of base-pairs was recorded, and the total number of mutations of each type was calculated. It is clear that mutations most frequently involved the loss of sequence, as decreases in length accounted for 65% of the observed events. The most common decrease in length was 2 bp, which probably resulted from the loss of a single repeat unit in the dinucleotide microsatellites, as this occurred in 55% of the mutations analysed. Losses of 4 and 6 bp were also observed, and probably represent the deletion of 2 and 3 repeat units respectively. 25% of the mutation events seen involved an increase in length of 2 bp, again probably representing the addition of a single repeat unit. One allele was also seen to increase in length by 4 bp, which probably involved the addition of 2 repeat units. Finally, a single event in which an allele increased by 1 bp may not have involved replication slippage. Overall, this study revealed that inactivation of *MLH1* in *T. brucei* results in the mutator phenotype, MSI, indicating that the protein is involved in the repair of slipped replication intermediates, as is the case in yeast and higher eukaryotes. These data confirm that *T. brucei MLH1* is required for nuclear MMR.

#### 4.2.7 Antigenic variation in *MLH1* knockout trypanosomes

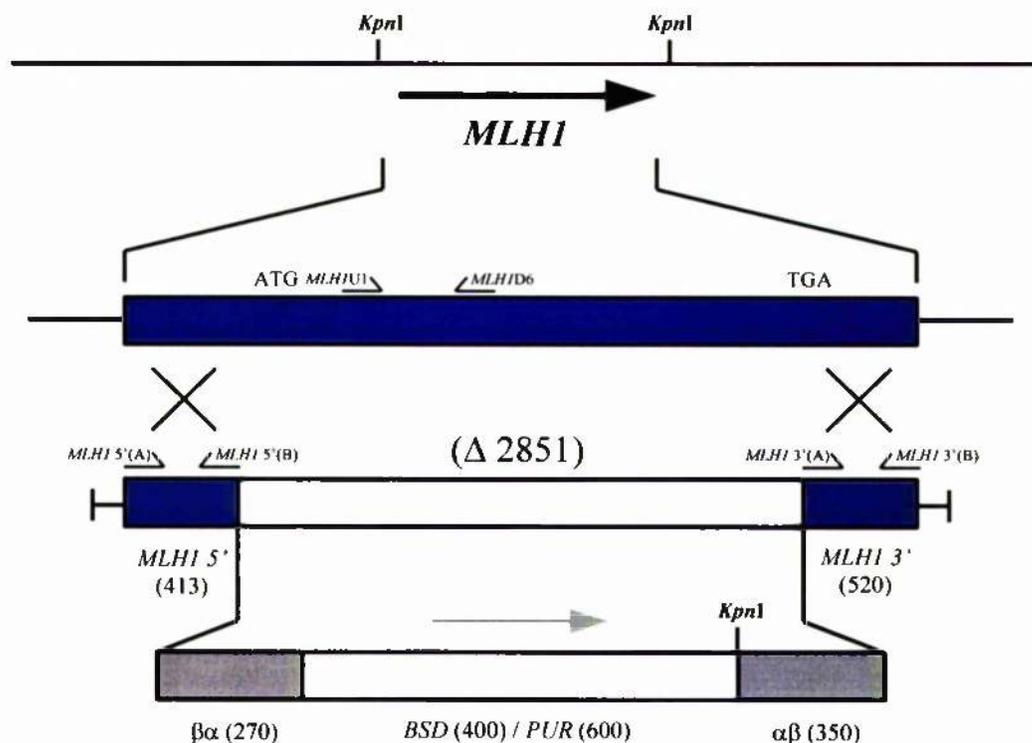
In order to assay for potential changes in the frequency and mechanisms used in antigenic variation in *MLH1*<sup>-/-</sup> mutant trypanosomes, the *MLH1* knockout mutants were generated in the transgenic 3174 background (McCulloch *et al.*, 1997; McCulloch and Barry, 1999). The experimental approach used was essentially the same as that described in section 4.2.8 and Figure 4.22, and the exact experimental conditions used during this study are given in section 2.15. This approach was used to isolate switched variant trypanosomes, arising after immune selection, in two independent *MLH1*<sup>-/-</sup>, and two independent *MLH1*<sup>-/-</sup> cell lines. A number of repetitions were performed using entirely independently grown populations for each cell line. The data from each repetition was combined to allow the calculation of an average VSG switching frequency for each cell line, as shown in Table 4.21 and Figure 4.38, where it is compared with the switching frequency data for wild-type cells determined previously (section 4.2.8). As described previously, the estimated VSG switching frequency varied greatly between repetitions within the same cell line, due to a founder effect. *MLH1* wild-type cells switched their VSG coat at an average frequency of  $0.55 \times 10^{-6}$  switches/cell/generation. The average switching frequency of the two independent *MLH1*<sup>-/-</sup> mutants was about half that of the *MLH1* wild-type cells. In

contrast, the average switching frequency of the two independent *MLHI*<sup>-/-</sup> mutants was similar to the frequency observed for the *MLHI* wild-type cells. The differences observed between the cell lines probably reflects the influence of founder effects rather than real changes in the frequency of antigenic variation. These results suggest that inactivation of the *MLHI* gene, like inactivation of *MSH2*, has no substantial effect on the frequency of VSG switching in 3174 trypanosomes.

To determine the switching mechanisms used by each cell line, ten switched variants were analysed (where possible) from each repetition for *MLHI* wild-type cells, the two *MLHI*<sup>+/-</sup> cell lines and the two *MLHI*<sup>-/-</sup> mutants. The results of this analysis are shown in Table 4.22, and as a graph in Figure 4.39. Inactivation of *MLHI* appeared to have no substantial effect on the relative use of the different switching mechanisms, although this is again complicated by the influence of founder effects which caused variation in their frequency from experiment to experiment. Expression site gene conversion was the most commonly observed switching mechanism, except in the *MLHI*<sup>+/-</sup> 2.1 cell line, being used in 14% to 76% of the switched variants analysed. Every cell line also used the *in situ* transcriptional switching mechanism, which occurred in between 10% and 48% of the switched variants analysed. This mechanism was the most common mechanism used in the *MLHI*<sup>+/-</sup> 2.1 cell line. The final switching mechanism, *VSG* gene conversion, was the least frequently observed, and was not used to generate any switched variants in the *MLHI*<sup>-/-</sup> 2.1 cell line. However, it could be detected in 12% to 37% of the switched variants analysed in the remaining cell lines, including *MLHI* wild-type trypanosomes, *MLHI* heterozygous cells and *MLHI* homozygous mutants. The lack of a detectable alteration in the use of *VSG* switching mechanisms, allied to the lack of change in switching frequency, due to *MLHI* deletion is highly comparable to the result detailed for *MSH2* mutation. Taken together, they suggest that MMR plays little or no role in regulating *VSG* switching in *T. brucei*.

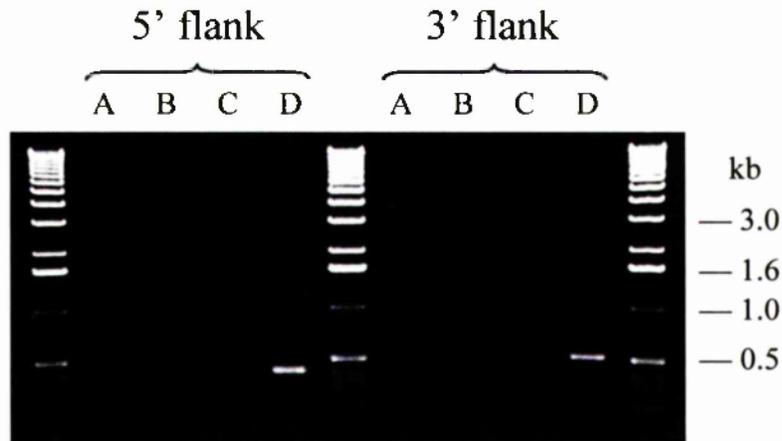
**Figure 4.25** A strategy for the generation of *T. brucei* *MLH1* knockout mutants

The *MLH1* locus is indicated as a blue box, and the constructs used to disrupt its expression are shown below. *MLH1* 5' and *MLH1* 3' are sequences flanking the *MLH1* ORF (ATG to TGA) which allow for integration of the knockout constructs, as indicated by crosses. Two constructs were used, one containing the 400 bp blasticidin S deaminase (*BSD*) gene and the other the 600 bp puromycin acetyltransferase gene (*PUR*); both are flanked by sequences derived from the tubulin locus ( $\beta\alpha$  and  $\alpha\beta$ ; grey boxes) that allow transcripts of the resistance markers to be trans-spliced and polyadenylated into mature mRNAs. In each construct, the 2851 bp *MLH1* ORF is precisely replaced by the resistance cassettes. Arrows represent the direction of transcription of the genes, numbers in brackets denote lengths in base-pairs, and half-arrows indicate the approximate positions of primers used in the analysis. *KpnI* restriction sites are shown as vertical bars.



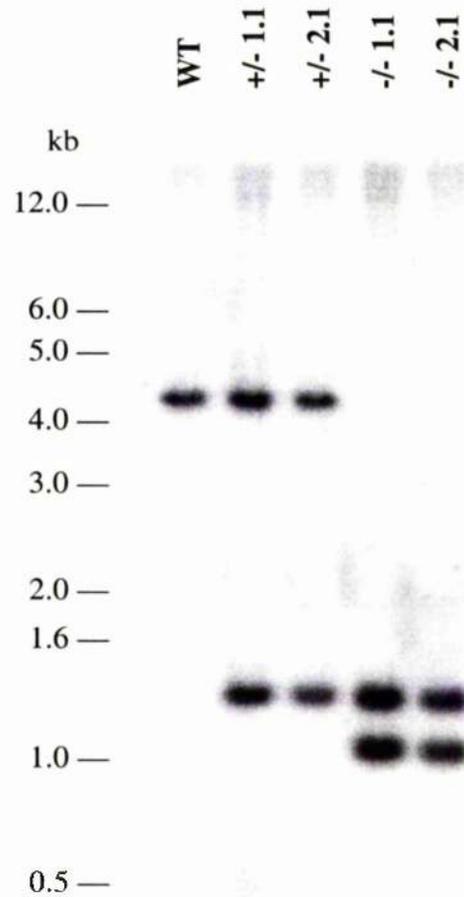
**Figure 4.26 PCR amplification of *MLH1* knockout construct flanks**

The *MLH1* integration flanks were amplified from MITat 1.2 genomic DNA using the primers *MLH1* 5'(A) and *MLH1* 5'(B) (5' flank), and *MLH1* 3'(A) and *MLH1* 3'(B) (3' flank). For each primer pair control reactions were performed with both primers in the absence of *T. brucei* DNA template (1), the (A) primer only (2) and the (B) primer only (3). Lane 4 shows the reaction containing both primers and template for each primer pair.



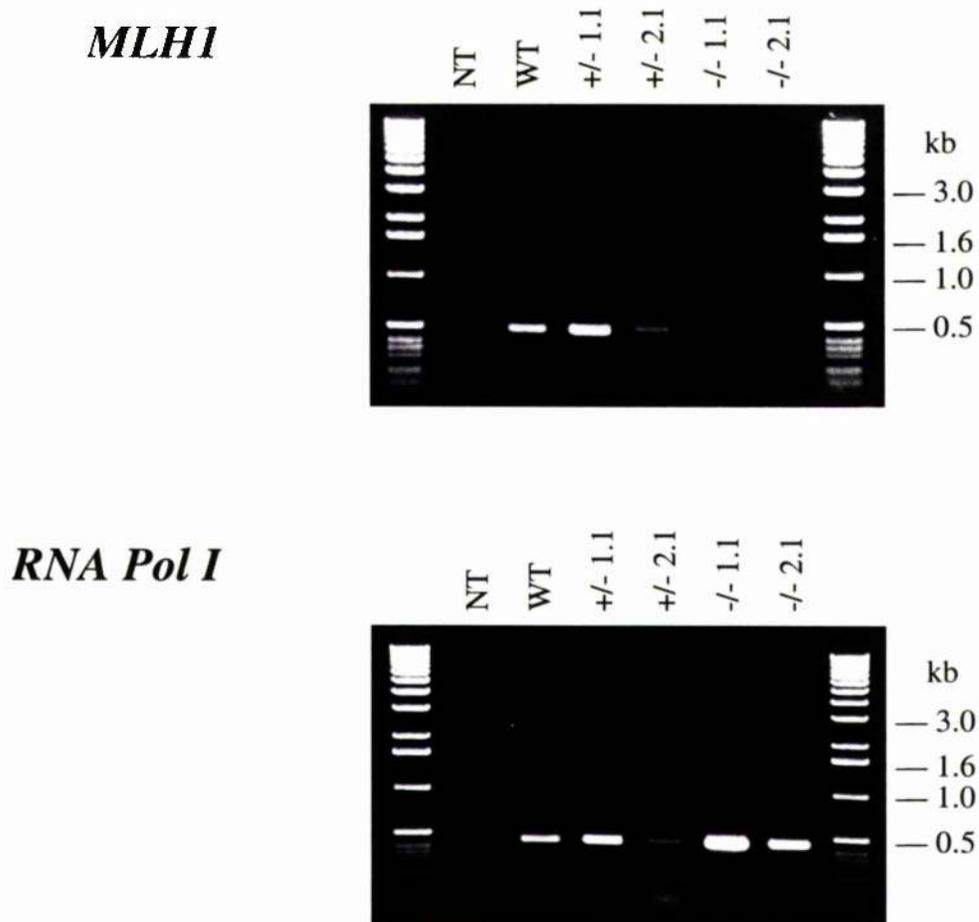
**Figure 4.27 Southern analysis of 3174 *MLH1* knockout cell lines**

Genomic DNA isolated from the 3174 *MLH1* knockout cell lines was digested with *KpnI*, separated on a 0.6% agarose gel, Southern blotted and probed with the *MLH1* 5' flank used as the 5' integration flank in the *MLH1* knockout construct. The following samples were analysed: (WT) untransformed 3174 cells; (+/-) 3174 cells transformed with  $\Delta MLH1::PUR$ ; (-/-) 3174 cells transformed with  $\Delta MLH1::PUR$  and  $\Delta MLH1::BSD$ .



**Figure 4.28 PCR analysis of the 3174 *MLH1* knockout cell lines**

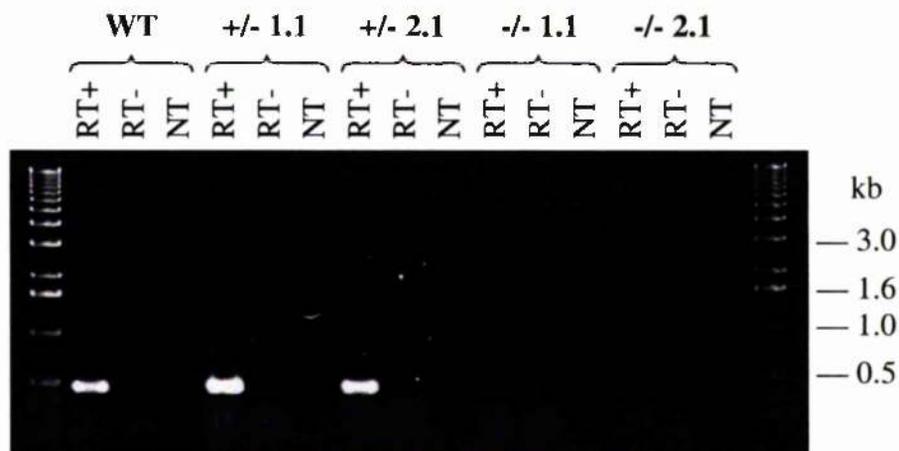
PCR was performed using genomic DNA isolated from the 3174 cell lines transformed with the *MLH1* knockout constructs. Internal primers complementary to the *T. brucei MLH1* gene were used (*MLH1*) and the integrity of the DNA was confirmed using primers directed against the large subunit of *T. brucei* RNA polymerase I (*RNA Pol I*; Rudenko *et al.*, 1996). For both sets of primers the following reactions were run: (NT) no template DNA; (WT) untransformed 3174 cells; (+/-) 3174 cells transformed with  $\Delta MLH1::PUR$ ; (-/-) 3174 cells transformed with both  $\Delta MLH1::PUR$  and  $\Delta MLH1::BSD$ .



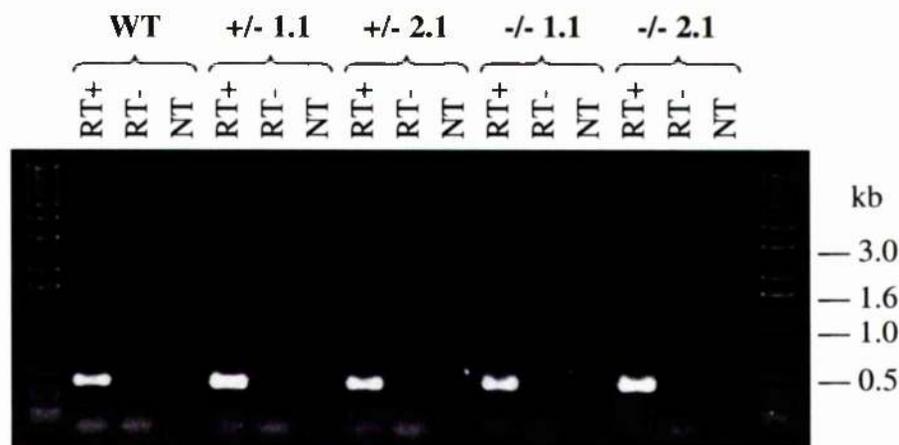
**Figure 4.29** Analysis of the expression of *MLH1* in the putative 3174 *MLH1* mutant cell lines

RT-PCR was performed using total RNA isolated from the 3174 cell lines transformed with the *MLH1* knockout constructs. Internal primers complementary to the *T. brucei MLH1* gene were used (*MLH1*) and the integrity of the DNA was confirmed using primers directed against the large subunit of *T. brucei* RNA polymerase I (*RNA Pol I*; Rudenko *et al.*, 1996). For both sets of primers RT positive (RT+), RT negative (RT-), and no template (NT) reactions were performed using total RNA isolated from the following cell lines: (WT) untransformed 3174 cells; (+/-) 3174 cells transformed with  $\Delta MLH1::PUR$ ; (-/-) 3174 cells transformed with both  $\Delta MLH1::PUR$  and  $\Delta MLH1::BSD$ .

### *MLH1*

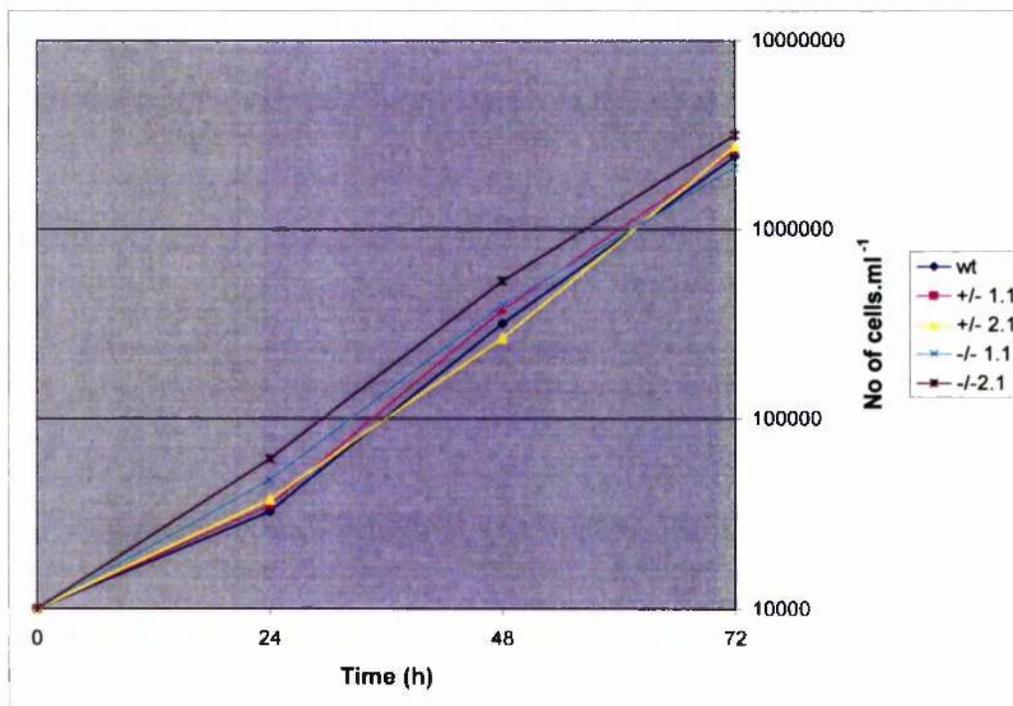


### *RNA Pol I*



**Figure 4.30 Growth of *MLH1* knockout trypanosomes**

Bloodstream trypanosomes that had been unaltered at the *MLH1* locus (WT); that had one *MLH1* allele disrupted (+/-); or that had both *MLH1* alleles disrupted (-/-) were grown in HMI-9. The concentration of trypanosomes in the medium was counted at given times.



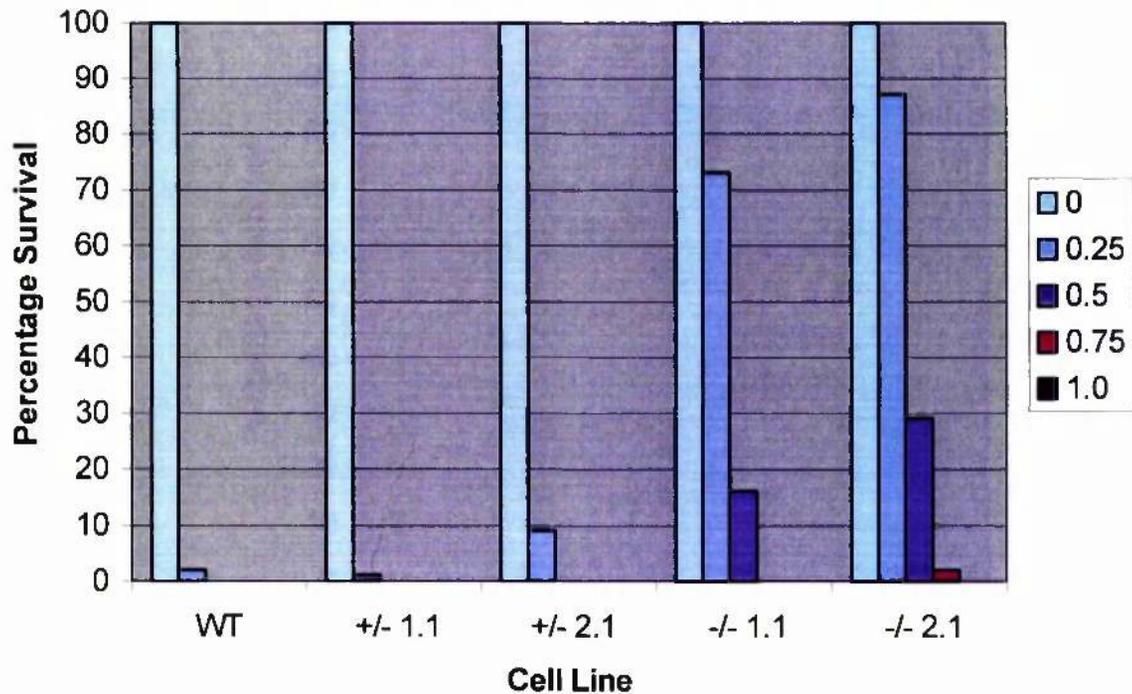
**Table 4.14 Survival of *MLH1* knockout trypanosomes in the presence of MNNG**

Bloodstream trypanosomes that had been unaltered at the *MLH1* locus (WT); that had one *MLH1* allele disrupted (+/-); or that had both *MLH1* alleles disrupted (-/-) were plated in HMI-9 containing increasing concentrations of MNNG (as shown) at a density of 2 cells per well in 96-well plates. For each cell line, at each concentration, 3 plates were prepared, and the number of wells containing growth were counted after 14 days. Percentage survival is presented relative to the number of wells growing in the absence of MNNG, which is given as 100%.

Cell Line	MNNG Conc. ( $\mu\text{g}\cdot\text{ml}^{-1}$ )	No. of wells growing				Percentage Survival
		Plate 1	Plate 2	Plate 3	Average	
WT	0	61/96	60/96	57/96	59/96	100
	0.25	2/96	2/96	0/96	1/96	2
	0.50	0/96	0/96	0/96	0/96	0
	0.75	0/96	0/96	0/96	0/96	0
	1.00	0/96	0/96	0/96	0/96	0
+/- 1.1	0	64/96	50/96	60/96	58/96	100
	0.25	1/96	0/96	1/96	1/96	1
	0.50	0/96	0/96	0/96	0/96	0
	0.75	0/96	0/96	0/96	0/96	0
	1.00	0/96	0/96	0/96	0/96	0
+/- 2.1	0	80/96	76/96	75/96	77/96	100
	0.25	6/96	7/96	8/96	7/96	9
	0.50	0/96	0/96	0/96	0/96	0
	0.75	0/96	0/96	0/96	0/96	0
	1.00	0/96	0/96	0/96	0/96	0
-/- 1.1	0	50/96	47/96	55/96	51/96	100
	0.25	32/96	35/96	44/96	37/96	73
	0.50	4/96	9/96	11/96	8/96	16
	0.75	0/96	0/96	0/96	0/96	0
	1.00	0/96	0/96	0/96	0/96	0
-/- 2.1	0	56/96	66/96	62/96	61/96	100
	0.25	54/96	50/96	56/96	53/96	87
	0.50	16/96	16/96	22/96	18/96	29
	0.75	1/96	2/96	0/96	1/96	2
	1.00	0/96	0/96	0/96	0/96	0

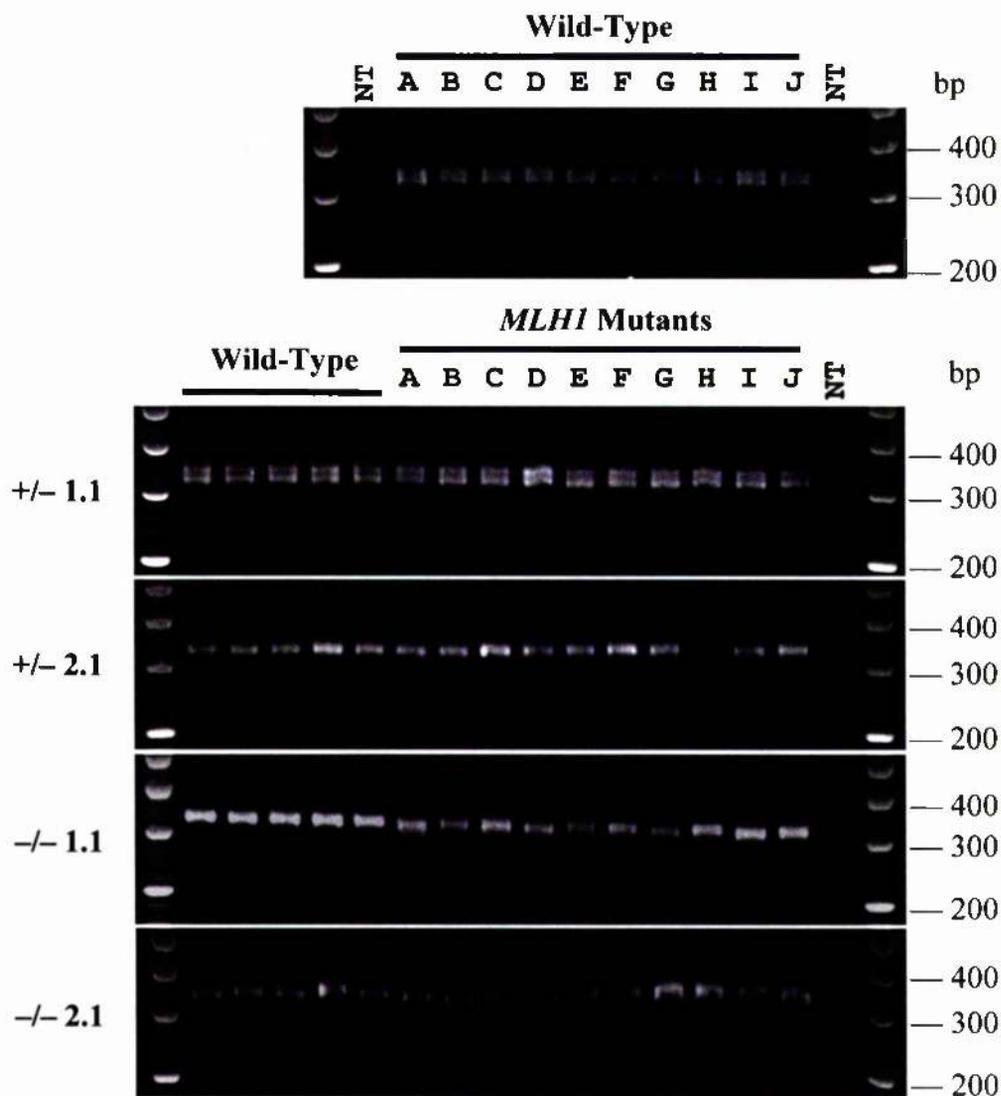
**Figure 4.31 Survival of *MLH1* knockout trypanosomes in the presence of MNNG**

The survival of bloodstream trypanosomes in increasing concentrations of MNNG (shown in  $\mu\text{g.ml}^{-1}$ ) that had been unaltered at the *MLH1* locus (WT); that had one *MLH1* allele disrupted (+/-); or that had both *MLH1* alleles disrupted (-/-) was determined as described in Table 4.14.



**Figure 4.32 PCR amplification of microsatellite JS-2 in 3174 *MLH1* mutant cell lines**

PCR amplification of the JS-2 locus, using the primers JS-2A and JS-2B (see section 2.16), was performed using genomic DNA isolated from 10 clones of each 3174 cell line (labelled A-J in each case): *MLH1* wild-type cells (Wild-Types); *MLH1* heterozygous mutants (+/-); and *MLH1* homozygous mutants (-/-). Control reactions with both primers but lacking template DNA were performed for each set of clones (NT). The PCR products were separated by electrophoresis on 3% agarose gels. In each case, the PCR products derived from five of the *MLH1* wild-type clones were run along side the products amplified from the *MLH1* mutant clones for ease of comparison.



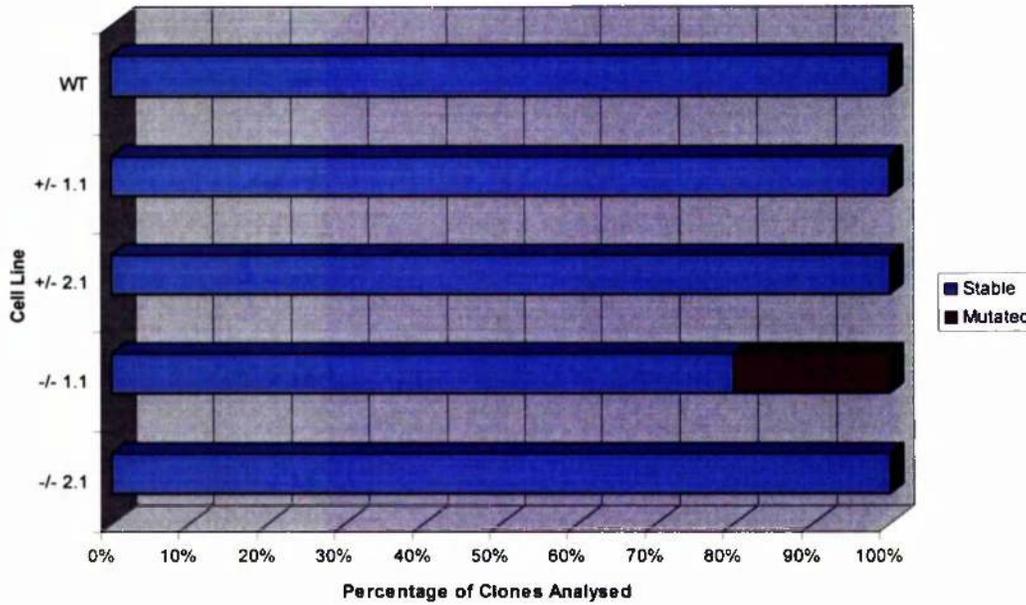
**Table 4.15** Synopsis of results for microsatellite JS-2 in 3174 *MLH1* knockout trypanosomes

The JS-2 microsatellite locus was analysed as shown in Figure 4.32. Ten clones each, derived from the 3174 *MLH1* wild-type cell line (WT); two heterozygous *MLH1* cell lines (+/-); and two *MLH1* homozygous mutants (-/-) were analysed. The amplified JS-2 alleles were compared between the 10 clones derived from the same *MLH1* mutant cell line, and with alleles amplified from 5 *MLH1* wild-type clones. Alleles which appeared identical between the majority of clones derived from the same cell line and with the products amplified from the *MLH1* wild-type clones were considered stable. If all the alleles amplified from all the clones derived from the same cell line appeared identical then they were also considered stable. Alleles which appeared different from the majority of alleles derived from the same cell line were considered mutant versions of the microsatellite. For each cell line, the number and percentage of clones in which the JS-2 locus was shown to be stable or mutated is given.

Cell line	No of clones analysed	Stable clones	Percentage	Mutated clones	Percentage
WT	10	10	100	0	0
+/- 1.1	10	10	100	0	0
+/- 2.1	10	10	100	0	0
-/- 1.1	10	8	80	2	20
-/- 2.1	10	10	100	0	0

**Figure 4.33** Analysis of microsatellite JS-2 in *MLH1* knockout trypanosomes

The percentage number of clones which were stable (shown in blue) and mutated (shown in red) at the JS-2 microsatellite locus was determined as described in Table 4.2. (WT) 3174 *MLH1* wild-type trypanosomes; (+/-) 3174 *MLH1* heterozygous mutants; (-/-) 3174 *MLH1* homozygous mutants.



**Table 4.16 Genescan analysis of microsatellite ChrI-7 in 3174 *MLH1* mutant cell lines**

PCR amplification of the ChrI-7 locus was performed using the primers ChrI-7A (labelled with 6-FAM) and ChrI-7B (see section 2.16) on genomic DNA isolated from 10 clones of each 3174 cell line (designated A-J in each case): (WT) *MLH1* wild-type trypanosomes; (+/-) *MLH1* heterozygous mutants; (-/-) *MLH1* homozygous mutants. The PCR products were separated by acrylamide-urea gel electrophoresis and the fluorescent bands detected by an automated sequencer. The sizes (shown in bp) of the PCR products were determined using the GeneScan software as described in Figure 4.13. The size of stable alleles, taking migration error into account to give a size range, was determined by comparing the size of alleles in every clone analysed, and assessing which sizes occurred most frequently. The most frequently observed size was considered the stable, or wild-type, allele length. Where a PCR product differed from this size by 1 bp or more it was considered to have mutated, and the change in length from the stable allele length was determined.

Cell Line	Clone	Size of Smaller Allele	Stable/Mutated	Size of Larger Allele	Stable/Mutated
WT	A	102.7	Stable	217.2	Stable
	B	102.7	Stable	217.2	Stable
	C	102.7	Stable	217.2	Stable
	D	102.7	Stable	217.3	Stable
	E	102.7	Stable	217.3	Stable
	F	102.7	Stable	217.3	Stable
	G	102.4	Stable	218.4	Stable
	H	102.5	Stable	218.3	Stable
	I	102.5	Stable	218.4	Stable
	J	102.5	Stable	218.5	Stable
+/- 1.1	A	102.6	Stable	217.2	Stable
	B	102.7	Stable	217.2	Stable
	C	102.6	Stable	217.7	Stable
	D	102.8	Stable	217.2	Stable
	E	102.7	Stable	217.2	Stable
	F	102.7	Stable	217.3	Stable
	G	102.4	Stable	218.4	Stable
	H	102.4	Stable	218.3	Stable
	I	102.4	Stable	218.3	Stable
	J	102.5	Stable	218.4	Stable
+/- 2.1	A	102.7	Stable	217.2	Stable
	B	102.6	Stable	217.2	Stable
	C	102.7	Stable	217.3	Stable
	D	102.7	Stable	217.1	Stable
	E	102.8	Stable	217.2	Stable
	F	102.8	Stable	217.2	Stable
	G	102.5	Stable	218.3	Stable
	H	102.4	Stable	218.3	Stable
	I	102.5	Stable	218.4	Stable
	J	102.5	Stable	218.4	Stable

Table 4.16 continued

Cell Line	Clone	Size of Smaller Allele	Stable/Mutated	Size of Larger Allele	Stable/Mutated
-/- 1.1	A	102.6	Stable	217.3	Stable
	B	102.7	Stable	217.2	Stable
	C	102.7	Stable	217.2	Stable
	D	102.6	Stable	<b>215.7</b>	<b>Mutated (-2)</b>
	E	102.8	Stable	217.3	Stable
	F	102.8	Stable	217.3	Stable
	G	102.4	Stable	218.3	Stable
	H	102.7	Stable	<b>215.7</b>	<b>Mutated (-2)</b>
	I	102.6	Stable	<b>215.8</b>	<b>Mutated (-2)</b>
	J	102.5	Stable	218.4	Stable
-/- 2.1	A	102.6	Stable	217.2	Stable
	B	<b>100.7</b>	<b>Mutated (-2)</b>	217.5	Stable
	C	102.6	Stable	<b>212.0</b>	<b>Mutated (-6)</b>
	D	102.8	Stable	217.3	Stable
	E	102.8	Stable	217.3	Stable
	F	102.6	Stable	<b>215.9</b>	<b>Mutated (-2)</b>
	G	102.6	Stable	<b>220.3</b>	<b>Mutated (+2)</b>
	H	<b>103.4</b>	<b>Mutated (+1)</b>	218.5	Stable
	I	102.6	Stable	<b>216.1</b>	<b>Mutated (-2)</b>
	J	102.6	Stable	<b>220.4</b>	<b>Mutated (+2)</b>
Size Range of Allele in Stable Clones (bp)		102.4 – 102.8		217.1 – 218.5	

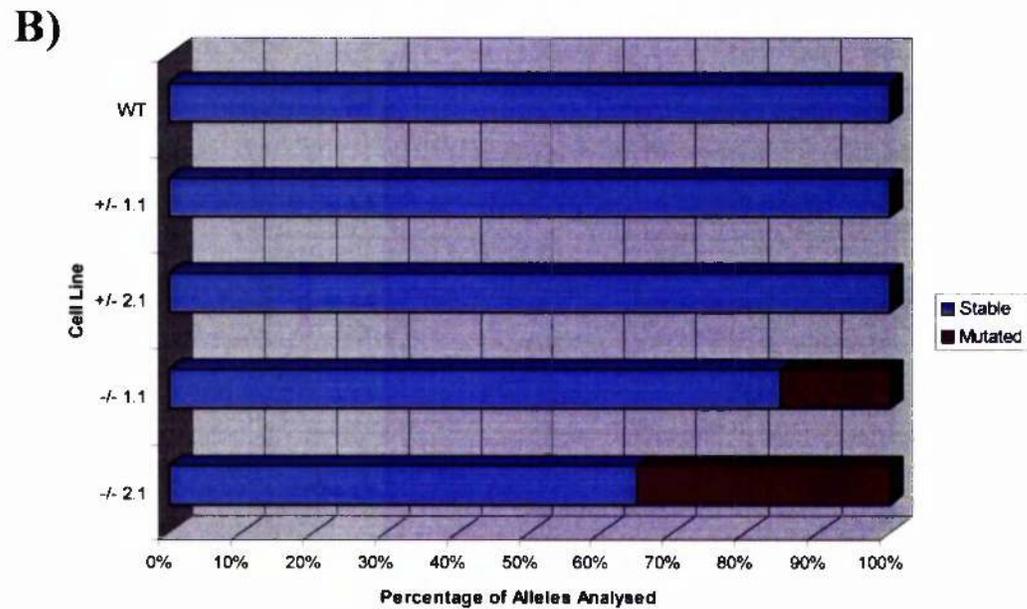
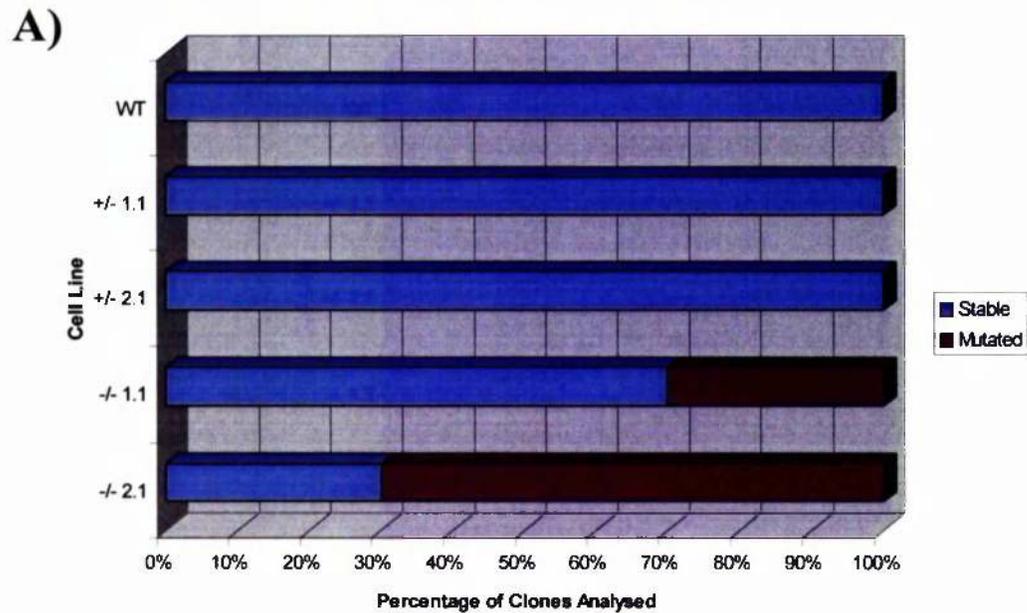
**Table 4.17 Synopsis of results for microsatellite Chr1-7 in 3174 *MLHI* knockout trypanosomes**

The Chr1-7 microsatellite locus was analysed as described in Table 4.16. Ten clones were analysed from each of the following 3174-derived cell lines: *MLHI* wild-type (WT); two heterozygous *MLHI* mutants (+/-); and two *MLHI* homozygous mutants (-/-). For each cell line, the number and percentage of clones and alleles which were stable or mutated at the Chr1-7 locus is given.

Cell line	Smaller allele					Larger allele				
	No. analysed	Stable	Percentage	Mutated	Percentage	No. analysed	Stable	Percentage	Mutated	Percentage
WT	10	10	100	0	0	10	10	100	0	0
+/- 1.1	10	10	100	0	0	10	10	100	0	0
+/- 2.1	10	10	100	0	0	10	10	100	0	0
-/- 1.1	10	10	100	0	0	10	7	70	3	30
-/- 2.1	10	8	80	2	20	10	5	50	5	50
<b>Total alleles</b>										
Cell line	No. analysed	Stable	Percentage	Mutated	Percentage	No. analysed	Stable	Percentage	Mutated	Percentage
WT	20	20	100	0	0	10	10	100	0	0
+/- 1.1	20	20	100	0	0	10	10	100	0	0
+/- 2.1	20	20	100	0	0	10	10	100	0	0
-/- 1.1	20	17	85	3	15	10	7	70	3	30
-/- 2.1	20	13	65	7	35	10	3	30	7	70

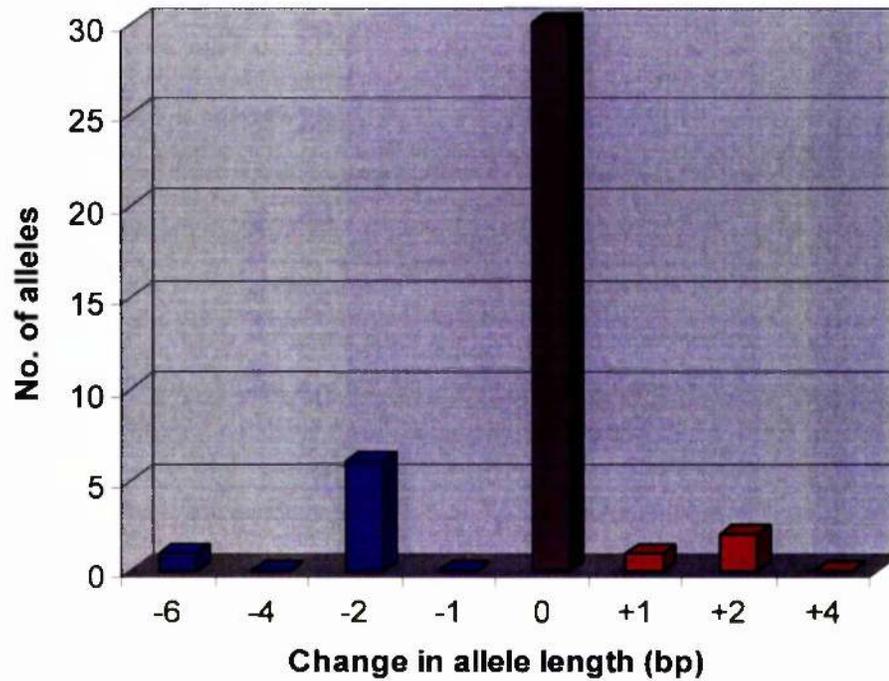
**Figure 4.34 Analysis of microsatellite ChrI-7 in *MLH1* knockout trypanosomes**

The percentage number of clones (A) and alleles (B) which were stable (shown in blue) and mutated (shown in red) at the ChrI-7 locus was determined as described in Table 4.17. (WT) 3174 *MLH1* wild-type trypanosomes; (+/-) 3174 *MLH1* heterozygous mutants; (-/-) 3174 *MLH1* homozygous mutants.



**Figure 4.35** Spectrum of mutations observed at microsatellite ChrI-7 in *MLH1* homozygous mutants

The ChrI-7 microsatellite locus was analysed as described in Table 4.16. The change in allele length from the stable allele length, indicated as either no change (shown in purple), addition (+; shown in red) or deletion (-; shown in blue), was calculated for each allele from 10 clones each derived from the *MLH1*<sup>-/-</sup> 1.1 and 2.1 cell lines.



**Table 4.18 Genescan analysis of microsatellite ChrI-15 in 3174 *MLHI* mutant cell lines**

PCR amplification of the ChrI-15 locus was performed using the primers ChrI-15A (labelled with 6-FAM) and ChrI-15B (see section 2.16) on genomic DNA isolated from 10 clones of each 3174 cell line (designated A-J in each case): (WT) *MLHI* wild-type trypanosomes; (+/-) *MLHI* heterozygous mutants; (-/-) *MLHI* homozygous mutants. The PCR products were separated by acrylamide-urea gel electrophoresis and the fluorescent bands detected by an automated sequencer. The sizes (shown in bp) of the PCR products were determined using the GeneScan software as described in Figure 4.13. The size of stable alleles, taking migration error into account to give a size range, was determined by comparing the size of alleles in every clone analysed, and assessing which sizes occurred most frequently. The most frequently observed size was considered the stable, or wild-type, allele length. Where a PCR product differed from this size by 1 bp or more it was considered to have mutated, and the change in length from the stable allele length was determined.

Cell Line	Clone	Size of Smaller Allele	Stable/Mutated	Size of Larger Allele	Stable/Mutated
WT	A	167.2	Stable	175.6	Stable
	B	167.2	Stable	175.5	Stable
	C	167.2	Stable	175.7	Stable
	D	167.2	Stable	175.5	Stable
	E	167.2	Stable	175.5	Stable
	F	167.2	Stable	175.5	Stable
	G	167.2	Stable	175.4	Stable
	H	167.0	Stable	175.3	Stable
	I	167.0	Stable	175.4	Stable
	J	167.2	Stable	175.4	Stable
+/- 1.1	A	167.2	Stable	175.5	Stable
	B	167.1	Stable	175.4	Stable
	C	167.2	Stable	175.5	Stable
	D	167.3	Stable	175.4	Stable
	E	167.3	Stable	175.6	Stable
	F	ND	ND	ND	ND
	G	167.2	Stable	175.4	Stable
	H	167.2	Stable	175.5	Stable
	I	167.2	Stable	175.4	Stable
	J	167.2	Stable	175.4	Stable
+/- 2.1	A	167.2	Stable	175.5	Stable
	B	167.1	Stable	175.5	Stable
	C	167.1	Stable	175.4	Stable
	D	167.1	Stable	175.4	Stable
	E	167.2	Stable	175.4	Stable
	F	167.3	Stable	175.6	Stable
	G	167.1	Stable	175.4	Stable
	H	167.2	Stable	175.4	Stable
	I	167.1	Stable	175.4	Stable
	J	167.2	Stable	175.5	Stable

Table 4.18 continued

Cell Line	Clone	Size of Smaller Allele	Stable/Mutated	Size of Larger Allele	Stable/Mutated
-/- 1.1	A	167.3	Stable	<b>177.6</b>	<b>Mutated (+2)</b>
	B	<b>169.2</b>	<b>Mutated (+2)</b>	<b>173.4</b>	<b>Mutated (-2)</b>
	C	167.3	Stable	175.6	Stable
	D	167.1	Stable	175.4	Stable
	E	167.3	Stable	<b>173.5</b>	<b>Mutated (-2)</b>
	F	<b>165.3</b>	<b>Mutated (-2)</b>	<b>177.6</b>	<b>Mutated (+2)</b>
	G	167.2	Stable	175.4	Stable
	H	167.2	Stable	<b>173.4</b>	<b>Mutated (-2)</b>
	I	167.1	Stable	175.4	Stable
	J	167.1	Stable	175.4	Stable
-/- 2.1	A	ND	ND	ND	ND
	B	167.3	Stable	175.6	Stable
	C	167.3	Stable	175.6	Stable
	D	167.3	Stable	<b>179.7</b>	<b>Mutated (+4)</b>
	E	167.3	Stable	175.4	Stable
	F	167.3	Stable	175.4	Stable
	G	167.1	Stable	<b>173.3</b>	<b>Mutated (-2)</b>
	H	<b>162.9</b>	<b>Mutated (-4)</b>	175.3	Stable
	I	167.2	Stable	175.4	Stable
	J	167.2	Stable	175.4	Stable
Size Range of Allele in Stable Clones (bp)		167.0 – 167.3		175.3 – 175.7	

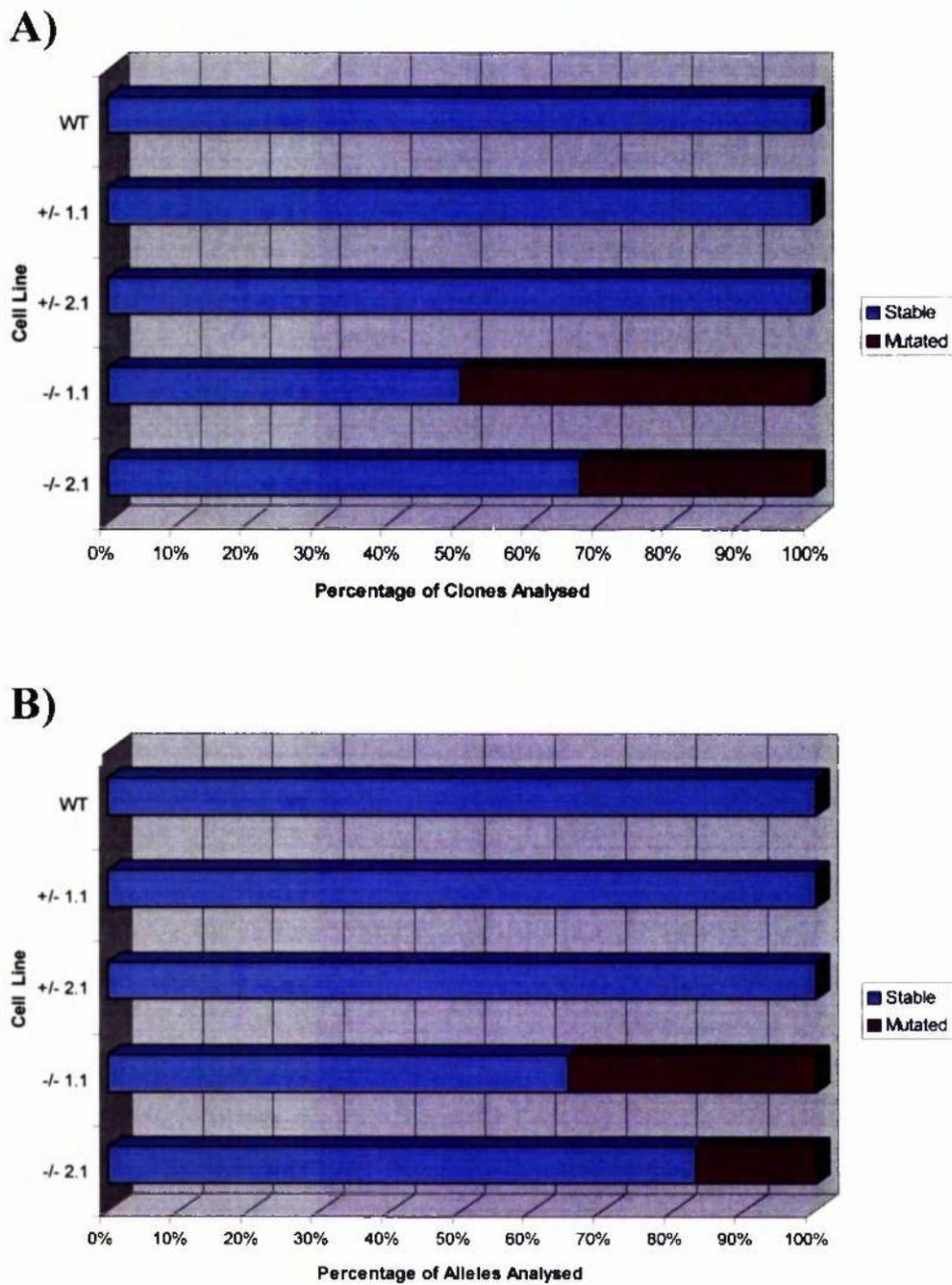
**Table 4.19** Synopsis of results for microsatellite ChrI-15 in 3174 *MLHI* knockout trypanosomes

The ChrI-15 microsatellite locus was analysed as described in Table 4.18. Ten clones (where possible) were analysed from each of the following 3174-derived cell lines: *MLHI* wild-type (WT); two heterozygous *MLHI* mutants (+/-); and two *MLHI* homozygous mutants (-/-). For each cell line, the number and percentage of clones and alleles which were stable or mutated at this locus is given.

Cell line	Smaller allele					Larger allele				
	No. analysed	Stable	Percentage	Mutated	Percentage	No. analysed	Stable	Percentage	Mutated	Percentage
WT	10	10	100	0	0	10	10	100	0	0
+/- 1.1	9	9	100	0	0	9	9	100	0	0
+/- 2.1	10	10	100	0	0	10	10	100	0	0
-/- 1.1	10	8	80	2	20	10	5	50	5	50
-/- 2.1	9	8	89	1	11	9	7	78	2	22
<b>Total alleles</b>										
Cell line	No. analysed	Stable	Percentage	Mutated	Percentage	No. analysed	Stable	Percentage	Mutated	Percentage
WT	20	20	100	0	0	10	10	100	0	0
+/- 1.1	18	18	100	0	0	9	9	100	0	0
+/- 2.1	20	20	100	0	0	10	10	100	0	0
-/- 1.1	20	13	65	7	35	10	5	50	5	50
-/- 2.1	18	15	83	3	17	9	6	67	3	33

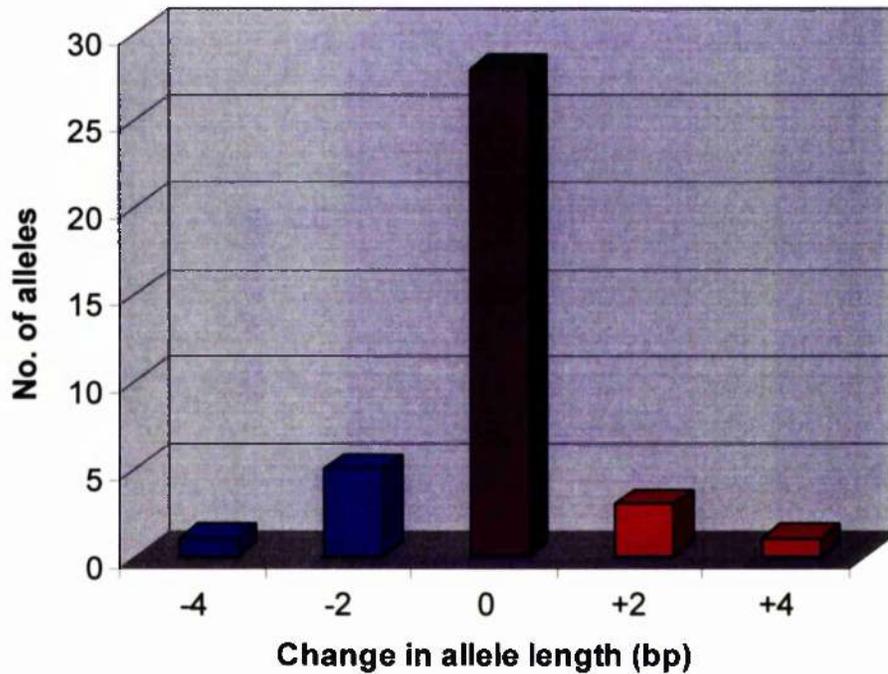
**Figure 4.36 Analysis of microsatellite ChrI-15 in *MLH1* knockout trypanosomes**

The percentage number of clones (A) and alleles (B) which were stable (shown in blue) and mutated (shown in red) at the ChrI-15 locus was determined as described in Table 4.19. (WT) 3174 *MLH1* wild-type trypanosomes; (+/-) 3174 *MLH1* heterozygous mutants; (-/-) 3174 *MLH1* homozygous mutants.



**Figure 4.37** Spectrum of mutations observed at microsatellite ChrI-15 in *MLH1* homozygous mutants

The ChrI-15 microsatellite locus was analysed as described in Table 4.18. The change in allele length from the stable allele length, indicated as either no change (shown in purple), addition (+; shown in red) or deletion (-; shown in blue), was calculated for each allele from 10 clones derived from the *MLH1*<sup>-/-</sup> 1.1 cell line, and 9 clones derived from the *MLH1*<sup>-/-</sup> 2.1 cell line.



**Table 4.20** Mutational spectrum of *MLH1*<sup>-/-</sup> mutants analysed by GeneScan in *T. brucei*

The microsatellite loci were amplified by PCR, using one fluorescently-labelled primer and an unmodified primer, and the PCR products separated by acrylamide-urea gel electrophoresis. The resulting fluorescent bands were detected using an automated sequencer and analysed using the GeneScan system. After comparing the size of alleles at a given locus in every sample, the most frequently observed size was determined to be the size of the wild-type or stable allele. Where a PCR product differed from this size by 1 bp or more it was considered to have mutated, and the change in length from the size of the stable allele was calculated. For each microsatellite locus the number of clones and alleles analysed is given, and the number of alleles which increased (+) or decreased (-) in length is shown. The total number of mutations of a given size at both loci has been calculated and expressed as a percentage of all the mutations observed.

Locus	Type of Repeat	No. of clones (alleles) analysed	Number of tracts with additions (+) or deletions (-) of base-pairs					
			-6	-4	-2	+1	+2	+4
ChrI-7	GT	20 (40)	1	0	6	1	2	0
ChrI-15	GT	19 (38)	0	1	5	0	3	1
<b>Number of Mutations (total number analysed)</b>			1/20	1/20	11/20	1/20	5/20	1/20
<b>Percentage of Mutations</b>			5	5	55	5	25	5

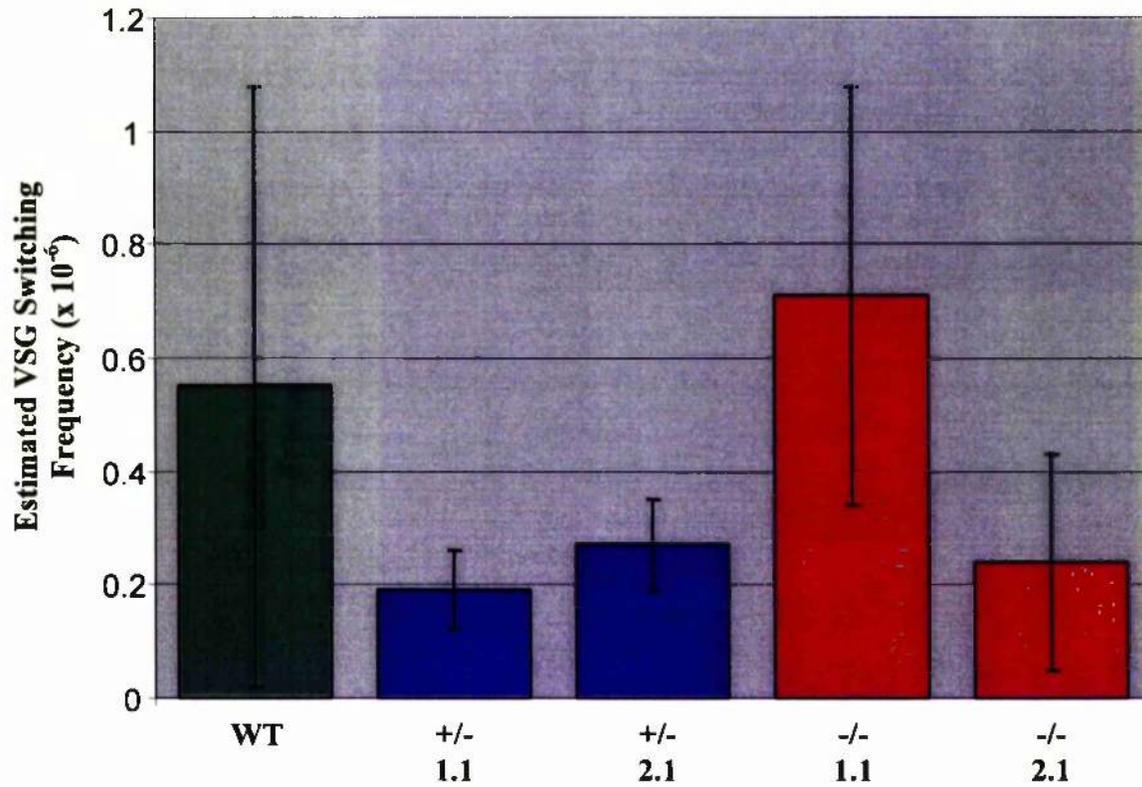
**Table 4.21 Effect of *MLH1* mutation on the frequency of VSG switching in the 3174 trypanosome strain**

The frequency of VSG switching was determined as described in Materials and Methods (see section 2.15) using 3174 bloodstream trypanosomes that had been unaltered at the *MLH1* locus (WT); that had one *MLH1* allele disrupted (+/-); or that had both *MLH1* alleles disrupted (-/-). In each independent experiment, switched trypanosomes were plated over 96-well plates in HMI-9 and the number of wells containing growth were counted after 10 days. The number of VSG switching events/cell/generation (estimated VSG switching frequency ( $\times 10^{-6}$ )) was determined as described in McCulloch *et al.* (1997).

Strain	No of wells growing	Estimated VSG switching frequency ( $\times 10^{-6}$ )
WT	64/96	1.00
	9/96	0.14
	3/96	0.05
	65/96	1.02
	Average 35/96	0.55
+/- 1.1	13/96	0.20
	7/96	0.11
	16/96	0.25
	Average 12/96	0.19
+/- 2.1	23/96	0.36
	14/96	0.22
	15/96	0.23
	Average 17/96	0.27
-/- 1.1	62/96	0.97
	28/96	0.44
	Average 45/96	0.71
-/- 2.1	3/96	0.05
	27/96	0.42
	16/96	0.25
	Average 15/96	0.24

**Figure 4.38** Effect of *MLH1* mutation on the frequency of VSG switching in the 3174 trypanosome strain

The average frequency of VSG switching was determined as described in Table 4.21 using 3174 bloodstream trypanosomes that had been unaltered at the *MLH1* locus (WT; shown in green); that had one *MLH1* allele disrupted (+/-; shown in blue); or that had both *MLH1* alleles disrupted (-/-; shown in red). Error bars depict standard deviation.



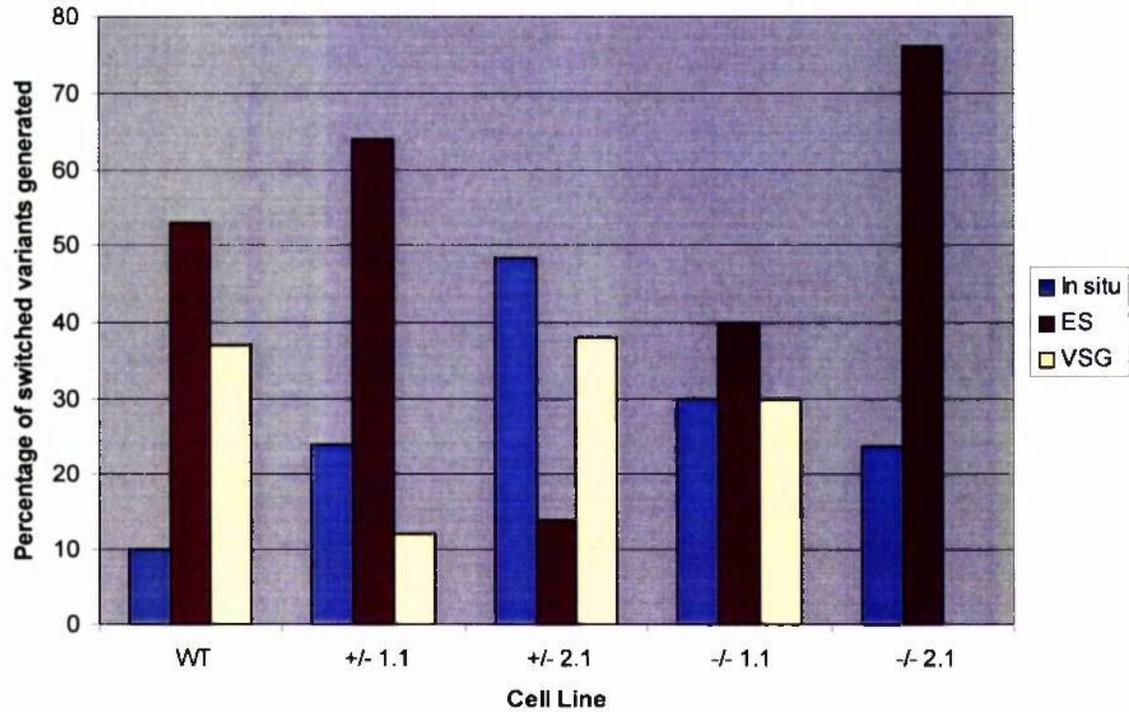
**Table 4.22** *VSG* switching mechanisms used by *MLH1* wild-type and *MLH1* mutant trypanosomes

The types of switching event used by untransformed 3174 bloodstream forms (WT); 3174 cells transformed with  $\Delta MLH1::PUR$  (+/-); and 3174 cells transformed with both  $\Delta MLH1::PUR$  and  $\Delta MLH1::BSD$  (-/-) was determined as described in Figure 4.22, Materials and Methods (see section 2.15) and McCulloch *et al.* (1997).

Strain (no. of switched variants analysed)	<i>In situ</i> switch	Expression site gene conversion	<i>VSG</i> gene conversion
<i>MLH1</i> wild-type (30)	3/30 10%	16/30 53%	11/30 37%
<i>MLH1</i> +/- 1.1 (25)	6/25 24%	16/25 64%	3/25 12%
<i>MLH1</i> +/- 2.1 (29)	14/29 48%	4/29 14%	11/29 38%
<i>MLH1</i> -/- 1.1 (20)	6/20 30%	8/20 40%	6/20 30%
<i>MLH1</i> -/- 2.1 (21)	5/21 24%	16/21 76%	0/21 0%

**Figure 4.39** *VSG* switching mechanisms used by *MLH1* knockout trypanosomes

The types of switching event used were determined as described in Table 4.22 for untransformed 3174 bloodstream forms (WT); 3174 cells transformed with  $\Delta MLH1::PUR$  (+/-); and 3174 cells transformed with both  $\Delta MLH1::PUR$  and  $\Delta MLH1::BSD$  (-/-). For each cell line the percentage of switched variants generated by *in situ* transcriptional switching (shown in blue), expression site gene conversion (shown in red), and *VSG* gene conversion (shown in yellow) are presented.



#### 4.4 Discussion

In order to confirm that the putative *MSH2* and *MLH1* genes encoded active components of a nuclear MMR system in *T. brucei*, a strategy to generate homozygous knockout mutants of each of these genes was developed. Since this study involved long-term functional assays, a well-established method of gene disruption was adopted, involving homology-directed gene targeting and stable integration of antibiotic resistance markers into the trypanosome genome (Ray and Hines, 1995; Li *et al.*, 1996; McCulloch and Barry, 1999). Homozygous knockout mutants were generated for each of the genes, indicating that it is possible to remove all intact transcripts of either *MSH2* or *MLH1*, and thus all expression of either the MSH2 or MLH1 proteins, and trypanosomes remain viable. However, to confirm that any phenotypes observed were due to the loss of MMR, and not to changes which arose during cloning and selection of the transformants, a copy of the *MSH2* ORF was re-integrated into one of the *MSH2* homozygous mutants. The equivalent re-expressor cell line was not generated for *MLH1* since the properties of *MLH1* mutant cell lines were investigated primarily to confirm the results obtained in *MSH2*-deficient trypanosomes. As predicted by studies of MMR in other eukaryotes, the phenotypes observed for *MLH1* mutants were essentially the same as those observed in *MSH2* knockout cell lines. As it seems highly unlikely that such similar phenotypes could arise in *MLH1* knockout trypanosomes as a consequence of the cloning and selection process used to generate the cell lines, it was deemed unnecessary to generate an *MLH1* re-expressor cell line.

Since neither gene is essential in other eukaryotes it was expected that both *MSH2* and *MLH1* knockout trypanosomes would be viable (Strand *et al.*, 1993; Durant *et al.*, 1999; Flores and Engels, 1999; Degtyareva *et al.*, 2002). Analysis of the *MSH2* mutant trypanosomes revealed no differences in growth compared with *MSH2* wild-type cells (Figure 4.7). This is similar to the results of *MSH2* insertion mutations in *S. cerevisiae*, where no obvious effects on cell growth were identified in diploid cells (Reenan and Kolodner, 1992a). Similarly, *MSH2*<sup>-/-</sup> mouse ES cells were indistinguishable from *MSH2* wild-type cells with respect to growth and plating efficiency, and *MSH2*<sup>-/-</sup> mice were healthy at birth and fertile, although they were predisposed to tumorigenesis (de Wind *et al.*, 1995). This study also revealed that *MLH1*-deficient trypanosomes showed no difference in growth in comparison with *MLH1* wild-type cells (Figure 4.30). Again, this bears comparison to other eukaryotes, since disruption of *S. cerevisiae MLH1* does not lead to growth rate changes in diploid cells (Prolla *et al.*, 1994a), and *MLH1*<sup>-/-</sup> mice are healthy at birth, but develop tumours at the same time of onset as *MSH2*<sup>-/-</sup> mutants (Baker *et al.*, 1996). *MLH1*<sup>-/-</sup> mutants reveal some differences compared to *MSH2*<sup>-/-</sup>, however.

Disruption of *S. cerevisiae* *MLH1* causes an increase in spore lethality and increased post-meiotic segregation (Prolla *et al.*, 1994a), and both female and male *MLH1*<sup>-/-</sup> mice are infertile (Baker *et al.*, 1996). Studies indicate that *MLH1*<sup>-/-</sup> spermatocytes arrest in meiosis I with prematurely separated chromosomes (Baker *et al.*, 1996; Edelman and Kucherlapati, 1996). While *T. brucei* possesses a non-obligatory sexual cycle which occurs between life cycle stages in the tsetse fly vector, it is unknown whether this process of genetic exchange involves meiosis, although segregation is Mendelian (Jenni *et al.*, 1986; Gibson and Stevens, 1999). Currently it would be technically challenging to determine whether *MLH1* is involved in genetic exchange in this organism, given the difficulties of generating knockout mutants in the pleomorphic cell lines used during mating experiments, and the low rates of *T. brucei* transmission through the tsetse fly vector. However, it is interesting to note that putative homologues of *MSH4* and *MSH5*, the components of the MutS-related heterodimer involved in meiosis, are also present in the *T. brucei* genome (data not shown), perhaps suggesting that meiosis is involved in the trypanosome sexual cycle, and that the roles of the MutS and MutL homologues associated with this process in other eukaryotes are conserved in this organism.

Loss of MMR in bacteria and some eukaryotes (see below) is associated with increased resistance to a number of cytotoxic drugs, including the alkylating agent, MNNG (Karran and Marinus, 1982; Branch *et al.*, 1993; Koi *et al.*, 1994; Hawn *et al.*, 1995). This study has revealed that inactivation of either *MSH2* or *MLH1* in *T. brucei* results in increased resistance to MNNG, and therefore the phenotype of methylation tolerance (Tables 4.1 and 4.14 and Figures 4.8 and 4.31). The results of these assays appear to suggest that inactivation of *MLH1* results in a greater increase in resistance to methylation than does inactivation of *MSH2*. As discussed in section 4.3.5, it seems likely that this is simply a consequence of the cells growing more poorly during the *MLH1* analysis that was performed later, probably due to altered growth conditions. Nevertheless, it may be interesting to repeat the MNNG survival assays for both the *MSH2* and *MLH1* mutants to determine if a difference in methylation tolerance does exist. In contrast to bacteria and mammalian cells, MMR-deficient yeast cells and fruit flies do not exhibit increased resistance to MNNG (Xiao *et al.*, 1995; Flores and Engels, 1999). It is not clear why the MMR systems of *S. cerevisiae* and *D. melanogaster* should differ in this respect, primarily due to a lack of complete understanding of the mechanism of methylation tolerance. However, this study reveals that the phenotype of methylation tolerance in eukaryotes is not limited solely to mammalian cells. Under normal cellular conditions in most eukaryotes, O<sup>6</sup>-meG lesions would be repaired by the highly conserved MGMT protein (described in section 4.2.6). This is also likely to be the case in trypanosomes, as BLAST

searching of the *T. brucei* sequencing databases revealed that a putative MGMT homologue is present in the genome (data not shown). Nothing is known about the expression levels of this gene in *T. brucei*, nor has it been functionally characterised, but it may be that insufficient levels of MGMT are available to repair the lesions formed after treatment with the concentrations of MNNG used here and thereby prevent cell death.

Another phenotype associated with the loss of MMR is microsatellite instability (described in section 4.2.7). In other organisms, MMR-deficient cells exhibit a mutator phenotype since mismatches arising in DNA during replication, or due to the actions of chemical mutagens, are not recognised and repaired (Levinson and Gutman, 1987; Strand *et al.*, 1993; de Wind *et al.*, 1995; Baker *et al.*, 1996). The mutation rate is elevated in such cells, and while mutations occur more frequently throughout the genome, the rate of mutation at microsatellite loci is more greatly increased, due to slippage of the DNA polymerases over repetitive sequences during replication (Strand *et al.*, 1993; Prolla *et al.*, 1994a; Strauss, 1999; Ellegren, 2000b). Since a number of microsatellite loci have been well characterised in *T. brucei* (Hope *et al.*, 1999; A. MacLeod, A. Tweedie, S. McLellan, C.M.R. Turner, and A. Tait, personal communication), an assay to investigate whether the frequency of mutation at such loci was elevated in MMR-deficient trypanosomes was utilised. Five microsatellite loci were investigated in the *MSH2*<sup>-/-</sup> mutants, and mutant alleles were detected at four of these loci (Figure 4.11 and Tables 4.3, 4.5, 4.7 and 4.9). Of these four loci, the three which showed the greatest frequency of mutation in *MSH2*<sup>-/-</sup> mutants were also analysed in the *MLH1*<sup>-/-</sup> mutants, and mutant alleles were observed in all three microsatellites (Figure 4.32 and Tables 4.16 and 4.18). The results of these assays clearly demonstrate that inactivation of either *MSH2* or *MLH1* results in the mutator phenotype, MSI. Since MSI results from a failure to repair replication errors (Claij and te Riele, 1999), this indicates that the MMR system is involved in the correction of replication slippage events in *T. brucei*.

In general, slightly fewer mutations were observed in the *MLH1*<sup>-/-</sup> mutants than in the *MSH2*<sup>-/-</sup> clones, but it is possible that these differences arose because the *MLH1*<sup>-/-</sup> mutant cell lines were cultured for a shorter period of time before cloning than the *MSH2*<sup>-/-</sup> mutants, allowing fewer generations to accumulate mutations. No rates of microsatellite instability can be inferred from these data as the number of generations the mutant cells were cultured for after their generation and prior to re-cloning was unknown, as was the sequence of the microsatellite loci in the parental clones. The assay was designed simply to demonstrate whether or not inactivation of MMR resulted in an MSI phenotype in *T. brucei*. However, the assay could be adapted to allow determination of the frequency of

microsatellite mutation in MMR-deficient trypanosomes. This would involve cloning of the cell lines prior to growth in culture for a given length of time. A sample of cells would be removed from an early passage and the genomic DNA isolated so that the microsatellite loci could be sequenced to determine the length of the parental or wild-type microsatellites. It would also be important to assay the growth rate of the cells so that the number of generations the cells had undergone since cloning could be estimated for the final genomic DNA samples. After the period in culture, the cells would be re-cloned to separate cells containing mutated microsatellite alleles from cells containing wild-type alleles, and genomic DNA would then be isolated from these cultures. GeneScan analysis of the microsatellite loci in the genomic DNA samples would reveal the number of clones containing mutated microsatellites. In conjunction with the estimate of the number of generations the cells had undergone since cloning, these data could be used to estimate the frequency of microsatellite mutation in MMR mutants. Moreover, if a large enough number of wild-type clones were analysed, it might also be possible to infer a mutational frequency for microsatellites in MMR-proficient *T. brucei* cells.

The first of the microsatellite loci investigated in both the *MSH2*<sup>-/-</sup> and *MLH1*<sup>-/-</sup> mutant cell lines, JS-2, was analysed by agarose gel electrophoresis. This revealed that greater changes in microsatellite length were apparent in *MSH2*<sup>-/-</sup> and *MLH1*<sup>-/-</sup> mutants than in cells expressing both these genes (Table 4.2 and Table 4.15). However, while this method is quick and simple, and can be used to differentiate relatively large changes in allele size only, it cannot be used to determine the length of alleles precisely. For this reason, a fluorescence-based PCR method, GeneScan analysis, was adopted.

GeneScan analysis allows the determination of the precise length of microsatellite alleles, and therefore mutations altering the length of these alleles can be accurately characterised. This study has revealed that in MMR-deficient cells, for all the microsatellites analysed, except ChrII-6 where no mutations were observed, there was a bias towards mutations which resulted in a decrease in microsatellite length. These events accounted for approximately two thirds of the 46 mutations analysed in the *MSH2*<sup>-/-</sup> and *MLH1*<sup>-/-</sup> mutant backgrounds (Tables 4.11 and 4.20). This is in contrast to the findings in other eukaryotes (see below; Wierdl *et al.*, 1997; Yamada *et al.*, 2002). Further analysis of the data gathered during this study shows that the majority of the mutational events appear to involve the addition or deletion of one repeat unit. Conversely, this is in agreement with the findings in yeast and mammals (Henderson and Petes, 1992; Weber and Wong, 1993; Wierdl *et al.*, 1997; Twerdi *et al.*, 1999). In yeast and mammalian cells there is a tendency for mutations to increase the length of a microsatellite, until the allele reaches a threshold size, after

which deletions become more common, thereby preventing infinite growth of the microsatellite sequence (Wierdl *et al.*, 1997). Similarly, a slight bias towards increasing microsatellite length has been reported for *D. melanogaster* (Flores and Engels, 1999). A probable explanation for this discrepancy lies in the size of the microsatellite alleles analysed in this study. In yeast it has been shown that GT-dinucleotide repeat tracts 51 bp (25 repeat units) or longer in length exhibit significantly more large deletions than similar tracts 33 bp (16 repeat units) or less in length (Wierdl *et al.*, 1997). Similar results have also been observed in human cells (Yamada *et al.*, 2002). Since the smallest allele containing GT-dinucleotide repeats analysed in this study contained 18 repeat units (the ChrI-7 smaller allele), it is probable that the mutational bias observed towards decreasing microsatellite length in this study of *MSH2*<sup>-/-</sup> and *MLH1*<sup>-/-</sup> mutant trypanosomes has arisen due to the large size of the microsatellite alleles analysed. If a mutational bias towards decreasing microsatellite length is really active in *T. brucei*, it would be difficult to explain how this organism could maintain its large array of microsatellite loci. In addition, as no mutations were observed in any of the microsatellites in MMR-proficient trypanosomes, it is impossible to determine if the mutations observed in the *MSH2*<sup>-/-</sup> and *MLH1*<sup>-/-</sup> mutants reflect the mutational dynamics of these loci in wild-type *T. brucei*.

Comparison of the GeneScan analysis results from both the *MSH2*<sup>-/-</sup> and *MLH1*<sup>-/-</sup> mutants, indicates that the frequency of replication slippage increases with the number of repeat units encoded within the microsatellite allele. This phenomenon is also observed in yeast and human microsatellites, where the number of mutations observed increases as the length of the allele increases (Wierdl *et al.*, 1997; Ellegren, 2000a; Yamada *et al.*, 2002). Although the ChrI-7 and ChrI-15 microsatellite loci are not directly comparable as they are found in distinct genomic locations, they both contain GT-dinucleotide repeats, and so the mutational pressures on these loci may be similar. Sequencing has shown that in stable clones the smaller allele of ChrI-7 contains 18 perfect repeats, while the larger allele comprises 70 imperfect repeat units (Figure 4.14). At least 50% of *MSH2*<sup>-/-</sup> and *MLH1*<sup>-/-</sup> clones were mutated at this locus: 8% of the clones analysed were mutated at the smaller allele, but the larger allele was mutated in 49% of the clones analysed (see Tables 4.4 and 4.17). For ChrI-15, the stable allele sizes differ by only 8 bp, and the smaller allele contains 30 perfect repeats, which is intermediate between the two alleles of ChrI-7 (Figure 4.14). (The larger allele of ChrI-15 was omitted from this analysis because its sequence is unknown, although it probably contains 34 GT-dinucleotide repeats.) Again, at least 50% of *MSH2*<sup>-/-</sup> and *MLH1*<sup>-/-</sup> clones were mutated at this locus, and overall, the smaller ChrI-15 allele was mutated in 26% of the clones analysed (see Tables 4.6 and 4.19). It should also be noted that studies in yeast have revealed that interruption of the microsatellite by variant

repeats, such as the imperfections seen in the ChrI-7 larger allele, have a strong stabilising effect on the microsatellite (Petes *et al.*, 1997). It is therefore possible that the mutation rate observed for the ChrI-7 larger allele may not be as high as it would have been if the microsatellite sequence was not interrupted.

In *E. coli*, artificial GC- and AT-dinucleotide repeats have been shown to mutate at a higher frequency than other, non-self-complementary dinucleotide repeats (Bichara *et al.*, 2000). It is also possible that different repeat sequences exhibit different mutation rates in *T. brucei*. Unfortunately, although microsatellites composed of repeat units with different sequences were included in the analysis performed using the *MSH2*<sup>-/-</sup> mutants, very little can be inferred about the effect of repeat unit sequence on microsatellite stability, because the data set generated during this study is very small. Both the ChrI-7 and ChrI-15 loci are composed of GT-dinucleotide repeats, while PLC contains TA-dinucleotide repeats, and ChrII-6 consists of CT-dinucleotide repeats (Figures 4.14 and 4.15). On average, the PLC locus contained mutations in 15% of the *MSH2*<sup>-/-</sup> clones (Table 4.8). By comparison, an average of only 5% of the same clones contained mutations in the smaller allele of the ChrI-7 locus (Table 4.4). Although these alleles are not directly comparable, since the smaller allele of ChrI-7 contains 18 perfect repeat units, while the PLC allele only contains 12 perfect repeat units, the greater frequency of mutation of the PLC allele may indicate that TA-dinucleotide repeats are more unstable in *T. brucei* than GT-dinucleotide repeats. No mutations were observed for the ChrII-6 microsatellite locus in any of the clones analysed, suggesting that the mutation rate at this locus is very low. This is not surprising given the small size and imperfect sequence of this microsatellite in 3174 trypanosomes. In order to assess the effect of repeat unit sequence on the frequency of microsatellite mutation in *T. brucei*, it would probably be necessary to develop a range of size-matched artificial microsatellite alleles which could be integrated into the genome. By targeting the alleles to different regions of the genome it would be possible not only to assess the effect of base composition on microsatellite stability but also the influence of genomic context.

Finally, an assay to determine the frequency of, and mechanisms used during, antigenic variation was employed to determine whether MMR was involved in the regulation of *VSG* switching in *T. brucei*. The frequency of *VSG* switching was similar in wild-type trypanosomes, *MSH2*<sup>+/-</sup> cells, *MSH2*<sup>-/-</sup> mutants and *MSH2*<sup>-/-/+</sup> cells (Table 4.12 and Figure 4.23), and there was no clearly detectable change in the relative levels of transcriptional versus recombinational *VSG* switching mechanisms between the cell lines (Table 4.13 and Figure 4.24), indicating that the product of the *MSH2* gene has little or no effect on antigenic variation. Extrapolation from data generated in other eukaryotes suggested that

*MSH2* may be involved in the inhibition of recombination between divergent sequences in *T. brucei* (de Wind *et al.*, 1995; Datta *et al.*, 1996; Nicholson *et al.*, 2000), and thus might be expected to down-regulate recombinational *VSG* switching. It was therefore hypothesised that inactivation of *MSH2* might increase the frequency of recombinational *VSG* switching. In contrast, other studies in eukaryotes have revealed that *MSH2* may also be required to facilitate resolution of recombination intermediates (Sugawara *et al.*, 1997; Marsischky *et al.*, 1999), and thus loss of *MSH2* might also be expected to decrease the frequency of recombination because the protein would not be available to perform these functions. If this were the case in *T. brucei*, loss of *MSH2* might not result in a detectable change in *VSG* switching, as the protein would perform both an active and suppressive role. However, this possibility was excluded by the data derived by performing the assay using *MLH1*<sup>+/-</sup> cells and *MLH1*<sup>-/-</sup> mutants. Again, all these cell lines underwent *VSG* switching at a similar frequency to the wild-type (Table 4.21 and Figure 4.38) and there was no detectable change in the mechanisms used during antigenic variation (Table 4.22 and Figure 4.39). *MLH1* has not been implicated in the resolution of recombination intermediates, but is known to be involved in the inhibition of homeologous recombination in other eukaryotes (Nicholson *et al.*, 2000). In *T. brucei*, *MLH1* would be predicted to suppress recombinational *VSG* switching, and so the frequency of expression site gene conversion and *VSG* gene conversion would be expected to increase in *MLH1*-deficient trypanosomes. Since this is not the case, and the results of the *VSG* switching assay indicate that neither *MSH2* or *MLH1* affect either the frequency of, or mechanisms used during, *VSG* switching in trypanosomes, it is probable that MMR exerts little or no influence over antigenic variation in *T. brucei*.

It has been hypothesised that the monomorphic cell line MITat 1.2a S427 used during this analysis is deficient in an active mechanism which promotes recombinational *VSG* switching in pleomorphic cell lines (Barry, 1997b; Robinson *et al.*, 1999). An enzymatic activity that promotes homologous recombination has been invoked to explain the discrepancy in *VSG* switching rates between monomorphic and pleomorphic trypanosomes strains, which differ by several orders of magnitude (Lamont *et al.*, 1986; Turner and Barry, 1989; Turner, 1997). However, in bacteria, loss of MMR can lead to a 1000-fold increase in recombination between divergent DNA sequences, allowing inter-species recombination to occur (Rayssiguier *et al.*, 1989). Therefore, it was possible, if highly unlikely, that pleomorphic trypanosomes achieved their high rates of *VSG* switching via down-regulation or inactivation of MMR. Since inactivation of *MSH2* or *MLH1* caused no substantial alterations to either the frequency of, or mechanisms used during, antigenic

variation in the monomorphic line used in this study, MMR cannot account for the huge differences in VSG switching rates between monomorphic and pleomorphic trypanosomes.

The VSG switching assay used in this study relies upon *in vivo* selection of switched trypanosomes, and it is necessary to make a number of assumptions about the data in order to calculate VSG switching frequencies. This reduces the sensitivity and reproducibility of the assay, perhaps to the point where subtle changes induced by inactivation of genes such as *MSH2* and *MLH1* cannot be detected. It is possible, therefore, that MMR may play a small regulatory role in antigenic variation, but that the changes caused by inactivation of *MSH2* or *MLH1* are beneath the threshold which can be detected by the approach adopted. Furthermore, this assay is not able to detect changes in the order of VSG gene expression, since it looks only at one relapse peak, rather than a long-term infection. It has been demonstrated that different VSGs are activated at different points during the course of an infection (Gray, 1965; Capbern *et al.*, 1977; Barry, 1986a). Some VSGs are activated early in infection, some are expressed later during infection, and others are expressed towards the end of infection. It is possible that the probability of a given VSG being activated by recombinational switching is dependent on the level of homology between its flanking sequences, and the sequence flanking the active VSG. Given that MMR is known to regulate recombination between non-identical DNA sequences (Shen and Huang, 1989; de Wind *et al.*, 1995; Datta *et al.*, 1996), it is possible that MMR might regulate recombinational switching, and thus the hierarchy of VSG expression in *T. brucei*.

The functions of a number of other genes in trypanosome antigenic variation have now been assessed using the same VSG switching assay. The first knockout mutants analysed in this way were inactivated for the *RAD51* gene, which encodes a primary component of the homologous recombination machinery in trypanosomes (McCulloch and Barry, 1999). *RAD51*<sup>-</sup> mutant cell lines displayed a reduced frequency of antigenic variation and were impaired in their ability to undergo both transcriptional and recombinational VSG switching, indicating that the gene was either involved in the regulation or catalysis of these processes. However, VSG switching was not completely ablated in *RAD51*<sup>-</sup> cell lines and switched variants were isolated which had arisen by both the transcriptional and recombinational VSG switching mechanisms in these cells. More recently, the roles of the *KU70* and *MRE11* genes have been investigated using the VSG switching assay.

Extrapolation from other eukaryotes suggests that *KU70* encodes a component of the NHEJ machinery, catalysing a double-strand break repair pathway that competes with homologous recombination in other eukaryotes (Pâques and Haber, 1999; Cromie *et al.*, 2001). While any potential contribution of *KU70* to DNA recombination has not been

directly investigated in *T. brucei*, Conway *et al.* (2002a) have shown that inactivation of this gene has no effect on antigenic variation. This is in keeping with the supposition that antigenic variation occurs by homologous recombination. Surprisingly, however, inactivation of *T. brucei* *MRE11* also had no discernible effect on VSG switching (Robinson *et al.*, 2002). In other eukaryotes, MRE11 forms a complex with RAD50, and either XRS2 in yeast, or NBS1 in mammals. This complex is thought to be involved in the nucleolytic processing of DSBs, and DNA damage checkpoint signalling, and is also believed to play a structural role in the alignment of DNA ends in both homologous recombination and NHEJ (D'Amours and Jackson, 2002; Hopfner *et al.*, 2002). The fact that *MRE11* mutation does not affect antigenic variation may suggest that it occurs via a pathway of homologous recombination that is different to general DNA damage repair pathways. Recently, analysis of general recombination in *RAD51*<sup>-/-</sup> cells has revealed that trypanosomes possess at least two, and perhaps three, pathways of homologous recombination (Conway *et al.*, 2002b). It is also known that at least two of these pathways can promote VSG switching, in either a RAD51-dependent or -independent manner (McCulloch and Barry, 1999). It is clear that we are far from fully understanding the controls and mechanisms that underlie VSG switching, and it remains possible that the process involves a distinct RAD51-dependent pathway which does not require MRE11, and which is not influenced by MMR. This hypothesis may be supported by the recent finding that VSG switching events, in which the upstream limit of gene conversion maps to the 70-bp repeats, can utilise as little as 40 bp of homology between the donor repeat sequence and the 70-bp repeats within the recipient expression site (P. Burton and J. D. Barry, personal communication). It is possible that most, if not all, VSG switching events utilise small regions of homology, thereby avoiding MMR surveillance and inhibition. It would be valuable, therefore, to determine the role of MMR and sequence divergence in the regulation of recombination in *T. brucei*. To this end, an assay has been developed to investigate both homologous and homeologous recombination in wild-type and MMR-deficient trypanosomes, and is described in Chapter 5 of this work. Studies to date have suggested that transcriptional switching is at least as frequently used as recombinational switching during antigenic variation in monomorphic trypanosomes (Liu *et al.*, 1985; McCulloch and Barry, 1999), while recombinational VSG switching appears to predominate in the fast-switching pleomorphic cell lines (Robinson *et al.*, 1999). It is likely, therefore, that a full understanding of the role of MMR in VSG switching will not be elucidated until techniques for investigating antigenic variation in pleomorphic cell lines are readily available.

# CHAPTER 5

**An assay to determine the influence of  
mismatch repair on trypanosome  
homeologous recombination**

## 5.1 Introduction

Recombination between homeologous (divergent) DNA sequences occurs much less efficiently than recombination between homologous (identical) sequences in such diverse organisms as bacteria, yeast and mammals (reviewed in Evans and Alani, 2000). During heteroduplex formation between two independent but complementary single-stranded DNA molecules, mismatched base-pairs form where the single-strands differ in sequence. These mispairs seem to be subject to recognition by components of the MMR system since the frequency of recombination between divergent sequences is often greatly elevated in MMR-deficient cell lines (Shen and Huang, 1989; Alani *et al.*, 1994; de Wind *et al.*, 1995; Selva *et al.*, 1995; Datta *et al.*, 1996; Datta *et al.*, 1997). This evidence suggests that MMR forms the basis of an active anti-recombination activity within the cell. Indeed, it has been reported that even a single mismatch is sufficient to inhibit recombination in *S. cerevisiae* (Datta *et al.*, 1997). Studies in yeast have shown that *MSH2* mutants exhibit the greatest elevation in homeologous recombination frequencies, compared with wild-type strains, suggesting that this protein possesses the most potent anti-recombination activity (Nicholson *et al.*, 2000). Similar elevations in homeologous recombination frequency were observed with *MSH3 MSH6* double mutants suggesting that these proteins act in concert with *MSH2*, forming heterodimers with similar roles to those performed in MMR. However, analysis of *MLH1* and *PMS1* single mutants and *MLH1 PMS1* double mutants indicated that although these proteins appear to act as a heterodimer, as in MMR, they are involved in repressing only a portion of the homeologous recombination events inhibited by *MSH2*, since none of these mutants show the same elevation in recombination as *MSH2* mutants (Datta *et al.*, 1996; Nicholson *et al.*, 2000). This suggests that the MutS-related proteins can block some homeologous recombination events without the assistance of the *MLH1-PMS1* heterodimer, indicating that the mechanism underlying this anti-recombination activity has distinct characteristics compared with the mechanism of MMR.

It is clear that the combined results from *MSH2*- and *MLH1*-deficient trypanosomes confirm the existence of an active MMR system in *T. brucei*, and indicate that *MSH2* and *MLH1* represent genuine homologues of *mutS* and *mutL* respectively. As in other eukaryotes, inactivation of *MSH2* or *MLH1* results in both methylation tolerance and microsatellite instability, two phenotypes which are commonly detected in MMR-deficient cells, and can be considered diagnostic when investigating potential components of the MMR system. Furthermore, the observation that MMR exerts little or no influence on antigenic variation in this organism indicates that *VSG* switching may rely on a distinct recombination mechanism, which perhaps relies on only short stretches of sequence

homology, or which can successfully bypass sequence heterology. In this chapter, an assay will be described whose use will help elucidate the effect of sequence divergence on the frequency of recombination in both MMR-proficient and MMR-deficient trypanosomes.

## **5.2 Design of an assay to determine the frequency of integration of homologous and homeologous constructs into MMR-deficient trypanosomes**

In order to assess the impact of MMR on homologous and homeologous recombination in *T. brucei*, an assay was developed to determine the frequency of integration of increasingly divergent targeting constructs into the trypanosome genome. For explanatory purposes the assay system can be considered as three separate components: a *T. brucei* cell line containing an unique target integration site; a set of MMR mutant trypanosomes containing this integration site; and a set of targeting constructs that possess integration flanks for homologous recombination based on the sequence of this integration site, but which contain increasing numbers of heterologies compared with this sequence. Figure 5.1 shows an overview of this experimental approach.

It was necessary that the sequence chosen as the integration site fulfilled a number of criteria. Firstly, it was important that the exact sequence of the locus chosen as the integration site was known so that the level of divergence of the targeting constructs could be accurately calculated. It was also necessary that this sequence be unique in the genome, otherwise the targeting constructs might be able to integrate into related sequences. Furthermore, it would be interesting to determine the relative frequencies of integration, and therefore recombination, into different genomic locations, including chromosome-internal and telomeric sites, using the same targeting constructs so that the results would be directly comparable. Given these requirements, no native *T. brucei* sequence seemed adequate. Instead, the one type of targeting sequence which fulfilled all these requirements was a bacterial antibiotic resistance gene integrated into the *T. brucei* genome. Such an antibiotic resistance gene should not only have no natural *T. brucei* related sequences, but could be integrated anywhere within the genome by modification of associated targeting and processing flanks, and could be targeted by a single set of increasingly divergent homeologous constructs. For this reason, the hygromycin phosphotransferase (*HYG*) ORF was chosen as the integration site.

For this assay, the tubulin array was chosen as the chromosome-internal location into which the *HYG* ORF would be inserted, since a construct targeting the *HYG* gene to this locus was already available (Figure 5.1A and Figure 5.2). The plasmid pHygro-Tub

contains the *HYG* ORF flanked at the 5' end by 230 bp of  $\alpha$ - $\beta$  tubulin intergenic sequence and at the 3' end by 300 bp of  $\beta$ - $\alpha$  tubulin intergenic sequence. These intergenic sequences act as targeting flanks for the replacement of an  $\alpha$  tubulin gene, and also provide splicing signals for the *HYG* ORF allowing transcripts of the gene to be processed into mature mRNAs. The *IHYG* gene is expressed by endogenous tubulin transcription since the targeting construct contains no promoter elements. Transformation of MITat 1.2a S427 bloodstream form trypanosomes with the linearised pHygro-Tub construct would generate the HTUB cell line (see section 5.3).

Having generated a unique targeting site, the assay required that MMR mutants be generated in this HTUB background. Since extrapolation from other eukaryotes suggested that loss of *MSH2* should result in greater increases in the frequency of homologous and homeologous recombination than loss of *MLH1* (Nicholson *et al.*, 2000), it was decided that inactivation of *MSH2* would afford the best opportunity to assess the usefulness of this assay (Figure 5.1B). As described in section 4.2, knockout constructs have been developed for the deletion of both copies of the *MSH2* ORF in *T. brucei*. The HTUB cell line was therefore transformed with the *AMSH2::PUR* and *AMSH2::BSD* constructs to generate two independent *MSH2* knockout cell lines (see section 5.4).

The *HYG* ORF was used as parental sequence to generate targeting flanks for a number of antibiotic resistance constructs aimed at the *HYG* integration site in the trypanosome genome (Figure 5.1C and Figure 5.3). The flanks differed in each construct in that they were either perfectly homologous to the *HYG* ORF in the genome, or contained increasing numbers of base changes compared with the wild-type *HYG* gene sequence. In each construct, the antibiotic resistance cassette used contained the bleomycin resistance protein ORF (*BLE*) flanked at the 5' end by intergenic sequence derived from the actin locus and at the 3' end by intergenic sequence from the calmodulin locus. These sequences provided splicing and polyadenylation signals for the *BLE* ORF allowing transcripts of the gene to be processed into mature mRNAs, and since the constructs do not contain promoter elements the *BLE* gene is expressed by endogenous tubulin transcription. The actin and calmodulin loci are unlinked in the *T. brucei* genome (see The Wellcome Trust Sanger Institute *T. brucei* genome sequencing database; [http://www.sanger.ac.uk/Projects/T\\_brucei/](http://www.sanger.ac.uk/Projects/T_brucei/)), so homologous integration can only occur via the *HYG* targeting flanks. To allow this construction, the *HYG* targeting sequences were amplified by PCR around an *NdeI* restriction site within the *HYG* ORF. *AscI* restriction sites were introduced at the 5' and 3' termini of each amplified PCR product to allow linearisation of the final constructs

and to ensure the termini of all the constructs were equivalent. After cloning the *HYG* products into the pCR2.1-TOPO vector, the *BLE* cassette was introduced into the *NdeI* site. The generation of these targeting constructs is described fully in section 5.5.

### 5.3 Generation of the transgenic HTUB strain of *T. brucei*

The HTUB transgenic cell line was generated by direct transformation of MITat 1.2a S427 bloodstream form trypanosomes with the linearised pHygro-Tub construct, and the transformants were selected in HMI-9 containing  $5 \mu\text{g}\cdot\text{ml}^{-1}$  hygromycin B in 24-well plates (Figure 5.2; see section 2.17.2). Correct integration of the construct into the tubulin locus was determined by Southern analysis using the 914 bp PCR product shown in Figure 5.4 as a probe. This PCR product was amplified using Herculase polymerase with the primers *HYGFor* and *HYGRev* and the pHygro-Tub plasmid as a template (see section 2.17.1). Genomic DNA isolated from each transformant was digested with *KpnI* or *EcoRI*, separated by agarose gel electrophoresis, Southern blotted and probed with the 914 bp probe. Figure 5.5 shows the results of this analysis for the MITat 1.2a S427 cell line and a putative HTUB transformant. As predicted, no fragments were detected by the probe in the digestions of MITat 1.2a S427 genomic DNA as this cell line does not contain a copy of the *HYG* ORF. For the HTUB transformant, the probe hybridised to fragments which appeared smaller than the expected sizes for both restriction digestions: a *KpnI* fragment approximately 2.5 kb in length was detected, while two *EcoRI* fragments, the larger of which was approximately 1.7 kb in length, and the smaller approximately 0.6 kb in length, were revealed. The restriction map of the tubulin locus and the pHygro-Tub sequence suggests that the *KpnI* fragment should be 2.9 kb in length, and the *EcoRI* fragments 2.1 kb and 1.0 kb in length. However, these differences in size are probably a consequence of the agarose gel electrophoresis conditions used, as the gel shown in Figure 5.5 was run rapidly at 100 kV for 3 h, rather than at 20 kV for 18 h which would normally be used for a gel intended for Southern analysis. Since the fragments show the correct pattern of hybridisation, it is probable that the size differences are simply the result of a poor quality gel, and that the linearised pHygro-Tub construct has integrated into the genome correctly to generate the desired HTUB transformant. From this analysis it was impossible to determine whether only a single copy of the *HYG* ORF had integrated into the tubulin locus of the HTUB cell line, as more than one copy would give the same restriction pattern. However, given that all the cell lines used in this analysis were derived from these transgenic HTUB trypanosomes, the results of this assay were comparable.

#### 5.4 Generation of *MSH2* mutants in the HTUB trypanosome strain

Two independent *MSH2* mutants were generated by transformation of the HTUB transgenic cell line with the *MSH2* knockout constructs  $\Delta MSN2::PUR$  and  $\Delta MSN2::BSD$  described in section 4.2.2. Initially, the linearised  $\Delta MSN2::PUR$  construct was electroporated into two independent cell populations and the transformants selected in HMI-9 containing  $1 \mu\text{g}\cdot\text{ml}^{-1}$  puromycin dihydrochloride in 24-well plates to generate putative *MSH2* heterozygous ( $MSH2^{+/-}$ ) knockout cell lines (see section 2.1.3). Two independent first round transformants were then electroporated with the linearised  $\Delta MSN2::BSD$  construct and selected in HMI-9 containing  $2.5 \mu\text{g}\cdot\text{ml}^{-1}$  blasticidin S hydrochloride in 24-well plates to generate putative *MSH2* homozygous ( $MSH2^{-/-}$ ) mutants. Since the results described in Chapter 4 (section 4.2) illustrate that the phenotypes demonstrated by trypanosomes transformed with the  $\Delta MSN2::PUR$  and  $\Delta MSN2::BSD$  constructs arise due to loss of *MSH2* and not as a consequence of incidental changes associated with the generation of such transformants, it was considered unnecessary to generate an *MSH2* re-expressing ( $MSH2^{-/+}$ ) mutant in the HTUB background. Furthermore, integration of the  $\Delta MSN2::MSH2-BLE$  construct results in resistance to phleomycin, and selection of transformants which have integrated the divergent targeting constructs generated for this assay also relies on phleomycin resistance (see section 5.6), making it impossible to determine the frequency of integration of the targeting constructs in  $MSH2^{-/+}$  cells.

Correct integration of the constructs into the *MSH2* locus of the HTUB cell line was determined by Southern analysis (see section 2.11.1). Genomic DNA from untransformed HTUB trypanosomes, two independent putative  $MSH2^{+/-}$  mutants and two independent putative  $MSH2^{-/-}$  mutants was digested with either *SacI* or *KpnI*, separated by agarose gel electrophoresis and Southern blotted. The Southern blot of the *SacI* genomic digestions was probed with the *MSH2* 5' flank described in section 4.2.2, and for each transformant the probe hybridised to fragments of the expected size (Figure 5.6). A 5.8 kb fragment corresponding to the intact *MSH2* locus was detected in the *MSH2* wild-type cell line and the two putative *MSH2* heterozygous mutants, but was absent from the two putative *MSN2* homozygous mutants. Integration of the  $\Delta MSN2::PUR$  construct was indicated by the detection of a 2.85 kb fragment, which was present in all the  $MSH2^{+/-}$  and  $MSH2^{-/-}$  mutant cell lines. Finally, the probe hybridised to a 2.65 kb fragment corresponding to the  $\Delta MSN2::BSD$  allele in the *MSH2* homozygous mutants. For each cell line, Southern analysis indicated that the *MSH2* knockout constructs had integrated into the HTUB cell

line as expected to generate the desired *MSH2* mutant trypanosomes. In order to verify that the *HYG* ORF was present in all the putative *MSH2* mutants generated in the HTUB background, the Southern blot of the *Kpn*I digestions was probed with the 914 bp *HYG* probe described in section 5.3 (Figure 5.4). Figure 5.7 shows that the probe hybridises to a 2.9 kb fragment in each cell line, indicating that the *HYG* ORF is present in each transformant. Furthermore, this 2.9 kb fragment corresponds to the 2.5kb fragment detected in the HTUB transformant shown in Figure 5.5, confirming that the *HYG* ORF did indeed integrate as expected into the tubulin locus in this cell line.

In order to show that no copies of the *MSH2* ORF existed within the putative *MSH2* homozygous mutants, PCR using primers internal to the *MSH2* ORF (*MSH2D5* and *MSH2U2*; Figure 4.1) was performed using genomic DNA isolated from the putative HTUB *MSH2* mutant cell lines (Figure 5.8; see section 2.11.1). The integrity of the DNA template was confirmed using primers specific to the large subunit of *T. brucei* RNA polymerase I (PolI 5' and PolI 3'; Rudenko *et al.*, 1996). PCR amplification using the *MSH2* internal primers indicated that the *MSH2* ORF was present in the *MSH2* wild-type cell line, and the two putative *MSH2*<sup>+/-</sup> mutants. However, no PCR product could be identified for the two putative *MSH2*<sup>-/-</sup> mutants confirming the results of the Southern analysis and indicating that both copies of the *MSH2* ORF had been removed during the integration of the  $\Delta$ *MSH2*::*PUR* and  $\Delta$ *MSH2*::*BSD* constructs into these cell lines.

In order to confirm that no *MSH2* mRNA was being transcribed in the putative HTUB *MSH2* homozygous mutants, total RNA was isolated from HTUB *MSH2* wild-type cells, the two putative *MSH2*<sup>+/-</sup> cell lines and the two putative *MSH2*<sup>-/-</sup> mutants (see section 2.4). cDNA was generated from these samples by reverse transcription, and used as a template for PCR amplification using the *MSH2* internal primers described above (Figure 5.9; see section 2.11.1). As predicted, a PCR product of the expected size was generated from the untransformed HTUB trypanosomes and the *MSH2* heterozygous mutants since these cells all possess intact copies of the *MSH2* locus. However, no PCR product could be discerned for the putative *MSH2* homozygous mutants, confirming that no copies of the *MSH2* ORF were present in these cells. Amplification using primers specific to the large subunit of RNA polymerase I (PolI 5' and PolI 3'; Rudenko *et al.*, 1996) confirmed that the cDNA was intact for each transformant. Amplification with these primers using the samples generated from the putative *MSH2*<sup>-/-</sup> 2.1 cell line gave a product in both the RT positive and RT negative reactions. The product in the RT negative reaction probably arose because of incomplete DNase I treatment leaving some intact genomic DNA present in the sample. Since the PCR product seen in the RT positive reaction is much more abundant it

can be assumed that it was generated from cDNA derived from *MSH2* mRNA present in the total RNA sample. These results confirm that the putative *MSH2* knockout cell lines developed in the transgenic HTUB strain of trypanosomes are genuine *MSH2* mutants and that the *MSH2* knockout constructs integrated into the transformants as expected.

### 5.5 Generation of homologous and homeologous *HYG* targeting constructs

As outlined in section 5.2, the *HYG* ORF was used as the parental sequence to generate a set of targeting constructs possessing increasingly divergent integration flanks aimed at the *HYG* gene integrated into the HTUB cell line (Figure 5.3). *HYG* sequence for use as targeting flanks was amplified from the *HYG* ORF present in the plasmid pHygro-Tub, using a variety of PCR conditions (see 2.17.1). In every case, the primers *HYG*For and *HYG*Rev were used, both of which contain an *AscI* restriction site to allow linearisation of the final constructs prior to transformation into trypanosomes. To generate the perfectly homologous wild-type *HYG* flanks, PCR amplification was performed using high-fidelity Herculase polymerase (Figure 5.4). The resulting PCR product was cloned to give the plasmid p*HYGWT*, and sequencing revealed it to be 914 bp in length. Apart from the terminal 10 bp regions which include the *AscI* sites introduced by the primers, the sequence of p*HYGWT* was identical to the sequence of the *HYG* ORF in pHygro-Tub (see Appendix 6). This sequence was therefore used as the 0% divergent targeting flanks. Furthermore, the Herculase polymerase PCR product shown in Figure 5.4 was also used a probe for the *HYG* ORF during Southern analysis (see sections 5.3 and 5.4).

To generate *HYG* flanks containing base changes compared to the wild-type *HYG* sequence, a random mutagenesis approach was used (see section 2.17.1). This required two rounds of PCR amplification with *Taq* polymerase and the *HYG*For and *HYG*Rev primers using the pHygro-Tub plasmid as a template (Figure 5.10). The first round PCR reaction was supplemented with two nucleotide analogues: dPMP, which is incorporated in place of dTTP and to a lesser extent dCIP; and 8-oxo-dGTP which is incorporated in place of dTTP (Zaccolo *et al.*, 1996). At the end of cycles 5, 10, 15, 20, 25 and 30, a 1 µl sample of the reaction mix was removed and used as the template for a second, standard *Taq* PCR amplification reaction. The second reaction was required to remove the mismatches generated by incorporation of the nucleotide analogues, thereby completing the mutagenesis. The PCR products generated from the six second round amplifications were cloned, and a number of clones from each PCR were sequenced. The levels of divergence, or number of base-pair changes, compared with the p*HYGWT* sequence were calculated, and for this study clones showing 1% and 3% divergence were chosen, although many

other clones showing up to 11% divergence were observed (see Appendix 6; all percentages rounded to the nearest integer). The two clones chosen were named *pHYG01* and *pHYG03* respectively, and were both 914 bp in length. It is apparent in Figure 5.10 that a number of larger, non-specific PCR products were generated during the second round amplification reactions. Since these were approximately twice the size of the expected PCR products it is possible that these arose by annealing the *AscI* sites at opposite ends of the denatured PCR products, while at the other ends the correct primers annealed, thereby priming the next round of amplification to generate a double length PCR product.

A multiple alignment of the *pHYGWT*, *pHYG01* and *pHYG03* is shown in Figure 5.11. This highlights the positions of the heterologies compared to the sequence of the *pHYGWT* plasmid which is perfectly homologous to the wild-type *HYG* ORF (except for 10 bp at each terminus; see Appendix 6). The *pHYG01* sequence contains 7 base changes. Unfortunately, these are not divided equally between the flanks, as the 5' flank contains 6 base changes while the 3' flank only contains 1 mutation. Ideally the changes would be dispersed throughout the flanks, but this was the only clone sequenced which contained so few mutations. The *pHYG03* sequence contains 30 base changes from the wild-type *HYG* sequence, and these are distributed equally, with 15 changes each in the 5' and 3' flanks. At both the 5' and 3' termini of the flanks contained within the *pHYGWT*, *pHYG01* and *pHYG03* plasmids there are 10 bp of sequence, incorporating the *AscI* sites, which are not present in the wild-type *HYG* ORF. While these bases can be considered as heterologous from the wild-type *HYG* sequence, they are present in all the *HYG* flanks generated for this assay. Therefore, any influence these bases have on the frequency of integration will be equivalent between the constructs and so their effect can be disregarded.

The *BLE*-containing resistance cassette was cloned between the *HYG* flanks within the *pHYGWT*, *pHYG01* and *pHYG03* plasmids to generate the *HYGWT::BLE*, *HYG01::BLE*, and *HYG03::BLE* targeting constructs respectively (see section 2.17.1). Prior to transformation into transgenic HTUB trypanosomes, the constructs were linearised by restriction digestion with *AscI*, and resuspended to approximately  $3 \mu\text{g}\cdot\mu\text{l}^{-1}$  each (Figure 5.12), thereby ensuring that any differences in the frequency of integration of the three constructs were due to the mutations in the *HYG* flanks and not to differences in DNA concentration.

## 5.6 Assay to determine the effect of *MSH2* mutation on the frequency of integration of homologous and homeologous *HYG* targeting constructs into the HTUB cell line

The transgenic HTUB cell line, which contains a *HYG* ORF integrated into the tubulin array, was used to generate both heterozygous and homozygous HTUB *MSH2* mutant trypanosomes. Also, three constructs were developed which contained either perfectly homologous or increasingly divergent targeting flanks derived from the *HYG* ORF. These targeting constructs and the HTUB cell line were developed to assay for the effect of sequence divergence on the frequency of integration into both MMR-proficient and MMR-deficient cell lines. To this end, HTUB *MSH2* wild-type trypanosomes, two HTUB *MSH2*<sup>+/-</sup> cell lines and two HTUB *MSH2*<sup>-/-</sup> mutants were electroporated with either the linearised *HYGWT::BLE*, *HYG01::BLE* or *HYG03::BLE* constructs (see section 2.17.3). To ascertain the effect of mismatch repair on both homologous and homeologous recombination, the HTUB *MSH2* wild-type cell line, and each of the HTUB *MSH2* heterozygous and HTUB *MSH2* homozygous cell lines were electroporated with 6 µg of *AscI*-digested *HYGWT::BLE*, *HYG01::BLE* and *HYG03::BLE*. Three independent transformations were performed using each construct with each cell line. Transformants were selected in HMI-9 containing 2.5 µg.ml<sup>-1</sup> phleomycin in 24-well tissue culture plates, and the number of wells showing growth was determined after 7 days.

Table 5.1 shows the estimated frequency of integration, expressed as numbers of transformants per 10<sup>6</sup> cells put on selection, and these data are represented in graphical form in Figure 5.13. In almost all the *HYGWT::BLE* transformations for every cell line, all 24 wells showed growth. If it is assumed that the growth in each well arose from a single transformant, this represents an estimated 3.20 x 10<sup>-6</sup> integration events/cell. However, where every well, or indeed a large proportion of the wells, shows growth in a 24-well plate it is probable that some of the wells contained more than one transformant, making it impossible to derive an accurate estimate of the integration frequency from these transformations. For the purposes of this study, where all 24 wells showed growth the frequency has been shown as equal to or more than 3.20 x 10<sup>-6</sup> integration events/cell, although in reality the frequencies could be significantly higher. For transformations where up to 23 wells showed growth after selection, the frequency of construct integration has been calculated on the assumption that there was only one transformant in each well showing growth, even though here also this could be an underestimation. These problems arose because too many cells were placed on selection after electroporation with the targeting constructs, and would be avoided by selecting fewer transformants in future

experiments. Transformation with the *HYG01::BLE* construct resulted in similar integration frequencies in every cell line. These were slightly lower than the frequencies seen using the *HYGWT::BLE* construct, perhaps suggesting that the introduction of mutations into the targeting flanks reduced the ability of the construct to integrate into the *HYG* ORF. From these data loss of *MSH2* does not appear to affect the frequency of integration of this construct. However, the effect of *MSH2* deletion may have been masked owing to the fact that too many cells were placed on selection during these transformations, and the resulting problems in estimating integration frequencies. These experiments would have to be repeated and refined to address this issue. Finally, transformation with the *HYG03::BLE* construct gave similar integration frequencies in the HTUB *MSH2* wild-type cell line and the two *MSH2* heterozygous mutants. These were greatly reduced (by at least 1.5-fold) in comparison to the results observed for the *HYGWT::BLE* and *HYG01::BLE* constructs, suggesting that the probability of successful integration into the trypanosome genome decreases as the level of sequence divergence between the donor and recipient molecules increases. In the *MSH2* homozygous mutants, however, the frequency of integration of the *HYG03::BLE* construct was at least 1.3-fold higher than in cells expressing *MSH2*. Furthermore, in the *MSH2*<sup>-/-</sup> 1.1 cell line the frequency of integration of this construct was barely reduced in comparison with the *HYGWT::BLE* construct. These results suggest that the probability of recombination between homeologous sequences in *T. brucei* is elevated in *MSH2* null mutants, at least where the sequences have diverged by 3%. While this conclusion is perhaps at odds with the results of the transformations using the *HYG01::BLE* construct, and would need to be repeated to confirm these data, it appears that loss of *MSH2* does elevate the frequency of integration of the *HYG03::BLE* construct.

## 5.7 Discussion

In both bacteria and eukaryotes, MMR had been found to regulate recombination between homocologous DNA sequences. The aim of this present study was to determine if this was also the case in *T. brucei*. At present, very little is known about the sequence requirements of recombination in this organism. No illegitimate recombination event has ever been observed in wild-type trypanosomes during many transformations in several laboratories (Lee and Van der Ploeg, 1990; Ten Asbroek *et al.*, 1990; Eid and Sollner-Webb, 1991; Ten Asbroek *et al.*, 1993), although altered forms of homologous recombination are seen in *RAD51* null mutants (McCulloch and Barry, 1999; Conway *et al.*, 2002b). Also, it has been shown that targeting of an antibiotic resistance marker into a region conserved between the *VSG* expression sites resulted in integration into the expression site where the

sequence was identical to the targeting flanks rather than into a site where the sequence differed by 8% (Blundell *et al.*, 1996). Although the targeting construct did appear to integrate into a number of other expression sites the sequence of these sites was never determined. This chapter describes an assay which should yield quantifiable data about the effect of sequence divergence on the frequency of integration into the trypanosome genome in both MMR-proficient and MMR-deficient cells.

While the assay has greater potential than has been exploited in this pilot study (see below), it has nonetheless yielded several important observations about the mechanism of recombination in trypanosomes. Firstly, the frequency of integration of constructs decreases as the number of heterologies between the targeting flanks and the target sequence increases. Where the targeting flanks were diverged by 1%, the frequency of integration appeared to be slightly reduced, although the actual level could not be accurately calculated because too many transformants had been put on selection during these experiments, resulting in an underestimation of the integration frequencies. To determine the frequencies more accurately these experiments would need to be repeated placing fewer transformants on selection. It can be concluded, however, that where the targeting flanks have diverged by 3% from the target sequence, the integration frequency is reduced by at least 1.5-fold in *MSH2* wild-type cells. Secondly, although there was no discernible difference between the frequency of integration of the constructs bearing perfectly homologous or 1% divergent targeting flanks between the *MSH2* null mutants or cells expressing *MSH2*, there was a clear elevation in the integration frequency of the construct containing 3% divergent targeting flanks in the *MSH2* homozygous mutants. This supports the conclusion that a functional MMR system inhibits recombination between divergent sequences in *T. brucei*. The anti-recombination activity of MMR has been characterised most fully in *S. cerevisiae*. While it seems clear that mismatch recognition underlies the process, the mechanism by which MMR down-regulates homeologous recombination remains unknown. Furthermore, it has been observed that the overall frequency of recombination between identical DNA sequences is also elevated in MMR-deficient yeast cells, again without a clear understanding of the mechanistic basis (Datta *et al.*, 1997). This may also be the case in *T. brucei*, but too many transformants were placed on selection in these experiments to get an accurate measure of integration frequency, and any elevation in the frequency of integration in the *MSH2* homozygous mutants may have been masked. Again, these questions can only be addressed by repeating these experiments and placing fewer transformants on selection to collect more accurate data.

One problem evident from this pilot study is that there is no way to determine the number of copies of the *HYG* ORF that have integrated into the trypanosome genome using Southern analysis. Unfortunately, pulsed-field gel electrophoresis of whole chromosomes could not be used to resolve this issue either, since two copies of the *HYG* ORF could integrate into the tubulin array of a single chromosome again resulting in hybridisation to a single fragment. Given the generally low frequency of integration of constructs into bloodstream stage trypanosomes it seems unlikely that two copies of the *HYG* ORF would integrate into the same cell. Nevertheless, it would be still be useful to be able to determine the copy number of the *HYG* ORF within the cell lines used in the assay. It is possible to do this after transformation of the HTUB-derived cell lines with the *HYG::BLE* constructs. Genomic DNA could be isolated from the bleomycin resistant clones, and PCR amplification using primers specific to the *HYG* locus performed. In cells possessing only one copy of the locus a single allele would be amplified corresponding to the *HYG* ORF interrupted by the *BLE* resistance cassette (*HYG::BLE* allele). PCR amplification from any clones containing two or more copies of the *HYG* locus would give two products, the allele described above and a smaller product corresponding to the intact *HYG* ORF. It would be possible to determine the *HYG* alleles present in a number of *BLE* resistant transformants from all the HTUB-derived cell lines in order to confirm that only a single copy of the *HYG* ORF was present in the genomes of the HTUB cell line and its descendants. Furthermore, sequencing of the PCR products corresponding to the *HYG::BLE* allele generated in MMR-proficient cell lines might reveal details about the mechanisms of the recombination reactions used to create them. In MMR-proficient yeast cells it has been reported that during gene conversion reactions between homeologous substrates, more than 85% of the mismatches present in the heteroduplex intermediate are repaired in favour of the recipient (unbroken) strand (Leung *et al.*, 1997). This same study suggested that recombination reactions between linear and chromosomal DNA proceeded by complete assimilation of one of the two strands of linear DNA, rather than by two crossovers each close to the end of the linear fragment. Comparison of the sequences of the *HYG::BLE* allele from MMR-proficient transformants with both the wild-type *HYG* ORF and relevant *HYG::BLE* construct sequences would show whether the mismatches formed in the heteroduplex recombination intermediate were repaired in favour of the donor or recipient strand. Where base changes from the wild-type *HYG* ORF were observed this would suggest that the mismatch was repaired in favour of the donor (*HYG::BLE* construct) sequence. Furthermore, if such changes were observed in the sequences corresponding to the extreme termini of the *HYG::BLE* construct this would indicate that the gene conversion reaction involved the assimilation of the entire sequence of the linear

transformation construct as in yeast. The sequence of the *HYG::BLE* alleles generated in MMR-deficient transformants could not be used in this way as the mismatches present in the heteroduplex intermediate would not be repaired, and would segregate at mitosis. Therefore some clones would contain sequence derived from the wild-type *HYG* ORF and others the sequence derived from the *HYG::BLE* construct.

In the future, this assay could be extended to further investigate the sequence requirements of recombination in *T. brucei*. A broader range of constructs with divergent targeting flanks could be generated; as stated earlier, flanks which show up to 11% sequence divergence are already available (see Appendix 6). These span the 8% sequence divergence of the only previous study (Blundell *et al.*, 1996), and could be used to investigate the effect of increasing numbers of heterologies on recombination in *T. brucei* in detail. In an MMR-competent yeast strain a construct containing a single mismatch (0.3% divergence) showed a 4.1-fold reduction in the frequency of recombination compared to identical substrates (Datta *et al.*, 1997). Increasing the levels of divergence in the substrates led to a cumulative negative effect on recombination, with the integration frequencies decreasing further as the number of heterologies between the substrates increased. In MMR-deficient yeast cells, however, sequence divergence showed little effect on the frequency of integration until the sequences differed by about 10% (Datta *et al.*, 1996). This indicates that where substrates are greater than 90% identical the MMR machinery is responsible for all the inhibitory effects of sequence divergence upon recombination. Above this level of sequence divergence another factor strongly inhibits homeologous recombination, and it has been hypothesised that this repression may be due to an inability to form sufficiently stable base-paired strand exchange intermediates where high levels of mismatches are present (Selva *et al.*, 1995; Datta *et al.*, 1996; Datta *et al.*, 1997; Chen and Jinks-Robertson, 1998). It has been speculated that the inhibitory effect of MMR on homeologous recombination in *T. brucei* may be suppressed to allow gene conversion reactions between divergent *VSG* sequences (Blundell *et al.*, 1996). While this assay cannot be directly used to investigate recombination between the flanking sequences of *VSG* genes, it will reveal whether the sequence requirements for general recombination are less stringent in this organism.

The influence of MMR on the frequency of integration could be determined more fully using a range of divergent targeting constructs in both *MSH2* and *MLH1* null mutants (and perhaps eventually *MSH3*, *MSH8* and *PMS1* mutants). As discussed in section 5.1, experiments in *S. cerevisiae* have revealed that the frequency of homeologous recombination shows the greatest elevation in *MSH2* mutants, while *MLH1* is required in

the repression of only a portion of these events (Datta *et al.*, 1996; Nicholson *et al.*, 2000). Similar results in *T. brucei* would suggest that the mechanism of inhibition was conserved between these organisms, and that the models of heteroduplex rejection which have been expounded to explain the anti-recombination activities of MMR could be applied to trypanosomes as well.

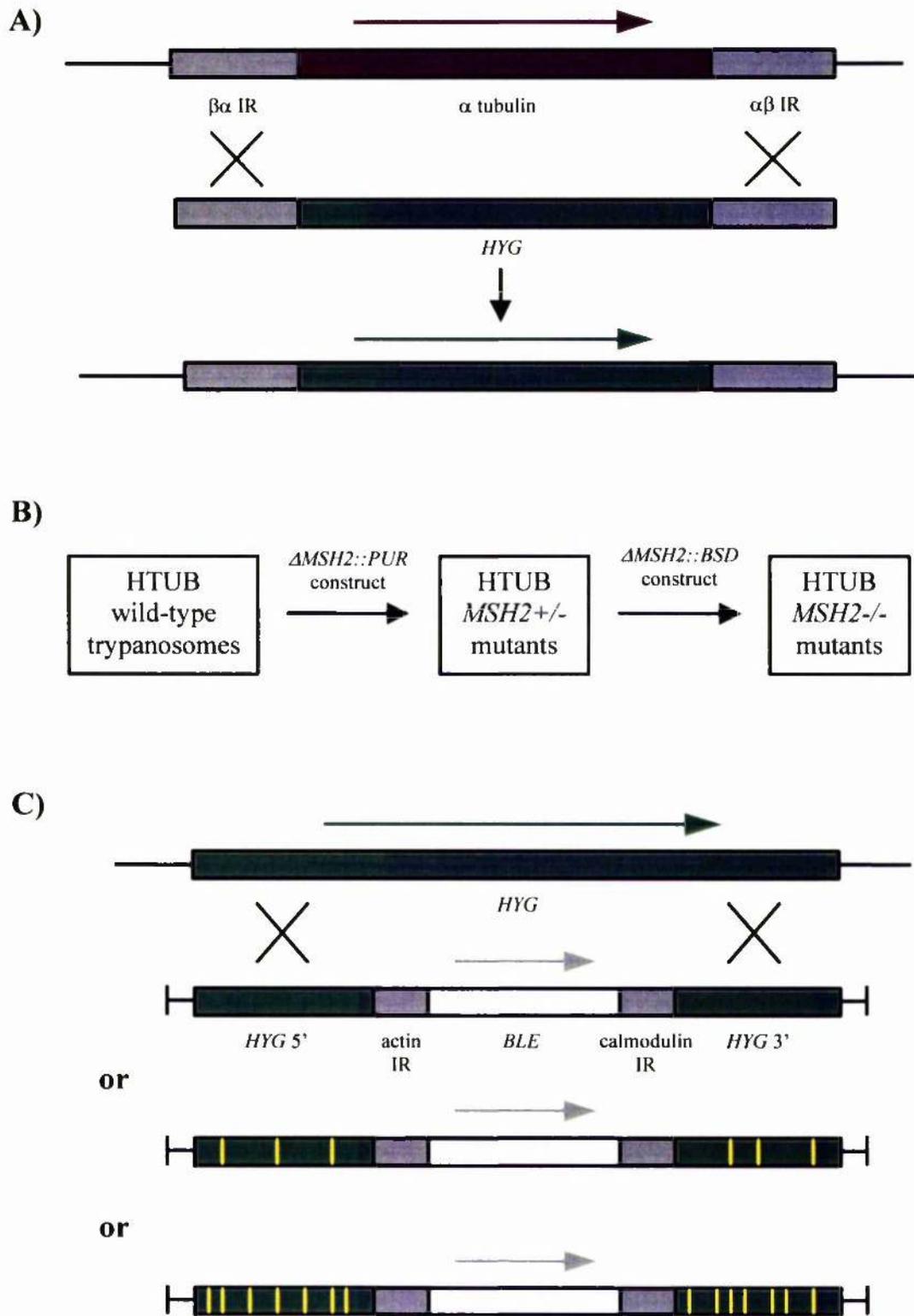
It is possible that different genomic locations exhibit different intrinsic recombinogenic potentials. For instance, the tubulin array in *T. brucei* is composed of tandem repeats of  $\alpha$  and  $\beta$  tubulin genes, so this locus may be more prone to intra- and inter- chromosomal recombination events between the numerous alleles. Also, regions of the genome which are being actively transcribed, and so are more accessible to the recombination machinery, appear to participate in recombination events more frequently than silent loci (Saxe *et al.*, 2000). It should also be noted that in a number of organisms which undergo antigenic variation, for instance *T. brucei*, *Pneumocystis carinii*, *Plasmodium* spp and *Borrelia* spp, the expressed surface antigen gene is positioned in a sub-telomeric location, where it frequently recombines, often with sequences from other telomeric regions, resulting in the expression of a novel surface antigen (Zhang *et al.*, 1997; Newbold, 1999; Barry and McCulloch, 2001; Stringer and Keely, 2001). This perhaps suggests that these regions of the genome are predisposed to recombination. This assay system could be used with transgenic cell lines containing the *HYG* ORF integrated into sites other than the tubulin array; for instance, the active expression site close to the *VSG*, or the telomeric region of a minichromosome. Assaying for the frequency of integration simply using the perfectly homologous *HYGWT::BLE* construct would indicate whether there were any differences in the recombinogenic potential of these loci. Moreover, assaying the effects of *MSH2* and *MLH1* mutants on integration into *HYG* within the active ES would address whether this site is some form of 'privileged domain' not scoured by the MMR system, perhaps explaining the lack of detectable effect these mutations had on *VSG* switching (see chapter 4). As the same sequences, all derived from the *HYG* ORF, would be used in each case, any differences seen would be due to the properties of the locus under investigation and not to sequence specific effects. Moreover, this system could be used to investigate the influence of other genes, such as *RAD51* or *MRE11*, on recombination in *T. brucei*, and since the data gathered would be comparable it would be possible to generate a 'league-table' suggesting the relative importance of numerous genes to recombination in *T. brucei*.

**Figure 5.1 Overview of experimental approach**

(A) A hygromycin phosphotransferase (*HYG*) gene ORF (derived from pHygro-1ub; see section 2.17.2; green boxes) was integrated into the tubulin array (only  $\alpha$  tubulin is shown) using integration and processing flanks derived from the tubulin locus ( $\beta\alpha$  and  $\alpha\beta$ ; grey boxes) and replaced an  $\alpha$  tubulin gene (purple box) to generate the HTUB cell line. (B) Transformation of the HTUB cell line with the  $\Delta$ *MSH2*::*PUR* and  $\Delta$ *MSH2*::*BSD* constructs generated *MSH2* heterozygous (+/-) and homozygous (-/-) knockout mutant cell lines (see section 2.11.1). (C) The *HYG* ORF was the parental sequence of the *HYG* 5' and *HYG* 3' integration flanks (green boxes) which were separated by a resistance cassette that contained the bleomycin resistance protein (*BLE*) ORF (white boxes) flanked by sequences encoding splicing and polyadenylation signals (actin IR and calmodulin IR; grey boxes). A number of constructs incorporating integration flanks which contained increasing numbers of base changes compared to the wild-type *HYG* sequence (shown as vertical yellow lines) were generated. The *HYG* 5' and *HYG* 3' flanks allowed the integration of the targeting constructs into the *HYG* ORF in the HTUB transgenic cell line (indicated by crosses). Transformants were selected on phleomycin in 96-well plates, the number of wells showing growth was recorded and the frequency of integration calculated. Arrows indicate the direction of transcription of the genes.

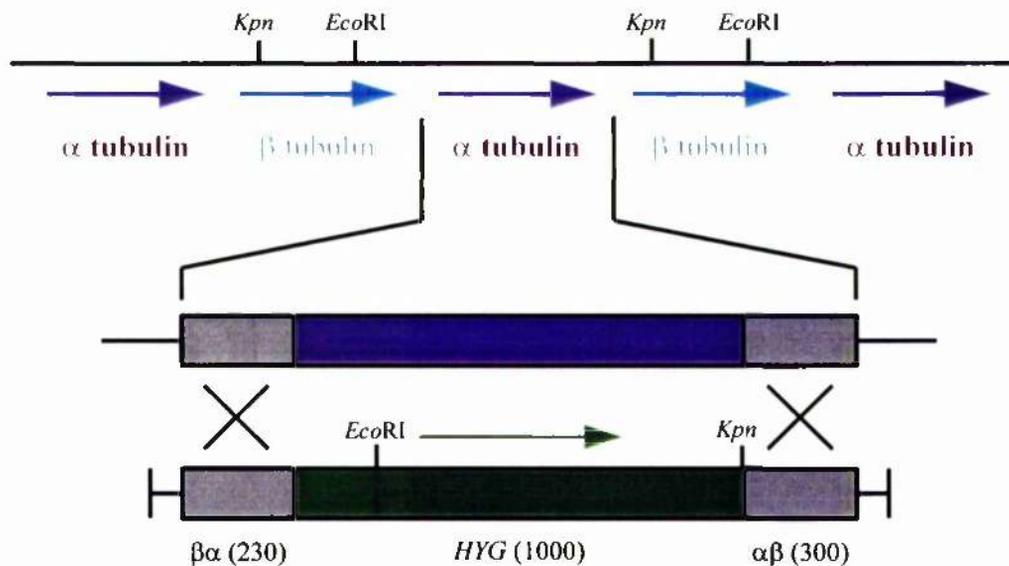
over page

Figure 5.1 continued



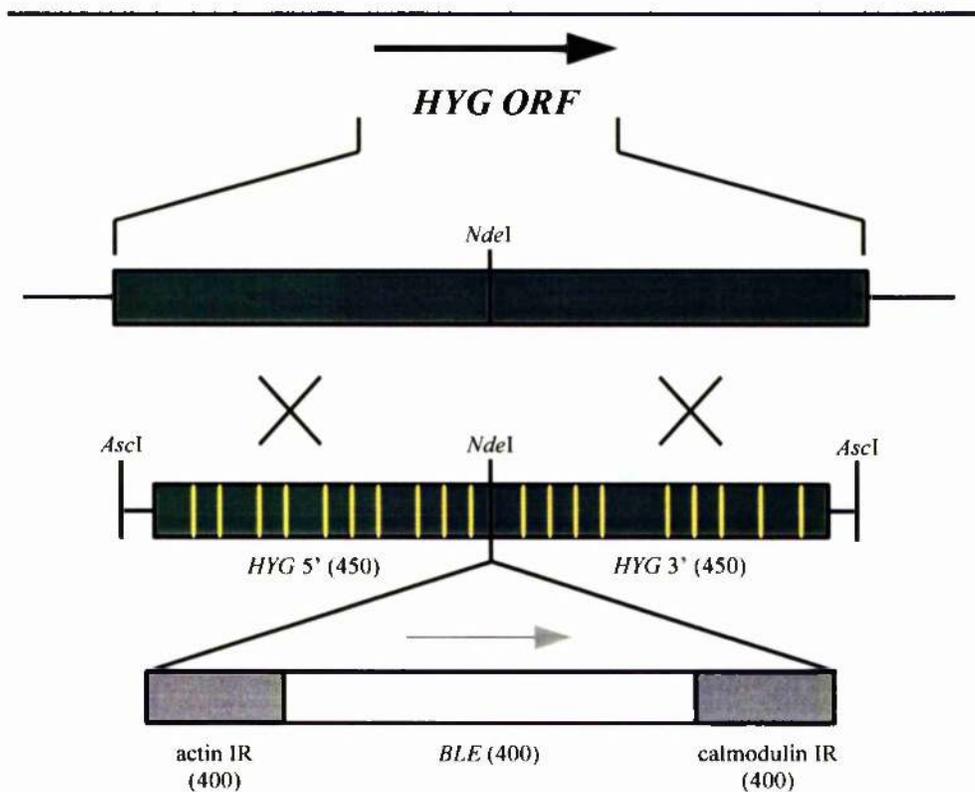
### Figure 5.2 Targeting a hygromycin phosphotransferase ORF into the tubulin array

A resistance cassette (derived from pHygro-Tub; see section 2.17.2) containing a hygromycin phosphotransferase ORF (*HYG*) will be integrated into the tubulin array (indicated by crosses) and replace an  $\alpha$  tubulin gene (shown in purple).  $\beta$  tubulin genes are shown in turquoise. The targeting flanks for this replacement reaction consist of the 5'  $\beta$ - $\alpha$  intergenic sequence ( $\beta\alpha$ ) and the 3'  $\alpha$ - $\beta$  intergenic sequence ( $\alpha\beta$ ) which flank the  $\alpha$  tubulin genes. These flanks will also provide signals allowing the transcripts of the resistance gene to be processed into mature mRNAs. Arrows denote the direction of transcription of the genes and numbers in brackets indicate lengths in base-pairs. Restriction sites are shown as black vertical bars.



**Figure 5.3 Design of constructs with homeologous integration flanks to target the *HYG* ORF present in the HTUB transgenic cell line**

*HYG* 5' and *HYG* 3' are the sequences derived from the hygromycin phosphotransferase gene (shown in green) that allow integration of the constructs into the *HYG* ORF in the HTUB transgenic cell line (indicated by crosses). A number of constructs incorporating integration flanks which contain increasing numbers of base changes compared to the wild-type *HYG* sequence (shown as vertical yellow lines) were generated. Between these flanks is an expression cassette that gives phleomycin resistance to select for integration; this contains the bleomycin resistance protein ORF (*BLE*) flanked at the 5' end by sequence derived from the actin locus (actin IR) and at the 3' end by sequence from the calmodulin locus (calmodulin IR). Arrows indicate the direction of transcription of the genes, numbers in brackets denote lengths in base-pairs and restriction sites are shown as vertical black lines.



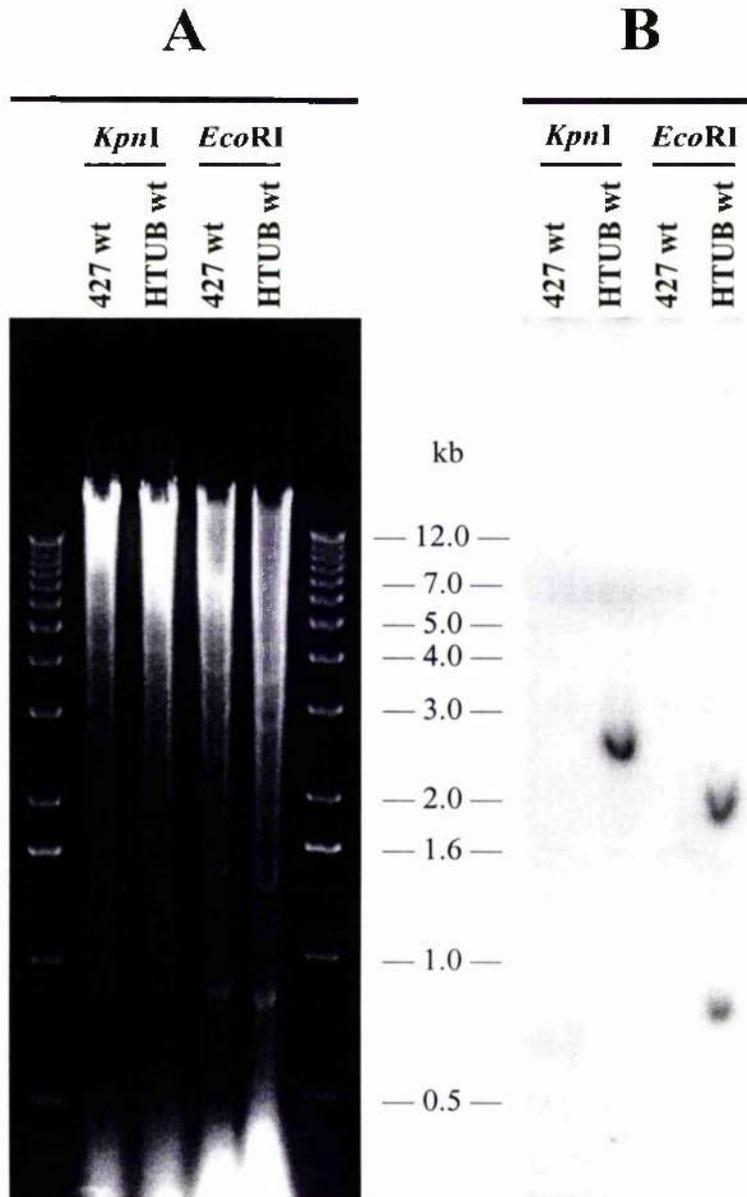
**Figure 5.4 PCR amplification of homologous *HYG* targeting flanks and *HYG* probe sequence**

PCR was performed using Herculase polymerase and the pHygro-Tub plasmid as a template. Primers complementary to the *HYG* ORF were used (*HYG*For and *HYG*Rev; see section 2.17.1) and the following reactions were run: (A) both primers with no template DNA; (B) *HYG*For primer only with plasmid template DNA; (C) *HYG*Rev primer only with plasmid template DNA; (D) both primers with plasmid template DNA.



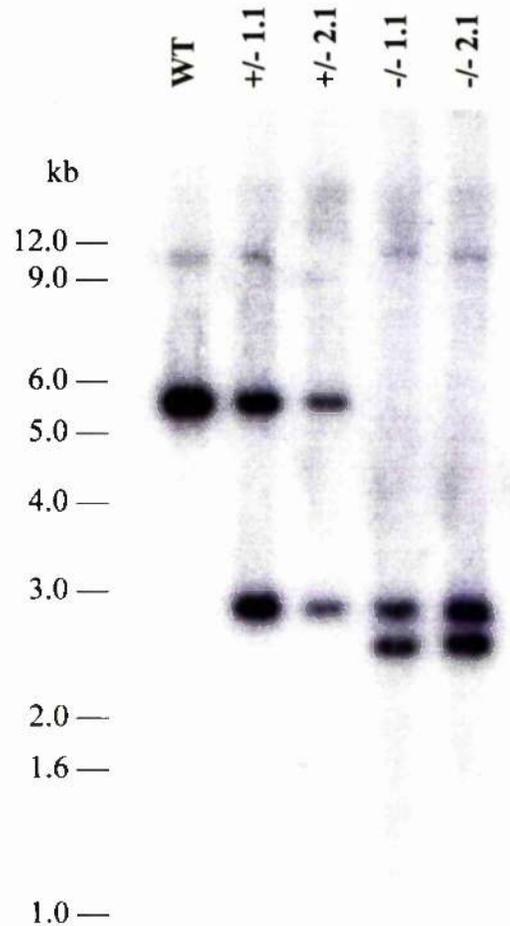
### Figure 5.5 Southern analysis of the HTUB transgenic cell line

(A) Genomic DNA isolated from the MITat 1.2a S427 cell line (427 wt) and the HTUB transgenic cell line (HTUB wt) was digested with *Kpn*I and *Eco*RI and separated on a 0.6% agarose gel. (B) Southern blot of the gel shown in panel (A) probed with the 914 bp *HYG* probe described in Figure 5.4.



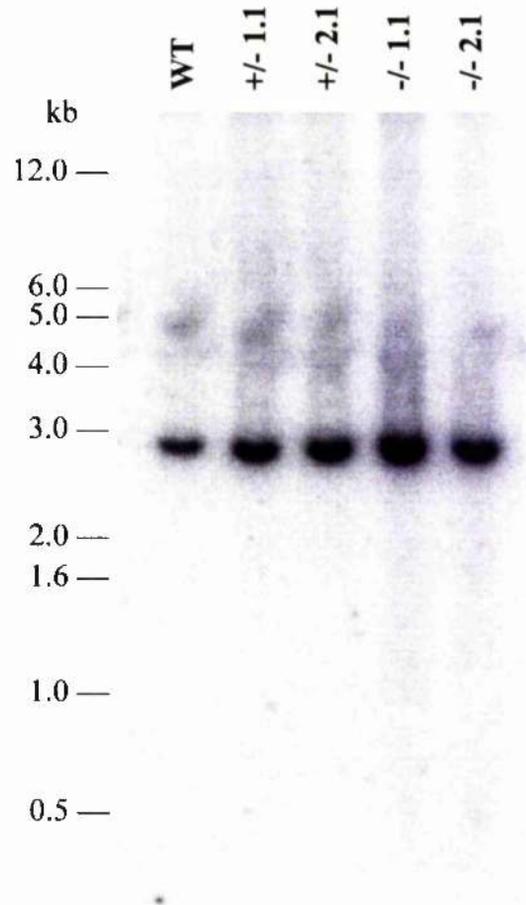
**Figure 5.6 Southern analysis of HTUB *MSH2* knockout cell lines**

Genomic DNA isolated from the HTUB *MSH2* knockout cell lines was digested with *SacI*, separated on a 0.6% agarose gel, Southern blotted and probed with the *MSH2* 5' flank used as the 5' integration flank in the *MSH2* knockout construct (see section 2.11.1). The following samples were analysed: (WT) untransformed HTUB cells; (+/-) HTUB cells transformed with  $\Delta MSH2::PUR$ ; (-/-) HTUB cells transformed with  $\Delta MSH2::PUR$  and  $\Delta MSH2::BSD$ .



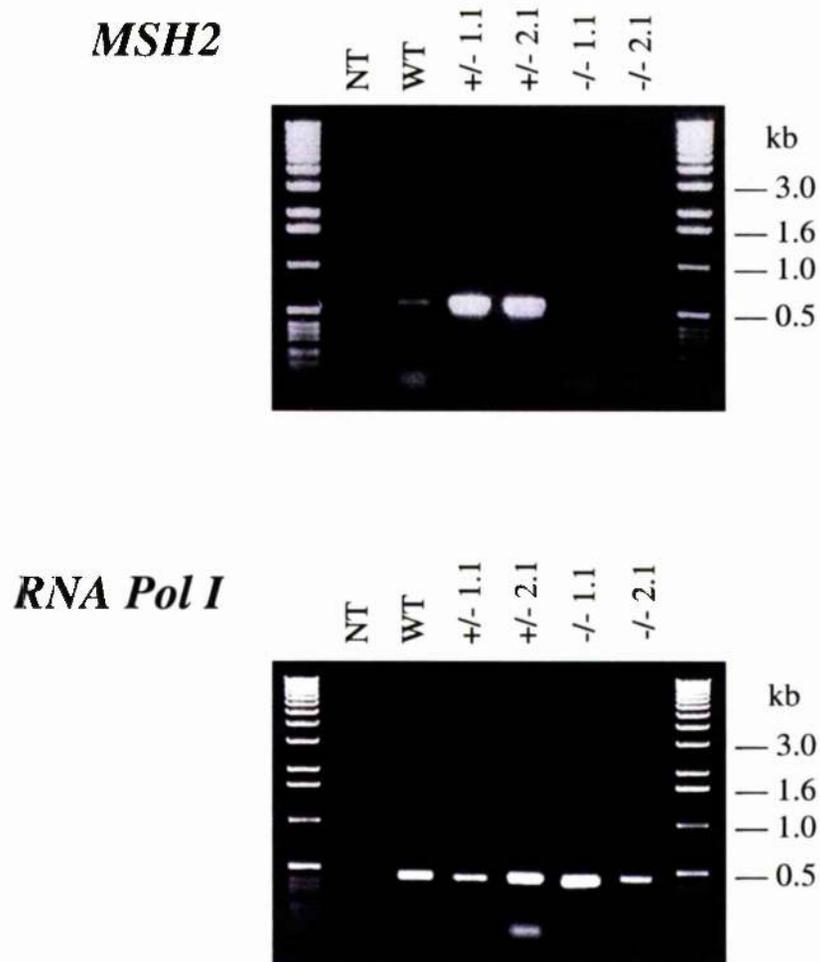
**Figure 5.7 Southern analysis of HTUB *MSH2* knockout cell lines**

Genomic DNA isolated from the HTUB *MSH2* knockout cell lines was digested with *KpnI*, separated on a 0.6% agarose gel, Southern blotted and probed with the 914 bp *HYG* probe described in Figure 5.4. The following samples were analysed: (WT) untransformed HTUB cells; (+/-) HTUB cells transformed with  $\Delta$ *MSH2*::*PUR*; (-/-) HTUB cells transformed with  $\Delta$ *MSH2*::*PUR* and  $\Delta$ *MSH2*::*BSD*.



### Figure 5.8 PCR analysis of the HTUB *MSH2* knockout cell lines

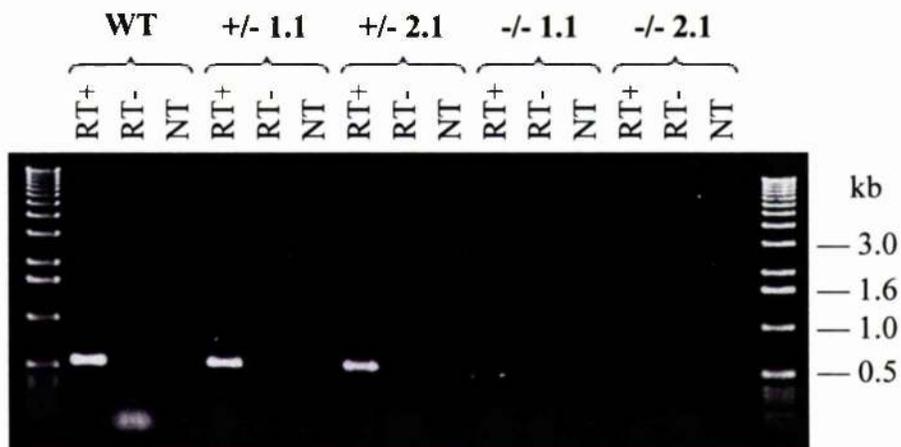
PCR was performed using genomic DNA isolated from the HTUB cell lines transformed with the *MSH2* knockout constructs. Internal primers complementary to the *T. brucei* *MSH2* gene were used (*MSH2*) and the integrity of the DNA was confirmed using primers directed against the large subunit of *T. brucei* RNA polymerase I (*RNA Pol I*; Rudenko *et al.*, 1996). For both sets of primers the following reactions were run: (NT) no template DNA; (WT) untransformed HTUB cells; (+/-) HTUB cells transformed with  $\Delta$ *MSH2*::*PUR*; (-/-) HTUB cells transformed with both  $\Delta$ *MSH2*::*PUR* and  $\Delta$ *MSH2*::*BSD*.



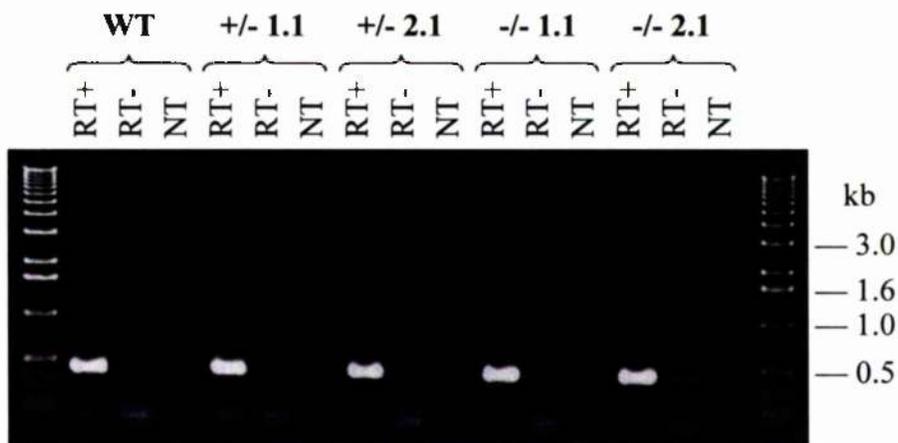
**Figure 5.9 Analysis of the expression of *MSH2* in the putative HTUB *MSH2* mutant cell lines**

RT-PCR was performed using total RNA isolated from the HTUB cell lines transformed with the *MSH2* knockout constructs. Internal primers complementary to the *T. brucei MSH2* gene were used (*MSH2*) and the integrity of the DNA was confirmed using primers directed against the large subunit of *T. brucei* RNA polymerase I (*RNA Pol I*; Rudenko *et al.*, 1996). For both sets of primers RT positive (RT+), RT negative (RT-), and no template (NT) reactions were performed using total RNA isolated from the following cell lines: (WT) untransformed HTUB cells; (+/-) HTUB cells transformed with  $\Delta$ *MSH2*::*PUR*; (-/-) HTUB cells transformed with both  $\Delta$ *MSH2*::*PUR* and  $\Delta$ *MSH2*::*BSD*.

### *MSH2*

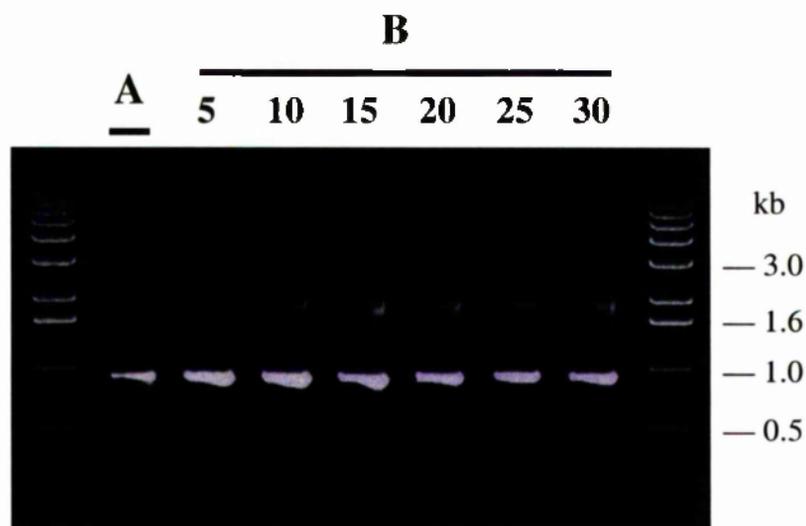


### *RNA Pol I*



### Figure 5.10 PCR amplification of homeologous *HYG* targeting flanks

Two rounds of PCR amplification were performed using *Taq* polymerase and the primers *HYG*For and *HYG*Rev (see section 2.17.1). In the first round of amplification, the pHygro-Tub plasmid was used as a template and the reaction was supplemented with two nucleotide analogues: dPTP, which *Taq* polymerase incorporates in place of dTTP and to a lesser extent dCTP; and 8-oxo-dGTP which is incorporated in place of dTTP (Zaccolo *et al.*, 1996). During amplification, at the end of cycles 5, 10, 15, 20, 25 and 30, a 1  $\mu$ l sample was removed and used as a template for a second round of PCR amplification using standard reaction conditions. (A) A sample of the first round PCR product taken after 30 cycles. (B) Samples of the second round PCR products taken after 30 cycles. Numbers refer to the cycle number of the first round of PCR from which the template was derived.



**Figure 5.11 Multiple alignment of the pHYGWT, pHYG01 and pHYG03 sequences**

Multiple alignment of the pHYGWT, pHYG01 and pHYG03 DNA sequences. Sequences were aligned using Multalin (<http://prodes.toulouse.inra.fr/multalin/multalin.html>; Corpet, 1988) and shaded using the BOXSHADE server ([http://www.ch.embnet.org/software/BOX\\_form.html](http://www.ch.embnet.org/software/BOX_form.html)): identical residues are shown in black; residues that differ from the pHYGWT sequence are shown in turquoise for the pHYG01 sequence and in red for the pHYG03 sequence. The sequence of the HYGFor primer is indicated in blue, and the sequence complementary to the HYGRev primer is indicated in green. The sequence corresponding to the *NdeI* restriction site is shown in purple.

```

pHYGWT 1 AAGGCGCGCCAGCCTGAACTCACCGCGACGTCTGTTCGAGAAGTTTCTGATCGAAAAGTTC
pHYG01 1 AAGGCGCGCCAGCCTGAACTCACCGCGACGTCTGTTCGAGAAGTTTCTGATCGAAAAGTTC
pHYG03 1 AAGGCGCGCCAGCCTGAACTCACCGCGACGTCTGTTCGAGAAGCTTCTGATCGAAAAGTTC

pHYGWT 61 GACAGCGTCTCCGACCTGATGCAGCTCTCGGAGGGCGAAGAATCTCGTGCTTTTCAGCTTC
pHYG01 61 GACAGCGTCTCCGACCTGATGCAGCTCTCGGAGGGCGAAGAATCTCGTGCTTTTCAGCTTC
pHYG03 61 GACAGCGTCTCCGACCTGATGCAGCTCTCGGAGGGCGAAGAATCTCGTGCTTTTCAGCTCC

pHYGWT 121 CATGTAGGAGGGCGTGGATATGTCCTGCGGGTAAATAGCTGCGCCGATGGTTTCTACAAA
pHYG01 121 CATGTAGGAGGGCGTGGATAATGTCCTGCGGGTAAATAGCTGCGCCGATGGTTTCTACAAA
pHYG03 121 CATGTAGGAGGGCGTGGATATGTCCTGCGGGTAAATAGCTGTGCGCGATGGTTTCTACAAA

pHYGWT 181 GATCGTTATGTTTATCGGCCACTTTGCATCGGCCGCGCTCCCGATTCCGGAAGTGCTTGAC
pHYG01 181 GATCGTTATGTTTATCGGCCACTTTGCATCGGCCGCGCTCCCGATTCCGGAAGTGCTTGCC
pHYG03 181 GATCGCTATGTTTATCGGCCACTTTGCATCGGCCGCGCTCCCGATTCCGGAAGTGCTTGAC

pHYGWT 241 ATTGGGGAATTCAGCGAGAGCCTGACCTATTGCATCTCCCGCCGTGCACAGGGTGTACAG
pHYG01 241 ATTGGGGAATTCAGCGAGAGCCTGACCTATTGCATCTCCCGCCGTGCACAGGGTGTACAG
pHYG03 241 ATCGGGAATTCAGCGAGAGCCTGACCTATTGCACCTCCCGCCGCGCACAGGGTGTACAG

pHYGWT 301 TTGCAAGACCTGCCTGAAACCGAACTGCCCGCTGTTCTGCAGCCGGTTCGCGGAGGCCATG
pHYG01 301 TTGCAAGACCTGCCTGAAACCGAACTGCCCGCTGTTCTGCAGCCGGTTCGCGGAGGCCATG
pHYG03 301 TTGCAAGACCTGCCTGAAACCGAACTGCCCGCTGTTCTGCAGCCGGTTCGCGGAGGCCATG

pHYGWT 361 GATGCGATCGCTGCGGCCGATCTTAGCCAGACGAGCGGGTTTCGGCCATTTCGGACCGCAA
pHYG01 361 GATGCGATCGCTGCGGCCGATCTTAGCCAGACGAGCGGGTTTCGGCCATTTCGGACCGCAA
pHYG03 361 GATGCGACCGCTGCGGCCGATCTTAGCCAGACGAGCGGGTTTCGGCCATTTCGGACCGCAA

pHYGWT 421 GGAATCGGTCAATACACTACATGGCGTGATTTTCATATGCGCGATTGCTGATCCCCATGTG
pHYG01 421 GGAATCGGTCAATACACTACATGGCGTGATTTTCATATGCGCGATTGCTGATCCCCATGTG
pHYG03 421 GGAATCGGTCAATACACACACGGCGTGATTTTCATATGCGCGATTGCTGATCCCCATGTG

pHYGWT 481 TATCACTGGCAAACCTGTGATGGACGACACCGTCAGTCCGTCCGTCCGCGCAGGCTCTCGAT
pHYG01 481 TATCACTGGCAAACCTGTGATGGACGACACCGTCAGTCCGTCCGTCCGCGCAGGCTCTCGAT
pHYG03 481 TATCACTGGCAAACCTGTGACGGACGACACCGTCAGTCCGTCCGTCCGCGCAGGCTCTCGAT

pHYGWT 541 GAGCTGATGCTTTGGGCCGAGGACTGCCCGAAGTCCGGCACCTCGTGACCGCGGATTTTC
pHYG01 541 GAGCTGATGCTTTGGGCCGAGGACTGCCCGAAGTCCGGCACCTCGTGACCGCGGATTTTC
pHYG03 541 GAGTTGATGCTTTGGGCCGAGGACTGCCCGAAGTCCGGCACCTCGTGACCGCGGATTTTC

```

Figure 5.11 continued

*pHYGWT* 601 GGCTCCAACAATGTCCTGACGGACAATGGCCGCATAACAGCGGTCATTGACTGGAGCGAG  
*pHYG01* 601 GGCTCCAACAATGTCCTGACGGACAATGGCCGCATAACAGCGGTCATTGACTGGAGCGAG  
*pHYG03* 601 GGCTCCAACAATGTCCTGACGGACAATGGCCGCATAACAGCGGTCATTGACTGGAGCGAG

*pHYGWT* 661 GCGATGTTCCGGGATTCCCAATACGAGGTCGCCAACATCTTCTTCTGGAGGCCGTGGTTG  
*pHYG01* 661 GCGATGTTCCGGGATTCCCAATACGAGGTCGCCAACATCTTCTTCTGGAGGCCGTGGTTG  
*pHYG03* 661 GCGATGTTCCGGGATTCCCAATACGAGGTCGCCAACATCTTCTTCTGGAGGCCGTGGTTG

*pHYGWT* 721 GCTTGTATGGAGCAGCAGACGCGCTACTTCGAGCGGAGGCATCCGGAGCTTGCAGGATCG  
*pHYG01* 721 GCTTGTATGGAGCAGCAGACGCGCTACTTCGAGCGGAGGCATCCGGAGCTTGCAGGATCG  
*pHYG03* 721 GCTTGTATGGAGCAGCAGACGCGCTACTTCGAGCGGAGGCATCCGGAGCTTGCAGGATCG

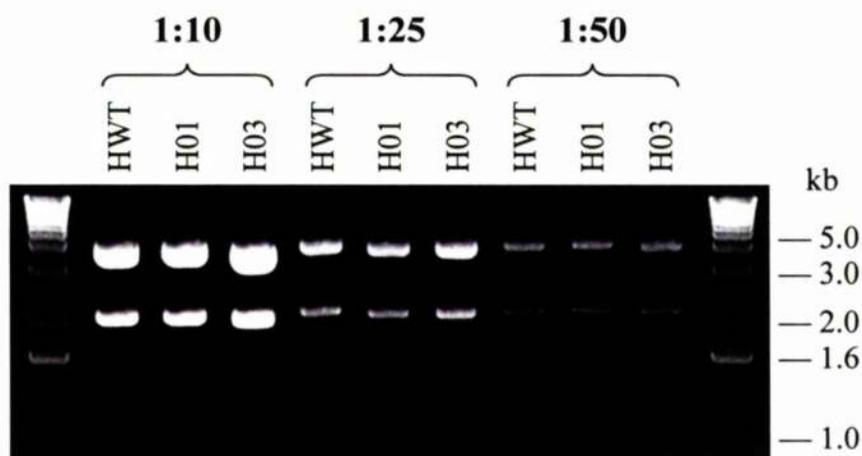
*pHYGWT* 781 CCGCGGCTCCGGGCGTATATGCTCCGCATTGGTCTTGACCAACTCTATCAGAGCTTGTT  
*pHYG01* 781 CCGCGGCTCCGGGCGTATATGCTCCGCATTGGTCTTGACCAACTCTATCAGAGCTTGTT  
*pHYG03* 781 CCGCGGCTCCGGGCGTATATGCTCCGCATTGGTCTTGACCAACTCTATCAGAGCTTGTT

*pHYGWT* 841 GACGGCAATTTCAATGATGCAGCTTGGGCGCAGGGTCGATGCGACGCAATCGTCCGATCC  
*pHYG01* 841 GACGGCAATTTCAATGATGCAGCTTGGGCGCAGGGTCGATGCGACGCAATCGTCCGATCC  
*pHYG03* 841 GACGGCAATTTCAATGATGCAGCTTGGGCGCAGGGTCGATGCGACGCAATCGTCCGATCC

*pHYGWT* 901 GGAGGGCGCGCCAT  
*pHYG01* 901 GGAGGGCGCGCCAT  
*pHYG03* 901 GGAGGGCGCGCCAT

**Figure 5.12** Estimation of the concentration of the linearised *HYGWT::BLE*, *HYG01::BLE*, and *HYG03::BLE* constructs

The *HYGWT::BLE* (HWT), *HYG01::BLE* (H01), and *HYG03::BLE* (H03) constructs were linearised by restriction digestion with *AscI* (see section 2.17.1). The purified restriction fragments were diluted 1 in 10 (1:10), 1 in 25 (1:25) and 1 in 50 (1:50) in distilled H<sub>2</sub>O and separated by electrophoresis on a 1.0% agarose gel to estimate their concentrations. By comparison with the 1.6 kb marker band (which contained 50 ng of DNA) the concentration of the undiluted linearised constructs was estimated to be approximately 3  $\mu\text{g}\cdot\mu\text{l}^{-1}$  each.



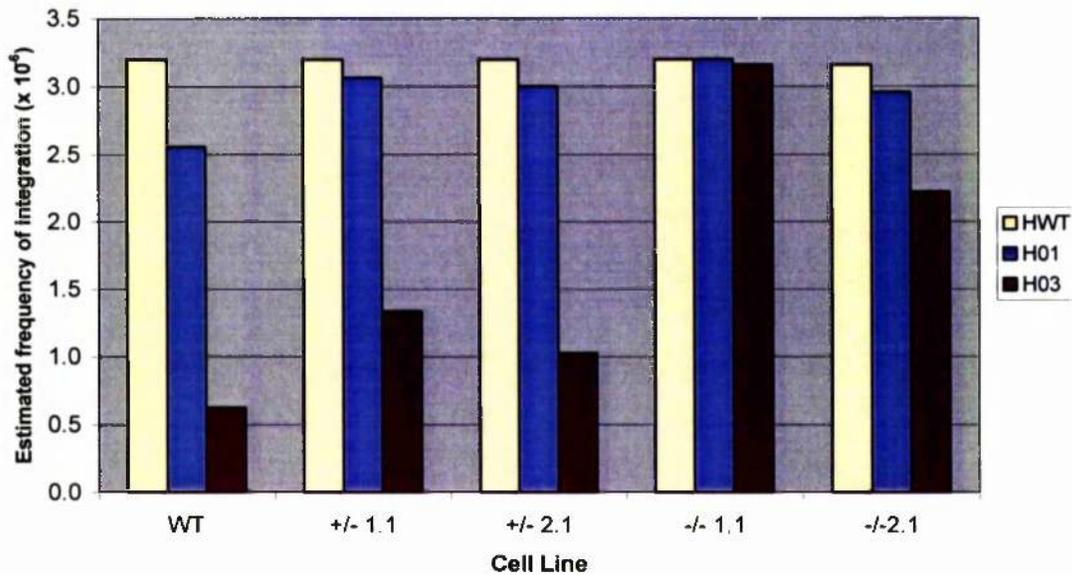
**Table 5.1 Effect of *MSH2* mutation on the frequency of integration of constructs containing increasing numbers of mismatches into the HTUB trypanosome strain**

Bloodstream HTUB trypanosomes that had been unaltered at the *MSH2* locus (WT); that had one *MSH2* allele disrupted (+/-); or that both *MSH2* alleles disrupted (-/-) were transformed with either the *HYGWT::BLE*, *HYG01::BLE* or *HYG03::BLE* constructs (see section 2.17.3). For each transformation  $7.5 \times 10^6$  cells were plated over a 24-well plate on selection with  $2.5 \mu\text{g}\cdot\text{ml}^{-1}$  phleomycin, and three independent transformations were performed using each construct for each cell line. After 7 days the number of wells which showed growth were counted, and it was assumed that in each well growth arose from a single transformant. The estimated frequency of construct integration ( $\times 10^{-6}$ ) was calculated.

Cell Line	Estimated Frequency of Construct Integration ( $\times 10^{-6}$ )		
	<i>HYGWT::BLE</i>	<i>HYG01::BLE</i>	<i>HYG03::BLE</i>
WT	$\geq 3.20$	3.00	0.80
	$\geq 3.20$	2.60	0.60
	$\geq 3.20$	2.07	0.47
<b>Average</b>	$\geq 3.20$	<b>2.56</b>	<b>0.62</b>
<i>MSH2 +/- 1.1</i>	$\geq 3.20$	2.93	1.27
	$\geq 3.20$	3.13	1.33
	$\geq 3.20$	3.13	1.40
<b>Average</b>	$\geq 3.20$	<b>3.06</b>	<b>1.33</b>
<i>MSH2 +/- 2.1</i>	$\geq 3.20$	3.13	1.67
	$\geq 3.20$	2.87	0.53
	$\geq 3.20$	3.00	0.87
<b>Average</b>	$\geq 3.20$	<b>3.00</b>	<b>1.02</b>
<i>MSH2 -/- 1.1</i>	$\geq 3.20$	$\geq 3.20$	$\geq 3.20$
	$\geq 3.20$	$\geq 3.20$	$\geq 3.20$
	$\geq 3.20$	$\geq 3.20$	3.07
<b>Average</b>	$\geq 3.20$	$\geq 3.20$	$\geq 3.16$
<i>MSH2 -/- 2.1</i>	3.07	2.87	2.27
	$\geq 3.20$	3.00	2.13
	$\geq 3.20$	3.00	2.27
<b>Average</b>	$\geq 3.16$	<b>2.96</b>	<b>2.22</b>

**Figure 5.13** The frequency of integration of constructs containing increasing numbers of mismatches into HTUB *MSH2* knockout cell lines

The average frequency of integration of the *HYGWT::BLE* (HWT), *HYG01::BLE* (H01), and *HYG03::BLE* (H03) constructs was determined as described in Table 5.1 for HTUB *MSH2* wild-type trypanosomes (WT), HTUB *MSH2* heterozygous mutants (+/-) and HTUB *MSH2* homozygous mutants (-/-).



# **CHAPTER 6**

## **Final Discussion**

## 6.1 Final discussion

The aim of this study was to investigate the MMR system in *T. brucei*, and assess the role of this highly conserved DNA repair pathway in antigenic variation. Initially, it was important to identify as many putative MutS and MutL homologues as possible, so that an overview of the MMR system in this organism could be attained. During the course of this work, five putative MutS homologues were identified along with two putative MutL-related proteins. The *mutS* family members included putative homologues of the *MSH2*, *MSH3* and *MSH6/7* genes (designated *MSH2*, *MSH3* and *MSH8* respectively) which probably encode components of the nuclear post-replicative MMR system (see chapter 3), and also putative homologues of the meiotic *mutS*-related genes *MSH4* and *MSH5*. The *mutL*-related genes in *T. brucei* appear to encode homologues of the MLH1 and PMS1 proteins from *S. cerevisiae* (see chapter 3).

In order to confirm the presence of an active MMR system in *T. brucei*, functional analyses were performed on the putative *T. brucei* *MSH2* and *MLH1* genes (see chapter 4). These genes were chosen because the MSH2 and MLH1 proteins are central to the MMR process in that they are required to form all known MutS- and MutL-related heterodimers active in MMR respectively, and thus were likely to give the most easily observable phenotypes upon deletion analysis. Homologous integration constructs, containing a selectable marker, were used to generate stable transformants from which both copies of either the *MSH2* or *MLH1* ORFs were deleted. A cell line re-expressing the *MSH2* gene from its original genomic location was then generated from an *MSH2* homozygous mutant in order to confirm that any phenotypes observed were due to the loss of MMR in these trypanosomes, and not to incidental changes which arose during the generation of the knockout mutants. Two assays were performed to confirm that *MSH2* and *MLH1* encoded genuine components of the *T. brucei* MMR system. Firstly, in bacteria and mammals, loss of MMR leads to an increase in resistance to alkylating agents such as MNNG. Using increasing concentrations of MNNG, the relative survival of the *MSH2* and *MLH1* mutants compared with cells expressing these genes was assayed. It was clear from the data generated that the MMR-deficient mutants showed increased resistance to the cytotoxic effects of MNNG compared with wild-type cells, indicating that deletion of *MSH2* or *MLH1* resulted in the loss of MMR activity. In both bacteria and eukaryotes MMR is involved in the correction of replication slippage events at microsatellite loci. Loss of MMR results in the second phenotype characteristic of MMR mutants, termed microsatellite instability, because slippage of DNA polymerase at these loci is not corrected. In *T. brucei*, analysis of PCR products amplified from several microsatellite loci

revealed that clones derived from both *MSH2* and *MLH1* homozygous mutants showed a higher frequency of microsatellite mutation than clones derived from cells expressing both of these genes. Observation of alkylation tolerance and microsatellite instability in *T. brucei* *MSH2* or *MLH1* homozygous mutants indicates that these genes encode genuine and active components of the MMR pathway in this organism.

In order to elucidate whether MMR regulates antigenic variation in trypanosomes, an assay to determine the frequency of, and mechanisms used during, VSG switching in the *MSH2* and *MLH1* mutant cell lines was performed (see chapter 4). These data indicated that there was no alteration in either the frequency or profile of the VSG switching mechanisms used in the MMR mutants compared with cells expressing both *MSH2* and *MLH1*, revealing that MMR has little or no influence over antigenic variation. This clearly shows that MMR activity does not underlie the difference in switch rates between monomorphic and pleomorphic cell lines. However, this assay is not sensitive enough to detect any subtle influence that MMR might exert on VSG switching, nor can it be used to detect changes in the hierarchy of VSG expression. In order to detect a subtle regulatory effect imposed by MMR it would be advantageous to develop an *in vitro* VSG switching assay, which would be more reproducible than the current *in vivo* assay. Here the switched variants would be isolated by clearing cells expressing *VSG221* using a specific-antiserum followed by complement-mediated lysis.

Recent advances have made it possible to generate knockout mutants in pleomorphic cells, allowing investigation into the roles of MMR in antigenic variation in non-adapted, high-switching trypanosome strains. Currently, an assay is under development which will allow investigation of the hierarchy of VSG switching in pleomorphic cells (L. Morrison and J. D. Barry, personal communication). Ten single copy *VSG* genes representing telomeric, minichromosomal, and internal (basic-copy) loci have been characterised from *T. brucei* strain ILTat 1.2, and antibodies against these VSG coats raised. The appearance of these VSG coats during chronic infections in mice is being monitored by *in vitro* immune lysis, and the position of each VSG in the VSG switching hierarchy assessed. Comparison of the results of this assay for wild-type and MMR mutant trypanosomes could therefore reveal whether MMR had any influence on the timing of activation, or hierarchy, of these *VSG* genes.

In both bacteria and eukaryotes, MMR has been shown to constrain recombination so that it occurs only between highly related DNA sequences. An assay to investigate whether MMR regulates homologous recombination in trypanosomes was therefore developed (see

chapter 5). A unique sequence was integrated into the *T. brucei* genome, and used as the integration site for a set of antibiotic resistance constructs that possess integration flanks for homologous recombination based on the sequence of this unique site, but which contain increasing numbers of heterologies compared with this sequence. A pilot study using this system was performed to determine the frequency of integration of these increasingly divergent targeting constructs into wild-type and *MSH2* mutant trypanosomes. In wild-type cells, a construct containing 3% divergent flanks integrated at a greatly reduced frequency compared to a construct containing identical flanks, indicating that the probability of integration into the trypanosome genome decreases upon the introduction of sequence heterologies. However, the integration frequency of the 3% divergent construct was greater in *MSH2* mutant cells than wild-type trypanosomes, suggesting that the probability of recombination between homeologous sequences is elevated in MMR-deficient cells, and revealing that MMR does indeed regulate homologous and homeologous recombination in this organism. In the future this assay can be used to further investigate the effect of sequence divergence on the frequency of recombination in *T. brucei*, as well as the anti-recombination activities of the MMR components. Furthermore, this assay could be extended to include constructs containing a greater variety of mismatches, as well as investigating the influence of other DNA repair and recombination factors on trypanosome homologous recombination. To attempt to tie this with the process of *VSG* switching by recombination, it would be interesting to develop an assay to examine rates of recombination between homologous or homeologous markers already integrated in the *T. brucei* genome. This would provide a more realistic measure of the contribution MMR makes to *VSG* translocation reactions.

During this study the functions of only the *T. brucei* *MSH2* and *MLH1* genes have been investigated. However, as stated above, a number of other *mutS*- and *mutL*-related genes are present in the genome. At present, nothing is known about the roles of the polypeptides encoded by these loci. Functional analyses, similar to those described above, would reveal whether the putative *T. brucei* *MSH3*, *MSH8* and *PMS1* genes encode active MMR components. Further to this, the assay to assess the frequency of homeologous recombination could be used to investigate the types of mismatches recognised by the *MSH3* and *MSH8* gene products. Further constructs containing insertions or deletions which would form IDLs upon heteroduplex formation could be generated for this purpose. If the putative *T. brucei* *MSH3* gene encodes a genuine orthologue of *MSH3* from yeast or mammals, inactivation of this gene would be expected to increase the frequency of integration of constructs containing large insertions or deletions. In contrast, deletion of the putative *MSH8* gene should result in an increase in the frequency of integration of

constructs containing simple base changes, but not large insertions or deletions, if this gene encodes an orthologue of MSH6. This assay could also be used to investigate the role of *MLH1* and *PMS1* in homeologous recombination in *T. brucei*.

In eukaryotes, the MutS and MutL homologues function as heterodimers which interact to form ternary complexes. It would therefore be interesting to look for interactions between the putative *T. brucei* polypeptides using yeast two-hybrid analysis. Using this system, it would be possible to test for heterodimer formation between *T. brucei* MSH2 and the putative MSH3 or MSH8 polypeptides, as well as between MLH1 and the putative PMS1 polypeptide. In addition it would be possible to examine the interactions between the putative *T. brucei* MutS and MutL homologues and thereby the potential for ternary complex formation. Furthermore, a putative PCNA gene has recently been identified in *T. brucei* (J. Bell and R. McCulloch, unpublished), and in other eukaryotes this protein also interacts with components of the MMR machinery (Umar *et al.*, 1996; Gu *et al.*, 1998; Kleczkowska *et al.*, 2001). It would therefore be possible to test for interactions between the *T. brucei* MMR polypeptides and PCNA.

Currently, antibodies specific to the *T. brucei* MSH2 and MLH1 proteins are unavailable. The generation of such antibodies would allow further investigation of the functions of MMR in this organism. Initially, it would be possible to determine the expression levels of the MMR proteins in both bloodstream form and procyclic trypanosomes by Western blotting. It would also be interesting to determine whether the MMR proteins are expressed only at certain stages of the cell cycle using fluorescent *in situ* hybridisation (FISH). In trypanosomes, the kinetoplast and nuclear genomes each have a distinct S phase, followed by segregation, prior to cell division (Matthews and Gull, 1994), and fluorescent staining of these genomes can be used to determine the approximate cell cycle stage. Furthermore, it would be possible to investigate whether the MMR proteins localise to newly replicated DNA labelled with 5-bromodeoxyuridine (BrdU).

Far-Western analysis using purified recombinant MMR proteins could be used to confirm the protein-protein interactions seen using yeast two-hybrid analysis. Increasing amounts of each recombinant protein could be spotted on to a nitrocellulose membrane and allowed to hybridise to a second recombinant protein which could be detected using the antibody specific to that protein. Alternatively, recombinant proteins could be used in band-shift assays. Two recombinant MMR components could be mixed and the sizes of the products present in the mixture determined by polyacrylamide gel electrophoresis followed by Western blotting. Identification of a band of increased size, compared to either of the

recombinant proteins, indicates that these proteins form a complex. This approach could also be used to determine the mismatch specificities of the complexes identified, by mixing the protein complexes with radiolabelled DNA duplexes containing a mismatch or IDL, and assaying for an increase in the size of the radiolabelled band. Addition of a MutS-related heterodimer to a mismatch-containing DNA duplex, followed by addition of a MutL-heterodimer might allow the definition of ternary complexes in *T. brucei*. Theoretically, such approaches could be carried out without recombinant protein, using simply *T. brucei* cell extracts.

Several phylogenetic analyses have been performed on the known MutS homologues (Kolodner, 1996; Eisen, 1998; Culligan *et al.*, 2000). These studies suggest that the progenitor of the eukaryotic MutS homologues was transferred from the genome of an endosymbiont into the host cell nucleus early in eukaryote evolution (Figure 6.1; Culligan *et al.*, 2000). Duplication events then led to the evolution of both the MMR- and meiosis-specific MutS homologues. Although the phylogenetic relationships between the MutL homologues have yet to be fully investigated, it is speculated that similar transfer and duplication events occurred during their evolution (Kolodner, 1996; Culligan *et al.*, 2000). In both cases, these events are believed to have occurred prior to the radiation of animals, plants and fungi (Culligan *et al.*, 2000). This knowledge raises several questions about the putative *T. brucei* *MSH3* and *MSH8* genes. Sequence analysis suggests that the *MSH8* gene encodes a member of the MSH6/MSH7 subfamily of MutS homologues. However, in comparison with MSH6 and MSH7 proteins from other eukaryotes, the predicted MSH8 polypeptide appears to exhibit a large N-terminal truncation (Figure 3.21). Furthermore, neither the *T. brucei* MSH3 or MSH8 polypeptides appear to possess the conserved PCNA binding site characteristic of members of the MSH3/MSH6 subfamily (Figure 3.22; Flores-Rozas *et al.*, 2000). Without functional analysis it is impossible to determine the effect these features have on the MMR system in *T. brucei*, but, it would nevertheless be interesting to determine when these sequence differences arose in trypanosome evolution. There are two possibilities: trypanosomes diverged prior to the acquisition of the conserved PCNA binding sites and N-terminal extension; or the trypanosome MSH3 and MSH8 polypeptides once possessed these features but have subsequently lost them.

Investigation of the MutS and MutL homologues present in early-diverging eukaryotes (Baldauf *et al.*, 1996; Philippe and Adoutte, 1998), which include many unicellular parasites, may shed light on the evolution of these features, and indeed the timing of the duplication events which occurred to generate the variety of MutS- and MutL-related proteins found in eukaryotes. To this end, we searched the genome sequencing databases

of a number of unicellular eukaryotic parasites in an attempt to determine which MutS and MutL family members are present in these organisms (Table 6.1). These database searches could not be exhaustive, as not all of the sequencing databases available for each organism were searched, and none of the genome sequencing projects are complete, except that of *Encephalitozoon cuniculi* (Katinka *et al.*, 2001). Furthermore, the identities assigned to the sequences retrieved during these database searches are putative and based solely on their similarity to sequences present in the NCBI protein database. Every species examined contains at least one gene encoding a putative MMR component, suggesting that all these organisms possess the MMR pathway. All three kinetoplastids, including *T. brucei*, *Leishmania major* and *Trypanosoma cruzi*, contain putative homologues of both *MSH3* and *MSH6*. Unfortunately, the ORFs of the *L. major* and *T. cruzi* genes are incomplete, making it impossible to determine if these genes encode polypeptides which also lack the PCNA binding motifs. Neither is it possible to determine whether the *MSH6* homologues from these species possess a similar N-terminal truncation to *T. brucei* *MSH8*. These three species also encode orthologues of *MSH2*, and a comparison of the predicted *T. brucei* and *T. cruzi* polypeptides reveals that they share 65% sequence identity (72% similarity). This suggests that a high degree of sequence conservation will also be found between the other MMR components of these species when their full sequences become available.

*E. cuniculi* has recently been shown to be an atypical member of the fungal kingdom, rather than an early-diverging eukaryote as was previously believed (Katinka *et al.*, 2001). As the genome sequence of this organism is complete, it is evident that this species possesses a highly streamlined MMR system, probably consisting of one MutS-related heterodimer and one MutL-related heterodimer. It is possible that a similar streamlining has taken place among the apicomplexan parasites, as putative homologues of only *MSH2*, *MSH6*, *MLH1* and *PMS1* were identified in any of the databases searches of *Babesia bovis*, *Plasmodium falciparum*, *Theileria annulata* and *Toxoplasma gondii*. Searches of the genome sequencing databases compiled for *P. chabaudi*, *P. knowlesi* and *P. yoelii* also failed to identify any other MutS or MutL homologues (data not shown). While these databases are all incomplete, it seems unlikely that these genes have simply been missed by the sequencing so far performed on these seven related genomes. The earliest-diverging eukaryote examined here, *Giardia lamblia* (Baldauf *et al.*, 1996), also appears to possess only putative homologues of the *MSH2*, *MSH6*, *MLH1* and *PMS1* proteins. However, no conclusion can be drawn about the evolution of the MMR system in this organism until the genome sequencing project is complete, because this was the only diplomonad species examined and more MMR components may yet come to light.

*Entamoeba histolytica* is also thought to have diverged from the main eukaryote lineage prior to the *Kinetoplastida* (Baldauf *et al.*, 1996). While still incomplete, the genome sequence of this organism contains putative homologues of the same MMR components as *T. brucei*, with the exception that, as yet, no gene encoding MSH3 can be identified. Once the genome sequence is complete it will be interesting to determine whether MSH3 is truly absent from this organism, and whether the gene which looks most similar to *MSH6* encodes the N-terminal extension found in MSH6 orthologues from higher eukaryotes. If the *E. histolytica* and *G. lamblia* MSH6 homologues also lack the N-terminal extension and PCNA binding motif, this might indicate that these features evolved after the divergence of the *Kinetoplastida* from the main eukaryote lineage. Furthermore, the absence of genes encoding an MSH3 homologue from the complete *G. lamblia* and *E. histolytica* genome sequences might indicate that *MSH3* arose from an event which duplicated the *MSH6* gene after the divergence of *E. histolytica* but prior to the divergence of the *Kinetoplastida*. Alternatively, if the genomes of both *G. lamblia* and *E. histolytica* encode homologues of MSH3 and MSH6 which possess PCNA binding motifs, and the MSH6 orthologues have the N-terminal extension found in fungi, animals and plants, then this would indicate that the *T. brucei* *MSH3* and *MSH8* genes had lost these features after the divergence of this organism from the main eukaryote line.

*T. brucei* possesses a non-obligatory sexual cycle which occurs in the tsetse fly (Jenni *et al.*, 1986). Stocks undergo genetic exchange during co-transmission through the vector and crossing over events have been detected in the hybrid progeny. However, very little is known about the molecular events underlying this sexual cycle. The genes putatively encoding the meiosis-specific MutS homologues MSH4 and MSH5 in *T. brucei* might represent the first components of the genetic exchange machinery identified in trypanosomes. Unfortunately, it would be impossible to determine the roles of these proteins in knockout mutants as the frequency of genetic exchange in the tsetse fly is not quantifiable. However, cloning of these genes from a number of stocks, including both polymorphic and monomorphic trypanosome lines, might reveal differences between the loci, as mutations may have disrupted the ORFs in monomorphic stocks which cannot complete the life cycle and therefore no longer require the meiotic recombination machinery. In stocks able to undergo genetic exchange, for instance TREU 927/4, the *MSH4* and *MSH5* loci should still be intact, providing circumstantial evidence of the role of MSH4 and MSH5 in the sexual cycle of *T. brucei*. It is interesting to note that during searches of the genome sequencing databases of other unicellular eukaryotic parasites, putative homologues of MSH4 and MSH5 were identified only in *E. histolytica*. This organism is believed to undergo genetic exchange (Sargeant *et al.*, 1988). Not

surprisingly, the meiosis-specific MutS-homologues are absent from *E. cuniculi* reflecting the asexual life cycle of this organism (Vivarès and Méténier, 2001). Unexpectedly, however, these genes also appear to be absent from *P. falciparum*, which has an obligate sexual cycle (Walliker *et al.*, 1987). Searches of the genome sequencing databases compiled for *P. chabaudi*, *P. knowlesi* and *P. yoelii* also failed to identify the meiosis-specific MutS homologues (data not shown). The absence of the MSH4 and MSH5 proteins is not unprecedented in sexual organisms, however, as *D. melanogaster* also lacks these conserved elements (Sekelsky *et al.*, 2000).

Given the complexity of the kinetoplast DNA, it might be expected that this organellar genome would possess its own dedicated DNA repair mechanisms. In yeast, the MSH1 protein appears to recognise and correct premutagenic mismatches in the mitochondrial genome. *MSH1* has also been identified in *A. thaliana*, and may perform a similar function in this organism, but no *MSH1* homologues have been identified in the *C. elegans*, *D. melanogaster* or *H. sapiens* genomes, suggesting that this protein is absent from the animal kingdom. During searches of the *T. brucei*, *L. major* and *T. cruzi* genome sequencing databases, no MSH1 homologues were identified (Table 6.1). However, the complete genome sequences are not yet available for these organisms and it is possible that further MutS and MutL homologues remain to be discovered in these species. Given the high levels of divergence of the putative *T. brucei* *MSH3* and *MSH8* genes, it is also possible that any MSH1 homologue encoded in this organism is sufficiently different to other MutS-related proteins that it is not recognised during BLAST searches.

Alternatively, it may be encoded in the kinetoplast DNA rather than the nuclear genome, and so has yet to be recognised. This is not unprecedented as the mitochondrial DNA of the coral *Sarcophyton glaucum* encodes a MutS homologue which is thought to have been transferred to this organelle from the nucleus (Pont-Kingdon *et al.*, 1998).

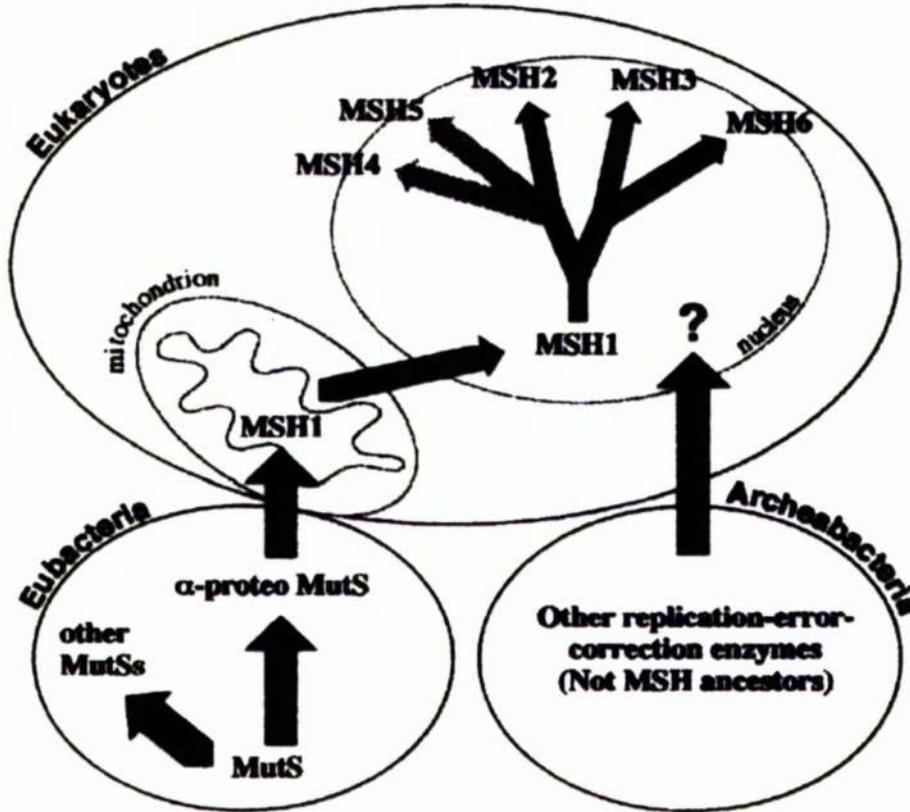
Until the genome sequences of these organisms are complete it will be impossible to draw any firm conclusions about the evolution of the MMR system in early-diverging eukaryotes. Even so, the available data reveals a complex picture of acquisition and loss of MMR components during eukaryote evolution. On a final note, it is interesting that MutL homologues other than MLH1 and PMS1 have yet to be identified in any organisms other than yeast and mammals. This perhaps suggests that the MLH2, MLH3 and human PMS1 proteins evolved after the divergence of the primitive eukaryotes.

The findings of this investigation into the MMR system in *T. brucei* have not only advanced our knowledge of DNA repair in this organism, but have also shed light on the

development of the MMR system during eukaryote evolution. This work has revealed that many of the components involved in MMR in higher eukaryotes had already evolved prior to the divergence of the Kinetoplastids. *T. brucei* contains an active MMR system which functions to repair mismatches formed during replication, and also provides an anti-recombination activity, which constrains homologous recombination to occur between highly-related DNA sequences. However, although antigenic variation in this organism is thought to involve homologous recombination, MMR does not appear to regulate these reactions. Further investigation into the roles of the MMR components, as well as other DNA repair and recombination proteins, may begin to unravel the complex inter-relationships between the DNA repair and recombination systems present in this organism.

**Figure 6.1** Scheme of MutS/MSH evolution

It is thought that *mutS* genes arose and evolved in the eubacteria. The progenitor of the *MSH1* gene was introduced into the eukaryotes by a mitochondrial endosymbiont. This gene was later transferred into the host cell nucleus where it gave rise to all eukaryotic *MSH* genes. (After Culligan *et al.*, 2000)



**Table 6.1 MutS and MutL homologues in unicellular eukaryotic parasites**

BLAST searching of the genome sequencing databases for the organisms shown was performed using MutS- and MutL-related protein sequences from *A. thaliana*, *S. cerevisiae* and *T. brucei* (see section 2.8.5). Putative protein identities was assigned solely on the basis of similarity of the sequences retrieved as above to known proteins as assessed by BLAST searching of the NCBI protein database (<http://www.ncbi.nlm.nih.gov/>). Where published sequences homologous to MutS and MutL were identified in the organisms, the accession numbers are shown in brackets. Since the genome sequencing projects are on-going (except that of *Encephalitozoon cuniculi*, which is complete), further putative components of the MMR system may be identified in these organisms, and the assignment of homologues listed here may need revision.

Organism	MutS Homologues	MutL Homologues
<i>Babesia bovis</i>	MSH2	
<i>Encephalitozoon cuniculi</i>	MSH2 (CAD26200) MSH6 (CAD25790)	MLH1 (CAD26547) PMS1 (CAD26036)
<i>Entamoeba histolytica</i>	MSH2 MSH6 MSH4 MSH5	MLH1 PMS1
<i>Giardia lamblia</i>	MSH2 MSH6	MLH1 PMS1
<i>Leishmania major</i>	MSH2 MSH3 MSH6	MLH1 PMS1
<i>Plasmodium falciparum</i>	MSH2 MSH6	MLH1 PMS1
<i>Theileria annulata</i>	MSH2 MSH6	MLH1
<i>Toxoplasma gondii</i>	MSH2 MSH6	
<i>Trypanosoma brucei</i>	MSH2 (AAK08648) MSH8 (AAK51796) MSH3 MSH4 MSH5	MLH1 (AAK29067) PMS1 (AAK08649)
<i>Trypanosoma cruzi</i>	MSH2 (AAG00261 ) MSH3 MSH6	

## APPENDICES

## Appendix 1 Sequence map showing the *T. brucei* *MSH2* ORF

A map showing the DNA sequence of the 4.5 kb *EcoRI* fragment sub-cloned from the ILTat 1.2 genomic lambda clone 'Plaque 1' into pBluescript II KS to form the plasmid pJB100, which contains the entire *T. brucei* *MSH2* ORF along with its processing flanks (see sections 2.9.1 and 3.1.3). The conceptual translation of the *MSH2* ORF is shown in red beneath the DNA sequence and restriction sites are shown above the sequence in black. Half arrows denote primer sequences: those used to amplify the *MSH2* 5' targeting flank (*MSH2* 5'(A) and *MSH2* 5'(B)) are shown in blue; those used to amplify the *MSH2* 3' flank (*MSH2* 3'(A) and *MSH2* 3'(B)) are shown in green; those used to amplify the *MSH2* internal fragment (*MSH2D5* and *MSH2U2*) are indicated in turquoise; and the degenerate primers used to amplify the conserved ATP-binding site (*MSH2* 5'(1) and *MSH2* 3'(1)) are shown as dashed purple half arrows.

```

EcoRI
|
1  GAATTCTAGAGCCGTTTGTCTAGGGCTTGGCGGCGTAGAAAAAGTGAACGTCGGTTAAC 60
   -----+-----+-----+-----+-----+-----+-----+
   CTTAAGATCTCGGCAAACAGCATCCCGAACCGCCGCATCTTTTCACTTGCAGCCAATTG

61  ATGCAACTGAGAAGGCGGACGATTTGGCAACGCAAGAAGGTAAGAAAACATTTCGTATCT 120
   -----+-----+-----+-----+-----+-----+-----+
   TACGTTGACTCTCCGCTGCTAAACCGTTGCGTTCTTCCATTCTTTGTAAAGCATAGA

121 CAGCATGTCTAACGGAGCAAGATGCAAAGCTAAGTGGGAAATGGTGAATGAACGACTGT 180
   -----+-----+-----+-----+-----+-----+-----+
   GTCGTACAGATTGCCTCGTTCACGTTTCGATTCACCCCTTTACCACTTACTTGCTGACA

181 AATTGCCTCTGAGGACTCAGGAGCCTTTACGCATGCTTGTACGCCTGTTGTTATTGTTG 240
   -----+-----+-----+-----+-----+-----+-----+
   TTAACGGAGACTCCTGAGTCTCGGAAATGCGTACGAAACAGTGGGACAACAATAACAAC

241 GATTTTTTCTGTTTTGGTCCCAACAGGAGCGCTTTTTGCCCTGATCCATTATAGCTTCTT 300
   -----+-----+-----+-----+-----+-----+-----+
   CTAAAAAAGACAAAACCAGGGTTGTCTCGCGAAAAACGGGACTAGGTAATATCGAAGAA

301 CTCATAAACGAGCCCGTTGTAGGCAAGTTGTGTGAGTTATGTGGAGGGTGAACCCATTT 360
   -----+-----+-----+-----+-----+-----+-----+
   GAGTATTTGCTCGGGCAACATCCGTTCAACACAGTCAATACACCTCCACATTGGGTAAG

361 TCTGAGCGTAAGGTTCAACTAACTTTAAAAGACTTAACGAAAAGGGGAGGGAAGCTTCTA 420
   -----+-----+-----+-----+-----+-----+-----+
   AGACTCGCATCCAAGTTGATTGAAATTTTCTGAATTGCTTTTCCCCTCCCTTCGAAGAT

421 TGGTGCGTATATGTAACATATATACGAACATGCACAAAAACTGGTCACCCTGTTAAAAG 480
   -----+-----+-----+-----+-----+-----+-----+
   ACGCACGCATATACATTGTATATATGCTTGTACGTGTTTTTGACCAGTGGGACAATTTTC

481 CAATATGAGAGCTTTGTCGCTATCAAATGCTTCTACCCTTTTTTTTCCACCCTTCAGTT 540
   -----+-----+-----+-----+-----+-----+-----+
   GTTATACTCTCGAAACAGCGATAAGTTTACGAAGATGGGAAAAAAAAGGTGGGAAGTCAA

541 GTGTTTTTCATGGGCTAAAGGTAATATTGACGAACATCACGGGGACACATGATACTTTAA 600
   -----+-----+-----+-----+-----+-----+-----+
   CACAAAAGTACCCGATTTCCATTTATAACTGCTTGTAGTGCCCTGTGTACTATGAAATT

```

## Appendix 1 continued

## MSH2 5' (A)

601 ATAGCAGCGAGTCCAGGAGAAAATAGCGCCTCCTCTGATGACGCTGCACAGATGTGGAAT 660  
 -----+-----+-----+-----+-----+-----+-----+  
 TATCGTCGCTCAGGTCTCTTTTATCGCGGAGGAGACTACTGCGACGTGTCTACACCTTA  
  
 661 GAGGGAATCTTCGATGTTCTTGTGTCACCTCCACTGTCCGCACCAGGGCGCCGTGGAGT 720  
 -----+-----+-----+-----+-----+-----+-----+  
 CTCCCTTAGAAGCTACAAGAACAACGGTGGAGGTGACAGGCGTGGTCCC GCGGCACCTCA  
  
 721 TCTTAACGTTTTGCTCGCAATTTGGTGGATATTTAGGAGGCCTGCCATGAGGGAAATAGT 780  
 -----+-----+-----+-----+-----+-----+-----+  
 AGAATTGCAAAACGAGCGTTAAACCACCTATAAATCCTCCGACGGTACTCCCTTTATCA  
  
 781 GCATCGTGGAGGTCATCAGATAACGGCACTGTTTATCAGTGTGAACTACCGTCGACACG 840  
 -----+-----+-----+-----+-----+-----+-----+  
 CGTAGCACCTCCAGTAGTCTATTGCCGTGACAAATAGTCACAACCTGATGGCAGCTGTGC  
  
 841 ATGTCGGTCTGATGCATGCGATGCCGAACAGCGGAAGGCAGCTACAGGCAAAGGTAGTG 900  
 -----+-----+-----+-----+-----+-----+-----+  
 TACAGCCAGACTACGTACGCTACGGCTTGTCGCCCTCCGTCGATGTCCGTTTCCATCAC  
  
 901 GATGAAACGCTTCTCTGCCGAGTCCTTTCTTCTGTGCCGAAGTGCTTGAGCGAGGCTAGG 960  
 -----+-----+-----+-----+-----+-----+-----+  
 CTACTTTGCGAAGAGACGGCTCAGGAAAGAAGACACGGCTTCACGAACTCGCTCCGATCC  
  
 961 TCAACACTGTGCTCTTCTCCGCTCTACCTCAACGTTAGGCGTACTGTTTCCCTTCTT 1020  
 -----+-----+-----+-----+-----+-----+-----+  
 AGTTGTGACACGAGAAGAGGCGCAGATGGAGTTGCAATCCGCAATGACAAAGGGAAGGAA  
  
 1021 TCAAACCCTTTTCACTGGTCTATCTCTTTATCCCAACGGGTGTGTAGTCAGCAACAAC 1080  
 -----+-----+-----+-----+-----+-----+-----+  
 AGTTTGGGAAAAGTGACCAAGATAGAGAAATAGGGTTGCCACACATCAGTCGTTGTTGA

SalI

|

## MSH2 5' (B)

1081 GTAAAAAGTATTATGAGTGACGACCGCACC CGGCGGTGGTGCAGGCCTTTAATGGTGCC 1140  
 -----+-----+-----+-----+-----+-----+-----+  
 CATTTTTTCATAATACTCACTGCTGGCGCTGGGCCGCCACCAGTCCGAAAATTACCACGG  
                   M S D D R D P A V V Q A F N G A  
  
 1141 GGTGGCGATGATACTTCTGCTTGGGTTGTTCTCGCGTGCATCGGCTGGCTGTTTTATT 1200  
 -----+-----+-----+-----+-----+-----+-----+  
 CCACCGCTACTATGAAGGACGAACGCCAACAAAGAGCGCACGTAGCCGACCGACAAAATAA  
                   G G D D T S C L R L F S R A S A G C F I  
  
 1201 CTAGGCAGCTGGGCTTCACTCGTGGCGGTGAATATGTCAAGTCCACTGCCGTTCTGAAG 1260  
 -----+-----+-----+-----+-----+-----+-----+  
 GATCCGTCGACCCGAAGTGAGCACCGCGCACTTATACAGTTCAGGTGACGGCAAGACTTC  
                   L G S W A S L V A R E Y V K S T A V L K  
  
 1261 AACTGGAGCGGAGTGGATGCCGTCGCCGTTAATGACAGCATCACCCGCGAAGTCATACGT 1320  
 -----+-----+-----+-----+-----+-----+-----+  
 TTGACCTCGCCTCACCTACGGCAGCGGCAATTACTGTCGTAGTGGGCGCTTCAGTATGCA  
                   N W S G V D A V A V N D S I T R E V I R

Appendix 1 continued

1321 GACTGCCTGCTTCGACGCGGTGTTTCTGTTGAACAATACGACCGACAGACGAGCGGGCGGC 1380  
 -----+-----+-----+-----+-----+-----+-----+-----+  
 CTGACGGACGAAGCTGCGCCACAAAGACAACCTGTTATGCTGGCTGTCTGCTCGCCGCCG  
 D C L L R R G V S V E Q Y D R Q T S G G

1381 AGATATGTTTGCATGCGGCGCGGCTCCCCTGGGCACATTGCAGACTTCGAAGCAATGCTC 1440  
 -----+-----+-----+-----+-----+-----+-----+-----+  
 TCTATACAAACGTACGCCGCGCCGAGGGGACCCGTGAACGTCTGAAGCTTCGTTACGAG  
 R Y V C M R R G S P G H I A D F E A M L

1441 TTTGCCTTCGAAGACGCAGAGATACAACCTCATGGCAATTGGGAGTGTGGTTATTGACGAT 1500  
 -----+-----+-----+-----+-----+-----+-----+-----+  
 AAACGGAAGCTTCTGCGTCTCTATGTTGAGTACCGTTAACCCCTCACACCAATAACTGCTA  
 F A F E D A E I Q L M A I G S V V I D D

1501 AAAGCAAATAGGGTTAATGGCCCAGGTGGACAGCACGTCCGGGTGGGTTACGCAGCGCTC 1560  
 -----+-----+-----+-----+-----+-----+-----+-----+  
 TTTGTTTTATCCCAATTACCGGGTCCACCTGTGTCGAGGCCACCCAATGCGTCGCGAG  
 K A N R V N G P G G Q H V R V G Y A A L

1561 AATACAACACTACGGACACTTACGTACGCGGAATATCATGACACACCACAGTTAACAAAC 1620  
 -----+-----+-----+-----+-----+-----+-----+-----+  
 TTATGTTGTGATGCCTGTGAATGCATGCGCCTTATAGTACTGTGTGGTGTCAATTGTTTG  
 N T T L R T L T Y A E Y H D T P Q L T N

1621 CTTGATGTGCTTATGGCGCAGTGTAACTGAAGCAGCTGCTGTACTCTAACACTGACTTT 1680  
 -----+-----+-----+-----+-----+-----+-----+-----+  
 GAACTACACGAATACCGGTCACATTGGACTTCGTCGACGACATGAGATTGTGACTGAAA  
 L D V L M A Q C N L K Q L L Y S N T D F

1681 TCGATGAATAATACGGGTGAGAAGGCTGCAGACTCTGACGAAAGTAGGGAACAGAGTGAT 1740  
 -----+-----+-----+-----+-----+-----+-----+-----+  
 AGCTACTTATTATGCCCACTCTCCGACGTCTGAGACTGCTTTCATCCCTTGTCTCACTA  
 S M N N T G E K A A D S D E S R E Q S D

1741 TTGTTACGAGCACTCAAGCAGCTCTGTGAACGAGCGAACATTACGCTCCAAGAGCGTGGA 1800  
 -----+-----+-----+-----+-----+-----+-----+-----+  
 AACAAATGCTCGTGAGTTTCGTCGAGACACTTGCTCGCTTGTAAATGCGAGGTTCTCGCACCT  
 L L R A L K Q L C E R A N I T L Q E R G

1801 CAAAGTAATCTACCTCATGGAAAGCAGAAGTCCAGGGCCACGAAGCGTAACAGCACTGGC 1860  
 -----+-----+-----+-----+-----+-----+-----+-----+  
 GTTTCATTAGATGGAGTACCTTTTCGCTTTCAGGTCGCTTCGCATTGTGCTGACCG  
 Q S N L P H G K Q K S R A T K R N S T G

1861 CCGAATGGAGAGCTACTTAGCACGCTCGAAGGGATTCTTCGCGTGCCGGAAGATAGACAT 1920  
 -----+-----+-----+-----+-----+-----+-----+-----+  
 GGCTTACCTCTCGATGAATCGTGCAGACTTCCCTAAGAAGCGCACGGCCTTCTATCTGTA  
 P N G E L L S T L E G I L R V P E D R H

1921 GGTTTAAATAGCTTCCCTCTTGCTTCCAGGGCTTTGGAAAGCCTTCTAGAATCCGCCATA 1980  
 -----+-----+-----+-----+-----+-----+-----+-----+  
 CCAAATTTATCGAAGGGAGAACGAAGTCCCAGAACCTTTCGGAAGATCTTAGCGGTAT  
 G L N S F P L A S R A L E S L L E S A I

1981 GATCCATTGATTCCTAATCAACATACATTTTACCTTAAGCATGTTATCCCTTCTACG 2040  
 -----+-----+-----+-----+-----+-----+-----+-----+  
 CTAGGTAAGCTAAGGTGATTAGTTGTATGTAATAAGGAATTCGTACAATAGGGAAGATGC  
 D P F D S T N Q H T F Y L K H V I P S T

## Appendix 1 continued

2041 TTCATGAAGATGGATGCGGCAGCCATCGAGGCGCTGCATATAATCCACCGAAAGCCAGAG 2100  
 -----+-----+-----+-----+-----+-----+-----+  
 AAGTACTTCTACCTACGCCGTCGGTAGCTCCGCGACGTATATTAGGTGGCTTTCGGTCTC  
 F M K M D A A A I E A L H I I H R K P E

2101 GCACGAGGATCGATGCCAACGTCTATTTACTCGTGGTTGAACCGCTGTACTACGGGTATG 2160  
 -----+-----+-----+-----+-----+-----+-----+  
 CGTGCTCCTAGCTACGGTTGCAGATAAATGAGCACCAACTTGGCGACATGATGCCCATAC  
 A R G S M P T S I Y S W L N R C T T G M

2161 GGGTCGCGACTGATGCAGCAGTGGCTACTCCAACCGTTGCGGAGCATCGAAGATATAAAT 2220  
 -----+-----+-----+-----+-----+-----+-----+  
 CCCAGCGCTGACTACGTCGTACCGATGAGGTTGGCAACGCCTCGTAGCTTCTATATTTA  
 G S R L M Q Q W L L Q P L R S I E D I N

2221 CAACGGCTCTCGCTTGTACAGATAATGGTGGAGAGTCCAATCCTAAGGGATGCGCTGATC 2280  
 -----+-----+-----+-----+-----+-----+-----+  
 GTTGCCGAGAGCGAACATGTCTATTACCACCTCTCAGGTTAGGATTCCTACGCGACTAG  
 Q R L S L V Q I M V E S P I L R D A L I

MSH2D5 

2281 ACGCAGGTACTGCGGGCTGCACTGACATGGACAGGTTGAACCGAAAGTTGCAACGCCGC 2340  
 -----+-----+-----+-----+-----+-----+-----+  
 TGCGTCCATGACGCCGCGACGTGACTGTACCTGTCCAACCTGGCTTTC AACGTTGCGGGC  
 T Q V L R R C T D M D R L N R K L Q R R

2341 ACAGTGGCTCTCAAGGACCTGCAATCTATTCTTGTCTTCGCTAATACTGTACCCCTAGCG 2400  
 -----+-----+-----+-----+-----+-----+-----+  
 TGTCACCGAGAGTTCCTGGACGTTAGATAAGAACAGAAGCGATTATGACATGGGGATCGC  
 T V A L K D L Q S I L V F A N T V P L A

2401 GTCGATGTGCTGCGAACATATCATGGCGGGCAGCAGCAGCTTACTTTTGAAGGGATAC 2460  
 -----+-----+-----+-----+-----+-----+-----+  
 CAGCTACACGACGCTTGTATAGTACCGCCCGTGTGTCGTCGAATGAAAACCTCCCTATG  
 V D V L R T Y H G G H D S S L L L K G Y

2461 GTGACACCTCTGGAAGACATTAGCGAGCATCTTCAAATCTGCGCACGTTGATAAATGCA 2520  
 -----+-----+-----+-----+-----+-----+-----+  
 CACTGTGGAGACCTTCTGTAATCGCTCGTAGAAAGTTTAGACGCGTCAACTATTTACGT  
 V T P L E D I S E H L S N L R T L I N A

2521 ACCGTTGACTTGTGAGATGAGAATACCGTGCCTATTAACCCTGAATTTGACGATGACCTC 2580  
 -----+-----+-----+-----+-----+-----+-----+  
 TGGCAACTGAACAGTCTACTCTTATGGCAGCATAATTGGGACTTAAACTGCTACTGGAG  
 T V D L S D E N T V R I N P E F D D D L

2581 AGCTTCCTGGAGCGGCAGCGTCAGAATCTTGTGAAGGCGATTGAAAAAGAAAACCACCGT 2640  
 -----+-----+-----+-----+-----+-----+-----+  
 TCGAAGGACCTCGCCGTCGCAGTCTTAGAACACTCCGCTAACTTTTCTTTTGGTGGCA  
 S F L E R Q R Q N L V K A I E K E N H R

2641 GTGCTGAAACAGTGCAGGATGGACTGAAAAACAAATGAAGTGTGAATATCACGCTAGTTAC 2700  
 -----+-----+-----+-----+-----+-----+-----+  
 CACGACTTTGTGACGCCTACCTGACTTTTTGTTTACTTACACTTATAGTGCATCAATG  
 V L K Q C G W T E K Q M K C E Y H A S Y

Appendix 1 continued

2701 GGT TATGTGTTTCGCGTTCCTCGGAAAGATGACCACCAGGTGCGCACGAGCAAAGAGTTC 2760  
 -----+-----+-----+-----+-----+-----+-----+  
 CCAATACACAAAGCGCAAGGAGCCTTTCTACTGGTGGTCCACGCGTGCTCGTTTCTCAAG  
 G Y V F R V P R K D D H Q V R T S K E F

2761 ATAACGGTAAGCACGGCAAAGGATGGCGTACGGTTCGTATCCGGGCAATTATCATCGCTG 2820  
 -----+-----+-----+-----+-----+-----+-----+  
 TATTGCCATTCTGTGCCGTTTCTACCGCATGCCAAGCATAGGCCCGTTAATAGTAGCGAC  
 I T V S T A K D G V R F V S G Q L S S L

2821 AGCGAGCAGTATAAGGGCATTACTGAGGACTACAAGACACGTCAGCAGGTCTCAAAAAA 2880  
 -----+-----+-----+-----+-----+-----+-----+  
 TCGCTCGTCATATTCCTGTAATGACTCCTGATGTTCTGTGCAGTCGTCCAGGAGTTTTTT  
 S E Q Y K G I T E D Y K T R Q Q V L K K

MSH212

2881 AAGCTCGTTGATACTGTAGCTACATACCTCCCGGTCCTTGATGATGCGAAAGAGCTACTT 2940  
 -----+-----+-----+-----+-----+-----+-----+  
 TTCGAGCAACTATGACATCGATGTATGGAGGGCCAGGAAGTACTACGCTTTCTCGATGAA  
 K L V D T V A T Y L P V L D D A K E L L

2941 GCAGCATTGGATGTGTTTCGCGGCGTGGGCACTCGTAGTGAAGGATTCGTGCGGCGCGATG 3000  
 -----+-----+-----+-----+-----+-----+-----+  
 CGTCGTAACCTACACAAGCGCCGACCCGTGAGCATCACTTCCTAAGCAGCGCCGGCTAC  
 A A L D V F A A W A L V V K D S S R P M

3001 GTGCGTCCCACCGTGAGGGCAACACAGAGTGAGGAGGTGAAAGGAAAACGTTGACAACAAC 3060  
 -----+-----+-----+-----+-----+-----+-----+  
 CACGCAGGGTGGCACTCCCGTTGTGTCTCACTCCTCCACTTTCCTTIGCAACTGTTGTG  
 V R P T V R A T Q S E E V K G N V D N N

3061 AGCGATGGTGCTATTTTGACCATTGTGAACGCCCGACACCCCTTGTGAACTGCGGCAG 3120  
 -----+-----+-----+-----+-----+-----+-----+  
 TCGCTACCACGATAAAACTGGTAACACTTGCGGGCTGTGGGGGAACAACCTGACGCCGTC  
 S D G A I L T I V N A R H P L V E L R Q

3121 CCCGCCTTACACCGAATACTGTACAACCTACCAACGAGGCCAATGCCCTTATAATAACT 3180  
 -----+-----+-----+-----+-----+-----+-----+  
 GGGCGGAAGTGTGGCTTATGACATGTTGAATGGTTGCTCCGGTTACGGGAATATTATTGA  
 P A F T P N T V Q L T N E A N A L I I T

MSH2 5' (1)

3181 GGGCCAAATATGGGAGGTAAGTCAACTTTTCATGAGGAGCATTGGTGTGTTGTGTCGCACTC 3240  
 -----+-----+-----+-----+-----+-----+-----+  
 CCCGGTTTATACCTCCATTTCAGTTGAAAGTACTCCTCGTAACCACAAACACAGCGTGAG  
 G P N M G G K S T F M R S I G V C V A L

SacII

3241 GCCCAAGCTGGGTGCTTTGTTCCCGCGGATTCAGCGGATATTGTTGTCCGTGACCGGATC 3300  
 -----+-----+-----+-----+-----+-----+-----+  
 CGGGTTTCGACCCACGAAACAAGGGCGCCTAAGTCGCCTATAACAACAGGCACTGCGCTAG  
 A Q A G C F V P A D S A D I V V R D A I

## Appendix 1 continued

3301 ATGTGCCGTGTTGGCGCGACAGACCACCTTGCGCAAGGAGTTTCTACCTTCATGGTGGAA 3360  
 -----+-----+-----+-----+-----+-----+  
 TACACGGCACAACCGCGCTGTCTGGTGGAAACGCGTTCCTCAAAGATGGAAGTACCACCTT  
 M C R V G A T D H L A Q G V S T F M V E

SacII  
 |

3361 ATGCTCGAGTCCGCGGCTATGCTCAACTCCGCCACCCAACAGACTCTGGCCATTGTAGAT 3420  
 -----+-----+-----+-----+-----+-----+  
 TACGAGCTCAGGCGCCGATACGAGTTGAGGCGGTGGGTTGTCTGAGACCGGTAACATCTA  
 M L E S A A M L N S A T Q Q T L A I V D

3421 GAACTGGGCCGGGGACGTCAACGTACGATGGATTCCGGTCTTGCCTGGGCCATTGCACAG 3480  
 -----+-----+-----+-----+-----+-----+  
 CTTGACCCGGCCCCCTGCAGTTGCATGCTACCTAAGCCAGAACGGACCCGGTAACGTGTC  
 E L G R G T S T Y D G F G L A W A I A Q


  
 MSH2 3' (1)

3481 GAGGTGGCAGTCAACGCAAAGTCTGCACTTCTCTTTCAACTCATTTCACGAAATGACA 3540  
 -----+-----+-----+-----+-----+-----+  
 CTCCACCGTCAAGTTGCGTTTCAGACGTGAAGAGAAAAGTTGAGTAAAGGTGCTTTACTGT  
 E V A V N A K S A L L F S T H F H E M T

3541 CAACTTGC GGCCCGACATACAAACGTGCGGAACGTTCAATTCGGCGCTGATGTGGATACT 3600  
 -----+-----+-----+-----+-----+-----+  
 GTTGAACGCCGGCTGTATGTTGCACGCCTTGCAAGTAAAGCCGGGACTACACCTATGA  
 Q L A A R H T N V R N V H F G A D V D T

3601 GCCGCTCGGACCTTACGTTTCTCTTACCAACTTCAACCAGGGCCATGCGGGCGCAGCTAC 3660  
 -----+-----+-----+-----+-----+-----+  
 CGGCGAGCCTGGAATGCAAAGAGAATGGTTGAAGTTGGTCCCGGTACGCCCGCGTCGATG  
 A A R T L R F S Y Q L Q P G P C G R S Y

3661 GGATTGTACGTGGCGCAACTCGCGCACATTCCCGACGACGTGCTCGATTTGCGCGGGCAG 3720  
 -----+-----+-----+-----+-----+-----+  
 CCTAACATGCACCGGTTGAGCGGTGTAAGGGCTGCTGCACGAGCTAAAGCGCGCCGTC  
 G L Y V A Q L A H I P D D V L D F A R Q

3721 AAGGCTGTGGAGTTGGAGGATTCGGGGGAGACGAAACCAAGAACC GCGCTCAGGTGCTC 3780  
 -----+-----+-----+-----+-----+-----+  
 TTCCGACACCTCAACCTCCTAAAGCCCCCTCTGCTTTGGTTCTTGGCGGAGTCCACGAG  
 K A V E L E D F G G D E T K N R A Q V L

3781 TTTTCCACCGCCACGCCGAGGTTGTGCAACGGGTGACGGAGTATGCAAAGCGCATCCGC 3840  
 -----+-----+-----+-----+-----+-----+  
 AAAAGGTGGCGGTGCGGCCTCCAACACGTTGCCACTGCCTCATACGTTTCGCGTAGGCG  
 F S T A T P E V V Q R V T E Y A K R I R

3841 GAGTTGAAAGTGGGGAGGGCGACGGTGA CTGAGGGAGGCTGCCCGTCGCCGGCTGTGC 3900  
 -----+-----+-----+-----+-----+-----+  
 CTCAACCTTTCACCCCTCCCGCTGCCACTGAGCTCCCTCCGACGGGCGAGCGGCCGACAG  
 E L E S G E G D G D S R E A A R R R L C


  
 MSH2 3' (A)

3901 AGCGAGATTAAGGAAGATGCATTGTTGTGTCGTCATTGGTGGAGGTGTGAGAAGGTAGAAG 3960  
 -----+-----+-----+-----+-----+-----+  
 TCGCTCTAATTCCTTCTACGTAACAACAGCAGTAACCACCTCCACACTCTTCCATCTTCC  
 S E I K E D A L L S S L V E V

## Appendix 1 continued

3961 ACTTAAAGCGGGAGTCAGAGAAATTTGTTGTTGGGTTTTTCTTTAAAAGCTGCTTTATCA 4020  
 -----+-----+-----+-----+-----+-----+-----+  
 TGAATTCGCCCTCAGTCTCTTTAAACAACAACCCAAAAAGAAATTTTCGACGAAATAGT

4021 TTATTTTTCTATTTTTCTTTTTTTTTCTGTTGTCGTTGTTTTGGGCTTTCCCGCCCCC 4080  
 -----+-----+-----+-----+-----+-----+-----+  
 AATAAAAAGATAAAAAAGAAAAAAAAGACAACAGCAACAAAACCCGAAAGGGCGGGGGG

4081 CCCAGTCCCTTCTCTCTTGATATACCCCTCATCGCATTTGACTTCACCAGCCTTTCGC 4140  
 -----+-----+-----+-----+-----+-----+-----+  
 GGGGTCAGGGAAAGAGAGAATATATGGGGAGTAGCGTAAACTGAAGTGGTCGGAAAGCG

4141 ATTGGATTGGCACGTTAATTTCTTTTCGCACCCGCGTACGACTCACATGATTTGGTGTGC 4200  
 -----+-----+-----+-----+-----+-----+-----+  
 TAACCTAACCGTCAATTAAGGAAAGCGTGGGCGCATGCTGAGTGTACTAAACCACACG

4201 CAGTGTGCTGAAAGAAATGTCACGGCTGAGTCCGAAGGGGGCCGATTGGTCGTTAGTAAT 4260  
 -----+-----+-----+-----+-----+-----+-----+  
 GTCACACGACTTTCTTTACAGTGCCGACTCAGGCTTCCCCGGCTAACCAAGCAATCATT

4261 AGTTCCTTTTTATTATTCTTTTGCTATCCTTACTTCAGTACCTCGTAGAATCTAAGGGGG 4320  
 -----+-----+-----+-----+-----+-----+-----+  
 TCAAGGAAAAATAATAAGAAAACGATAGGAATGAAGTCATGGAGCATCTTAGATTCCCCC

 MSH2 3' (B)

4321 AATGAGGAGGGGGGATGAAAATAACTGAAGAACCAAGGGGGAGCTACAAACGCAAACGG 4380  
 -----+-----+-----+-----+-----+-----+-----+  
 TTACTCCTCCCCCTACTTTTATGACTTCTTGGTTTCCCCCTCGATGTTTGCCTTGCC

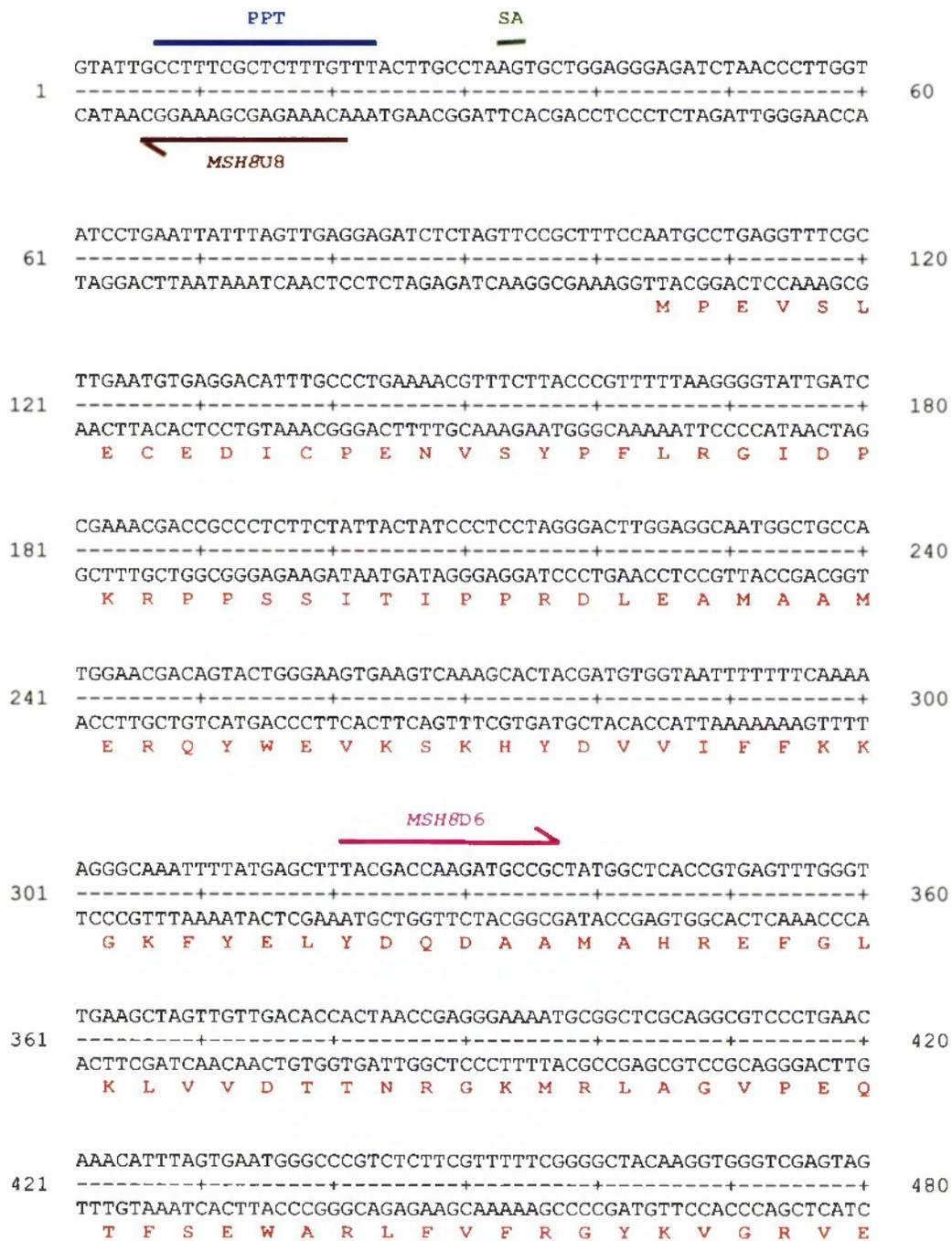
4381 GAAGTTTCTAAAAGGATGTAGAGAGACCAGGGTGCCAAATAACATGGATACTCAGCTGG 4440  
 -----+-----+-----+-----+-----+-----+-----+  
 CTTCAAAGATTTTTCTACATCTCTCTGGTCCCACGGTTTATTTGTACCTATGAGTCGACC

4441 CGGCTGAGGACCTGTACAGGGACAACAGTGGGGGAAATTTTTTTTTGAAAAAAAAAATGT 4500  
 -----+-----+-----+-----+-----+-----+-----+  
 GCCGACTCCTGGACATGTCCCTGTGTCAACCCCTTTAAAAAAAACCTTTTTTTTTAACA

4501 EcoRI  
 |  
 GGCTATGAAGAATTC 4515  
 -----+-----  
 CCGATACTTCTTAAG

## Appendix 2 Sequence map showing the putative *T. brucei* *MSH8* ORF

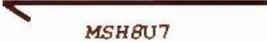
A map showing the DNA sequence of the entire putative *T. brucei* *MSH8* ORF (see sections 2.9.2 and 3.2.3). The conceptual translation of the *MSH8* ORF is shown in red beneath the DNA sequence. The putative 5' trans-splicing signals for the *MSH8* gene are shown: a polypyrimidine tract (PPT) is highlighted in blue and the splice acceptor (SA) site in green. Half arrows denote primer sequences; those used to amplify a control fragment to check the integrity of the cDNA used to investigate the 5' splicing of *MSH8* (*MSH8D6* and *MSH8U6*) are shown in pink; those used in conjunction with the splice-leader specific primer (*MSH8U7* and *MSH8U8*) are indicated in dark red; those used to amplify the *MSH8* internal fragment and 540 bp *MSH8* probe (*MSH8D3* and *MSH8U1*) are indicated in turquoise; and those used to generate the insert of the plasmid pJB301 and confirm the linkage of the initial putative *MSH8* contigs (*MSH8* 5' and *MSH8* 3') are shown in purple.



## Appendix 2 continued

AGCAAATGAAGGAGGAAGGGGAGTCATCAAAAAATGCTCGTCCGAAAGTCGTGCCACGAG  
 481 -----+-----+-----+-----+-----+-----+-----+-----+-----+-----+ 540  
 TCGTTTACTTCCCTCCTTCCCCTCAGTAGTTTTTTACGAGCAGGCTTTCAGCACGGTGCTC  
 Q M K E E G E S S K N A R P K V V P R E

AGCTAGTTGAGATACTTACCCCTGGCACCATAACCGATCCCATGATGATCAGCGGCTATG  
 541 -----+-----+-----+-----+-----+-----+-----+-----+-----+-----+ 600  
 TCGATCAACTCTATGAATGGGGACCGTGGTATTGGCTAGGGTACTACTAGTCGCCGATAC  
 L V E I L T P G T I T D P M M I S G Y G

 MSH8U7

GTGCAGTGTTTCGTACTTGCACGTGTATCCAATGGGGAGCGGTAGTGTGGATGGAATGGCCG  
 601 -----+-----+-----+-----+-----+-----+-----+-----+-----+-----+ 660  
 CAGGTCACAAGCATGAACGTGACATAGGTTACCCCTCGCCATCACACCTACCTTACCGGC  
 A V F V L A L Y P M G S G S V D G M A V

TCGATCTTTCCCGTCGCGTTGTTTTCCACTGTCCCTGCGGTACGAATGGCAAGGAGAGCG  
 661 -----+-----+-----+-----+-----+-----+-----+-----+-----+-----+ 720  
 AGCTAGAAAGGGCAGCGCAACAAAAGGTGACAGGGACGCCATGCTTACCGTTCCTCTCGC  
 D L S R R V V F H C P C G T N G K E S A

CCGCAGGCTTTGTTGAAGAGGTGCTTAATGAAGTTTCTGCGCTCCTTCAACAAATCCGTC  
 721 -----+-----+-----+-----+-----+-----+-----+-----+-----+-----+ 780  
 GGGCTCCGAAACAACCTTCTCCACGAATTACTTCAAAGACGCGAGGAAGTTGTTTAGGCAG  
 A G F V E E V L N E V S A L L Q Q I R P

CCCGTGAAATAATAATTCCGCGTGGGGCTGTTGATGCGCCTGGTGAAGAGCCCAAGGGAT  
 781 -----+-----+-----+-----+-----+-----+-----+-----+-----+-----+ 840  
 GGGCACTTTATTATTAAGGCGCACCCCGACAACCTACGCGGACCACTCCTCGGGTTCCTTA  
 R E I I I P R G A V D A P G E E P K G S

CATTTGGGAGACGTCTATTTGAATGGGTAGAAGGGGAGGGTTTCCAGGTTGAACTGGTTG  
 841 -----+-----+-----+-----+-----+-----+-----+-----+-----+-----+ 900  
 GTAAACCCTCTGCAGATAAAGTACCCATCTTCCCCTCCCAAAGGTCCAACCTGACCAAC  
 F G R R L F E W V E G E G F Q V E L V E

AAGAGGTTGGTACCTCACTTCGCAAACTCCCCCTTGAGGAGAGGAGCTTAAAGGAGGCGG  
 901 -----+-----+-----+-----+-----+-----+-----+-----+-----+-----+ 960  
 TTCTCCAACCATGGAGTGAAGCGTTTGAGGGGGAACCTCCTCTCCTCGAATTTCTCCGCC  
 E V G T S L R K L P L E E R S L K E A G

 MSH8U6

GTCGTTTCTTAGCTCAGTATTTTAGATCGTTAAAGTTGAGTAATGTTGACTCTATACTTT  
 961 -----+-----+-----+-----+-----+-----+-----+-----+-----+-----+ 1020  
 CAGCAAAGAATCGAGTCATAAAATCTAGCAATTTCAACTCATTACAACCTGAGATATGAAA  
 R F L A Q Y F R S L K L S N V D S I L L

TGGAGGCTCGACCATACAACTTTCACCTTTTAAAGCAGCAGGTAACCAGCGGTGTGCCGT  
 1021 -----+-----+-----+-----+-----+-----+-----+-----+-----+-----+ 1080  
 ACCTCCGAGCTGGTATGTTGAAAGTGGAAAATTTGCTCGTCCATTGGTCGCCACACGGCA  
 E A R P Y N F H L L K Q Q V T S G V P S

MSH8 5'

CGAATGATCGAAGTCGTTGCTCCGATTCAACTCTTCTGTGGTACGAACGTGCGGAAGACC  
 1081 -----+-----+-----+-----+-----+-----+-----+-----+-----+-----+ 1140  
 GCTTACTAGCTTTCAGCAACGAGGCTAAGTTGAGAAGACACCATGCTTGCAGCGCTTCTGG  
 N D R S R C S D S T L L W Y E R R E D P

## Appendix 2 continued


  
 1141 CCGGGTTGGTTCCTTGATGCCGCGACAGTAAGTAATCTTGAACCTCGTGGGTAACCTGCGGG 1200  
 -----+-----+-----+-----+-----+  
 GGCCCAACCAAGAAGTACGGCGCTGTCATTCATTAGAACTTGAGCACCCATTGGACGCCC  
 G L V L D A A T V S N L E L V G N L R D

1201 ACGATAGTGAACGCGGATCTCTTTTAAATCTGATCAACCGTTGTTGCACCAACGGTGGAA 1260  
 -----+-----+-----+-----+-----+  
 TGCTATCACTTGCGCCTAGAGAAAAATTAGACTAGTTGGCAACAACGTGGTTGCCACCTT  
 D S E R G S L F N L I N R C C T N G G K

1261 AGCGTCTTTTCCGTTCTTGGATTTTGGACCTTCCGCGTCCCCGCGTGTATCAATGCGA 1320  
 -----+-----+-----+-----+-----+  
 TCGCAGAAAAGGCAAGAACCTAAAACGCTGGAAGGCGCAGGGGCGACAATAGTTACGCT  
 R L F R S W I L R P S A S P R V I N A R

1321 GGCAGGAAGCGGTACGCTTTATCATCGAAAATAATTGAATGATCTCTGGGCCAAAACGG 1380  
 -----+-----+-----+-----+-----+  
 CCGTCCTTCGCCATGCGAAATAGTAGCTTTTATTAACTTACTAGAGACCCGGTTTGGC  
 Q E A V R F I I E N N L N D L W A K T E

1381 AGGAGTCTGCCGATGTTACGACACCAATATGCACCCTAATTCGAGCACTCGTACTTCCG 1440  
 -----+-----+-----+-----+-----+  
 TCCTCAGACGGCTACAATGCTGTGTTTATACGTGGGGATTAAGCTCGTGAGCATGAAGGC  
 E S A D V T T P I C T P N S S T R T S E

1441 AGGGCCCTACCCAAGAATTCACTCAGGCGAGTGGTACGCAAGTGTGGATCGAAGCGGGGA 1500  
 -----+-----+-----+-----+-----+  
 TCCCGGGATGGGTTCTTAAGTGAGTCCGCTCACCATGCGTCACACCTAGCTTCGCCCCCT  
 G P T Q E F T Q A S G T Q C G S K R G R

1501 GGACAACAAACACATTTGAATCTCGATTTACAAACCTTTTCGCCACAGATTTGAGCGCA 1560  
 -----+-----+-----+-----+-----+  
 CCTGTGTTTGTGTAACTTAGAGCTAAATGTTTGGAAAAGCGGTGTCTAAAGCTCGCGT  
 T T N T F E S R F T N L F A T D F E R N

1561 ATCTTTCACGGTTGCGGATTTGAAGGGTGAAGTCTCGCAACAAATAGCCTTCGTAGATCCTC 1620  
 -----+-----+-----+-----+-----+  
 TAGAAAGTGCCAACCGGCTAAACTTCCCACTGAGCGTTGTTTATCGGAAGCATCTAGGAG  
 L S R L A D L K G D S Q Q I A F V D P L

1621 TCGTGCAGTACAAGAAGCATCTTCAACTTATTATTTCAACCGTTGTTGCATTGAGGAGA 1680  
 -----+-----+-----+-----+-----+  
 AGCACGTATGTTCTTCGTAGAAGTTGAATAATAAAGTTGGCAACAACGTAAGCTCCTCT  
 V Q Y K K H L Q L I I S T V V A F E E M

1681 TGCTGGATTGGTCTAACAACGTCCAAAAGGAGTGTGCCCCCTCTTTTGCAGGAAGTGT 1740  
 -----+-----+-----+-----+-----+  
 ACGACCTAACAGATTGTTGCAGGTTTTCCTCACACGGGGGGAGAAAACGTCCTTGACA  
 L D W S N N V Q K E C A P P L L Q E L W

1741 GGGGAACCATGGGCGCAGTGGCTCCTGCTGTGCGCTCAATCAAGGCCTGTTTCGATCGAA 1800  
 -----+-----+-----+-----+-----+  
 CCCCTTGGTACCCGCGTCACCGAGGACGACAGCGGAGTTAGTTCCGGACAAAAGCTAGCTT  
 G T M G A V A P A V A S I K A C F D R K

## Appendix 2 continued

1801 AAGCTGCTGAGGTGTCAGGGGTGATTGTACCATCGCAGGGAGCATGTCCC GCGTACGACG  
 -----+-----+-----+-----+-----+-----+-----+-----+-----+  
 TTCGACGACTCCACAGTCCCCACTAACATGGTAGCGTCCCTCGTACAGGGCGCATGCTGC 1860  
 A A E V S G V I V P S Q G A C P A Y D E

1861 AAGCTACGGAGTGCTTAGATATCATAGAGAAAAAATTGGATGAAGTTTAAAGGGAGCTGC  
 -----+-----+-----+-----+-----+-----+-----+-----+-----+  
 TTCGATGCCTCACGAATCTATAGTATCTCTTTTAAACCTACTTCAAATTCCTCGACG 1920  
 A T E C L D I I E K K L D E V L R E L R

1921 GCGACAACATTTTCAACGGTGCTGCAATCACCTACTCTCATATCGGTCGTGAAAATTTTC  
 -----+-----+-----+-----+-----+-----+-----+-----+-----+  
 CGTGTTGTAAAAGTTGCCACGACGTTAGTGGATGAGAGTATAGCCAGCACTTTAAAAG 1980  
 D N I F N G A A I T Y S H I G R E N F L

1981 TGGTTGAGGTTCCATTGATGGAGGCACCTAAGAGATGTCCCCCTGGGTTTCATCGAGCGAT  
 -----+-----+-----+-----+-----+-----+-----+-----+-----+  
 ACCAACTCCAAGGTAACCTACCTCCGTGGATTCTCTACAGGGGGACCCAAGTAGCTCGCTA 2040  
 V E V P L M E A P K R C P P G F I E R S

2041 CCCGTA CTTCAGCGTGTGT CAGATATACAGTCGCAGGTTTGGAACTCTGGTAGAGGAGC  
 -----+-----+-----+-----+-----+-----+-----+-----+-----+  
 GGGCATGAAGTCGCACACAGTCTATATGT CAGCGTCCAACTTGGAGACCATCTCCTCG 2100  
 R T S A C V R Y T V A G L E P L V E E H

2101 ATAAGCGTGCCAAAACGAAGAAAGCAGATGCCCTCTTACTTGTGTTGAAACATTGCCT  
 -----+-----+-----+-----+-----+-----+-----+-----+-----+  
 TATTCGCACGGTTTTGCTTCTTTTCGTCTACGGGAGAATGAACAACAAGCTTTGTACGGGA 2160  
 K R A K T K K A D A L L L V V R N I A S

2161 CGCATATATTTAATTACTTCCCCGCTCTTTATGAAGCCACAGCGGCTCTTTGCTATTTTG  
 -----+-----+-----+-----+-----+-----+-----+-----+-----+  
 GCGTATATAAATTAATGAAGGGGCAGGAAATACTTCGGTGTGCGCGAGAAACGATAAAAC 2220  
 H I F N Y F P V L Y E A T A A L C Y F D


  
 MSHRD3

2221 ATTGTTTGCTCAGTCTGGCGTCGCTACATAACAGTGCTGTCGCTACTTGCTACCCCGTTG  
 -----+-----+-----+-----+-----+-----+-----+-----+-----+  
 TAACAAACGAGTCAGACCGCAGCGATGTATGGTCACGACAGCGATGAACGATGGGGCAAC 2280  
 C L L S L A S L H T S A V A T C Y P V V

2281 TACAAGAATGTGACCCGGTGCCTACCTGCTTGC GGAAGAACTACGACATCCATTTCTCA  
 -----+-----+-----+-----+-----+-----+-----+-----+-----+  
 ATGTTCTTACACTGCGGCCACGCATGGACGAACGCCTTCTTGATGCTGTAGGTAAGAGT 2340  
 Q E C D A G A Y L L A E E L R H P F L K

2341 AAAGCGATTCTGTCCCAAACACTGTCAATCTTGACGCTACACATGGACGCATTTTGGTGC  
 -----+-----+-----+-----+-----+-----+-----+-----+-----+  
 TTTGCTAAGACAGGGTTTGTGACAGTTAGA ACTGCGATGTGTACCTGCGTAAAACCAGC 2400  
 S D S V P N T V N L D A T H G R I L V L

2401 TGACCGGACCGAATATGGCTGGAAAGAGTACGCTGATGCGCACAGTTGCCGTGAATGTTA  
 -----+-----+-----+-----+-----+-----+-----+-----+-----+  
 ACTGGCCTGGCTTATACCGACCTTCTCATGCGACTACGCGTGTCAACGGCACTTACAAT 2460  
 T G P N M A G K S T L M R T V A V N V I

## Appendix 2 continued

2461 TAATAGCCCAAATGGGTGGACCAGTGTGTTGCTACGTCTATGCGCTTGGCGCCCGTTACCC 2520  
 -----+-----+-----+-----+-----+-----+-----+-----+  
 ATTATCGGGTTTACCCACCTGGTCAAAACGATGCAGATACGCGAACC GCGGGCAATGGG  
 I A Q M G G P V F A T S M R L A P V T R

2521 GTGTTTTCAAAAGGATTGGCGCTCGCGATGCAACGCACAAAGGACAGAGTACCCTTTATG 2580  
 -----+-----+-----+-----+-----+-----+-----+-----+  
 CACAAAAGTGTTCCTAACC GCGAGCGCTACGTTGCGTGTTCCTGTCTCATGGGAAATAC  
 V F T R I G A R D A T H K G Q S T L Y V

2581 TCGAGTTAAGTGAGACAGCCGAGATTGTGCGGTTTGCCGGTCCATGGAGTCTTTGTTTAG 2640  
 -----+-----+-----+-----+-----+-----+-----+-----+  
 AGCTCAATTCACTCTGTGCGCTCTAACACGCCAAACGGCCAGGTACCTCAGAAACAAATC  
 E L S E T A E I V R F A G P W S L C L V

2641 TGGATGAACTCGGCAGGGGGACATCTACGCACGACGGCTACACAATAGCGCACGCAATGC 2700  
 -----+-----+-----+-----+-----+-----+-----+-----+  
 ACCTACTTGAGCCGTCCCCCTGTAGATGCGTGCTGCCGATGTGTTATCGCGTGCGTTACG  
 D E L G R G T S T H D G Y T I A H A M L

2701 TCGCCGCCATGAAGAAAAGGCACCCAGTGCCGCCCTTCTACTCTTCTCTACCCACTACC 2760  
 -----+-----+-----+-----+-----+-----+-----+-----+  
 AGCGGCGGTACTTCTTTTCCGTGGGTACGCGGGGAAGATGAGAAGAGATGGGTGATGG  
 A A M K K R H P V P P L L L F S T H Y H  
 ← MSHR1

2761 ACGCGCTCGCGCAGGAGGAGCACAAGTCCATGCAGAAATCCACTTCTCAGCAGCATCTG 2820  
 -----+-----+-----+-----+-----+-----+-----+-----+  
 TGCGCGAGCGCTCCTCCTCGTGTTCAGGTACGTCTTTAGGTGAAGGAGTCTCGTAGAC  
 A L A Q E E H K S M Q K S T S S A A S E

2821 AAActGGGGGGTTCAGCTCGGTTACATGGACTTTGCCGTTTCTGCTGCAAGTGACAGCA 2880  
 -----+-----+-----+-----+-----+-----+-----+-----+  
 TTTGACCCCCCAAGTTCGAGCCAATGTACCTGAAACGGCAAAGACGACGTTCACTGTGCT  
 T G G V Q L G Y M D F A V S A A S D S N

2881 ATATACCGACCATTACTTTTCTTTATCGCCTTGTCGCCGAATCTGCGCGAGAAGCTATG 2940  
 -----+-----+-----+-----+-----+-----+-----+-----+  
 TATATGGCTGTAATGAAAAGAAATAGCGGAACCGGGCCTTAGACGCGCTCTTCGATAC  
 I P T I T F L Y R L V P G I C A R S Y G

2941 GCGTTGAGGTGGCGCTGCTTGCCGGTATTTCCCCGGGGTTGTAAACACAGCACGGGTTA 3000  
 -----+-----+-----+-----+-----+-----+-----+-----+  
 CGCAACTCCACCGCAGCAACGGCCATAAAGGGGGCCCCAACATTTGTGTGCTGCCCAAT  
 V E V A L L A G I S P G V V N T A R V K  
 →

3001 AGTCTCTGGAGCTCGCCAAGTGGTACGAACGACAAAGGGACCTGGGTACGGTTCGGGGGT 3060  
 -----+-----+-----+-----+-----+-----+-----+-----+  
 TCAGAGACCTCGAGCGGTTACCATGCTTGCTGTTTCCCTGGACCCATGCCAAGCCCCCA  
 S L E L A K W Y E R Q R D L G T V R G F

MSH8 3'

3061 TCATCACGCCAGCGGGACACAGTTCTCGCATCGGTAGAAACCCAGGTGACGCCTTTACC 3120  
 -----+-----+-----+-----+-----+-----+-----+-----+  
 AGTAGTGCGGGTCGCCCTGTGTCAAGAGCGTAGCCATCTTTGGGTCCACTGCGGAAATGG  
 I T P S G T Q F S H R

**Appendix 2 continued**

3121 AATGAGTCCGAGGACTTGGCGGAGAGGAACATAGATCAAATTTACCCCTGTTCTGATGC  
-----+-----+-----+-----+-----+-----+  
TTACTCAGGCTCCTGAACCGCCTCTCCTTGTATCTAGTTTAAAGTGGGGACAAGACTACG 3180

### Appendix 3 Sequence map showing the putative *T. brucei* *MSH3* ORF

A map showing the DNA sequence of the putative *T. brucei* *MSH3* ORF, derived from The Wellcome Trust Sanger Institute *T. brucei* genome sequencing database contig TRYP9.0.000864 ([http://www.sanger.ac.uk/Projects/T\\_brucei/](http://www.sanger.ac.uk/Projects/T_brucei/); see sections 2.8.4 and 2.9.3). The conceptual translation of the *MSH3* ORF is shown in red beneath the DNA sequence.

```

1  ATGAGCAAACGCCGCCGTGGTGACGATGAAAGCGATATTTCCGGAAGGCTATCGAGCTG 60
   -----+-----+-----+-----+-----+-----+-----+
   TACTCGTTTGGCGGGCCACCACTGCTACTTTCGCTATAAAAGGCCTCCGATAGCTCGAC
   M S K R R R R G D D E S D I F R K A I E L

61  CTGTGTCACCGCCGGGCCACAACCTGAAGAAGTTCAGTTGCAACTAAGCTCACGCCTTGG 120
   -----+-----+-----+-----+-----+-----+-----+
   GACACAGTGGCGGCCCGGTGTTGACTTCTTGACAGTCAACGTTGATTCGAGTGCAGAAAC
   L C H R R A T T E E L S V A T K L T P L

121 GAAAGGCAGGTGGTGTGCCTCAAGGAGTCCCTTCCGCCCGATGTTATTCTGATGGTAGCG 180
   -----+-----+-----+-----+-----+-----+-----+
   CTTTCCGTCCACCACACGGAGTTCCTCAGGGAAGGGGGCTACAATAAGACTACCATCGC
   E R Q V V C L K E S L P P D V I L M V A

181 TGTGGTTACCGTGTGAAGTTTACGGTCGTGACAGTCGCGTCGTGAGCCCGGCTTTGGG 240
   -----+-----+-----+-----+-----+-----+-----+
   ACACCAATGGCACACTTCAAATGCCAGCACTGTCAGCGCAGCAGTCGGCGGCAAAACCC
   C G Y R V K F Y G R D S R V V S R R F G

241 ATAATGTGCATTCAAGCAACTCCATTGAGTATAGCAGTGTTCCTTACACAGGGGTCAAT 300
   -----+-----+-----+-----+-----+-----+-----+
   TATTACACGTAAGTTCGTTGAGGTAAGCTCATATCGTCACAAGGAATGTGTCCCCAGTTA
   I M C I Q A T P F E Y S S V P Y T G V N

301 ATTTATGTGCGCCGCTGGTAGCGATGGGTTATCGTGTGCTTTTGGGATCAAGAGAGC 360
   -----+-----+-----+-----+-----+-----+-----+
   TAAATACACGCGGGGACCATCGCTACCCAATAGCACAACGAAAACGCTAGTTCTCTCG
   I Y V R R L V A M G Y R V A F A D Q E S

361 GCTTCTATTCGGTCGACCAGTGGTAACTCGAAGGGGTTGTTTTCCCGAGAGATTGGGCGA 420
   -----+-----+-----+-----+-----+-----+-----+
   CGAAGATAAGCCAGCTGGTCACCATTGAGCTTCCCAACAAAAGGGCTCTTAACCCGCT
   A S I R S T S G N S K G L F S R E I G R

421 GTGTACAGCCGAGGAACGATGTTGCCTGATGAGGTGGTTGTCACCGCTGGAACCCCTCAA 480
   -----+-----+-----+-----+-----+-----+-----+
   CACATGTCGGCTCCTTGCTACAACGGACTACTCCACCAACAGTGGCGACCTTGGGGAGTT
   V Y S R G T M L P D E V V V T A G T P Q

481 GAAGGAAGCGCTACCGGTCAGGGGGATGAAGGGGCACCGGAAGGAGGAGATCCCGTTGGT 540
   -----+-----+-----+-----+-----+-----+-----+
   CTTCCCTTCGCGATGGCCAGTCCCTTACTTCCCGTGGCCTTCCCTCCTTAGGGCAACCA
   E G S A T G Q G D E G A P E G G D P V G

541 GTGGAGGAACCTCTCCCTTGAAGGAGGGATCTTCGGAATTGTTTATTTGCTTCTTGTGG 600
   -----+-----+-----+-----+-----+-----+-----+
   CACCTCCTTGAAGAGGGGAACCTCCTCCCTAGAAGCCTTAACAAATAAACGAAGAACACC
   V E E L L P L K E G S S E L F I C F L W

```

## Appendix 3 continued

601 CCCTCAACTGGGTGGGCGGGTGTGAGCCTGGTCGAAAGTGACGCTTCTCAGTTTCGTACAG  
 -----+-----+-----+-----+-----+-----+-----+-----+  
 GGGAGTTGACCCAGCCGCCACACTCGGACCAGCTTCACTGCGAAGAGTCAAAGCAGTGC  
 P S T G S A G V S L V E V T L L S F V T 660

661 TACTCTCTTTCTCGTCATCACCTGTCTTCAGGGGGAGAGCTTTTGGACATACTCAATCGC  
 -----+-----+-----+-----+-----+-----+-----+  
 ATGAGAGAAAGAGCAGTAGTGGACAGAAGTCCCCCTCTCGAAAACCTGTATGAGTTAGCG  
 Y S L S R H H L S S G G E L L D I L N R 720

721 TACAATATTGTTGAGGTGGTCTGTCTTTATGATGAAGACGGAGCGGGAGCTCGACCGG  
 -----+-----+-----+-----+-----+-----+-----+  
 ATGTTATAACAACCTCCACCAAGACAAAATACTACTTCTGCCTCGGCCCTCGAGCTGGCC  
 Y N I V E V V L F Y D E D G A R E L D R 780

781 CATCAGAACGCAGCAGCTACCGACGGCACTTTGCCTCCTTTTCGCTTGTGCGGTTTACCG  
 -----+-----+-----+-----+-----+-----+-----+  
 GTAGTCTTGCCTCGTTCGATGGCTGCCGTGAAACGGAGGAAAAGCGAACACGCCAAATGCC  
 H Q N A A A T D G T L P P F R L C G L P 840

841 GAGGCGTTTTTACACACCCTTAAACACAATTCTATCGCTTCATCATGGACCGACGGTTAAC  
 -----+-----+-----+-----+-----+-----+-----+  
 CTCCGCAAAATGTGTGGGAATTTGTGTTAAGATAGCGAAGTAGTACTGGCTGCCAATTG  
 E A F Y T P L N T I L S L H H G P T V N 900

901 GGTGAAGAGGATAATAATTCTGTAAGTGTCTGCACCTCGCCATTCATCGGTTCTATCGAT  
 -----+-----+-----+-----+-----+-----+-----+  
 CCACTTCTCCTATTATTAAGACATTGACAGACGTGGAGCGGTAAGTAGCCAAGATAGCTA  
 G E E D N N S V T V C T S P F I G S I D 960

961 GACTCAATTGCAGAGTATTTAAAGCCTAGTCGATTTGATACGACCTTCAGGAAAATGTGC  
 -----+-----+-----+-----+-----+-----+-----+  
 CTGAGTTAACGTCTCATAAATTTTCGATCAGCTAAACTATGCTGGAAGTCTTTTACACG  
 D S I A E Y L K P S R F D T T F R K M C 1020

1021 TCTAAATCTCCGCCGTTACTTGC CGGAGCACGGCCGGTGGGGTGGCTGAGTTGGTAATG  
 -----+-----+-----+-----+-----+-----+-----+  
 AGATTTAGAGCGCGCAATGAACGCGCCTCGTGCCGGCCACCCACCGACTCAACCATTAC  
 S K S P P L L A R S T A G G V A E L V M 1080

1081 GAGATGCCCGGCACCACGATGAGCGGTTGGACATATTTACAGCAGCATTGGGCTAAAA  
 -----+-----+-----+-----+-----+-----+-----+  
 CTCTACGGGCCGTGGTGTACTCGCGCAACCTGTATAAAGTGTCTGTAACCCGATTTT  
 E M P G T T M S A L D I F H S S I G L K 1140

1141 GGTTCTCTGCTGGCTCTATTGGACCACAGCCTTACGGTGCCTGGTCTGCGGCGGTTGCGT  
 -----+-----+-----+-----+-----+-----+-----+  
 CCAAGAGACGACCGAGATAACCTGGTGTGCGAATGCCACGGACAGACGCCGCCAACGCA  
 G S L L A L L D H S L T V P G L R R L R 1200

1201 TCGTGGCTGGCGGCCCGCTCTGCGACTTACGGGCCATTAATTCTCGGCGTGAGGCGGTG  
 -----+-----+-----+-----+-----+-----+-----+  
 AGCACCGACCGCGGGGCGAGACGCTGAATGCCCGGTAATTAAGAGCCGCACTCCGCCAC  
 S W L A A P L C D L R A I N S R R E A V 1260

1261 GCGTTTTTACTTTCGCGCGAAGGGGGTGAAGTCAAGTGGTGGGCTGCTGCGAGAGTTTGCC  
 -----+-----+-----+-----+-----+-----+-----+  
 CGCAAAAATGAAGCGCCGCTTCCCCCACTGAGTCACCACCCGGACGACGCTCTCAAACGG  
 A F L L R G E G G D S V V G L L R E F A 1320

## Appendix 3 continued

1321 AAGTTTGGTGATCTCGAGGCAACACTGGGGAAGCTGCGGGCGCAGCGGTGCACTGTAACC 1380  
 -----+-----+-----+-----+-----+-----+-----+  
 TTCAAACCACTAGAGCTCCGTTGTGACCCCTTCGACGCCCGCGTCGCCACGTGACATTGG  
 K F G D L E A T L G K L R A Q R C T V T

1381 GATTATCTTCGTTTGTGGCGTGGGTGAAAGTGACGCATAAACTCGCTCTTGACATTTTA 1440  
 -----+-----+-----+-----+-----+-----+-----+  
 CTAATAGAAGCAAACAACGCACGCCACTTTCCTGCGTATTGAGCGAGAAGTGTAAAAT  
 D Y L R L L R A V K V T H K L A L D I L

1441 TCACTTTGTGGCGAAGGCATGTGCGATCAGATTCGTGATGCTCTTGTGGCGTCACGTCT 1500  
 -----+-----+-----+-----+-----+-----+-----+  
 AGTGAACACCGCTTCCGTACACGCTAGTCTAAGCACTACGAGAACAACGGCAGTGCAGA  
 S L C G E G M C D Q I R D A L V A V T S

1501 GAGAATGTCGAGTTGTTCCCTCAATCGTGTAAGTGTGAGTAAACTTGGATGCGGATTCC 1560  
 -----+-----+-----+-----+-----+-----+-----+  
 CTCTTACAGCTCAACAAGGAAGTTAGCACATTACACTCAATTTGAACCTACGCCTAAGG  
 E N V E L F L Q S C K C E L N L D A D S

1561 CCTCAGGAGTACTACGCAGCACTTGGTTCCCCCTACCCGACCTTTTGCAGACTCATGCT 1620  
 -----+-----+-----+-----+-----+-----+-----+  
 GGAGTCCTCATGATGCGTCGTGAACCAAGGGGGATGGGCTGGAAAACGTCTGAGTACGA  
 P Q E Y Y A A L G S P L P D L L Q T H A

1621 AAGGAGCGCGACGAAGTGCTACGTGCGTTGGACGTGGAGCTGGAATGCATTGAAAGACA 1680  
 -----+-----+-----+-----+-----+-----+-----+  
 TTCCTCGCGTCTTACGATGCACGCAACCTGCACCTCGACCTTACGTAAGCTTCTGT  
 K E R D E V L R A L D V E L E C I R K T

1681 CTCAAGTTACCTGCGTTGGAGTACCGTACAATCGCTGGCACCACATTCATTGTGGACGTG 1740  
 -----+-----+-----+-----+-----+-----+-----+  
 GAGTTCAATGGACGCAACCTCATGGCATGTTAGCGACCGTGGTGTAAAGTAACACCTGCAC  
 L K L P A L E Y R T I A G T T F I V D V

1741 CCCAACGTAAGGGCCAATGATGCACCGAAAGAGTGGATTGTGCTGACAAGAACAAGACT 1800  
 -----+-----+-----+-----+-----+-----+-----+  
 GGGTTGCATTCCCGGTTACTACGTGGCTTTCTCACCTAACACGACTGTTCTTGTCTGA  
 P N V R A N D A P K E W I V L T R T K T

1801 CATGTGCGATTTCACTCCAAGAATTGTCAATCTTACCGTGGAGCTCTGCTCCGCCAAG 1860  
 -----+-----+-----+-----+-----+-----+-----+  
 GTACACGCTAAAGTGTGAGGTTCTTAACAGTTAGAATGGCACCTCGAGACGAGGCGGTTTC  
 H V R F H T P R I V N L T V E L C S A K

1861 GAACGCCTTGCCATTGCAGCAAATGAGGCATGGTTAGCGAAGCAAGCCGAGTTGGAAGGT 1920  
 -----+-----+-----+-----+-----+-----+-----+  
 CTTGCGGAACGGTAACGTCGTTTACTCCGTACCAATCGCTTCGTTCCGCTCAACCTTCCA  
 E R L A I A A N E A W L A K Q A E L E G

1921 TCTGTGACACGATGGAGATATTTAAGAGTGTCAATTAACCTCTGTTGCTGTGCTTGACGCA 1980  
 -----+-----+-----+-----+-----+-----+-----+  
 AGACAGCTGTGCTACCTCTATAAATTTCTCACAGTAATTGAGACAACGACACGAAGTGCCT  
 S V D T M E I F K S V I N S V A V L D A

## Appendix 3 continued

1981 CTTCACTGTCTTGCTGTGCGCATCTTCGGCCCCCTGGATACACTGCTCCATCCATCTCAGAC 2040  
 -----+-----+-----+-----+-----+-----+-----+-----+  
 GAAGTGACAGAACGACAGCGTAGAAGCCGGGGACCTATGTGACGAGGTAGGTAGAGTCTG  
 L H C L A V A S S A P G Y T A P S I S D

2041 GATGAGCAGTCGATTGTGATCCGCGACGGACGCCACCCGATGCTTGAGTCGCTCATGCCGT 2100  
 -----+-----+-----+-----+-----+-----+-----+-----+  
 CTACTCGTCAGCTAACACTAGGCGCTGCCTGCGGTGGGCTACGAACTCAGCGAGTACGCA  
 D E Q S I V I R D G R H P M L E S L M R

2101 GGGGGCTACGTGGGGTGTGACGTTTCCCTTGTCAAGCACGGCGCGTGGATTCTCACAGGG 2160  
 -----+-----+-----+-----+-----+-----+-----+-----+  
 CCCCCGATGCACCCACACTGCAAAGGGAACAGTTCGTGCCGCGCACCTAAGAGTGTCCC  
 G G Y V G C D V S L V K H G A W I L T G

2161 CCCAACATGGGAGGTAAATCAGCTCTGATGCGTATGGTAGGTACGTTTGTGGTATTGGCA 2220  
 -----+-----+-----+-----+-----+-----+-----+-----+  
 GGGTTGTACCTCCATTTAGTCGAGACTACGCATACCATCCATGCAAACACCATAACCGT  
 P N M G G K S A L M R M V G T F V V L A

2221 CAACTGGGTTGCTACGTGCCTGCCAAGTCTGCGCAACTGCCCTTGTPTGGGGCTGTATAC 2280  
 -----+-----+-----+-----+-----+-----+-----+-----+  
 GTTGACCCAACGATGCACGGACGTTTCAGACGCGTTGACGGGAACAAACCCGACATATG  
 Q L G C Y V P A K S A Q L P L F G A V Y

2281 TGTCGAATGGGGTCGAGCGATTCTCTGCTGGAGGGAAGTTCACCTTTTGAAGAGATG 2340  
 -----+-----+-----+-----+-----+-----+-----+-----+  
 ACAGCTTACCCAGCTCGCTAAGAGACGACCTCCCTTCAAGGTGGA AAAACTTTCTCTAC  
 C R M G S S D S L L E G S S T F L K E M

2341 GAGGAAACAAGTCGGATCCTCCGCTCTGAGATTGTGTCCTCTTCGCTTGTACTACTGGAT 2400  
 -----+-----+-----+-----+-----+-----+-----+-----+  
 CTCCTTTGTTTACGCTTAGGCGAGACTCTAACACAGGAGAAGCGAACATGATGACCTA  
 E E T S R I L R S E I V S S S L V L L D

2401 GAACTCGGCAGGGGAACGAGCAGTTATGATGGTATTGCTATTGCCGCTGCGACACTGGAG 2460  
 -----+-----+-----+-----+-----+-----+-----+-----+  
 CTTGAGCCGTCCTTGTCTGTC AATACTACCATAACGATAACGGCGACGCTGTGACCTC  
 E L G R G T S S Y D G I A I A A A T L E

2461 TATCTGCTTCGGAAGGGTGAACAACATTCTTCGTGACGCACTATTCGCAACTCTGTGAG 2520  
 -----+-----+-----+-----+-----+-----+-----+-----+  
 ATAGACGAAGCCTTCCCACGTTGTTGTAAGAAGCACTGCGTGATAAGCGTTGAGACACTC  
 Y L L R K G A T T F F V T H Y S Q L C E

2521 CCCTATGTCAATTCCAGCAACAACGGACTGGTGTGTCATGCTATTATATGGGATTTTCATGAG 2580  
 -----+-----+-----+-----+-----+-----+-----+-----+  
 GGGATACAGTTAAGGTCGTTGTTGCCTGACCACAGTACGATAATATACCCTAAAAGTACTC  
 P Y V N S S N N G L V S C Y Y M G F H E

2581 GAAAAGATTGTGTCTAGAGAGGGCGAGGGGGAGGTAAAAATTGTATTTACTTACAAACCA 2640  
 -----+-----+-----+-----+-----+-----+-----+-----+  
 CTTTTCTAACACAGATCTCTCCCGCTCCCCCTCCATTTTAAACATAAATGAATGTTTGGT  
 E K I V S R E G E G E V K I V F T Y K P

2641 ACACTTGGGGTCACGCCTTCTAGCTTTGGTGC GCGCGTGGCCCGCATGGCTGGGTTGCC 2700  
 -----+-----+-----+-----+-----+-----+-----+-----+  
 TGTGAACCCAGTCGCGAAGATCGAAACCACGCGCGCACCGGGCGTACCGACCCAACGGG  
 T L G V T P S S F G A R V A R M A G L P

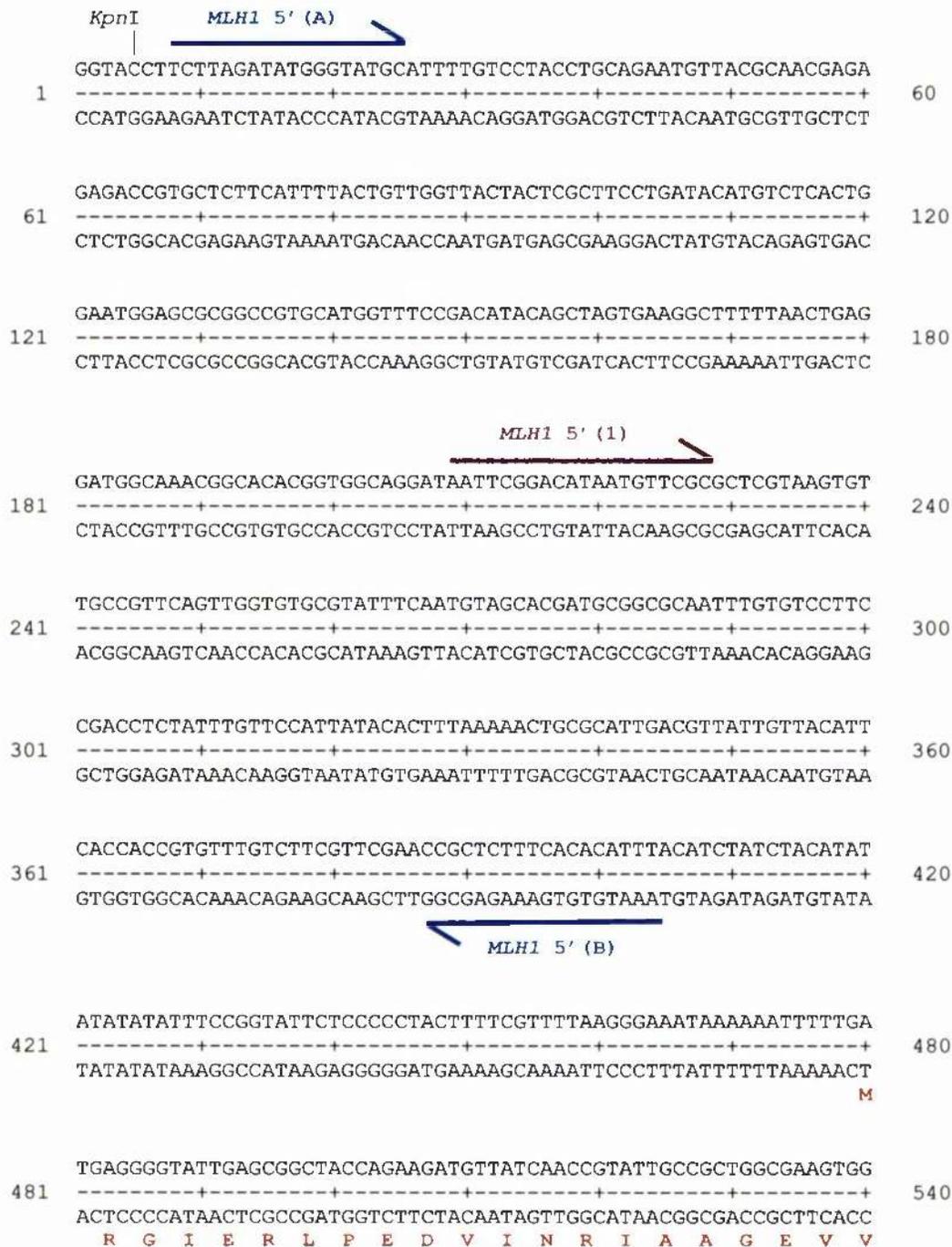
## Appendix 3 continued

2701 TCTGCAGTTGTAAGCGCAGCTGATCAGCGAGGAGGCGGAACGCACTCACGACTGG 2760  
-----+-----+-----+-----+-----+  
AGACGTCAACATTGACTTCGCGTCGACTAGTCGCTCCTCCGCCTTGCGTGAGTGCTGACC  
S A V V T E A Q L I S E E A E R T H D W

2761 TGCTTGTCTCTTCTTGAATTGCGGCGCTTCGTCAAAAACAATTCTT 2806  
-----+-----+-----+-----+  
ACGAACAGAGAAGAAGAACTTAACGCCGCGAAGCAGTTTTTGTTAAGAA  
C L S L L E L R R F V K N N S

#### Appendix 4 Sequence map showing the *T. brucei* *MLHI* ORF

A map showing the DNA sequence of the 4.0 kb *KpnI* fragment sub-cloned from the ILTat 1.2 genomic lambda clone 'Plaque 6' into pBluescript II KS to form the plasmid pJB200, which contains the entire *T. brucei* *MLHI* ORF along with its processing flanks (see section 2.10.1 and 3.3.2). The conceptual translation of the *MLHI* ORF is shown in red beneath the DNA sequence and restriction sites are shown above the sequence in black. Half arrows denote primer sequences: those used to amplify the *MLHI* 5' targeting flank (*MLHI* 5'(A) and *MLHI* 5'(B)) are shown in blue; those used to amplify the *MLHI* 3' flank (*MLHI* 3'(A) and *MLHI* 3'(B)) are shown in green; those used to amplify the 472 bp *MLHI* probe or *MLHI* internal fragment (*MLHIU1* and *MLHID6*) are shown in turquoise; and those used to confirm the linkage of the two putative *MLHI* contigs (*MLHI* 5'(1) and *MLHI* 3'(1)) are indicated in purple.





## Appendix 4 continued

1201 AAGATAACTGCACGCCAGAGAGTGTGCTGGCAACGTCTGGACCATCGGGTGAAGGGCGTT 1260  
 -----+-----+-----+-----+-----+-----+-----+-----+  
 TTCTATTGACGTGCGGTCTCTCACACGACCGTTGCAGACCTGGTAGCCCACTCCCAGAA  
 D N C T P E S V L A T S G P S G E G R F

1261 TCCTTATCACAGGTTACACCAGCGATATAACACTCGCTAGCCGTAATCGTATCTTTGCG 1320  
 -----+-----+-----+-----+-----+-----+-----+-----+  
 AGGAATAGTGTCCAATGTGGTCGCTATATTGTGAGCGATCGGCATTTAGCATAGAAACGC  
 L I T G Y T S D I T L A S R K S Y L C V

1321 TATTTCGTCAACAATCGGCTGGTGGACAGTACGGCCATTGCGCGGGCTCTTGACGCAGTGT 1380  
 -----+-----+-----+-----+-----+-----+-----+-----+  
 ATAAGCAGTTGTTAGCCGACCACCTGTGATGCCGGTAAGCGGCCCGAGAAGTGCCTCACA  
 F V N N R L V D S T A I R R A L D A V Y

1381 ACAGTGGTGTGTTGGTGGCGGGGAACCGCCCGTTTACCGTCCTCTTCGTACGGTGCCAC 1440  
 -----+-----+-----+-----+-----+-----+-----+-----+  
 TGTCAACCACACAACCACGCGCCCTTGGCGGGCAAATGGCAGGAGAAGCAGTGCCACGGTG  
 S G V L V R G N R P F T V L F V T V P P

1441 CTGACAGGGTGGACGTGAATATTCACCCAACAAAACATGAGGTCTGTCTTCTTGATGAAG 1500  
 -----+-----+-----+-----+-----+-----+-----+-----+  
 GACTGTCCCACCTGCACCTATAAGTGGGTTGTTTTGTACTCCAGACAGAAGAAGTACTTTC  
 D R V D V N I H P T K H E V C L L D E E

1501 AAATAATTGTATCGCAACTGTCCGAATGCGTTCAGGGTGTCTTCAGGCGTCTGCGGCTC 1560  
 -----+-----+-----+-----+-----+-----+-----+-----+  
 TTTATTAACATAGCGTTGACAGGCTTACGCAAGTCCCACGAGAAGTCCGCAGACGCCGAG  
 I I V S Q L S E C V Q G A L Q A S A A R

1561 GCAGGCAGATGGATATTCGGCAGATTCACTCCAAGGCGGTTATGCTAGGCGATAGGGAGA 1620  
 -----+-----+-----+-----+-----+-----+-----+-----+  
 CGTCCGCTCTACCTATAAGCCGTCTAAGTGAGGTTCCGCCAATACGATCCGCTATCCCTCT  
 R Q M D I R Q I H S K A V M L G D R E S

1621 GTCAACGGTCCAACCAGCCGATGCAGCCTCATTCCTCAACCTCGCCGTTTAACCCATTAC 1680  
 -----+-----+-----+-----+-----+-----+-----+-----+  
 CAGTTGCCAGGTTGGTTCGGCTACGTTCGGAGTAAGGAGTTGGAGCGGCAAATTGGGTAATG  
 Q R S N Q P M Q P H S S T S P F N P L P

1681 CCACTGGTGCCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1740  
 -----+-----+-----+-----+-----+-----+-----+-----+  
 GGTGACCACGGGACCACCACACCCGACGACACCGTGAACAAGTGAACACGCCACCTCG  
 T G A R G G V A A V A P C S L V R V E P

1741 CACAGCGAGGCGCGCTCGACGCCTTCGTGCGCCGGCCGAAGCCGACCGCTGAGGGCAATG 1800  
 -----+-----+-----+-----+-----+-----+-----+-----+  
 GTGTCGCTCCGCGGAGCTGCGGAAGCACGCGGCCGCTTCGGCTGGCGACTCCCCTTAC  
 Q R G A L D A F V R R P K P T A E G N G

1801 GAGATGCTCCGCTGCGTAGTGAGTCCATTGAGGAGCGCGCAGGTGGAGGCGTTGACGCAC 1860  
 -----+-----+-----+-----+-----+-----+-----+-----+  
 CTCTACGAGGCGACGCATCACTCAGGTAACCTCCTCGCGCTCCACCTCCGCAACTGCGTG  
 D A P L R S E S I E E R A G G G V D A Q

## Appendix 4 continued

1861 AGGTCCGCAGTGC GGGTACTGGGT CGTCCATCAAATCTCAAGATACAGCTACTGGGGGAA 1920  
 -----+-----+-----+-----+-----+-----+-----+-----+  
 FCCAGGCGTCACGCCCATGACCCAGCAGGTAGTTTAGAGTTCTATGTCGATGACCCCTT  
 V R S A G T G S S I K S Q D T A T G G N

1921 ATTTTCAGGTTGCGAGAACC GCCGATGATGGTAAGGCTCCCTCGGGTTTGACGTGTGCTG 1980  
 -----+-----+-----+-----+-----+-----+-----+-----+  
 TAAAGTCCAACGCTCTTGCGGCTACTACCATTCCGAGGGAGCCAAACTGCACACGAC  
 F Q V A R T A D D G K A P S G L T C A A

1981 CATCATCCAGCCCTGCAGTAACCAACCGTGACTGATGCCAAGCCAGCAGCCGCACTGCGG 2040  
 -----+-----+-----+-----+-----+-----+-----+-----+  
 GTAGTAGGTCGGGACGTCATTGGTGGCACTGACTACGGCTTCGGTCGTGCGCCCTCACCCC  
 S S S P A V T T V T D A E A S S G S G A

2041 CCAGAGGCTGGTCCGAAAACCAAAGCACTGGGACTCTTTCCATGACCCAGTTTTCCTGT 2100  
 -----+-----+-----+-----+-----+-----+-----+-----+  
 GGTCTCCGACCAGCCTTTTGGTTTCGTGACCTGAGAAAGGTAAGTGGGGTCAAACGACA  
 R G W S E N Q S T G T L S M T P V L L L

2101 TAGATACCACCGATGAGGATGGTGAAGAAGTGGAGTACACAATGGAGCATTTCAGAAAC 2160  
 -----+-----+-----+-----+-----+-----+-----+-----+  
 ATCTATGGTGGCTACTCCTACCCTTCTTACCTCATGTGTTACCTCGTAAAGTTCTTTG  
 D T T D E D G E E V E Y T M E H F K K H

2161 ATAGAAAGGAGGTTTCAGGATGTGGTGTGATCTGTAATAGATACCGCAGGGGTGGGTGTGG 2220  
 -----+-----+-----+-----+-----+-----+-----+-----+  
 TATCTTTCCCTCAAAGTCCACACCACTCAAGACATTATCTATGGCGTCCCCACCCACACC  
 R K E V Q D V V S S V I D T A G V G V G

2221 GCGCGCGCGCTCGGATTATATCGCCGCTGAAGATAACAAGGCGGCTGCATCGACGGGGG 2280  
 -----+-----+-----+-----+-----+-----+-----+-----+  
 CCGCGCCGCGCAGCCTAATATAGCGCGACTTCTATTGTTCCGCGGACGTAGCTGCCCCC  
 R G A S D Y I A A E D N K A A A S T G D

2281 ATGCCGCTGPGCGCATGCTGGAGGGCGCTGATTCCGGCTGGGAGCCAGGAGGAGGCGGGAT 2340  
 -----+-----+-----+-----+-----+-----+-----+-----+  
 TACGGCGACACGCGTACCACCTCCGCGACTAAGCCGACCCTCGGTCCCTCCTCCGCCCTA  
 A A V R M V E G A D S A G S Q E E A G F

*SalI*  
 |

2341 TTCTTTTGCTCACTAGTGTGTCGACCATTGTATCGAACATTCCGCGGGGTACATCACAAA 2400  
 -----+-----+-----+-----+-----+-----+-----+-----+  
 AAGAAAACGAGTGATCACACAGCTGGTAACATAGCTTGTAAAGCGCGCCCATGTAGTGTT  
 T I L T S V S T I V S N I R A G T S Q T

2401 CTGCTCAATCCCTTTTTCAGAATTTGGCCFACGTTGGTGTCTTAAGGGGCACCTCTTCT 2460  
 -----+-----+-----+-----+-----+-----+-----+-----+  
 GACGAGTTAGGGAAAAGTCTTAAACCGGATGCAACCACPAAGAATCCCCGTGGAGAAGA  
 A Q S L F Q N L A Y V G V L K G H L F F

2461 TCGCCCAATCGGGTACGACACTTACGTTGTTGATTCCCTACGGCTCGTCCGGCACGTGCG 2520  
 -----+-----+-----+-----+-----+-----+-----+-----+  
 AGCGGGTTAGCCCATCCTCTGAAATGCAACAACCTAAGGGATGCCGAGCAGGCCGTGCAGC  
 A Q S G T T L Y V V D S L R L V R H V V

## Appendix 4 continued

2521 TGTATCAGCGAATTTTCCTTCGGTGGGCCACCCCTCTCTCCACTGTGCCTCAGTTGT 2580  
 -----+-----+-----+-----+-----+-----+-----+  
 ACATAGTCGCTTAAAAGGAAGCCACCCGGTGGGGGAGAGAGAGGTGACACGGAGTCAACA  
 Y Q R I F L R W A T P S L S T V P Q L S

2581 CGTTTGAGGAGCCTATACACTTGTGAGATTGCTTTCTTTGCACTACAGAATGACGTTTC 2640  
 -----+-----+-----+-----+-----+-----+-----+  
 GCAAACCTCCTCGGATATGTGAACAGTCTAAACGAAAGGAAACGTGATGTCTTACTGCAAG  
 F E E P I H L S D L L S F A L Q N D V Q

2641 AGCTTCCACCTTCACAAAAGCGCGCTGACGGAGGACCGGGTTCTTTCTGTCTCGCTTGG 2700  
 -----+-----+-----+-----+-----+-----+-----+  
 TCGAAGGTGGAAGTGTTCGCGCGACTGCCTCCTGGCCCAAGAGAAGACAGAGCGAACC  
 L P P S Q K R A D G G P G S L L S R L G

2701 GGGCGCGCTCTGCAACTGGCGTTACATGTTGCAGGATTACTTTGCCGTGGAAATCAGTG 2760  
 -----+-----+-----+-----+-----+-----+-----+  
 CCGCGGGCGGAGACGTTGACCGCAATGTACAACGTCTAATGAAACGGCACCTTTAGTCAC  
 R R L C N W R Y M L Q D Y F A V E I S A

2761 CTGATGGCCACCTGATCGCGCTTCCACTTTCTATGGGCACTTCGTGGCCACCCCGCTTC 2820  
 -----+-----+-----+-----+-----+-----+-----+  
 GACTACCGGTGGACTAGCGCGAAGGTGAAAGATACCCGTGAAGCACCGGTGGGGCGAAG  
 D G H L I A L P L S M G T S W P P P L R

2821 GGGCTGTGCCTCTTTTTATATGGCGGCTTGCAGCAGAAGTCCGTATAATGCCGGGGAGA 2880  
 -----+-----+-----+-----+-----+-----+-----+  
 CCCGACACGGAGAAAAATATACCGCCGAACGTCTTCAAGGCATATTACGGCCCTCT  
 A V P L F I W R L A A E V P Y N A G E I

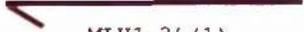
2881 TTGAGTGTTCCTACTGCCATAGCTCGGCACATTGCAGAGACATTGTATGGTGTGCAGCTGC 2940  
 -----+-----+-----+-----+-----+-----+-----+  
 AACTCACAAAGTGACGGTATCGAGCCGTGTAACGTCTCTGTAACATACCACACGTGACG  
 E C F T A I A R H I A E T L Y G V Q L H

2941 ACAGCTCGTGGCTGCCGAATGTAATAAAGGATGGTATTCGGCAAGATGATGTTCTCCAT 3000  
 -----+-----+-----+-----+-----+-----+-----+  
 TGTGAGCACCGACGGCTTACATTATTTCTACCATAAGCCGTCTACTACAAGGAGGTA  
 S S W L P N V I K D G I R Q D D V P P F

3001 TTTGTGATGCCATTCGCTTTGGCCTTTTGCCGTGCGCGACGAACTCAACCTTCTTTGTGC 3060  
 -----+-----+-----+-----+-----+-----+-----+  
 AAACACTACGGTAAGCGAAACCGGAAAACGGCACGCGCTGCTTGAGTTGGAAGAAACACG  
 C D A I R F G L L P C A T N S T F F V P

3061 *Sal*I  
 |  
 CTCCGTGTGACGCGCTGGTTCGACGGAACCGTGCAGGCGGTAGTTTCTGTGGATGAGCTGT 3120  
 -----+-----+-----+-----+-----+-----+-----+  
 GAGGCACACTGCGCGACCAGCTGCCTTGGCACGTCCGCCATCAAAGACACCTACTCGACA  
 P C D A L V D G T V Q A V V S V D E L Y

3121 ACAAGGTCTTCGAGCGTTGCTGAGCTTTTCGTTTCCCTCTTTTCTTTTCTTAGTTTTC 3180  
 -----+-----+-----+-----+-----+-----+-----+  
 TGTTCCAGAAGCTCGCAACGACTCGAAAAGCAAAAGGGAGAAAAGAAAAGAATCAAAG  
 K V F E R C


 MLH1 3' (1)

## Appendix 4 continued

3181 TTTTTTTGACTTTCACCTTGTGGTTCATTTTACTTTTCTGTGCGTTGAAAAGCGCTTAATT 3240  
 -----+-----+-----+-----+-----+-----+-----+  
 AAAAAAAGTAAAAGTGAACAACCAAGTAAAATGAAAAGACAGCAACTTTTCGCGAATTAA

MLH1 3' (A) 

3241 GCGGTGCATTTTGGGCTTATATATCCTCTTCTTCTGTGCGTTAACATTGAACTCTTC 3300  
 -----+-----+-----+-----+-----+-----+-----+  
 CGCCACGTAAAACCCGAATATATAGGAGAAGAAAGACAGCACCAATTGTAAGTTGAGAAG

3301 AAGACCTCGTTACCCGGGTGGTGTGGAAGTTACTATGCTCAAGAGAAGGGCGGGATGC 3360  
 -----+-----+-----+-----+-----+-----+-----+  
 TTCTGGAGCAATGGGCCCAACCACACCTTTCAATGATACGAGTTCTCTTCCCGCCCTACG

3361 GCGTCTTCAGAAGAAAGGGATAGTGAACCGAAGTTACCAAGGGCGGGAAATGAATAA 3420  
 -----+-----+-----+-----+-----+-----+-----+  
 CGCAGAAGTCTTCTTTCCCTATCACCTTGGCTTCCAATGGTTCCCGCCCTTTACTTATT

3421 TTGTGCTCAATGCACACCAACAGACCGAGTAAACCTCGTGTGCTTTCTTCCCCACTCCT 3480  
 -----+-----+-----+-----+-----+-----+-----+  
 AACACGAGTTACGTGTGGTTGTCTGGCTCATTGGAGCACACGAAAGAAGGGGGTGAGGA

3481 GCTTTGCGGTTCTCTTTTTTACACGTTTTCCCTCCTGTCATACCTTTATTCTATGCTGAG 3540  
 -----+-----+-----+-----+-----+-----+-----+  
 CGAAAGCCCAAGAGAAAAAATGTGCAAAAGGGAGGACAGTATGGAAATAAGATACGACTC

3541 TTTCTTTTACTATTTTTTCTTGTGCATCACGCGTATTTCTTTCTTTTGCAGCGCACCTTC 3600  
 -----+-----+-----+-----+-----+-----+-----+  
 AAAGAAAATGATAAAAAAGAACAGTAGTGCGCATAAAAGAAAGAAAACGTGCGGTGGAAG

3601 GTAATGCAGCATTAAAGTTTAGTGTATTGCTTTCCTATTAAGTGGCAGCAATTTCCCCAC 3660  
 -----+-----+-----+-----+-----+-----+-----+  
 CATTACGTCGTAATTCAAATCACATAACGAAAGTATAATTACCGTCGTTAAAGGGGTG

3661 CACCCCGCTGTATACATATATATATATATATATATCCCTCGTTTTTTTTGTTTCTGGTTAAGCT 3720  
 -----+-----+-----+-----+-----+-----+-----+  
 GTGGGGCGACATATGTATATATATATATATAAGGGAGCAAAAAACAAAGACCAATTCGA

3721 TCTGGTTTCATCCTCACACCGGCGTGTGTGGCTGTTTTCTCCCCACTTGAAATTTGA 3780  
 -----+-----+-----+-----+-----+-----+-----+  
 AGACCAAAGTAGGAGTGTGGCCGCACAAACACCGAACAAAGGAGGGGTGAACTTTAAACT

 MLH1 3' (B)

3781 TACAAATTTCCCCCTTTCAACAGGAGACAGCGACTCCATTTACATGCTCACACACCCAC 3840  
 -----+-----+-----+-----+-----+-----+-----+  
 ATGTTTAAAGGGGGAAAGTTGTCCTCTGTGCTGAGGTAAATGTACGAGTGTGTGGGTG

3841 AACTGTGCGGAAAATATACACCGCTTCCCCCTTTTTCATTGTGTTGGTTTTCAGCAAATAT 3900  
 -----+-----+-----+-----+-----+-----+-----+  
 TTGACAGCGCTTTAATATGTGGCGAAGGGGGAAAAGTAACACAACCAAAGTCGTTTATA

3901 ATGTATATGCATATATTACACATTTCTGTGACATGTGACCGTTTGGGGGATGTAAGGGG 3960  
 -----+-----+-----+-----+-----+-----+-----+  
 TACATATACGTATATAATGTGTAAGACACTGTACACTGGCAAACCTCCCTACATTCCCC

## Appendix 4 continued

```

3961 ATCTTTTTTTTTTTTTGGCATTGGGAACTCGAAGCCGTTGGAGACAAAGGTTTTGTACT 4020
-----+-----+-----+-----+-----+-----+
TAGAAAAAAAAAAAAACCCTAAACCCTTGAGCTTCGGCAACCTCTGTTTCCAAAACATGA

4021 CGTTTTTTGTAAGTTAGTGGGGCGAGCAATTTTTGCACCTTCCTCGTCACTATTGCTGCT 4080
-----+-----+-----+-----+-----+-----+
GCAAAAAACATTCAATCACCCCGCTCGTTAAAAACGTGGAAGGAGCAGTGATAACGACGA

4081 AGTCGTCGCCACCTATAAGCATATGTACTTTTTTTTACTAATATTCTTTTTTCCACCGCC 4140
-----+-----+-----+-----+-----+-----+
TCAGCAGCGGTGGATATTTCGTATACATGAAAAAATGATTATAAGAAAAAGGTGGCGG

4141 GGAACCAGGTAGTCATCAACATATATGCGACGTGGAACATAACGACTGACGACTCCGCG 4200
-----+-----+-----+-----+-----+-----+
CCTTGGTCCATCAGTAGTTGTATATACGCTGCACCTTGATTATGCTGACTGCTGAGGCGC

4201 GCAGCACCCAGCAGCGTCGCCGTGACTCCCCGGGAGTTAACCAGGGGACGATA 4253
-----+-----+-----+-----+-----+
CGTCGTGGGTTCGTCGCAGCGGCACTGAGGGGCCCTCAAATTGGCCCCTGCTAT

```

## Appendix 5 Sequence map showing the putative *T. brucei* *PMS1* ORF

A map showing the DNA sequence of the 5.5 kb *EcoRI* fragment sub-cloned from an ILTat 1.2 genomic lambda clone into pBluescript II KS to form the plasmid pJB400, which contains the entire putative *T. brucei* *PMS1* ORF along with its processing flanks (see sections 2.10.2 and 3.4.3). The conceptual translation of the *PMS1* ORF is shown in red beneath the DNA sequence and restriction sites are shown above the sequence in black. Half arrows denote primer sequences: those used to amplify the *PMS1* internal fragment (*PMSIU1* and *PMSID1*) are indicated in turquoise; those used to amplify the 460 bp *PMS1* probe (*PMSIU2* and *PMSI 3'(2)*) are indicated in dark red; and those used to confirm the linkage of the putative *PMS1* contigs (*PMSI 5'(2)* and *PMSI 3'(2)*) are shown in purple.

```

EcoRI
|
1  GAATTCCTGGCGTTGTTCCATGAGGCACATGTCACGTCGGTCACGTCGATGGGATCCCTG
   -----+-----+-----+-----+-----+-----+-----+-----+
60  CTTAAGGACCGCAACAAGGTACTCCGTGTACAGTGCAGCCAGTGCAGCTACCCTAGGGAC

   GTGCCACCTGTCTCCCACTTCCCCGTTTGGCAGCCCTTTGTTCATCTTCCCCTCGTTGAAC
61  -----+-----+-----+-----+-----+-----+-----+-----+
120  CACGGTGGACAGAGGGTGAAGGGGCAAACGCTCGGGAACAGTAGAAGGGGAGCAACTTG

   TGGTCAAAAAACCCGCCCTACCGTCTGTAGCATAATGATGAGTTGTCTCTCTGCGACGTC
121  -----+-----+-----+-----+-----+-----+-----+-----+
180  ACCAGTTTTTGGGCGGGATGGCAGACATCGTATTACTACTCAACAGAGAGACGCTGCAG

   ATAGAACTGCCTTACTCGCGTTTGTGATGTGTAACCTCGTGGGTGTTTTGCTCTGTGAT
181  -----+-----+-----+-----+-----+-----+-----+-----+
240  TATCTTGACGGAATGAGCGCAAACACTACACATTTGAGCACCCACAAAAACGAGACACTA

   TTCCTAGGAAGCAAACCTGTTTCCCACTCCCTCTTACATTATTTACGGCTCCCTCTACTG
241  -----+-----+-----+-----+-----+-----+-----+-----+
300  AAGGATCCTTCGTTTGACAAAGGGTGAGGGAGAATGTAATAAAGTGCCGAGGGAGATGAC

   TGTAACCGAAGTGTCTTAGGTGTGAGCTATGGTTGGCATCAAGTGCAGGAACCGCCGAA
301  -----+-----+-----+-----+-----+-----+-----+-----+
360  ACATTGGCTTCACAAAGATCCACACTCGATACCAACCGTAGTTCACGTCCTTGGCGGCTT

   AGGCCCGTCGCGCACACTTCCAAGCGCCAGTCATGTCCGCCGCATCCTCATGAGTGCCC
361  -----+-----+-----+-----+-----+-----+-----+-----+
420  TCCGGGCAGCGCGTGTGAAGGTTCCGCGGGTCAGTACAGGCGGCGTAGGAGTACTCACGGG

   CACTCTTAAGGAGCTACGCGCCAAGTACAACGTGCGCTCGATGCCTGTGCGCAAGGACG
421  -----+-----+-----+-----+-----+-----+-----+-----+
480  GTGAGAGATTCCGATGCGCGGTTTATGTTGCACGCGAGCTACGGACACGCGTTCCTGC

   ACGAGGTGCGTGTAAAGCGTGGGAAGTTCAAGGGCCGTGAGGGCAGAGTCAACGCATGCT
481  -----+-----+-----+-----+-----+-----+-----+-----+
540  TGCTCCACGCACAATTGACACCCTCAAGTTCACGGCACTCCCGTCTCAGTGGCGTACGA

   ACCGCCTCAAGTGGGTTATTACATTTGACAAGGTGAGTTGCGAGAAGGCGAACGGCACCA
541  -----+-----+-----+-----+-----+-----+-----+-----+
600  TGGCGGAGTTCACCCAATAAGTGTAACTGTTCCACTCAACGCTCTCCGCTTGCCGTGGT

```

## Appendix 5 continued

601 CCGTTCCCGTCGGTGTACACACCTCCAACGTGAAATCACGAACTGAAGCTCAACACCA 660  
 -----+-----+-----+-----+-----+-----+-----+  
 GGCAAGGGCAGCCACATGTGTGGAGGTTGCACCTTTAGTGCTTTGACTTCGAGTTGTGGT

661 GGCGGAAGGCAATCCTCGAGCGCAAGGACCGCAGCACGAAGACCGACAAATCAAAGGGAA 720  
 -----+-----+-----+-----+-----+-----+-----+  
 CCGCCTCCGTTAGGAGCTCGCGTTCCTGGCGTCGTGCTTCTGGCTGTTTAGTTCCCTT

721 AGGTGACCGCTGCTGAGAAGGCCATGCAGCAGATGGACTAAAGAAGACGGGGGTCTACGG 780  
 -----+-----+-----+-----+-----+-----+-----+  
 TCCACTGGCGACGACTCTTCCGGTACGTGCTACCTGATTTCTTCTGCCCCCAGATGCC

781 AGCCACCGAGGTTTACCGACTTCTTTACCTGCTGTTTCACTTAGGAACACCTCGTGATAT 840  
 -----+-----+-----+-----+-----+-----+-----+  
 TCGGTGGCTCCAATGGCTGAAGAAATGGACGACAAAGTGAATCCTTGTGGAGCACTATA

PMS1 5' (2)

---

841 TTCTGACGTTTATTCTTCATTTTCTGTGCTAACGTCTCTTTTACGACTCGATGTGGAG 900  
 -----+-----+-----+-----+-----+-----+-----+  
 AAGACTGCAAATAAGAAGTAAAAAGACACGATTGCAGAGAAAAATGCTGAGCTACACCTC

901 CCACACTGCTGCCCCCCCCCACACGTGCACCATTGCGCGCTTGTGCCGAACCCTGGTC 960  
 -----+-----+-----+-----+-----+-----+-----+  
 GGTGTGACGACGGGGGGGGTGTGCACGTGGTAACGCGCAACAACGGCTTGGGACCAG

961 TCGGTTTGAGAAAGAGGATCAAACATTTATTTTCGTTTCCCCTGATCTTCTTTCCTTT 1020  
 -----+-----+-----+-----+-----+-----+-----+  
 AGCCAAACTCTTCTCCTAGTTTGTAAATAAAAGCAAAGGGGGACTAGAAAGAAAGGAAA

1021 CCTCCAAGCTTCTATTTCTTCCGCCACTCAGTCGGAGAAACCTTCTATGATCCTGAGA 1080  
 -----+-----+-----+-----+-----+-----+-----+  
 GGAGGTTTCGAAAGATAAAGAAGGCGGTGAGTCAGCCTCTTTGGAAAGATACTAGGACTCT

1081 CACCAAAGACTTCAATTTCTCCTTGCAGAAATGAAACAGCGTATAAATTTGAAGATAGA 1140  
 -----+-----+-----+-----+-----+-----+-----+  
 GTGGTTTCTGAAGTTAAAGGAGGAACGCCTTAACTTTGTGCGATATTTAACTTCTATCT

1141 GCAACGCCCATAGTGGGGGGAGAAGGTTGTGAAGAAAGAGGTAAGCAAGGGGCATGATTA 1200  
 -----+-----+-----+-----+-----+-----+-----+  
 CGTTGCGGGTATCACCCCTCTTCCAACACTTCTTCTCCATTCGTTCCCCGTAATAAT  
M I T

1201 CCCTTTTGGACGAGGGAAGTTCGCGCAAGTTGAGTGCAGGGCAGGTAATTACGAATCTTT 1260  
 -----+-----+-----+-----+-----+-----+-----+  
 GGGAAAACCTGCTCCCTTCAAGCGGTTCAACTCACGTCCCGTCCATTAATGCTTAGAAA  
L L D E G S S R K L S A G Q V I T N L S

1261 CCAGCGTCGTGAAAGAGTTGGTAGAGAATAGCCTAGATGCCGGTGCCCGTACCGTTGCAA 1320  
 -----+-----+-----+-----+-----+-----+-----+  
 GGTCGCGACTTTCTCAACCATCTCTTATCGGATCTACGGCCACGGGCATGGCAACGTT  
S V V K E L V E N S L D A G A R T V A I

## Appendix 5 continued

1321 TACGTGTGGAGGATAGCGGCGCTGGAATATTACCGTGGAAGATGATGGAAGTGGCATGG 1380  
 -----+-----+-----+-----+-----+-----+-----+-----+  
 ATGCACACCTCCTATCGCCGCGACCTTTATAATGGCACCTTCTACTACCTTCACCGTACC  
   R V E D S G A G N I T V E D D G S G M D

1381 ACTTATCTTATCTGCTAGACAGCGAGGGCCGACTGAAAGAGGATGCTTCCCTGCCCTGC 1440  
 -----+-----+-----+-----+-----+-----+-----+  
 TGAATAGAATAGACGATCTGTGCTCCCGGCTGACTTTCTCCTACGAAGGGACGGGGACG  
   L S Y L L D S E G R L K E D A S L P L L

1441 TAGCCTCGAGGGCGACAACAAAACGAAGGGGGGCGACAGTGGTCTTTCGTGCGAGGCGG 1500  
 -----+-----+-----+-----+-----+-----+-----+  
 ATCGGAGCTCCCGCTGTTGTTTTGCTTCCCCCGCTGTCACCAGAAAGCAGCGTCCGCC  
   A S R A T T K R R G G D S G L S S Q A A

1501 CGCAAACACTGGGTTTTTCGTGGCGAAGCGCTCCACTCTTTGGCTCACCTTAGCGAGTTGT 1560  
 -----+-----+-----+-----+-----+-----+-----+  
 GCGTTTGTGACCCAAAAGCACCGCTTCGCGAGGTGAGAAACCGAGTGAATCGCTCAACA  
   Q T L G F R G E A L H S L A H L S E L S

1561 CCATATGCACAATGTCAGAATCAACCCGACCTACTGCACTACTGATCGCGTACGACAGTA 1620  
 -----+-----+-----+-----+-----+-----+-----+  
 GGTATACGTGTTACAGTCTTAGTTGGGCTGGATGACGTGATGACTAGCGCATGCTGTCAT  
   I C T M S E S T R P T A L L I A Y D S N

1621 ATTCCCGCCGTACGACTGTAAAGGTAACGAGTGAACGGAGGGACGTAGGCACAACCGTTG 1680  
 -----+-----+-----+-----+-----+-----+-----+  
 TAAGGGCGGCATGCTGACATTTCCATTGCTCACTTGCCCTCCCTGCATCCGTGTTGGCAAC  
   S R R T T V K V T S E R R D V G T T V V

1681 TCGTTAGCAAACCTTTGCGGCTCTTCCGGTGCCTCACAAGGATTTTCGTACGGGGTAGGA 1740  
 -----+-----+-----+-----+-----+-----+-----+  
 AGCAATCGTTTGAGAAACGCCGAGAAGGCCACGCAGTGTTCCTAAAGCATGCCCCATCCT  
   V S K L F A A L P V R H K D F V R G R K

1741 AGAAGCAGCTTCTCGCTGCTACCCTACTGATGAAGCAGTACGCCCTTTCACACCCCCACG 1800  
 -----+-----+-----+-----+-----+-----+-----+  
 TCTTCGTGGAAGAGCGACGATGGGATGACTACTTCGTGATGCGGGAAAGTGTGGGGGTGC  
   K Q L L A A T L L M K Q Y A L S H P H V

1801 TGCGCTTGTGATGACCCATCGTGCGGGACCTGACAGTGCCCCCGTTACTCTGGTGTGCT 1860  
 -----+-----+-----+-----+-----+-----+-----+  
 ACGCGAACAACACTACTGGGTAGCACGCCCTGGACTGTACGGGGGCAATGAGACCACAGCA  
   R L L M T H R A G P D S A P V T L V S L

1861 TGACGGGGACGGGGACCCCCAGCGAGCCCTTGCTGAGGCGTACGGTGGACGTGTTATCG 1920  
 -----+-----+-----+-----+-----+-----+-----+  
 ACTGCCCTGCCCGCTGGGGTTCGCTCGGGAACGACTCCGCATGCCACCTGCACAATAGC  
   T G T G D P Q R A L A E A Y G G R V I A

1921 CCAACATGGAGCGTGTGGAGTGGGAACGACTTTTGGGACAATTACGGGTTATGTTTCGA 1980  
 -----+-----+-----+-----+-----+-----+-----+  
 GGTGTGACCTCGCACACCTCACCTTGACTGAAAACCTGTTAATGCCCAATACAAAGCT  
   N M E R V E W E L T F G T I T G Y V S K

1981 AGGGGAATGCCGGCAGACTTTCCTCCGATATGCAAGTGTTCGACTGGATGGGAGGCTTG 2040  
 -----+-----+-----+-----+-----+-----+-----+  
 TCCCTTACGGCGTCTGAAAGGAGGCTATACGTTCAAAACGTGACCTACCCTCCGAAC  
   G N A G R L S S D M Q V F A L D G R L V

## Appendix 5 continued

2041 TTGACCTCCCCATGATGGCCAAGGCGGTTAACGATGCCTACGCCGAATCCCTGCCAAACG 2100  
 -----+-----+-----+-----+-----+-----+-----+  
 AACTGGAGGGGTACTACCGGTTCCGCCAATTGCTACGGATGCGGCTTAGGGACGGTTTGC  
   D L P M M A K A V N D A Y A E S L P N A


  
 EMS101

2101 CCGCGCAGCGCACATTTCTGCCTTCTTCTGCATGTCAGCTCTGGTGAGTCACTTCCCT 2160  
 -----+-----+-----+-----+-----+-----+-----+  
 GGCGCGTTCGGTGTAAAGGACGGAAGAAGGACGTACAGTTCGAGACCACTCAGTGAAGGGA  
   A Q R T F P A F F L H V S S G E S L P Y

2161 ACGATGTGAATTTAGTTCCTGACAAACGCAAAGTTCTCATTTCCGGATGAGGAGAGACATG 2220  
 -----+-----+-----+-----+-----+-----+-----+  
 TGCTACACTTAAATCAAGGACTGTTTGCCTTTCAGAGTAAAGCCTACTCCTCTCTGTAC  
   D V N L V P D K R K V L I S D E E R H A

2221 CTGGGGAGGTTTCGTACATGCGGTTTTCGAACGTTCCAGGCATCAACTGACGGAATAGATC 2280  
 -----+-----+-----+-----+-----+-----+-----+  
 GACCCCTCCAAGCATGTACGCCAGAAGCTTGAAGGTCGGTAGTTGACTGCCTTATCTAG  
   G E V R T C G L R T F Q A S T D G I D L

2281 TGCCCGTGCACAACGAAGGCGGATGGAGACACGTACCGGAGCGCAGGAACACACAGGAGA 2340  
 -----+-----+-----+-----+-----+-----+-----+  
 ACGGGCACGCGTTGCTTCCGCCACCTCTGTGCATGGCCTCGCGTCTTGTGTGCCTCT  
   P V R N E G G W R H V P E R R N T Q E T

2341 CAATGCCTACACAAACGCCTCTTTCTGCTACATCTATTGCACAGTTCATCTATCAGCGGC 2400  
 -----+-----+-----+-----+-----+-----+-----+  
 GTTACGGATGTGTTGCGGAGAAAGACGATGTAGATAACGTGTCAAGTAGATAGTCGCCG  
   M P T Q T P L S A T S I A Q F I Y Q R R

2401 GGGAAACATCGCAGGGAGACAACCTTATTGACAATGCTGCGGTGGCGCAGGTGCAGCCCT 2460  
 -----+-----+-----+-----+-----+-----+-----+  
 CCCTTTGTAGCGTCCCTCTGTTGGAATAACTGTTACGACGCCACCGCGTCCACGTCCGGGA  
   E T S Q G D N L I D N A A V A Q V Q P S

2461 CCGTTTGCTTAAGTCAACTGTTGTCTGGAAGCTCGCCTATTGGACGGACATCCTCCCCTG 2520  
 -----+-----+-----+-----+-----+-----+-----+  
 GGCAAACGAATTCAGTTGACAACAGACCTTCGAGCGGATAACCTGCCTGTAGGAGGGGAC  
   V C L S Q L L S G S S P I G R T S S P D

2521 ATGCTACCGCTTCCCCAACCCAGCACCACCAACCGTGCACCATCGGAGCGTTCTGCAGGTT 2580  
 -----+-----+-----+-----+-----+-----+-----+  
 TACGATGGCGAAGGGGTTGGTCGTTGGTTGGCACGTGGTAGCCTCGCAAGACGTCCAA  
   A T A S P T S T T N R A P S E R S A G S


  
 EMS101

2581 CTGTTGAATCAGTCTTGGAGTATCCCTTGACCTTTGAGCCAACCCAGAAGAGGCAGCGGT 2640  
 -----+-----+-----+-----+-----+-----+-----+  
 GACAACTTAGTCAGAACCTCATAGGAACTGGAAACTCGGTTGGGTCTTCTCCGTCGCCA  
   V E S V L E Y P L T F E P T Q K R Q R L

2641 TGGAGTCTCAGCGGAAGAAGGAAATACCGGTGACACTGATGGTAGCAGCTGGGGTGAAG 2700  
 -----+-----+-----+-----+-----+-----+-----+  
 ACCTCAGGAGTCGCTTCTTCTTTATGGCCACTGTGACTACCATCGTCGACCCCACTTC  
   E S S A E E G N T G D T D G S S W G E E

## Appendix 5 continued

2701 AAGATCCCCGGGAGCTCGGACCGGAAGAGATGATGGGGACTGATGATGCAGTTGTGAATT 2760  
 -----+-----+-----+-----+-----+-----+-----+-----+  
 TTCTAGGGGGCCCTCGAGCCTGGCCTTCTCTACTACCCTGACTACTACGTCAACACTTAA  
 D P R E L G P E E M M G T D D A V V N Y

2761 ACGTTACCGACGATACGAATCAACAGCGACCCGGTCCCCCTCGCTCTTCTGTGAGGTTTC 2820  
 -----+-----+-----+-----+-----+-----+-----+  
 TGCAATGGCTGCTATGCTTAGTTGTGCTGGGCCAGGGGAGCGAGAAGACACTCCAAAG  
 V T D D T N Q Q R P G P P R S S V R F P

2821 CTCCATTTTCAGTTTTGGCTGAGATGCCACTGGTTCACTCCCTTGGAGAATGGGCGGCTC 2880  
 -----+-----+-----+-----+-----+-----+-----+  
 GAGGTAAGTCAAAACCGACTCTACGGTGACCAAGTGAGGGAACCTCTTACCCGCCGAG  
 P F S V L A E M P L V H S L G E W A A P

2881 CGTCCCAACCTTCCGATGGGGGAGGTGTTGCAAAGTTTTTCGAGGCTTCAGAAACAAACCG 2940  
 -----+-----+-----+-----+-----+-----+-----+  
 GCAGGGTTGGAAGGCTACCCCTCCACAAGCTTTCAAAAGCTCCGAAGTCTTTGTTGGC  
 S Q P S D G G G V R K F S R L Q K Q T E

2941 AGGAGGAGCTCACGTTGTATTTGGGGAAGGAGTCGTTCAAAAATATGGTTGTACATGGTC 3000  
 -----+-----+-----+-----+-----+-----+-----+  
 TCCTCCTCGAGTGCAACATAAACCCCTTCCTCAGCAAGTTTTATACCAACATGTACCAG  
 E E L T L Y L G K E S F K N M V V H G Q

3001 AGTTCAACCATGGCTTCATTGTAACCTCACTGGACGACAACATCTTTGTGATCGACCAGC 3060  
 -----+-----+-----+-----+-----+-----+-----+  
 TCAAGTTGGTACCGAAGTAACATTGAAGTGACCTGCTGTTGTAGAAACACTAGCTGGTCG  
 F N H G F I V T S L D D N I F V I D Q H

3061 ACGTGCCGATGAGAAAGGCAACTATGAACATCTAATGAGTCATTATGTAGCCAGGCCAC 3120  
 -----+-----+-----+-----+-----+-----+-----+  
 TGCGACGGCTACTCTTCCGTTGATACTTGTAGATTACTCAGTAATACATCGGTCCGGTG  
 A A D E K G N Y E H L M S H Y V A R P Q

3121 AACCCCTCTTTCTCCAGTACCTGTGTCGATGGAGCCCCAGGCTGTGGATTGGCTGTTG 3180  
 -----+-----+-----+-----+-----+-----+-----+  
 TTGGGGAGAAAAGAGGTCATGGACACAGCTACCTCGGGTCCGACACCTAAACCGACAAC  
 P L F S P V P V S M E P Q A V D L A V D

3181 ATCATGCTGAAGAACTGCGGCAACACGGCTTCATCGTGCAGCGGAGTGACGACCCAATA 3240  
 -----+-----+-----+-----+-----+-----+-----+  
 TAGTACGACTTCTTGACGCCGTTGTGCCGAAGTAGCACGTGCCTCACTGCTGTGTTAT  
 H A E E L R Q H G F I V Q R S D D T N K

PMSIU2 

3241 AGTTGCTGGTGCTATCAGTGCCTGTGATTCCCTTACGAGGTTGTGGACCCGAGAACGTGCG 3300  
 -----+-----+-----+-----+-----+-----+-----+  
 TCAACGACCACGATAGTCACGGACACTAAGGAATGCTCCAACACCTGGGCGTCTGCAGC  
 L L V L S V P V I P Y E V V D P Q N V V

3301 TAGAGCTGATTAGGCAGTTAGTGCATTACAATACGATCAGTAAACCGATGCGATGTGTGT 3360  
 -----+-----+-----+-----+-----+-----+-----+  
 ATCTCGACTAATCCGTCAATCACGTAATGTTATGCTAGTCATTTGGCTACGCTACACACA  
 E L I R Q L V H Y N T I S K P M R C V W

## Appendix 5 continued

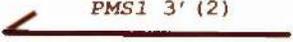
3361 GGCACCTCAATGGCAACTAAGGCATGCCGTTCTAGCATTATGGTTGGAACGATGCTTAGTG 3420  
 -----+-----+-----+-----+-----+-----+-----+-----+  
 CCGTGAGTTACCGTTGATTCCGTACGGCAAGATCGTAATACCAACCTTGCTACGAATCAC  
   H S M A T K A C R S S I M V G T M L S E

3421 AGAAGAAAATGCGGTGAGTCCGTGGACCGAATGGGTGAGCTGGAACAGCCGTGGAAGTGC 3480  
 -----+-----+-----+-----+-----+-----+-----+  
 TCTTCTTTTACGCCAGTCAGCACCTGGCTTACCCACTCGACCTGTGCGGCACCTTGACAG  
   K K M R S V V D R M G E L E Q P W N C P

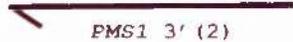
3481 CACACGGGCGACCGACCGTGCGGCAGTCTCCAAGATTTCTTCGTTGGTTTCCTTAATGA 3540  
 -----+-----+-----+-----+-----+-----+-----+  
 GTGTGCCCGTGGCTGGCAGCGCGTGCAGAGTTCTAAAGAAGCAACCAAAGGAATTACT  
   H G R P T V R H V S K I S S L V S L M T

3541 CGAAGTCCCGCCGTGCGTGAGTTTCACTTATTTATCTGGGCACCACCAAACCTTATGTTT 3600  
 -----+-----+-----+-----+-----+-----+-----+  
 GCTTCAGGGCGGCACGCACCTCAAAGTGAATAAATAGACCCGTGGTGGTTTGAATACAAA  
   K S R R A

3601 GCTTTTTTCTTAGTTTTTCGTCATTGCTGTTGTTGTTTCTTTGATATGCTTTTTTCCCGT 3660  
 -----+-----+-----+-----+-----+-----+-----+  
 CGAAAAAAGAATCAAAGCAGTAACGACAACAACAAGAACTATACGAAAAAAGGGCA


  
 PMS1 3' (2)

3661 GCTTTTGGTTCGCTTTTTCTACACGGAACCTAATGCTCGTCATTGTTCTCTTGCTCCCGT 3720  
 -----+-----+-----+-----+-----+-----+-----+  
 CGAAAACGCAGCGAAAAAGATGTGCCTTGGATTACGAGCAGTAACAAGAGAACGAGGGCA


  
 PMS1 3' (2)

3721 GCTTTTCCACTCACTGATTTCCGGTATTGTTGTTCTTCATTTTTCTCTAAACTGATTTCC 3780  
 -----+-----+-----+-----+-----+-----+-----+  
 CGAAAAGGTGAGTGACTAAAGCCATAACAACAAGAAGTAAAAGAGATTTGACTAAAAGG

3781 ATTGCGGACTTGATCTATTTCTTTCCTTGGTTCATCTGCCTGACGGTAAAGGGAAGGA 3840  
 -----+-----+-----+-----+-----+-----+-----+  
 TAACGCGCTGAACTAGATAAAAGAAAGGAACCAAGTAGACGGACTGCCATTTCCCTTCT

3841 AGGAAACGAAATCCAGCGGCAGCAGCTGTGCGGTGTCGGTGGCGACCACCTGAACAAACC 3900  
 -----+-----+-----+-----+-----+-----+-----+  
 TCCTTTGCTTTAGGTGCGCGTGTGACACGCCACGCCACCGCTGGTGGACTTGTGTTGG

3901 TTGTAGGTTTCATTTGAGGAAGACGGCGTAGTTTGGCATGGATGTGTGTGAGGGATTGAG 3960  
 -----+-----+-----+-----+-----+-----+-----+  
 AACATCAAAGTAAACTCCTTCTGCCGATCAAACCGTACCTACACACTCCCTAACTC

3961 GGTTCAGTTGGACATATGTCAGGAACTGGTGGGTTGGGGTTGACGGCAGGGGGTGGAGG 4020  
 -----+-----+-----+-----+-----+-----+-----+  
 CCAAGGTCAACCTGTATACAGTCCCTTGACCACCAACCCCAACTGCCGTCCCCACCTCC

4021 AACCGGACTAGTAGTGTTTTTTCCGACCTCAGGGAGCTTAACGAGGCAAAGGTGTGCGAA 4080  
 -----+-----+-----+-----+-----+-----+-----+  
 TTGGCGCTGATCATCACAAAAAGGCTGGAGTCCCTCGAATTGCTCCGTTTCCACAGCTT

## Appendix 5 continued

4081 CATTCTTTCATACCTAGATGACGTTGTTTCTACGAGACCCCGTTTCCCGCATCATTTGG 4140  
 -----+-----+-----+-----+-----+-----+-----+  
 GTAAGAAAGTATGGATCTACTGCAACAAAGATGCTCTGGGGGCAAAGGGCGTAGTAAACC

4141 CCCTTCAGCGTGCCGTGACGTATACATGCCAGCAAATATTTCCGGATGGTGGTGTGGCGTT 4200  
 -----+-----+-----+-----+-----+-----+-----+  
 GGAAGTGCACGGCACTGCATATGTACGGTCGTTTATAAAGCCTACCACCACACCGCAA

4201 GGCACCGTCGGCTGACCTACTACAGCCTTCTACGCAGTTACCACCGTTTGGCGGTTCCGGC 4260  
 -----+-----+-----+-----+-----+-----+-----+  
 CCGTGGCAGCCGACTGGATGATGTCCGAAGATGCGTCAATGGTGGCAAACCGCCAAGCCG

4261 TTCATGTCTTACGTGCGACATAATAGGGTGTGGACCCGACGTTCCGAACGATGTTTACCA 4320  
 -----+-----+-----+-----+-----+-----+-----+  
 AAGTACAGAATGCAGCGTGTATTATCCCACACCTGGGCTGCAAGCGTTGCTACAAATGGT

4321 CGGTATCAAGGCTAAAATATCTGCCTTACAGTTCAGCAATGATGAGCTGAGGGCTGAAAA 4380  
 -----+-----+-----+-----+-----+-----+-----+  
 GCCATAGTCCGATTTTATAGACGGAATGTCAAGTCGTTACTACTCGACTCCCGACTTTT

4381 CGAAGAACTGAAGGAGCGAGTCAAGATGGCGCGGGAGAGGGAAGCGGAGCGTCTTGAAG 4440  
 -----+-----+-----+-----+-----+-----+-----+  
 GCTTCTTGACTTCTCGCTCAGTTCTACCGCGCCCTCTCCCTTCGCTCGCAGAACCTTC

4441 CCAGGAAGCCCGGGCTCGCAACGAGCTGGAAGGGCTCCGCAAGAAGCTACGGGAGACGGA 4500  
 -----+-----+-----+-----+-----+-----+-----+  
 GGTCTTCGGCGCCGAGCGTTGCTCGACCTTCCCAGGGCGTCTTCGATGCCCTCTGCCT

4501 GCGCAACTACGAACGCGTGGTGCAAGAGTTTCAGCGCGAAAGGTGCGAGTTGACCCAGGC 4560  
 -----+-----+-----+-----+-----+-----+-----+  
 CGCGTTGATGCTTGCACACCAGTTCTCAAAGTCGCGCTTCCAGCGTCAACTGGGTCCG

4561 AGTGGAGTCTGTTACGTCTCAACTACGGCAAGAGATGTCTCGCGGGGAGGAGGAGATCGC 4620  
 -----+-----+-----+-----+-----+-----+-----+  
 TCACCTCAGACAATGCAGAGTTGATGCCGTTCTCTACAGAGCCGCCCTCCTCCTTAGCG

4621 CAGACTTGAGTCTGCGAACCGGACCGCCATAGCGCAACTTAAGACCCGGTGGCAAGCTCA 4680  
 -----+-----+-----+-----+-----+-----+-----+  
 GTCTGAACTCAGACGCTTGCCTGGCGGTATCGCGTTGAATTCTGGGCCACCGTTCGAGT

4681 GGAGAAGGCTGCTCGCGAGAAGTGGAGGATAGCCGAGGCGAAACGCATAAAGGAGAACAC 4740  
 -----+-----+-----+-----+-----+-----+-----+  
 CCTCTCCGACGAGCGCTCTCACCTCCTATCGGCTCCGCTTTCGCTATTTCTCTTGTG

4741 CCTCCAGTCGTTAGAGCCTGATATCGTGTGTGTTGAACAGGCACAAGGCGGAAAAGGC 4800  
 -----+-----+-----+-----+-----+-----+-----+  
 GGAGGTCAGCAATCTCGGACTATAGCACAACAACAATTGTCCGTGTTCCGCCTTTTCCG

4801 GCGTATGCGGGAAGAGTTCGAGAACGAACTACGnCAGCGGGATGAGGTTATTGCGGCAAA 4860  
 -----+-----+-----+-----+-----+-----+-----+  
 CGCATACGCCCTTCTCAAGCTCTTGCTTGATGChGTCGCCCTACTCCAATAACGCCGTTT

4861 GGAGGCGCCACTCGCAGAGTCCAGGGCGCGGTTGGAGCGTGAGGCATCTGCAACTCGGGC 4920  
 -----+-----+-----+-----+-----+-----+-----+  
 CCTCCGCGGTGAGCGTCTCAGGTCCCAGCGCAACCTCGCACTCCGTAGACGTTGAGCCCG

**Appendix 5 continued**

```
                EcoRI
                |
4921  TCGAGAGCAACAGGAGTTCGGGATCGAGAATTC
-----+-----+-----+----- 4954
      AGCTCTCGTTGTCCTCAAGGCCCTAGCTCTTAAG
```

## Appendix 6 Multiple alignment of the pHygro-Tub, pHYGWT, pHYG01, pHYG02, pHYG03, pHYG05, pHYG07, pHYG09 and pHYG11

Multiple alignment of the pHygro-Tub, pHYGWT, pHYG01, pHYG02, pHYG03, pHYG05, pHYG07, pHYG09 and pHYG11 DNA sequences. Only the region of the pHygro-Tub sequence which aligns with the pHYGWT sequence is shown. Numbers in italics indicate the percentage of sequence divergence of the *HYG* flanks of that construct from the pHYGWT sequence rounded to the nearest integer. Sequences were aligned using Multalin (<http://prodes.toulouse.inra.fr/multalin/multalin.html>; Corpet, 1988) and shaded using the BOXSHADE server ([http://www.ch.embnet.org/software/BOX\\_form.html](http://www.ch.embnet.org/software/BOX_form.html)): identical residues are shown in black; residues that differ from the pHYGWT sequence are shown in red. The sequence of the *HYG*For primer is indicated in blue, and the sequence complementary to the *HYG*Rev primer is indicated in green. The sequence corresponding to the *Nde*I restriction site is shown in purple. Missing bases as shown as dots.

```

pHygro-Tub      1  . . . . . AGCCTGAACCTCACCGCGACGTCTGTGCGAGAAGTTTCTGATCGAAAAGTTC
pHYGWT          1  AAGGCGCGCCAGCCTGAACCTCACCGCGACGTCTGTGCGAGAAGTTTCTGATCGAAAAGTTC
pHYG01          1  AAGGCGCGCCAGCCTGAACCTCACCGCGACGTCTGTGCGAGAAGTTTCTGATCGAAAAGTTC
pHYG02          1  AAGGCGCGCCAGC .TGAACCTCACCGCGACGTCTGTGCGAGAAGTTTCTGATCGAAAAGTTC
pHYG03          1  AAGGCGCGCCAGCCTGAACCTCACCGCGACGTCTGTGCGAGAAGTTTCTGATCGAAAAGTTC
pHYG05          1  AAGGCGCGCCAGCCTGAACCTCACCGCGACGTCTGTGCGAGAAGTTTCTGATCGAAAAGTTC
pHYG07          1  AAGGCGCGCCAGCCTGAACCTCACCGCGACGCTGTGCGAGAAGTTTCTGTTGAAAGGTTCC
pHYG09          1  AAGGCGCGCCAGCCTGAACCTCACCGCGACGTCTGTGCGAGGCGCTCCTGACCGAGAAGCTC
pHYG11          1  AAGGCGCGCCAGCCTGAACCTCACCGCGACGTCTGTGCGAGAAGTCCCTGATTGAAAAGTTC

```

```

pHygro-Tub      51  GACAGCGTCTCCGACCTGATGCAGCTCTCGGAGGGCGAAGAATCTCGTGCTTTTCAGCTTC
pHYGWT          61  GACAGCGTCTCCGACCTGATGCAGCTCTCGGAGGGCGAAGAATCTCGTGCTTTTCAGCTTC
pHYG01          61  GACAGCGTCTCCGACCTGATGCAGCTCTCGGAGGGCGAAGAATCTCGTGCTTTTCAGCTTC
pHYG02          60  GACAGCGTCTCCGACCTGATGCAGCTCTCGGAGGGCGAAGAATCTCGTGCTTTTCAGCTTC
pHYG03          61  GACAGCGTCTCCGACCTGATGCAGCTCTCGGAGGGCGAAGAATCTCGTGCTTTTCAGCTTC
pHYG05          61  GACAGCGTCTCCGACCTGATGCAGCTCTCGGAGGGCGAAGAATCTCGTGCTTTTCAGCTTC
pHYG07          61  GACAGCGTCTCCGACCTGATGCAGCTCTCGGAGGGCGAGGACCTCGTGCTTTTCAGCTTC
pHYG09          61  GACAGCGTCTCCGACCTGATGCAGCTCTCGAAGGGCGAAGAATCTCGTGCTTTTCAGCTTC
pHYG11          61  GCGAGCGTCTCCGACCTGATGCAGCTCTCGAAGGGCGCGAATCTCGTGCTTTTCAGCTTC

```

```

pHygro-Tub      111 CATGTAGGAGGGCGTGGATATGTCTCGCGGGTAAATAGCTGCGCCGATGGTTTCTACAAA
pHYGWT          121 CATGTAGGAGGGCGTGGATATGTCTCGCGGGTAAATAGCTGCGCCGATGGTTTCTACAAA
pHYG01          121 GATGCAGGAGGGCGTGGATAGTCTCGCGGGTAAATAGCTGCGCCGATGGTTTCTACAAA
pHYG02          120 GATGTAGGAGGGCGTGGATATGTCTCGCGGGTAAATAGCTGCGCCGATGGTTTCTACAAA
pHYG03          121 GATGTAGGAGGGCGTGGATATGTCTCGCGGGTAAATAGCTGTGCCGATGGTTTCTACAAA
pHYG05          121 GATGTAGGAGGGCGTGGATATGTCTCGCGGGTAAATAGCTGCGCCGATGGTTTCTACGAA
pHYG07          121 GATGTAGGAGGGCGTGGATATGTCTCGCGGGTAAATAGCTGCGCCGATAGTTTCTGCAAG
pHYG09          121 GGTGTAGGAGGGCGTAGATATGTCTCGCGGGTAAATAGCTGCGCCGATGGTTTCTACGAA
pHYG11          121 GATGTAGGAGGGCGTGGATAGTCTCGCGGGTAAATAGCTGCGCCGATGGTTTCTACAGA

```

```

pHygro-Tub      171 GATCGTTATGTTTATCGGCACCTTGCATCGGCCCGCGCTCCCGATTCCGGAAGTGCTTGAC
pHYGWT          181 GATCGTTATGTTTATCGGCACCTTGCATCGGCCCGCGCTCCCGATTCCGGAAGTGCTTGAC
pHYG01          181 GATCGTTATGTTTATCGGCACCTTGCATCGGCCCGCGCTCCCGATTCCGGAAGTGCTTGAC
pHYG02          180 GATCATTACGTTTATCGGCACCTTGCATCGGCCCGCGCTCCCGATTCCGGAAGTGCTTGAC
pHYG03          181 GATCGTTATGTTTATCGGCACCTTGCATCGGCCCGCGCTCCCGATTCCGGAAGTGCTTGAC
pHYG05          181 GATCGTTATGTTTATCGGCACCTTGCATCGGCCCGCGCTCCCGATTCCGGAAGTGCTTGAC
pHYG07          181 GATCGTTATGTTTATCGGCACCTTGCATCGGCCCGCGCTCCCGATTCCGGAAGTGCTTGAC
pHYG09          181 GACCGTTATGTTTATCGGCACCTTGCACCGGCCG .GCTCCCGATTCCGGAAGTGCTTGAC
pHYG11          181 GTCGTTATGTTTATCGGCACCTTGCACCGGCCG .GCTCCCGATTCCGGAAGTGCTTGAC

```

## Appendix 6 continued

pHygro-Tub 231 ATTGGGGAAATTCAGCGAGAGCCTGACCTATTGCATCTCCCGCCGTGCACAGGGTGTCCAG  
 pHYGWT 241 ATTGGGGAAATTCAGCGAGAGCCTGACCTATTGCATCTCCCGCCGTGCACAGGGTGTCCAG  
 pHYG01 241 ATTGGGGAAATTCAGCGAGAGCCTGACCTATTGCATCTCCCGCCGTGCACAGGGTGTCCAG  
 pHYG02 240 ATTGGGGAGCTCAGCGAGAGCCTGACCTATTGCATCTCCCGCCGTGCACAGGGTGTCCAG  
 pHYG03 241 ATCGGGGAATTCAGCGAGAGCCTGCCCTATTGCACCTCCCGCCGGCACAGGGTGTCCAG  
 pHYG05 241 ATTGGGGAACTCAGCGAGGGCCTGACCTATTGCATCTCCCGCCGGCACAGGGTGTCCAG  
 pHYG07 241 GTTGGGGAGTTCAGCGAGAGCCTGACCTGTTGCATCTCCCGCCGTGCACAGGGTGTCCAG  
 pHYG09 240 AGTGGGGAAATTCAGCGGAGCCTGACTTATTGGCTCTCCCGCCGTGCACAGGGTGTCCAG  
 pHYG11 241 ATTGGGGAACTCAGCGAGGGCCGACCTACTGCCTCTCCCGCCGTGCACAGGGTGTCCAG

pHygro-Tub 291 TTGCAAGACCTGCCTGAAACCGAACTGCCCGCTGTTCTGCAGCCGGTTCGCGGAGGCCATG  
 pHYGWT 301 TTGCAAGACCTGCCTGAAACCGAACTGCCCGCTGTTCTGCAGCCGGTTCGCGGAGGCCATG  
 pHYG01 301 TTGCAAGACCTGCCTGAAACCGAACTGCCCGCTGTTCTGCAGCCGGTTCGCGGAGGCCATG  
 pHYG02 300 TTGCAAGACCTGCCTGAAACCGAACTGCCCGCTGTTCTGCAGCCGGTTCGCGGAGGCCATG  
 pHYG03 301 TTGCAAGACCTGCCTGAAACCGAACTGCCCGCTGTTCTGCAGCCGGTTCGCGGAGGCCATG  
 pHYG05 301 TTGTAAGACCTGCCTGAAACCGAACTGCCCGCTGCTCTGCAGCCGGTTCGCGGGGCCATG  
 pHYG07 301 TTGCAAGACCTGCCTGAAACCGAACTGCCCGCTGCTCTGCAGCCGGTTCGCGGAGGCCATG  
 pHYG09 300 TTGTAAGACCTGCCTGAAACCGAACTGCCCGCTGCTCTGCAGCTGGTTCGCGGAGGCCATG  
 pHYG11 301 CTGCAAGACCTGCCTGAGACCGAACTGCCCGCTGCTTTCAGCCGGTTCGCGGAGGCCATG

pHygro-Tub 351 GATGCGATCGCTGCGGCCGATCTTAGCCAGACGAGCGGGTTCGGCCCATTCGGACCGCAA  
 pHYGWT 361 GATGCGATCGCTGCGGCCGATCTTAGCCAGACGAGCGGGTTCGGCCCATTCGGACCGCAA  
 pHYG01 361 GATGCGATCGCTGCGGCCGATCTTAGCCAGACGAGCGGGTTCGGCCCATTCGGACCGCAA  
 pHYG02 360 GATGCGACCGCTGCGGCCGATCTTAGCCAGACGAGCGGGTTCGGCCCATTCGGACCGCAA  
 pHYG03 361 GATGCGACCGCTGCGGCCGATCTTAGCCAGACGAGCGGGTTCGGCCCATTCGGACCGCAA  
 pHYG05 361 GACCGGGTTCGCTGCGGCCGCTTAGCTAGACGAGCGGGTTCGGCCCATTCGGACCGCAA  
 pHYG07 361 GGTGCGACCGCTGCGGCCGCTTAGCTAGCCAGACGAGCGGATTCGGCCCATTCGGACCGCAA  
 pHYG09 360 GATGCGATCGCTGCGGCCGATCTTAGCCAGACGAGCGGGTTCGGCCCATTCGGACCGCAA  
 pHYG11 361 GATGCGACCGCTGCGGCCGATCTTAGCCAGACGAGCGGGTTCGGCCCATTCGGACCGTAA

pHygro-Tub 411 GGAATCGGTCAATACACTACATGGCGTGATTTTCATATGCGCGATTGCTGATCCCCATGTG  
 pHYGWT 421 GGAATCGGTCAATACACTACATGGCGTGATTTTCATATGCGCGATTGCTGATCCCCATGTG  
 pHYG01 421 GGAATCGGTCAATACACTACATGGCGTGATTTTCATATGCGCGATTGCTGATCCCCATGTG  
 pHYG02 420 GGAATCGGTCAATACACTACATGGCGTGATTTTCATATGCGCGATTGCTGATCCCCATGTG  
 pHYG03 421 GGAATCGGTCAATACACTACATGGCGTGATTTTCATATGCGCGATTGCTGATCCCCATGTG  
 pHYG05 421 GGAATCGGTCAATACACTACATGGCGTGATTTTCATATGCGCGATTGCTGATCCCCATGTG  
 pHYG07 421 GGAATCGGTCAATATACTACATGGCATGATTTTCATATGCGCGATTGCTGATCCCCATGTG  
 pHYG09 420 GGAATCGGTCAATACACTACATGGCGTGATTTTCATATGCGCGATTGCTGATCCCCATGTG  
 pHYG11 421 GGAATCGGTCAATACACTACATGGCGTGATTTTCATATGCGCGATTGCTGATCCCCATGTG

pHygro-Tub 471 TATCACTGGCAAACCTGTGATGGACGACACCGTCAGTGCCTCCGTCCGCGAGGCTCTCGAT  
 pHYGWT 481 TATCACTGGCAAACCTGTGATGGACGACACCGTCAGTGCCTCCGTCCGCGAGGCTCTCGAT  
 pHYG01 481 TATCACTGGCAAACCTGTGATGGACGACACCGTCAGTGCCTCCGTCCGCGAGGCTCTCGAT  
 pHYG02 480 TATCACTGGCAGACTGTGATGGACGACACCGTCAGTGCCTCCGTCCGCGAGGCTCTCGAT  
 pHYG03 481 TATCACTGGCAAACCTGTGATGGACGACACCGTCAGTGCCTCCGTCCGCGAGGCTCTCGAT  
 pHYG05 481 TATCACTGGCAAACCTGTGATGGACAACACCGTCAGTGCCTCCGTCCGCGAGGCTCTCGAT  
 pHYG07 481 TATCACTGGCAGGCTGTGATGGACGACACCGTCAGTGCCTCCGTCCGCGAGGCTCTCGAT  
 pHYG09 480 TATCACTAGCAAACCTGTGATGGACGACACCGTCAGTGCCTCCGTCCGCGAGGCTCTCGAT  
 pHYG11 481 CATCACTGGCAAACCTGTGATGGACGACACCGTCAGTGCCTCCGTCCGCGAGGCTCTCGAT

pHygro-Tub 531 GAGCTGATGCTTTGGGCCGAGGACTGCCCCGAAGTCCGGCACCTCGTGCACGCGGATTTTC  
 pHYGWT 541 GAGCTGATGCTTTGGGCCGAGGACTGCCCCGAAGTCCGGCACCTCGTGCACGCGGATTTTC  
 pHYG01 541 GAGCTGATGCTTTGGGCCGAGGACTGCCCCGAAGTCCGGCACCTCGTGCACGCGGATTTTC  
 pHYG02 540 GAGCTGATGCTTTGGGCCGAGGACTGCCCCGAAGTCCGGCACCTCGTGCACGCGGATTTTC  
 pHYG03 541 GAGTTGATGCTTTGGGCCGAGGACTGCCCCGAAGTCCGGCACCTCGTGCACGCGGATTTTC  
 pHYG05 541 GAGCTGATGCTTTGGGCCGAGGACTGCCCCGAGTCCGGCACCTCGTGCACGCGGATTTTC  
 pHYG07 541 GGGCTGCTGCTTTGGGCCGAGGACTGCCCCGAGTCCGGCACCTCGTGCACGCGGATTTTC  
 pHYG09 540 GGGTTAATGCTTTGGGCCGAGGACTGCCCCGAGTCCGGCACCTCGTGCACGCGGATTTTC  
 pHYG11 541 GAGCTGACGCTTTGGGCCGAGGACTGCCCCGAAGTCCGGCACCTCGTGCATGCGGACTTTTC

## Appendix 6 continued

pHygro-Tub 591 GGCTCCAACAATGTCTGACGGACAATGGCCGCATAACAGCGGTCAATGACTGGAGCGAG  
 pHYGWT 601 GGCTCCAACAATGTCTGACGGACAATGGCCGCATAACAGCGGTCAATGACTGGAGCGAG  
 pHYG01 601 GGCTCCAACAATGTCTGACGGACAATGGCCGCATAACAGCGGTCAATGACTGGAGCGAG  
 pHYG02 600 GGCTCCAACAATGTCTGACGGACAATGGCCGCATAACAGCGGTCAATGACTGGAGCGAG  
 pHYG03 601 GGCTCCAACAATGTCTGACGGACAATGGCCGCATAACAGCGGTCAATGACTGGAGCGAG  
 pHYG05 601 GGCTCCAACAATGTCTGACGGACAATGGCCGCATAACAGCGGTCAATGACTGGAGCGAG  
 pHYG07 601 GGCTCCAACAATGTCTGACGGACAATGGCCGCATAACAGCGGTCAATGACTGGAGCGAG  
 pHYG09 600 GGCTCCAACAATGTCTGACGGACAATGGCCGCATAACAGCGGTCAATGACTGGAGCGAG  
 pHYG11 601 GGCTCCAACAATGTCTGACGGACAATGGCCGCATAACAGCGGTCAATGACTGGAGCGAG

pHygro-Tub 651 GCGATGTTCCGGGATTCCCAATACGAGGTCGCCAACATCTTCTTCTGGAGGCCGTGGTTG  
 pHYGWT 661 GCGATGTTCCGGGATTCCCAATACGAGGTCGCCAACATCTTCTTCTGGAGGCCGTGGTTG  
 pHYG01 661 GCGATGTTCCGGGATTCCCAATACGAGGTCGCCAACATCTTCTTCTGGAGGCCGTGGTTG  
 pHYG02 660 GCGATGTTCCGGGATTCCCAATACGAGGTCGCCAACATCTTCTTCTGGAGGCCGTGGTTG  
 pHYG03 661 GCGATGTTCCGGGATTCCCAATACGAGGTCGCCAACATCTTCTTCTGGAGGCCGTGGTTG  
 pHYG05 661 GCGATGTTCCGGGATTCCCAATACGAGGTCGCCAACATCTTCTTCTGGAGGCCGTGGTTG  
 pHYG07 661 GCGATGTTCCGGGATTCCCAATACGAGGTCGCCAACATCTTCTTCTGGAGGCCGTGGTTG  
 pHYG09 660 GCGATGTTCCGGGATTCCCAATACGAGGTCGCCAACATCTTCTTCTGGAGGCCGTGGTTG  
 pHYG11 661 GCGATGTTCCGGGATTCCCAATACGAGGTCGCCAACATCTTCTTCTGGAGGCCGTGGTTG

pHygro-Tub 711 GCTTGTATGGAGCAGCAGACCGCTACTTCGAGCGGAGGCATCCGGAGCTTGCAGGATCG  
 pHYGWT 721 GCTTGTATGGAGCAGCAGACCGCTACTTCGAGCGGAGGCATCCGGAGCTTGCAGGATCG  
 pHYG01 721 GCTTGTATGGAGCAGCAGACCGCTACTTCGAGCGGAGGCATCCGGAGCTTGCAGGATCG  
 pHYG02 720 GCTTGTATGGAGCAGCAGACCGCTACTTCGAGCGGAGGCATCCGGAGCTTGCAGGATCG  
 pHYG03 721 GCTTGTATGGAGCAGCAGACCGCTACTTCGAGCGGAGGCATCCGGAGCTTGCAGGATCG  
 pHYG05 721 GCTTGTATGGAGCAGCAGACCGCTACTTCGAGCGGAGGCATCCGGAGCTTGCAGGATCG  
 pHYG07 721 GCTTGTATGGAGCAGCAGACCGCTACTTCGAGCGGAGGCATCCGGAGCTTGCAGGATCG  
 pHYG09 720 GCTTGTATGGAGCAGCAGACCGCTACTTCGAGCGGAGGCATCCGGAGCTTGCAGGATCG  
 pHYG11 721 GCTTGTATGGAGCAGCAGACCGCTACTTCGAGCGGAGGCATCCGGAGCTTGCAGGATCG

pHygro-Tub 771 CCGCGCTCCGGGCGTATATGCTCCGCATTGGTCTTGACCAACTCTATCAGAGCTTGGTT  
 pHYGWT 781 CCGCGCTCCGGGCGTATATGCTCCGCATTGGTCTTGACCAACTCTATCAGAGCTTGGTT  
 pHYG01 781 CCGCGCTCCGGGCGTATATGCTCCGCATTGGTCTTGACCAACTCTATCAGAGCTTGGTT  
 pHYG02 780 CCGCGCTCCGGGCGTATATGCTCCGCATTGGTCTTGACCAACTCTATCAGAGCTTGGTT  
 pHYG03 781 CCGCGCTCCGGGCGTATATGCTCCGCATTGGTCTTGACCAACTCTATCAGAGCTTGGTT  
 pHYG05 781 CCGCGCTCCGGGCGTATATGCTCCGCATTGGTCTTGACCAACTCTATCAGAGCTTGGTT  
 pHYG07 781 CCGCGCTCCGGGCGTATATGCTCCGCATTGGTCTTGACCAACTCTATCAGAGCTTGGTT  
 pHYG09 780 CCGCGCTCCGGGCGTATATGCTCCGCATTGGTCTTGACCAACTCTATCAGAGCTTGGTT  
 pHYG11 781 CCGCGCTCCGGGCGTATATGCTCCGCATTGGTCTTGACCAACTCTATCAGAGCTTGGTT

pHygro-Tub 831 GACGGCAATTTCAATGATGCAGCTTGGGCGCAGGGTCGATGCGACGCAATCGTCCGATCC  
 pHYGWT 841 GACGGCAATTTCAATGATGCAGCTTGGGCGCAGGGTCGATGCGACGCAATCGTCCGATCC  
 pHYG01 841 GACGGCAATTTCAATGATGCAGCTTGGGCGCAGGGTCGATGCGACGCAATCGTCCGATCC  
 pHYG02 840 GACGGCAATTTCAATGATGCAGCTTGGGCGCAGGGTCGATGCGACGCAATCGTCCGATCC  
 pHYG03 841 GACGGCAATTTCAATGATGCAGCTTGGGCGCAGGGTCGATGCGACGCAATCGTCCGATCC  
 pHYG05 841 GACGGCAATTTCAATGATGCAGCTTGGGCGCAGGGTCGATGCGACGCAATCGTCCGATCC  
 pHYG07 841 GACGGCAATTTCAATGATGCAGCTTGGGCGCAGGGTCGATGCGACGCAATCGTCCGATCC  
 pHYG09 840 GACGGCAATTTCAATGATGCAGCTTGGGCGCAGGGTCGATGCGACGCAATCGTCCGATCC  
 pHYG11 841 GACGGCAATTTCAATGATGCAGCTTGGGCGCAGGGTCGATGCGACGCAATCGTCCGATCC

pHygro-Tub 891 GGAG . . . . .  
 pHYGWT 901 GGAGGGCGCGCCAT  
 pHYG01 901 GGAGGGCGCGCCAT  
 pHYG02 900 GGAGGGCGCGCCAT  
 pHYG03 901 GGAGGGCGCGCCAT  
 pHYG05 901 GGAGGGCGCGCCAT  
 pHYG07 901 GGAGGGCGCGCCAT  
 pHYG09 900 GGAGGGCGCGCCAT  
 pHYG11 901 GGAGGGCGCGCCAT

## LIST OF REFERENCES

- Aaltonen, L. A., Peltomäki, P., Leach, F. S., Sistonen, P., Pylkkänen, L., Mecklin, J. P., Järvinen, H., Powell, S. M., Jen, J., Hamilton, S. R., Petersen, G. M., Kinzler, K. W., Vogelstein, B., and de la Chapelle, A. (1993) Clues to the pathogenesis of familial colorectal cancer. *Science* **260**, 812-816
- 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. *Proceedings of the National Academy of Sciences USA* **93**, 13629-13634
- Adé, J., Belzile, F., Philippe, H., and Doutriaux, M.-P. (1999) Four mismatch repair paralogues coexist in *Arabidopsis thaliana*: *AtMSH2*, *AtMSH3*, *AtMSH6-1* and *AtMSH6-2*. *Molecular and General Genetics* **262**, 239-249
- Aebi, S., Fink, D., Gordon, R., Kim, H. K., Zheng, H., Fink, J. L., and Howell, S. B. (1997) Resistance to cytotoxic drugs in DNA mismatch repair-deficient cells. *Clinical Cancer Research* **3**, 1763-1767
- Alani, E. (1996) The *Saccharomyces cerevisiae* Msh2 and Msh6 proteins form a complex that specifically binds to duplex oligonucleotides containing mismatched DNA base pairs. *Molecular and Cellular Biology* **16**, 5604-5615
- Alani, E., Lee, S., Kane, M. F., Griffith, J., and Kolodner, R. D. (1997a) *Saccharomyces cerevisiae* MSH2, a mispaired base recognition protein, also recognizes Holliday junctions in DNA. *Journal of Molecular Biology* **265**, 289-301
- Alani, E., Reenan, R. A. G., and Kolodner, R. D. (1994) Interaction between mismatch repair and genetic recombination in *Saccharomyces cerevisiae*. *Genetics* **137**, 19-39
- Alani, E., Sokolsky, T., Studamire, B., Miret, J. J., and Lahue, R. S. (1997b) Genetic and biochemical analysis of Msh2p-Msh6p: role of ATP hydrolysis and Msh2p-Msh6p subunit interactions in mismatch base pair recognition. *Molecular and Cellular Biology* **17**, 2436-2447
- Aline, R., MacDonald, G., Brown, E., Allison, J., Myler, P. J., Rothwell, V., and Stuart, K. (1985) (TAA)<sub>n</sub> within sequences flanking several intrachromosomal variant surface glycoprotein genes in *Trypanosoma brucei*. *Nucleic Acids Research* **13**, 3161-3177
- Aline, R. F. and Stuart, K. (1989) *Trypanosoma brucei*: conserved sequence organization 3' to telomeric variant surface glycoprotein genes. *Experimental Parasitology* **68**, 57-66
- Aquilina, G. and Bignami, M. (2001) Mismatch repair in correction of replication errors and processing of DNA damage. *Journal of Cellular Physiology* **187**, 145-154
- Aronshtam, A. and Marinus, M. G. (1996) Dominant negative mutator mutations in the *mutL* gene of *Escherichia coli*. *Nucleic Acids Research* **24**, 2498-2504

Bahl, A., Brunk, B., Coppel, R. L., Crabtree, J., Diskin, S. J., Fraunholz, M. J., Grant, G. R., Gupta, D., Huestis, R. L., Kissinger, J. C., Labo, P., Li, L., McWeeney, S. K., Milgram, A. J., Roos, D. S., Schug, J., and Stoeckert, C. J. (2002) PlasmoDB: the *Plasmodium* genome resource. An integrated database providing tools for accessing, analyzing and mapping expression and sequence data (both finished and unfinished). *Nucleic Acids Research* **30**, 87-90

Baker, S. M., Plug, A. W., Prolla, T. A., Bronner, C. E., Harris, A. C., Yao, X., Christie, D.-M., Monell, C., Arnheim, N., Bradley, A., Ashley, T., and Liskay, R. M. (1996) Involvement of mouse *Mlh1* in DNA mismatch repair and meiotic crossing over. *Nature Genetics* **13**, 336-342

Baldauf, S. L., Palmer, J. D., and Doolittle, W. F. (1996) The root of the universal tree and the origin of eukaryotes based on elongation factor phylogeny. *Proceedings of the National Academy of Sciences USA* **93**, 7749-7754

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

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

Barbour, A. G. (2002) Antigenic variation of relapsing fever *Borrelia* and other pathogenic bacteria. In: *Mobile DNA II*, (eds. Craig, N. L., Craigie, R., Gellert, M., and Lambowitz, A.), pp. 972-994. ASM Press, Washington, DC.

Barbour, A. G. and Restrepo, B. I. (2000) Antigenic variation in vector-borne pathogens. *Emerging Infectious Diseases* **6**, 449-457

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

Barry, J. D. (1986b) The molecular biology of African trypanosomes. *Tropical Diseases Bulletin* **83**, R1-R25

Barry, J. D. (1997a) The biology of antigenic variation in African trypanosomes. In: *Trypanosomiasis and Leishmaniasis: biology and control*, (eds. Hide, G., Mottram, J. C., Coombs, G. H., and Holmes, P. H.), pp. 89-107. British Society for Parasitology/CAB International, Oxford.

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

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

- Barry, J. D. and McCulloch, R. (2001) Antigenic variation in trypanosomes: enhanced phenotypic variation in a eukaryotic parasite. *Advances in Parasitology* **49**, 1-70
- Bastin, P., Ellis, K., Kohl, L., and Gull, K. (2000) Flagellum ontogeny in trypanosomes studied via an inherited and regulated RNA interference system. *Journal of Cell Science* **113**, 3321-3328
- Baumann, P., Benson, F. E., and West, S. C. (1996) Human Rad51 protein promotes ATP-dependent homologous pairing and strand transfer reactions *in vitro*. *Cell* **87**, 757-766
- Beckmann, J. S. and Weber, J. L. (1992) Survey of human and rat microsatellites. *Genomics* **12**, 627-631
- Beeson, J. G. and Brown, G. V. (2002) Pathogenesis of *Plasmodium falciparum* malaria: the roles of parasite adhesion and antigenic variation. *Cellular and Molecular Life Sciences* **59**, 258-271
- Benson, F. E., Stasiak, A., and West, S. C. (1994) Purification and characterization of the human Rad51 protein, an analogue of *E. coli* RecA. *EMBO Journal* **13**, 5764-5771
- Berardini, M., Mazurek, A., and Fishel, R. (2000) The effect of O<sup>6</sup>-methylguanine DNA adducts on the adenosine nucleotide switch function of hMSH2-hMSH6 and hMSH2-hMSH3. *Journal of Biological Chemistry* **275**, 27851-27857
- Bergcrat, A., de Massy, B., Gadelle, D., Varoutas, P.-C., Nicolas, A., and Forterre, P. (1997) An atypical topoisomerase II from archaea with implications for meiotic recombination. *Nature* **386**, 414-417
- Bernards, A., De Lange, T., Michels, P. A. M., Liu, A. Y. C., Huisman, M. J., and Borst, P. (1984) Two modes of activation of a single surface antigen gene of *Trypanosoma brucei*. *Cell* **36**, 163-170
- Bertrand, P., Tishkoff, D. X., Filosi, N., Dasgupta, R., and Kolodner, R. D. (1998) Physical interaction between components of DNA mismatch repair and nucleotide excision repair. *Proceedings of the National Academy of Sciences USA* **95**, 14278-14283
- Bichara, M., Pinet, I., Schumacher, S., and Fuchs, R. P. P. (2000) Mechanisms of dinucleotide repeat instability in *Escherichia coli*. *Genetics* **154**, 533-542
- Bignami, M., O'Driscoll, M., Aquilina, G., and Karran, P. (2000) Unmasking a killer: DNA O<sup>6</sup>-methylguanine and the cytotoxicity of methylating agents. *Mutation Research - Reviews in Mutation Research* **462**, 71-82
- 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. *Journal of Molecular Biology* **305**, 805-816

- Blackwell, L. J., Wang, S. T., and Modrich, P. (2001) DNA chain length dependence of formation and dynamics of hMutS $\alpha$ ·hMutL $\alpha$ -heteroduplex complexes. *Journal of Biological Chemistry* **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. *Molecular and Biochemical Parasitology* **76**, 215-229
- Borst, P., Bitter, W., Blundell, P. A., Chaves, I., Cross, M., Gerrits, H., van Leeuwen, F., McCulloch, R., Taylor, M., and Rudenko, G. (1998) Control of *VSG* gene expression sites in *Trypanosoma brucei*. *Molecular and Biochemical Parasitology* **91**, 67-76
- 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. *Journal of Molecular Biology* **306**, 957-968
- Branch, P., Aquilina, G., Bignami, M., and Karran, P. (1993) Defective mismatch binding and a mutator phenotype in cells tolerant to DNA damage. *Nature* **362**, 652-654
- Bringaud, F., Bitcau, N., Donelson, J. E., and Baltz, T. (2001) Conservation of metacyclic variant surface glycoprotein expression sites among different trypanosome isolates. *Molecular and Biochemical Parasitology* **113**, 67-78
- Bringaud, F., Peyruchaud, S., Baltz, D., Giroud, C., Simpson, L., and Baltz, T. (1995) Molecular characterization of the mitochondrial heat shock protein 60 gene from *Trypanosoma brucei*. *Molecular and Biochemical Parasitology* **74**, 119-123
- Bucci, C., Lavitola, A., Salvatore, P., Del Giudice, I., Massardo, D. R., Bruni, C. B., and Alifano, P. (1999) Hypermutation in pathogenic bacteria: frequent phase variation in meningococci is a phenotypic trait of a specialized mutator biotype. *Molecular Cell* **3**, 435-445
- Buermeyer, A. B., Deschênes, S. M., Baker, S. M., and Liskay, R. M. (1999) Mammalian DNA mismatch repair. *Annual Review of Genetics* **33**, 533-564
- Campbell, D. A., van Brec, M. P., and Boothroyd, J. C. (1984) The 5'-limit of transposition and upstream barren region of a trypanosome *VSG* gene: tandem 76 base-pair repeats flanking (TAA)<sub>90</sub>. *Nucleic Acids Research* **12**, 2759-2774
- Capbern, A., Giroud, C., Baltz, T., and Mattern, P. (1977) *Trypanosoma equiperdum*: étude des variations antigéniques au cours de la trypanosomose expérimentale du lapin. *Experimental Parasitology* **402**, 6-13
- Carethers, J. M., Hawn, M. T., Chauhan, D. P., Luce, M. C., Marra, G., Koi, M., and Boland, C. R. (1996) Competency in mismatch repair prohibits clonal expansion of cancer cells treated with *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine. *Journal of Clinical Investigation* **98**, 199-206

- Carruthers, V. B. and Cross, G. A. M. (1992) High-efficiency clonal growth of bloodstream- and insect-form *Trypanosoma brucei* on agarose plates. *Proceedings of the National Academy of Sciences USA* **89**, 8818-8821
- Chan, S. N., Harris, L., Bolt, E. L., Whitby, M. C., and Lloyd, R. G. (1997) Sequence specificity and biochemical characterization of the RusA Holliday junction resolvase of *Escherichia coli*. *Journal of Biological Chemistry* **272**, 14873-14882
- 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 Journal* **18**, 4846-4855
- Chaves, I., Zomerdijk, J., Dirks-Mulder, A., Dirks, R. W., Raap, A. K., and Borst, P. (1998) Subnuclear localization of the active variant surface glycoprotein gene expression site in *Trypanosoma brucei*. *Proceedings of the National Academy of Sciences USA* **95**, 12328-12333
- Chen, C., Merrill, B. J., Lau, P. J., Holm, C., and Kolodner, R. D. (1999) *Saccharomyces cerevisiae* pol30 (proliferating cell nuclear antigen) mutations impair replication fidelity and mismatch repair. *Molecular and Cellular Biology* **19**, 7801-7815
- Chen, W. and Jinks-Robertson, S. (1998) Mismatch repair proteins regulate heteroduplex formation during mitotic recombination in yeast. *Molecular and Cellular Biology* **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. L., 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*. *Molecular Cell* **8**, 1117-1127
- Chi, N. W. and Kolodner, R. D. (1994) The effect of DNA mismatches on the ATPase activity of MSH1, a protein in yeast mitochondria that recognizes DNA mismatches. *Journal of Biological Chemistry* **269**, 29993-29937
- Claij, N. and te Riele, H. (1999) Microsatellite instability in human cancer: a prognostic marker for chemotherapy? *Experimental Cell Research* **246**, 1-10
- 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. *Journal of Biological Chemistry* **275**, 36498-36501
- Claverys, J. P. and Lacks, S. A. (1986) Heteroduplex deoxyribonucleic acid base mismatch repair in bacteria. *Microbiological Reviews* **50**, 133-165
- Connolly, B., Parsons, C. A., Benson, F. E., Dunderdale, H. J., Sharples, G. J., Lloyd, R. G., and West, S. C. (1991) Resolution of Holliday junctions *in vitro* requires the *Escherichia coli* *ruvC* gene product. *Proceedings of the National Academy of Sciences USA* **88**, 6063-6067

- Constantinou, A., Davies, A. A., and West, S. C. (2001) Branch migration and Holliday junction resolution catalyzed by activities from mammalian cells. *Cell* **104**, 259-268
- 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. *Journal of Biological Chemistry* **277**, 21269-21277
- Conway, C., Proudfoot, C., Burton, P., Barry, J. D., and McCulloch, R. (2002b) Multiple pathways of recombination in *Trypanosoma brucei*. *Molecular Microbiology*, submitted
- Corpet, F. (1988) Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Research* **16**, 10881-10890
- Cox, E. C. (1976) Bacterial mutator genes and the control of spontaneous mutation. *Annual Review of Genetics* **10**, 135-156
- Cromic, G. A., Connolly, J. C., and Leach, D. R. F. (2001) Recombination at double-strand breaks and DNA ends: conserved mechanisms from phage to humans. *Molecular Cell* **8**, 1163-1174
- Cross, G. A. M. (1975) Identification, purification and properties of clone-specific glycoprotein antigens constituting the surface coat of *Trypanosoma brucei*. *Parasitology* **71**, 393-417
- Cross, G. A. M. (1996) Antigenic variation in trypanosomes: secrets surface slowly. *Bioessays* **18**, 283-291
- 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 Research* **28**, 463-471
- D'Amours, D. and Jackson, S. P. (2002) The Mre11 complex: at the crossroads of DNA repair and checkpoint signalling. *Nature Reviews Molecular Cell Biology* **3**, 317-327
- D'Atri, S., Tentori, L., Lacal, P. M., Graziani, G., Pagani, E., Benincasa, E., Zambruno, G., Bonmassar, E., and Jiricny, J. (1998) Involvement of the mismatch repair system in temozolomide-induced apoptosis. *Molecular Pharmacology* **54**, 334-341
- 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 *Saccharomyces cerevisiae*. *Molecular and Cellular Biology* **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. *Proceedings of the National Academy of Sciences USA* **94**, 9757-9762

- De Lange, T., Kooter, J. M., Luirink, J., and Borst, P. (1985) Transcription of a transposed trypanosome surface antigen gene starts upstream of the transposed segment. *EMBO Journal* **4**, 3299-3306
- De Lange, T., Kooter, J. M., Michels, P. A. M., and Borst, P. (1983) Telomere conversion in trypanosomes. *Nucleic Acids Research* **11**, 8149-8165
- de Las Alas, M. M., de Bruin, R. A. M., Ten Eyck, L., Los, G., and Howell, S. B. (1998) Prediction-based threading of the hMSH2 DNA mismatch repair protein. *FASEB Journal* **12**, 653-663
- 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
- Degtyareva, N. P., Greenwell, P., Hofmann, E. R., Hengartner, M. O., Zhang, L., Culotti, J. G., and Petes, T. D. (2002) *Caenorhabditis elegans* DNA mismatch repair gene *msh-2* is required for microsatellite stability and maintenance of genome integrity. *Proceedings of the National Academy of Sciences USA* **99**, 2158-2163
- Donelson, J. E. (1995) Mechanisms of antigenic variation in *Borrelia hermsii* and African trypanosomes. *Journal of Biological Chemistry* **270**, 7783-7786
- Dong, C. M., Whitford, R., and Langridge, P. (2002) A DNA mismatch repair gene links to the *Ph2* locus in wheat. *Genome* **45**, 116-124
- Drotschmann, K., Aronshtam, A., Fritz, H.-J., and Marinus, M. G. (1998) The *Escherichia coli* MutL protein stimulates binding of Vsr and MutS to heteroduplex DNA. *Nucleic Acids Research* **26**, 948-953
- Drotschmann, K., Yang, W., Brownwell, F. E., Kool, E. T., and Kunkel, T. A. (2001) Asymmetric recognition of DNA local distortion: structure-based functional studies of eukaryotic Msh2-Msh6. *Journal of Biological Chemistry* **276**, 46225-46229
- Drummond, J. T., Genschel, J., Wolf, E., and Modrich, P. (1997) *DHFR/MSH3* amplification in methotrexate-resistant cells alters the hMutS $\alpha$ /hMutS $\beta$  ratio and reduces the efficiency of base-base mismatch repair. *Proceedings of the National Academy of Sciences USA* **94**, 10144-10149
- Drummond, J. T., Li, G. M., Longley, M. J., and Modrich, P. (1995) Isolation of an hMSH2-p160 heterodimer that restores DNA mismatch repair to tumor cells. *Science* **268**, 1909-1912
- Duckett, D. R., Drummond, J. T., Murchie, A. I. H., Reardon, J. T., Sancar, A., Lilley, D. M. J., and Modrich, P. (1996) Human MutS $\alpha$  recognizes damaged DNA base pairs containing O<sup>6</sup>-methylguanine, O<sup>4</sup>-methylthymine, or the cisplatin-d(GpG) adduct. *Proceedings of the National Academy of Sciences USA* **93**, 6443-6447
- 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. *Current Biology* **9**, 51-54

- Dutta, R. and Inouye, M. (2000) GHKL, an emergent ATPase/kinase superfamily. *Trends in Biochemical Sciences* **25**, 24-28
- Earley, M. C. and Crouse, G. F. (1998) The role of mismatch repair in the prevention of base pair mutations in *Saccharomyces cerevisiae*. *Proceedings of the National Academy of Sciences USA* **95**, 15487-15491
- Edelmann, W. and Kucherlapati, R. (1996) Role of recombination enzymes in mammalian cell survival. *Proceedings of the National Academy of Sciences USA* **93**, 6225-6227
- Eid, J. and Sollner-Webb, B. (1991) Stable integrative transformation of *Trypanosoma brucei* that occurs exclusively by homologous recombination. *Proceedings of the National Academy of Sciences USA* **88**, 2118-2121
- Eisen, J. A. (1998) A phylogenomic study of the MutS family of proteins. *Nucleic Acids Research* **26**, 4291-4300
- Ellegren, H. (2000a) Heterogeneous mutation processes in human microsatellite DNA sequences. *Nature Genetics* **24**, 400-402
- Ellegren, H. (2000b) Microsatellite mutations in the germline: implications for evolutionary inference. *Trends in Genetics* **16**, 551-558
- Esposito, M. S. (1978) Evidence that spontaneous mitotic recombination occurs at the two-strand stage. *Proceedings of the National Academy of Sciences USA* **75**, 4436-4440
- Evans, E. and Alani, E. (2000) Roles for mismatch repair factors in regulating genetic recombination. *Molecular and Cellular Biology* **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*. *Molecular Cell* **5**, 789-799
- Fabisiewicz, A. and Worth, L. (2001) *Escherichia coli* MutS,L modulate RuvAB-dependent branch migration between diverged DNA. *Journal of Biological Chemistry* **276**, 9413-9420
- Fanning, T. G. and Taubenberger, J. K. (1999) Phylogenetically important regions of the Influenza A H1 hemagglutinin protein. *Virus Research* **65**, 33-42
- Featherstone, C. and Jackson, S. P. (1999) Ku, a DNA repair protein with multiple cellular functions? *Mutation Research - DNA Repair* **434**, 3-15
- Fishel, R. (1998) Mismatch repair, molecular switches, and signal transduction. *Genes & Development* **12**, 2096-2101
- Fishel, R., Acharya, S., Berardini, M., Boecker, T., Charbonneau, N., Cranston, A., Gradia, S., Guerrette, S., Heinen, C. D., Mazurek, A., Snowden, T., Schmutte, C., Shim, K.-S., Tomblin, G., and Wilson, T. (2000) Signaling mismatch repair: The mechanics of an

adenosine-nucleotide molecular switch. *Cold Spring Harbor Symposia on Quantitative Biology* **65**, 217-224

Fishel, R. and Kolodner, R. D. (1995) Identification of mismatch repair genes and their role in the development of cancer. *Current Opinion in Genetics & Development* **5**, 382-395

Fishel, R. and Wilson, T. (1997) MutS homologs in mammalian cells. *Current Opinion in Genetics & Development* **7**, 105-113

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

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

Flores-Rozas, H. and Kolodner, R. D. (1998) The *Saccharomyces cerevisiae* *MLH3* gene functions in MSH3-dependent suppression of frameshift mutations. *Proceedings of the National Academy of Sciences USA* **95**, 12404-12409

Flores, C. and Engels, W. (1999) Microsatellite instability in *Drosophila spellchecker1* (MutS homolog) mutants. *Proceedings of the National Academy of Sciences USA* **96**, 2964-2969

Furuse, M., Nagase, Y., Tsubouchi, H., Murakami-Murofushi, K., Shibata, T., and Ohta, T. (1998) Distinct roles of two separable *in vitro* activities of yeast Mre11 in mitotic and meiotic recombination. *EMBO Journal* **17**, 6412-6425

Gardner, J. P., Pinches, R. A., Roberts, D. J., and Newbold, C. I. (1996) Variant antigens and endothelial receptor adhesion in *Plasmodium falciparum*. *Proceedings of the National Academy of Sciences USA* **93**, 3503-3508

Genschel, J., Littman, S. J., Drummond, J. T., and Modrich, P. (1998) Isolation of MutS $\beta$  from human cells and comparison of the mismatch repair specificities of MutS $\beta$  and MutS $\alpha$ . *Journal of Biological Chemistry* **273**, 19895-19901

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

Goldmacher, V. S., Cuzick, R. A., 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. *Journal of Biological Chemistry* **261**, 2462-2471

Gorbalenya, A. E. and Koonin, E. V. (1990) Superfamily of UvrA-related NTP-binding proteins. Implications for rational classification of recombination/repair systems. *Journal of Molecular Biology* **213**, 583-591

- Gorman, O. T., Bean, W. J., and Webster, R. G. (1992) Evolutionary processes in influenza viruses: divergence, rapid evolution, and stasis. *Current Topics in Microbiology and Immunology* **176**, 75-97
- 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., Subramanian, D., Wilson, T., Acharya, S., Makhov, A., Griffith, J., and Fishel, R. (1999) hMSH2-hMSH6 forms a hydrolysis-independent sliding clamp on mismatched DNA. *Molecular Cell* **3**, 255-261
- Gray, A. R. (1965) Antigenic variation in a strain of *Trypanosoma brucei* transmitted by *Glossina morsitans* and *G. palpalis*. *Journal of General Microbiology* **41**, 195-214
- Greene, C. N. and Jinks-Robertson, S. (1997) Frameshift intermediates in homopolymer runs are removed efficiently by yeast mismatch repair proteins. *Molecular and Cellular Biology* **17**, 2844-2850
- 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., Griffith, J., and Modrich, P. (1993) Bidirectional excision in methyl-directed mismatch repair. *Journal of Biological Chemistry* **268**, 11830-11837
- Gu, L. Y., 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 Research* **26**, 1173-1178
- Guerrette, S., Acharya, S., and Fishel, R. (1999) The interaction of the human MutL homologues in hereditary nonpolyposis colon cancer. *Journal of Biological Chemistry* **274**, 6336-6341
- Guerrette, S., Wilson, T., Gradia, S., and Fishel, R. (1998) Interactions of human hMSH2 with hMSH3 and hMSH2 with hMSH6: Examination of mutations found in hereditary nonpolyposis colorectal cancer. *Molecular and Cellular Biology* **18**, 6616-6623
- Gull, K., Alsford, S., and Ersfeld, K. (1998) Segregation of minichromosomes in trypanosomes: implications for mitotic mechanisms. *Trends in Microbiology* **6**, 319-323
- Haber, J. E. (1999) DNA recombination: the replication connection. *Trends in Biochemical Sciences* **24**, 271-275
- Haber, J. E. (2000) Partners and pathways: repairing a double-strand break. *Trends in Genetics* **16**, 259-264
- Haber, J. E. and Heyer, W.-D. (2001) The fuss about Mus81. *Cell* **107**, 551-554

- Haber, L. T. and Walker, G. C. (1991) Altering the conserved nucleotide binding motif in the *Salmonella typhimurium* MutS mismatch repair protein affects both its ATPase and mismatch binding activities. *EMBO Journal* **10**, 2707-2715
- Habraken, Y., Sung, P., Prakash, L., and Prakash, S. (1996) Binding of insertion/deletion DNA mismatches by the heterodimer of yeast mismatch repair proteins MSH2 and MSH3. *Current Biology* **6**, 1185-1187
- Habraken, Y., Sung, P., Prakash, L., and Prakash, S. (1997) Enhancement of MSH2-MSH3-mediated mismatch recognition by the yeast MLH1-PMS1 complex. *Current Biology* **7**, 790-793
- 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-MSH3 and MLH1-PMS1 protein complexes. *Journal of Biological Chemistry* **273**, 9837-9841
- Hajduk, S. L., Hager, K., and Esko, J. D. (1992) High-density lipoprotein-mediated lysis of trypanosomes. *Parasitology Today* **8**, 95-98
- Hall, M. C., Shcherbakova, P. V., and Kunkel, T. A. (2002) Differential ATP binding and intrinsic ATP hydrolysis by amino-terminal domains of the yeast Mlh1 and Pms1 proteins. *Journal of Biological Chemistry* **277**, 3673-3679
- Hamm, B., Schindler, A., Mecke, D., and Duszenko, M. (1990) Differentiation of *Trypanosoma brucei* bloodstream trypomastigotes from long slender to short stumpy-like forms in axenic culture. *Molecular and Biochemical Parasitology* **40**, 13-22
- Harfe, B. D. and Jinks-Robertson, S. (2000) DNA mismatch repair and genetic instability. *Annual Review of Genetics* **34**, 359-399
- Hawn, M. T., Umar, A., Carethers, J. M., Marra, G., Kunkel, T. A., Boland, C. R., and Koi, M. (1995) Evidence for a connection between the mismatch repair system and the G2 cell cycle checkpoint. *Cancer Research* **55**, 3721-3725
- Henderson, I. R., Owen, P., and Nataro, J. P. (1999) Molecular switches - the ON and OFF of bacterial phase variation. *Molecular Microbiology* **33**, 919-932
- Henderson, S. T. and Petes, T. D. (1992) Instability of simple sequence DNA in *Saccharomyces cerevisiae*. *Molecular and Cellular Biology* **12**, 2749-2757
- Hirumi, H. and Hirumi, K. (1989) Continuous cultivation of *Trypanosoma brucei* bloodstream forms in a medium containing a low concentration of serum protein without feeder cell layers. *Journal of Parasitology* **75**, 985-989
- Holland, I. B. and Blight, M. A. (1999) ABC-ATPases, adaptable energy generators fueling transmembrane movement of a variety of molecules in organisms from bacteria to humans. *Journal of Molecular Biology* **293**, 381-399

- 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 & Development* **9**, 1728-1739
- Holmes, J., Clark, S., and Modrich, P. (1990) Strand-specific mismatch correction in nuclear extracts of human and *Drosophila melanogaster* cell lines. *Proceedings of the National Academy of Sciences USA* **87**, 5837-5841
- Hope, M., MacLeod, A., Leech, V., Melville, S., Sasse, J., Tait, A., and Turner, C. M. R. (1999) Analysis of ploidy (in megabase chromosomes) in *Trypanosoma brucei* after genetic exchange. *Molecular and Biochemical Parasitology* **104**, 1-9
- Hopfner, K. P., Putnam, C. D., and Tainer, J. A. (2002) DNA double-strand break repair from head to tail. *Current Opinion in Structural Biology* **12**, 115-122
- 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
- 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 Journal* **15**, 1726-1733
- Iaccarino, I., Marra, G., Palombo, F., and Jiricny, J. (1998) hMSH2 and hMSH6 play distinct roles in mismatch binding and contribute differently to the ATPase activity of hMutSa. *EMBO Journal* **17**, 2677-2686
- Iaccarino, I., Palombo, F., Drummond, J., Totty, N. F., Hsuan, J. J., Modrich, P., and Jiricny, J. (1996) MSH6, a *Saccharomyces cerevisiae* protein that binds to mismatches as a heterodimer with MSH2. *Current Biology* **6**, 484-486
- Jean, M., Pelletier, J., Hilpert, M., Belzile, F., and Kunze, R. (1999) Isolation and characterization of *AtMLH1*, a *MutL* homologue from *Arabidopsis thaliana*. *Molecular and General Genetics* **262**, 633-642
- Jeggo, P. A. (1998) DNA breakage and repair. *Advances in Genetics* **38**, 185-218
- Jenni, L., Marti, S., Schweizer, J., Betschart, B., Le Page, R. W. F., 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
- Jiricny, J. (1998) Eukaryotic mismatch repair: an update. *Mutation Research - DNA Repair* **409**, 107-121
- 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. *Journal of Biological Chemistry* **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. *Journal of Biological Chemistry* **271**, 7285-7288
- Jonsson, Z. O. and Hubscher, U. (1997) Proliferating cell nuclear antigen: more than a clamp for DNA polymerases. *Bioessays* **19**, 967-975
- Karran, P. (2000) DNA double strand break repair in mammalian cells. *Current Opinion in Genetics & Development* **10**, 144-150
- Karran, P. and Bignami, M. (1992) Self-destruction and tolerance in resistance of mammalian cells to alkylation damage. *Nucleic Acids Research* **20**, 2933-2940
- Karran, P. and Bignami, M. (1994) DNA damage tolerance, mismatch repair and genome instability. *Bioessays* **16**, 833-839
- Karran, P. and Marinus, M. G. (1982) Mismatch correction at  $\theta^6$ -methylguanine residues in *E. coli* DNA. *Nature* **296**, 868-869
- Katinka, M. D., Duprat, S., Cornillot, E., Méténier, G., Thomarat, F., Prensier, G., Barbe, V., Peyretailade, E., Brottier, P., Wincker, P., Delbac, F., El Alaoui, H., Peyret, P., Saurin, W., Gouy, M., Weissenbach, J., and Vivarès, C. P. (2001) Genome sequence and gene compaction of the eukaryote parasite *Encephalitozoon cuniculi*. *Nature* **414**, 450-453
- Kearsey, J. M., Coates, P. J., Prescott, A. R., Warbrick, E., and Hall, P. A. (1995) Gadd45 is a nuclear cell cycle regulated protein which interacts with p21Cip1. *Oncogene* **11**, 1675-1683
- Kelman, Z. (1997) PCNA: Structure, functions and interactions. *Oncogene* **14**, 629-640
- Kelman, Z. and Hurwitz, J. (1998) Protein-PCNA interactions: a DNA-scanning mechanism? *Trends in Biochemical Sciences* **23**, 236-238
- Kissinger, J. C., Milgram, A. J., Fraunholz, M. J., Roos, D. S., Brunk, B. P., Crabtree, J., Diskin, S. J., Schug, J., Overton, G. C., Stoeckert, C. J., Coppel, R. L., and Huestis, R. L. (2001) PlasmoDB: An integrative database of the *Plasmodium falciparum* genome. Tools for accessing and analyzing finished and unfinished sequence data. *Nucleic Acids Research* **29**, 66-69
- 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 & Development* **15**, 724-736
- Kleff, S., Kemper, B., and Sternglanz, R. (1992) Identification and characterization of yeast mutants and the gene for a cruciform cutting endonuclease. *EMBO Journal* **11**, 699-704
- Koi, M., Umar, A., Chauhan, D. P., Cherian, S. P., Carethers, J. M., Kunkel, T. A., and Boland, C. R. (1994) Human chromosome 3 corrects mismatch repair deficiency and

microsatellite instability and reduces *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine tolerance in colon tumor cells with homozygous hMLH1 mutation. *Cancer Research* **54**, 4308-4312

Kolodner, R. (1996) Biochemistry and genetics of eukaryotic mismatch repair. *Genes & Development* **10**, 1433-1442

Kolodner, R. D. and Marsischky, G. T. (1999) Eukaryotic DNA mismatch repair. *Current Opinion in Genetics & Development* **9**, 89-96

Kondo, E., Horii, A., and Fukushima, 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 Research* **29**, 1695-1702

Kraus, E., Leung, W. Y., and Haber, J. E. (2001) Break-induced replication: A review and an example in budding yeast. *Proceedings of the National Academy of Sciences USA* **98**, 8255-8262

Kwong, P. D., Wyatt, R., Robinson, J., Sweet, R. W., Sodroski, J. G., and Hendrickson, W. A. (1998) Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature* **393**, 648-659

LaCount, D. J., Bruse, S., Hill, K. L., and Donelson, J. E. (2000) Double-stranded RNA interference in *Trypanosoma brucei* using head-to-head promoters. *Molecular and Biochemical Parasitology* **111**, 67-76

Lamers, M. H., Perrakis, A., Enzlin, J. H., Winterwerp, H. H. K., de Wind, N., and Sixma, T. K. (2000) The crystal structure of DNA mismatch repair protein MutS binding to a G-T mismatch. *Nature* **407**, 711-717

Lamont, G. S., Tucker, R. S., and Cross, G. A. M. (1986) Analysis of antigen switching rates in *Trypanosoma brucei*. *Parasitology* **92**, 335-367

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

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

Lee, M. G. S. and Van der Ploeg, L. H. T. (1990) Homologous recombination and stable transfection in the parasitic protozoan *Trypanosoma brucei*. *Science* **250**, 1583-1587

Lenardo, M. J., Rice-Ficht, A. C., Kelly, G., Esser, K. M., and Donelson, J. E. (1984) Characterization of the genes specifying two metacyclic variable antigen types in *Trypanosoma brucei rhodesiense*. *Proceedings of the National Academy of Sciences USA* **81**, 6642-6646

Leung, W.-Y., 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. *Proceedings of the National Academy of Sciences USA* **94**, 6851-6856

- Levinson, G. and Gutman, G. A. (1987) High frequencies of short frameshifts in poly-CA/TG tandem repeats borne by bacteriophage M13 in *Escherichia coli* K-12. *Nucleic Acids Research* **15**, 5323-5338
- Li, F., Hua, S. B., Wang, C. C., and Gottesdiener, K. M. (1996) Procyclic *Trypanosoma brucei* cell lines deficient in ornithine decarboxylase activity. *Molecular and Biochemical Parasitology* **78**, 227-236
- Lin, C.-T., I.yu, Y. I., Xiao, H., Lin, W.-H., and Whang-Peng, J. (2001) Suppression of gene amplification and chromosomal DNA integration by the DNA mismatch repair system. *Nucleic Acids Research* **29**, 3304-3310
- Lin, Y.-L., Shivji, M. K. 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. *Journal of Biological Chemistry* **273**, 1453-1461
- 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. *Nature Genetics* **24**, 27-35
- Liu, A. Y. C., Michels, P. A. M., Bernards, A., and Borst, P. (1985) Trypanosome variant surface glycoprotein genes expressed early in infection. *Journal of Molecular Biology* **182**, 383-396
- Liu, A. Y. C., Van der Ploeg, L. H. T., Rijsewijk, F. A. M., 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. *Journal of Molecular Biology* **167**, 57-75
- Liu, B., Parsons, R., Papadopoulos, N., Nicolaidis, N. C., Lynch, H. T., Watson, P., Jass, J. R., Dunlop, M., Wyllie, A., Peltomäki, P., de la Chapelle, A., Hamilton, S. R., Vogelstein, B., and Kinzler, K. W. (1996) Analysis of mismatch repair genes in hereditary non-polyposis colorectal cancer patients. *Nature Medicine* **2**, 169-174
- Longley, M. J., Pierce, A. J., and Modrich, P. (1997) DNA polymerase  $\delta$  is required for human mismatch repair *in vitro*. *Journal of Biological Chemistry* **272**, 10917-10921
- Malkov, V. A., Biswas, I., Camerini-Otero, R. D., and Hsieh, P. (1997) Photocross-linking of the NH<sub>2</sub>-terminal region of Taq MutS protein to the major groove of a heteroduplex DNA. *Journal of Biological Chemistry* **272**, 23811-23817
- Manolis, K. G., Nimmo, F. R., Hartsuiker, E., Carr, A. M., Jeggo, P. A., and Allshire, R. C. (2001) Novel functional requirements for non-homologous DNA end joining in *Schizosaccharomyces pombe*. *EMBO Journal* **20**, 210-221
- Marra, G., Iaccarino, I., Lettieri, T., Roscilli, G., Delmastro, P., and Jiricny, J. (1998) Mismatch repair deficiency associated with overexpression of the *MSH3* gene. *Proceedings of the National Academy of Sciences USA* **95**, 8568-8573

- Marra, G. and Schär, P. (1999) Recognition of DNA alterations by the mismatch repair system. *Biochemical Journal* **338**, 1-13
- 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 & Development* **10**, 407-420
- Marsischky, G. T., Lee, S., Griffith, J., and Kolodner, R. D. (1999) *Saccharomyces cerevisiae* *MSH2/6* complex interacts with Holliday junctions and facilitates their cleavage by phage resolution enzymes. *Journal of Biological Chemistry* **274**, 7200-7206
- Maslov, D. A., Podlipaev, S. A., and Lukeš, J. (2001) Phylogeny of the kinetoplastida: taxonomic problems and insights into the evolution of parasitism. *Memorias do Instituto Oswaldo Cruz* **96**, 397-402
- Matthews, K. R. and Gull, K. (1994) Cycles within cycles: the interplay between differentiation and cell division in *Trypanosoma brucei*. *Parasitology Today* **10**, 473-476
- Matthews, K. R. and Gull, K. (1997) Commitment to differentiation and cell cycle re-entry are coincident but separable events in the transformation of African trypanosomes from their bloodstream to their insect form. *Journal of Cell Science* **110**, 2609-2618
- 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 Research* **18**, 7219-7227
- McArthur, A. G., Morrison, H. G., Nixon, J. E. J., Passamaneck, N. Q. E., Kim, U., Hinkle, G., Crocker, M. K., Holder, M. E., Farr, R., Reich, C. I., Olsen, G. E., Aley, S. B., Adam, R. D., Gillin, F. D., and Sogin, M. L. (2000) The *Giardia* genome project database. *FEMS Microbiology Letters* **189**, 271-273
- McCulloch, R. and Barry, J. D. (1999) A role for RAD51 and homologous recombination in *Trypanosoma brucei* antigenic variation. *Genes & Development* **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. *Molecular and Cellular Biology* **17**, 833-843
- 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. *Chemistry & Biology* **3**, 579-589
- Meyer, R. R. and Laine, P. S. (1990) The single-stranded DNA-binding protein of *Escherichia coli*. *Microbiological Reviews* **54**, 342-380
- Michels, P. A. M., Liu, A. Y. C., 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*. *Journal of Molecular Biology* **166**, 537-556

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

Modrich, P. (1991) Mechanisms and biological effects of mismatch repair. *Annual Review of Genetics* **25**, 229-253

Modrich, P. and Lahue, R. (1996) Mismatch repair in replication fidelity, genetic recombination, and cancer biology. *Annual Review of Biochemistry* **65**, 101-133

Moore, J. K. and Haber, J. E. (1996) Cell cycle and genetic requirements of two pathways of nonhomologous end-joining repair of double-strand breaks in *Saccharomyces cerevisiae*. *Molecular and Cellular Biology* **16**, 2164-2173

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 Research* **59**, 2102-2106

Mortensen, U. H., Bendixen, C., Sunjevaric, I., and Rothstein, R. (1996) DNA strand annealing is promoted by the yeast Rad52 protein. *Proceedings of the National Academy of Sciences USA* **93**, 10729-10734

Mottram, J. C. and Smith, G. (1995) A family of trypanosome cdc2-related protein kinases. *Gene* **162**, 147-152

Mushegian, A. R., Bassett, D. E., Boguski, M. S., Bork, P., and Koonin, E. V. (1997) Positionally cloned human disease genes: Patterns of evolutionary conservation and functional motifs. *Proceedings of the National Academy of Sciences USA* **94**, 5831-5836

Myung, K., Datta, A., Chen, C., and Kolodner, R. D. (2001) SGS1, the *Saccharomyces cerevisiae* homologue of BLM and WRN, suppresses genome instability and homologous recombination. *Nature Genetics* **27**, 113-116

Nakagawa, T., Datta, A., and Kolodner, R. D. (1999) Multiple functions of MutS- and MutL-related heterocomplexes. *Proceedings of the National Academy of Sciences USA* **96**, 14186-14188

Nantulya, V. M., Musoke, A. J., and Moloo, S. K. (1986) Apparent exhaustion of the variable antigen repertoires of *Trypanosoma vivax* in infected cattle. *Infection and Immunity* **54**, 444-447

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

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

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

Newbold, C. I. (1999) Antigenic variation in *Plasmodium falciparum*: mechanisms and consequences. *Current Opinion in Microbiology* **2**, 420-425

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

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., Oki, E., Maehara, Y., and Sugimachi, K. (1997) Precise assessment of microsatellite instability using high resolution fluorescent microsatellite analysis. *Nucleic Acids Research* **25**, 3415-3420

Oliveira, R. P., Broude, N. E., Macedo, A. M., Cantor, C. R., Smith, C. L., and Pena, S. D. J. (1998) Probing the genetic population structure of *Trypanosoma cruzi* with polymorphic microsatellites. *Proceedings of the National Academy of Sciences USA* **95**, 3776-3780

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

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

Pang, Q. S., 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. *Molecular and Cellular Biology* **17**, 4465-4473

Pâques, F. and Haber, J. E. (1999) Multiple pathways of recombination induced by double-strand breaks in *Saccharomyces cerevisiae*. *Microbiology and Molecular Biology Reviews* **63**, 349-404

Pastink, A., Eeken, J. C. J., and Lohman, P. H. M. (2001) Genomic integrity and the repair of double-strand DNA breaks. *Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis* **480-481**, 37-50

Pastink, A. and Lohman, P. H. M. (1999) Repair and consequences of double-strand breaks in DNA. *Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis* **428**, 141-156

Paull, T. T. and Gellert, M. (1998) The 3' to 5' exonuclease activity of Mre11 facilitates repair of DNA double-strand breaks. *Molecular Cell* **1**, 969-979

- Paull, T. T. and Gellert, M. (2000) A mechanistic basis for Mre11-directed DNA joining at microhomologies. *Proceedings of the National Academy of Sciences USA* **97**, 6409-6414
- Paulson, T. G., Galipeau, P. C., and Reid, B. J. (1999) Loss of heterozygosity analysis using whole genome amplification, cell sorting, and fluorescence-based PCR. *Genome Research* **9**, 482-491
- Pays, E., Delauw, M. F., Van Assel, S., Laurent, M., Vervoort, T., Van Meirvenne, N., and Steinert, M. (1983a) Modifications of a *Trypanosoma b. brucei* antigen gene repertoire by different DNA recombinational mechanisms. *Cell* **35**, 721-731
- Pays, E., Guyaux, M., Aerts, D., Van Meirvenne, N., and Steinert, M. (1985a) Telomeric reciprocal recombination as a possible mechanism for antigenic variation in trypanosomes. *Nature* **316**, 562-564
- Pays, E., Houard, S., Pays, A., Van Assel, S., Dupont, F., Aerts, D., Huet-Duvillier, G., Gomés, V., Richet, C., Degand, P., Van Meirvenne, N., and Steinert, M. (1985b) *Trypanosoma brucei*: the extent of conversion in antigen genes may be related to the DNA coding specificity. *Cell* **42**, 821-829
- Pays, E., Lips, S., Nolan, D., Vanhamme, L., and Pérez-Morga, D. (2001) The *VSG* expression sites of *Trypanosoma brucei*: multipurpose tools for the adaptation of the parasite to mammalian hosts. *Molecular and Biochemical Parasitology* **114**, 1-16
- Pays, E. and Nolan, D. P. (1998) Expression and function of surface proteins in *Trypanosoma brucei*. *Molecular and Biochemical Parasitology* **91**, 3-36
- Pays, E., Van Assel, S., Laurent, M., Darville, M., Vervoort, T., Van Meirvenne, N., and Steinert, M. (1983b) Gene conversion as a mechanism for antigenic variation in trypanosomes. *Cell* **34**, 371-381
- Peltomäki, P., Vasen, H. F. A., Bisgaard, M. L., Buerstedde, J. M., Friedl, W., Grandjouan, S., Hutter, P., Kohonen-Corish, M., Kolodner, R., Kurzawski, G., Lindblom, A., Lynch, H. T., Piccoli, A., Ponz de Leon, M., Radice, P., Thibodeau, S., Weber, W., West, S., and Wijnen, J. (1997) Mutations predisposing to hereditary nonpolyposis colorectal cancer: database and results of a collaborative study. *Gastroenterology* **113**, 1146-1158
- Petes, T. D., Greenwell, P. W., and Dominska, M. (1997) Stabilization of microsatellite sequences by variant repeats in the yeast *Saccharomyces cerevisiae*. *Genetics* **146**, 491-498
- Petukhova, G., Stratton, S., and Sung, P. (1998) Catalysis of homologous DNA pairing by yeast Rad51 and Rad54 proteins. *Nature* **393**, 91-94
- Philippe, H. and Adoutte, A. (1998) The molecular phylogeny of Eukaryota: solid facts and uncertainties. In: *Evolutionary relationships among protozoa*, (eds. Coombs, G. H., Vickerman, K., Sleigh, M. A., and Warren, A.), pp. 25-56. Chapman & Hall, London.
- Pochart, P., Woltering, D., and Hollingsworth, N. M. (1997) Conserved properties between functionally distinct MutS homologs in yeast. *Journal of Biological Chemistry* **272**, 30345-30349

- Pont-Kingdon, G., Okada, N. A., Macfarlane, J. L., Beagley, C. T., Watkins-Sims, C. D., Cavalier-Smith, T., Clark-Walker, G. D., and Wolstenholme, D. R. (1998) Mitochondrial DNA of the coral *Sarcophyton glaucum* contains a gene for a homologue of bacterial MutS: A possible case of gene transfer from the nucleus to the mitochondrion. *Journal of Molecular Evolution* **46**, 419-431
- Prelich, G., Tan, C. K., Kostura, M., Mathews, M. B., So, A. G., Downey, K. M., and Stillman, B. (1987) Functional identity of proliferating cell nuclear antigen and a DNA polymerase  $\delta$  auxiliary protein. *Nature* **326**, 517-520
- Priest, J. W. and Hajduk, S. L. (1994) Developmental regulation of mitochondrial biogenesis in *Trypanosoma brucei*. *Journal of Bioenergetics and Biomembranes* **26**, 179-191
- Prolla, T. A., Christie, D.-M., and Liskay, R. M. (1994a) Dual requirement in yeast DNA mismatch repair for *MLH1* and *PMS1*, two homologs of the bacterial *mutL* gene. *Molecular and Cellular Biology* **14**, 407-415
- Prolla, T. A., Pang, Q. S., Alani, E., Kolodner, R. D., and Liskay, R. M. (1994b) *MLH1*, *PMS1*, and *MSH2* interactions during the initiation of DNA mismatch repair in yeast. *Science* **265**, 1091-1093
- Ramilo, C., Gu, L. Y., Guo, S., Zhang, X., Patrick, S. M., Turchi, J. J., and Li, G.-M. (2002) Partial reconstitution of human DNA mismatch repair *in vitro*: characterization of the role of human replication protein A. *Molecular and Cellular Biology* **22**, 2037-2046
- Ray, D. S. and Hines, J. C. (1995) Disruption of the *Crithidia fasciculata RNH1* gene results in the loss of two active forms of ribonuclease H. *Nucleic Acids Research* **23**, 2526-2530
- 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
- Reenan, R. A. G. and Kolodner, R. D. (1992a) Characterization of insertion mutations in the *Saccharomyces cerevisiae MSH1* and *MSH2* genes: evidence for separate mitochondrial and nuclear functions. *Genetics* **132**, 975-985
- Reenan, R. A. G. and Kolodner, R. D. (1992b) Isolation and characterization of two *Saccharomyces cerevisiae* genes encoding homologs of the bacterial HexA and MutS mismatch repair proteins. *Genetics* **132**, 963-973
- 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. *Molecular and Cellular Biology* **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*. *Journal of Biological Chemistry* **277**, 26185-26193

- Roditi, I. and Pearson, T. W. (1990) The procyclin coat of African trypanosomes. *Parasitology Today* **6**, 79-82
- 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
- Rossolillo, P. and Albertini, A. M. (2001) Functional analysis of the *Bacillus subtilis* *yshD* gene, a mutS paralogue. *Molecular and General Genetics* **264**, 809-818
- Roth, C. W., Longacre, S., Raibaud, A., Baltz, T., and Eisen, H. (1986) The use of incomplete genes for the construction of a *Trypanosoma equiperdum* variant surface glycoprotein gene. *EMBO Journal* **5**, 1065-1070
- 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*. *Molecular and Biochemical Parasitology* **80**, 65-75
- Rudolph, C., Fleck, O., and Kohli, J. (1998) *Schizosaccharomyces pombe exo1* is involved in the same mismatch repair pathway as *msh2* and *pms1*. *Current Genetics* **34**, 343-350
- Ruepp, S., Furger, A., Kurath, U., Renggli, C. K., Hemphill, A., Brun, R., and Roditi, I. (1997) Survival of *Trypanosoma brucei* in the tsetse fly is enhanced by the expression of specific forms of procyclin. *Journal of Cell Biology* **137**, 1369-1379
- Salmon, D., Geuskens, M., Hanocq, F., Hanocq-Quertier, J., Nolan, D., Ruben, L., and Pays, E. (1994) A novel heterodimeric transferrin receptor encoded by a pair of VSG expression site-associated genes in *T. brucei*. *Cell* **78**, 75-86
- Sambrook, J and Russell, D. N. (2001) *Molecular cloning: a laboratory manual* (3rd Edition). Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
- 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
- Sargeant, P. G., Jackson, T. F., Wiffen, S. R., and Bhojnani, R. (1988) Biological evidence of genetic exchange in *Entamoeba histolytica*. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **82**, 862-867
- Sasse, J. (1998) Ph.D. Thesis. University of Cambridge. The development of genetic markers for the *Trypanosoma brucei* genome.
- Saxe, D., Datta, A., and Jinks-Robertson, S. (2000) Stimulation of mitotic recombination events by high levels of RNA polymerase II transcription in yeast. *Molecular and Cellular Biology* **20**, 5404-5414
- Schär, P., Baur, M., Schneider, C., and Kohli, J. (1997) Mismatch repair in *Schizosaccharomyces pombe* requires the *mutL* homologous gene *pms1*: molecular cloning and functional analysis. *Genetics* **146**, 1275-1286

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 Research* **58**, 4537-4542

Schmutte, C., Sadoff, M. M., Shim, K.-S., Acharya, S., and Fishel, R. (2001) The interaction of DNA mismatch repair proteins with human exonuclease I. *Journal of Biological Chemistry* **276**, 33011-33018

Schofield, M. J., Brownwell, F. E., Nayak, S., Du, C. W., Kool, E. T., and Hsieh, P. (2001) The Phe-X-Glu DNA binding motif of MutS - The role of hydrogen bonding in mismatch recognition. *Journal of Biological Chemistry* **276**, 45505-45508

Sekelsky, J. J., Brodsky, M. H., and Burtis, K. C. (2000) DNA repair in *Drosophila*: insights from the *Drosophila* genome sequence. *Journal of Cell Biology* **150**, F31-F36

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

Shapiro, S. Z., Naessens, J., Liesegang, B., Moloo, S. K., and Magondi, J. (1984) Analysis by flow cytometry of DNA synthesis during the life cycle of African trypanosomes. *Acta Tropica* **41**, 313-323

Shea, C., Glass, D. J., Parangi, S., and Van der Ploeg, L. H. T. (1986) Variant surface glycoprotein gene expression site switches in *Trypanosoma brucei*. *Journal of Biological Chemistry* **261**, 6056-6063

Shen, P. and Huang, H. (1989) Effect of base pair mismatches on recombination via the RecBCD pathway. *Molecular & General Genetics* **18**, 358-360

Shi, H. F., Djikeng, A., Mark, T., Wirtz, E., Tschudi, C., and Ullu, E. (2000) Genetic interference in *Trypanosoma brucei* by heritable and inducible double-stranded RNA. *RNA* **6**, 1069-1076

Shibahara, K. and Stillman, B. (1999) Replication-dependent marking of DNA by PCNA facilitates CAF-I-coupled inheritance of chromatin. *Cell* **96**, 575-585

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

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

Sia, E. A., Dominska, M., Stefanovic, L., and Petes, T. D. (2001) Isolation and characterization of point mutations in mismatch repair genes that destabilize microsatellites in yeast. *Molecular and Cellular Biology* **21**, 8157-8167

Sia, E. A., Jinks-Robertson, S., and Petes, T. D. (1997a) Genetic control of microsatellite stability. *Mutation Research - DNA Repair* **383**, 61-70

- Sia, E. A., Kokoska, R. J., Dominska, M., Greenwell, P., and Petes, T. D. (1997b) Microsatellite instability in yeast: dependence on repeat unit size and DNA mismatch repair genes. *Molecular and Cellular Biology* **17**, 2851-2858
- Smith, A. B., Esko, J. D., and Hajduk, S. L. (1995) Killing of trypanosomes by the human haptoglobin-related protein. *Science* **268**, 284-286
- Smith, A. B. and Hajduk, S. L. (1995) Identification of haptoglobin as a natural inhibitor of trypanocidal activity in human serum. *Proceedings of the National Academy of Sciences USA* **92**, 10262-10266
- Sogin, M. L. (1989) Evolution of eukaryotic microorganisms and their small subunit ribosomal RNAs. *American Zoologist* **29**, 487-499
- Sogin, M. L., Elwood, H. J., and Gunderson, J. H. (1986) Evolutionary diversity of eukaryotic small-subunit rRNA genes. *Proceedings of the National Academy of Sciences USA* **83**, 1383-1387
- Spampinato, C. and Modrich, P. (2000) The MutL ATPase is required for mismatch repair. *Journal of Biological Chemistry* **275**, 9863-9869
- Štambuk, S. and Radman, M. (1998) Mechanism and control of interspecies recombination in *Escherichia coli*. I. Mismatch repair, methylation, recombination and replication functions. *Genetics* **150**, 533-542
- Steverding, D., Stierhof, Y. D., Chaudhri, M., Ligtenberg, M., Schell, D., Beck-Sickinger, A. G., and Overath, P. (1994) ESAG 6 and 7 products of *Trypanosoma brucei* form a transferrin binding protein complex. *European Journal of Cell Biology* **64**, 78-87
- Strand, M., Earley, M. C., Crouse, G. F., and Petes, T. D. (1995) Mutations in the *MSH3* gene preferentially lead to deletions within tracts of simple repetitive DNA in *Saccharomyces cerevisiae*. *Proceedings of the National Academy of Sciences USA* **92**, 10418-10421
- Strand, M., Prolla, T. A., Liskay, R. M., and Petes, T. D. (1993) Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. *Nature* **365**, 274-276
- Strauss, B. S. (1999) Frameshift mutation, microsatellites and mismatch repair. *Mutation Research - Reviews in Mutation Research* **437**, 195-203
- Stringer, J. R. and Keely, S. P. (2001) Genetics of surface antigen expression in *Pneumocystis carinii*. *Infection and Immunity* **69**, 627-639
- Studamire, B., Price, G., Sugawara, N., Habcr, J. E., and Alani, E. (1999) Separation-of-function mutations in *Saccharomyces cerevisiae MSH2* that confer mismatch repair defects but do not affect nonhomologous-tail removal during recombination. *Molecular and Cellular Biology* **19**, 7558-7567

- Studamire, B., Quach, T., and Alani, E. (1998) *Saccharomyces cerevisiae* Msh2p and Msh6p ATPase activities are both required during mismatch repair. *Molecular and Cellular Biology* **18**, 7590-7601
- Sugawara, N. and Haber, J. E. (1992) Characterization of double-strand break-induced recombination: homology requirements and single-stranded DNA formation. *Molecular and Cellular Biology* **12**, 563-575
- 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. *Molecular and Cellular Biology* **20**, 5300-5309
- Sugawara, N., Pâques, F., Colaiácovo, M., and Haber, J. E. (1997) Role of *Saccharomyces cerevisiae* Msh2 and Msh3 repair proteins in double-strand break-induced recombination. *Proceedings of the National Academy of Sciences USA* **94**, 9214-9219
- Sung, P. (1994) Catalysis of ATP-dependent homologous DNA pairing and strand exchange by yeast RAD51 protein. *Science* **265**, 1241-1243
- Sung, P. (1997) Yeast Rad55 and Rad57 proteins form a heterodimer that functions with replication protein A to promote DNA strand exchange by Rad51 recombinase. *Genes & Development* **11**, 1111-1121
- Svärd, S. G., Meng, T.-C., Hetsko, M. L., McCaffery, J. M., and Gillin, F. D. (1998) Differentiation-associated surface antigen variation in the ancient eukaryote *Giardia lamblia*. *Molecular Microbiology* **30**, 979-989
- Szostak, J. W., Orr-Weaver, T. L., Rothstein, R. J., and Stahl, F. W. (1983) The double-strand-break repair model for recombination. *Cell* **33**, 25-35
- Tait, A., Masiga, D., Ouma, J., MacLeod, A., Sasse, J., Melville, S., Lindegard, G., McIntosh, A., and Turner, C. M. R. (2002) Genetic analysis of phenotype in *Trypanosoma brucei*: a classical approach to potentially complex traits. *Philosophical Transactions of the Royal Society of London: Series B* **357**, 89-99
- Ten Asbroek, A. L. M. A., Mol, C. A. A. M., Kieft, R., and Borst, P. (1993) Stable transformation of *Trypanosoma brucei*. *Molecular and Biochemical Parasitology* **59**, 133-142
- Ten Asbroek, A. L. M. A., 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
- Thi, C. D. D., Aerts, D., Steinert, M., and Pays, E. (1991) High homology between variant surface glycoprotein gene expression sites of *Trypanosoma brucei* and *Trypanosoma gambiense*. *Molecular and Biochemical Parasitology* **48**, 199-210
- Thomas, D. C., Roberts, J. D., and Kunkel, T. A. (1991) Heteroduplex repair in extracts of human HeLa cells. *Journal of Biological Chemistry* **266**, 3744-3751

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 Research* **25**, 4876-4882

Thon, G., Baltz, T., Giroud, C., and Eisen, H. (1990) Trypanosome variable surface glycoproteins: composite genes and order of expression. *Genes & Development* **9**, 1374-1383

Timmers, H. T. M., De Lange, T., Kooter, J. M., and Borst, P. (1987) Coincident multiple activations of the same surface antigen gene in *Trypanosoma brucei*. *Journal of Molecular Biology* **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. *Proceedings of the National Academy of Sciences USA* **94**, 7487-7492

Toft, N. J., Winton, D. J., Kelly, J., Howard, L. A., Dekker, M., te Riele, H., Arends, M. J., Wyllie, A. H., Margison, G. P., and Clarke, A. R. (1999) *Msh2* status modulates both apoptosis and mutation frequency in the murine small intestine. *Proceedings of the National Academy of Sciences USA* **96**, 3911-3915

Tran, H. T., Gordenin, D. A., and Resnick, M. A. (1999) The 3'→5' exonucleases of DNA polymerases  $\delta$  and  $\epsilon$  and the 5'→3' exonuclease *Exo1* have major roles in postreplication mutation avoidance in *Saccharomyces cerevisiae*. *Molecular and Cellular Biology* **19**, 2000-2007

Tran, P. T. and Liskay, R. M. (2000) Functional studies on the candidate ATPase domains of *Saccharomyces cerevisiae* MutL $\alpha$ . *Molecular and Cellular Biology* **20**, 6390-6398

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

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

Turner, C. M. R., 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**, 269-276

Twerdi, C. D., Boyer, J. C., and Farber, R. A. (1999) Relative rates of insertion and deletion mutations in a microsatellite sequence in cultured cells. *Proceedings of the National Academy of Sciences USA* **96**, 2875-2879

Ullu, E., Tschudi, C., and Günzl, A. (1996) *Trans*-splicing in trypanosomatid protozoa. In: *Molecular biology of parasitic protozoa*, (eds. Smith, D. F. and Parsons, M.), pp. 115-133. Oxford University Press, Oxford.

- 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
- Vaidya, T., Bakhiet, M., Hill, K. L., Olsson, T., Kristensson, K., and Donelson, J. E. (1997) The gene for a T lymphocyte triggering factor from African trypanosomes. *Journal of Experimental Medicine* **186**, 433-438
- Vaisman, A., Varchenko, M., Umar, A., Kunkel, T. A., Risinger, J. I., Barrett, J. C., Hamilton, T. C., and Chaney, S. G. (1998) The role of hMLH1, hMSH3, and hMSH6 defects in cisplatin and oxaliplatin resistance: correlation with replicative bypass of platinum-DNA adducts. *Cancer Research* **58**, 3579-3585
- Van Den Abbeele, J., Claes, Y., van Bockstaele, D., Le Ray, D., and Coosemans, M. (1999) *Trypanosoma brucei* spp. development in the tsetse fly: characterization of the post-mesocyclic stages in the foregut and proboscis. *Parasitology* **118**, 469-478
- Van der Ploeg, L. H. T., Lui, A. Y. C., and Borst, P. (1984) Structure of the growing telomeres of trypanosomes. *Cell* **36**, 459-468
- Van der Ploeg, L. H. T., 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 Research* **10**, 5905-5923
- van Gent, D. C., Hoeijmakers, J. H. J., and Kanaar, R. (2001) Chromosomal stability and the DNA double-stranded break connection. *Nature Reviews Genetics* **2**, 196-206
- Vanhamme, L., Lecordier, L., and Pays, E. (2001) Control and function of the bloodstream variant surface glycoprotein expression sites in *Trypanosoma brucei*. *International Journal for Parasitology* **31**, 523-531
- 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. *Journal of Cell Science* **110**, 2661-2671
- Vassella, E., Van Den Abbeele, J., Bütikofer, P., Renggli, C. K., Furger, A., Brun, R., and Roditi, I. (2000) A major surface glycoprotein of *Trypanosoma brucei* is expressed transiently during development and can be regulated post-transcriptionally by glycerol or hypoxia. *Genes & Development* **14**, 615-626
- Vickerman, K. (1985) Developmental cycles and biology of pathogenic trypanosomes. *British Medical Bulletin* **41**, 105-114
- Vickerman, K., Tctley, L., Hendry, K. A. K., and Turner, C. M. R. (1988) Biology of the African trypanosomes in the tsetse fly. *Biology of the Cell* **64**, 109-119
- Viel, A., Fornasarig, M., Novella, E., Genuardi, M., Capozzi, E., Pedroni, M., Santarosa, M., De Leon, M. P., Della Puppa, L., Anti, M., and Boiocchi, M. (1998) Lack of *PMS2* gene-truncating mutations in patients with hereditary colorectal cancer. *International Journal of Oncology* **13**, 565-569

- Vivarès, C. P. and Méténier, G. (2001) The microsporidian *Encephalitozoon*. *Bioessays* **23**, 194-202
- Vulic, M., Lenski, R. E., and Radman, M. (1999) Mutation, recombination, and incipient speciation of bacteria in the laboratory. *Proceedings of the National Academy of Sciences USA* **96**, 7348-7351
- Wahlgren, M., Fernandez, V., Chen, Q., Svård, S., and Hagblom, P. (1999) Waves of malarial *var*-iations. *Cell* **96**, 603-606
- Walker, J. E., Saraste, M., Runswick, M. J., and Gay, N. J. (1982) Distantly related sequences in the  $\alpha$ - and  $\beta$ -subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO Journal* **1**, 945-951
- Walliker, D., Quakyi, I. A., Wellems, T. E., McCutchan, T. E., Szarfman, A., Corcoran, L. M., Burkot, T. E., and Carter, R. (1987) Genetic analysis of the human malaria parasite *Plasmodium falciparum*. *Science* **236**, 1661-1666
- 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. *Proceedings of the National Academy of Sciences USA* **96**, 13914-13919
- Wang, Z. F., Morris, J. C., Drew, M. E., and Englund, P. T. (2000) Inhibition of *Trypanosoma brucei* gene expression by RNA interference using an integratable vector with opposing T7 promoters. *Journal of Biological Chemistry* **275**, 40174-40179
- Warbrick, E. (2000) The puzzle of PCNA's many partners. *Bioessays* **22**, 997-1006
- Weber, J. L. and Wong, C. (1993) Mutation of human short tandem repeats. *Human Molecular Genetics* **2**, 1123-1128
- 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
- White, M. F. and Lilley, D. M. (1997) Characterization of a Holliday junction-resolving enzyme from *Schizosaccharomyces pombe*. *Molecular and Cellular Biology* **17**, 6465-6471
- Wierdl, M., Dominska, M., and Petes, T. D. (1997) Microsatellite instability in yeast: dependence on the length of the microsatellite. *Genetics* **146**, 769-779
- Wirtz, E., Leal, S., Ochatt, C., and Cross, G. A. M. (1999) A tightly regulated inducible expression system for conditional gene knock-outs and dominant-negative genetics in *Trypanosoma brucei*. *Molecular and Biochemical Parasitology* **99**, 89-101
- Worth, I., Bader, T., Yang, J., and Clark, S. (1998) Role of MutS ATPase activity in MutS,L-dependent block of *in vitro* strand transfer. *Journal of Biological Chemistry* **273**, 23176-23182

- Worth, L., Clark, S., Radman, M., and Modrich, P. (1994) Mismatch repair proteins MutS and MutL inhibit RecA-catalyzed strand transfer between diverged DNAs. *Proceedings of the National Academy of Sciences USA* **91**, 3238-3241
- Wu, T. H. and Marinus, M. G. (1994) Dominant negative mutator mutations in the *mutS* gene of *Escherichia coli*. *Journal of Bacteriology* **176**, 5393-5400
- Wyatt, R., Kwong, P. D., Desjardins, E., Sweet, R. W., Robinson, J., Hendrickson, W. A., and Sodroski, J. G. (1998) The antigenic structure of the HIV gp120 envelope glycoprotein. *Nature* **393**, 705-711
- Xiao, W., Rathgeber, L., Fontanie, T., and Bawa, S. (1995) DNA mismatch repair mutants do not increase *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine tolerance in O<sup>6</sup>-methylguanine DNA methyltransferase-deficient yeast cells. *Carcinogenesis* **16**, 1933-1939
- Xie, Y. L., Counter, C., and Alani, E. (1999) Characterization of the repeat-tract instability and mutator phenotypes conferred by a *Tn3* insertion in *RFC1*, the large subunit of the yeast clamp loader. *Genetics* **151**, 499-509
- 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
- Yamada, N. A., Smith, G. A., Castro, A., Roques, C. N., Boyer, J. C., and Farber, R. A. (2002) Relative rates of insertion and deletion mutations in dinucleotide repeats of various lengths in mismatch repair proficient mouse and mismatch repair deficient human cells. *Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis* **499**, 213-225
- Yamaguchi-Iwai, Y., Sonoda, E., Sasaki, M. S., Morrison, C., Haraguchi, T., Hiraoka, Y., Yamashita, Y. M., Yagi, T., Takata, M., Price, C., Kakazu, N., and Takeda, S. (1999) Mre11 is essential for the maintenance of chromosomal DNA in vertebrate cells. *EMBO Journal* **18**, 6619-6629
- Zaccolo, M., Williams, D. M., Brown, D. M., and Gherardi, E. (1996) An approach to random mutagenesis of DNA using mixtures of triphosphate derivatives of nucleoside analogues. *Journal of Molecular Biology* **255**, 589-603
- 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*. *Proceedings of the National Academy of Sciences USA* **94**, 9786-9791
- Zhang, J. R., Hardham, J. M., Barbour, A. G., and Norris, S. J. (1997) Antigenic variation in Lyme disease borreliae by promiscuous recombination of VMP-like sequence cassettes. *Cell* **89**, 275-285

