
ASPECTS OF GENE EXPRESSION AND REGULATION IN
PLASMODIUM FALCIPARUM GAMETOCYTOGENESIS

Daniel R Meyersfeld
Department of Molecular Medicine and Haematology, University of the
Witwatersrand, Johannesburg, South Africa

Declaration

I hereby declare that this thesis contains my own, unaided work. It is being submitted for the degree of Doctor of Philosophy in the University of the Witwatersrand, Johannesburg, South Africa. It has not been submitted before for any degree or examination at any other university.

For culturing malaria parasites in the blood of human volunteers, ethical clearance was obtained from the University of the Witwatersrand (clearance number M03-11-06).

D.R Meyersfeld

_____ day of _____ 2005

Abstract

Malaria is one of the most debilitating pathogenic infections known to man, responsible for approximately three million deaths annually, primarily children in sub-Saharan Africa. The parasite has evaded multiple attempts at eradication, predominantly through the complexity of its life cycle, the ability to elude host immune response, and gametocyte formation to ensure dissemination. The recent completion of the genome sequence has opened up a multitude of avenues for exploration and identification of novel drug and vaccine targets, as well as providing a glimpse into the complex mechanisms that have contributed to the success of this pathogen. The mechanisms of gene regulation, especially those governing gametocytogenesis, have, however, not yet been elucidated.

In this research, differential display has been used to identify some of the genes that are differentially expressed between the asexual parasite and gametocyte stages of *P. falciparum*. Numerous genes involved in diverse aspects of metabolism, protein synthesis and immune evasion were identified. A combination of BLASTN and BLASTX similarity searches was used to categorize and increase the confidence with which a transcript could be identified. Expression data for confidently identified genes were confirmed using reverse slot blot and available microarray data.

PfMyb2, a novel transcription factor which may regulate genes involved in gametocytogenesis, was characterized. The DNA binding domains of the protein were cloned and expressed as a histidine fusion protein. Mobility shift assays were used to assess the *in vitro* binding capability of the recombinant 6xHis-*PfMyb2*, which bound to oligonucleotides containing the consensus Myb regulatory element. Two of the oligonucleotides represent sequences located within promoters of *P. falciparum* genes (*Pferk1* and *Pfmap1*) known to play a role in regulating the cell cycle, a function ascribed to many members of the vertebrate Myb family. The identification of *PfMyb2* as a *bona fide* transcription factor is a first step into gaining some insight into the many regulatory processes that occur during the life cycle of this complex organism. A better understanding of the molecular mechanisms that govern its survival is essential for the ultimate eradication of this deadly parasite.

Publications arising from this work**Peer-reviewed Journals**

- A Reverse Slot Blot Technique for the Verification of Genes Identified through Differential Display

Daniel R. Meyersfeld and Thérèse L. Coetzer

Biotechniques. Feb; **34**(2):270-2 (2003)

- Integrated appraisal of *Plasmodium falciparum* differentially expressed transcripts

Daniel R. Meyersfeld, Raphael D. Isokpehi, and Thérèse L. Coetzer

Manuscript submitted to Molecular Biotechnology

- Characterisation of PfMyb2, a *Plasmodium falciparum* DNA binding protein

Daniel R. Meyersfeld and Thérèse L. Coetzer

Manuscript in preparation

Conference presentations:**Poster Presentations:**

- Identification of Stage Specific *Plasmodium falciparum* gene expression using Differential Display

IUBMB/SASBMB Special Meeting on the Biochemical and Molecular Basis of Disease, Cape Town, 19-23 November, 2001

- Characterisation of PfMyb2, a *Plasmodium falciparum* DNA binding protein

Molecular Parasitology Meeting, Woods Hole, Massachusetts, 11-15 September, 2005

Oral Presentations:

- Identification of Stage Specific *Plasmodium falciparum* gene expression using Differential Display

University of the Witwatersrand Health Sciences Faculty Research Day, August 2002

- Identification of stage specific *Plasmodium falciparum* gene expression with a view to understanding the regulatory mechanisms involved in sexual development

University of the Witwatersrand Health Sciences Faculty Research Day, August 2004

- Identification of Stage Specific *Plasmodium falciparum* gene expression using Differential Display

Molecular and Cell Biology Group (MCBG) Symposium, October 2004

To Romy

Acknowledgements

I owe a great debt of gratitude to my supervisor, Professor Thérèse L. Coetzer, for her unstinting support, enthusiasm and expertise. This would not have been possible without her guidance and eternal optimism.

My heartfelt thanks also go to the following:

- Sonja, Marcel, Roberto, Kuben and Pierre: you were a constant source of companionship during the long hours and late nights in the laboratory. It has been a pleasure working with all of you
- My family, for their continued love and support especially during the course of the last year
- My friends, who understood my absence far better than they understood my thesis
- Dr. Raphael Isokpehi for advice on bioinformatic analysis of differential display transcripts
- Dr. Catherine Vaquero and Mathieu Gissot for providing the transcription factor primer sequences, as well as for hosting me in their laboratory
- The Department of Pharmacy and Pharmacology, University of the Witwatersrand, Johannesburg South Africa for the stocks of *P. falciparum* strain 3D7

The following funding agencies:

- National Research Foundation
 - National Health Laboratory Services
 - Stella and Paul Loewenstein Charitable and Educational Trust
 - Medical Faculty Research Endowment Fund
 - University of the Witwatersrand
 - The Mellor Travel Fellowship
-

Table of Contents

Declaration	ii
Abstract	iii
Publications arising from this work	iv
Acknowledgements	vii
Table of Contents	v
List of Figures	xiv
List of Tables	xv
List of Abbreviations	xvi

CHAPTER 1- INTRODUCTION..... 1

1.1 The burden of malaria	1
1.2 <i>P. falciparum</i> life cycle	2
1.3 Sexual development in <i>P. falciparum</i>	4
1.3.1 Gametocyte morphology	4
1.3.2 Sexual stage-specific gene expression	6
1.3.3 Environmental factors influencing sexual development	9
1.3.4 Genetic factors regulating gametocytogenesis	12
1.3.5 Signalling pathways implicated in gametocytogenesis	13
1.4 <i>P. falciparum</i> gene regulation	15
1.4.1 Epigenetic gene regulation in <i>P. falciparum</i>	17
1.4.2 Transcriptional gene regulation	18
1.4.3 Post-transcriptional gene regulation	18
1.4.4 Translational mechanisms of gene regulation	19
1.5 Aims and Objectives	21

**CHAPTER 2- IDENTIFICATION OF *P. FALCIPARUM* STAGE
SPECIFIC GENE EXPRESSION 22**

2.1 Introduction	22
2.1.1 Differential display	23
2.1.2 Microarrays	26
2.1.3 Serial Analysis of Gene Expression	29
2.1.4 Suppression Subtractive Hybridisation	30
2.2 Materials and Methods	32
2.2.1 <i>Plasmodium falciparum</i> culture methods	32
2.2.1.1 Preparation of parasite culture from frozen stock	32
2.2.1.2 Freezing of asexual parasite cultures	33
2.2.1.3 Preparation of erythrocytes for malaria cultures	33
2.2.1.4 Continuous culture method	33
2.2.1.5 Sorbitol synchronisation of cultures	34
2.2.1.6 Preparation of gametocytes	34
2.2.1.7 Isolation of gametocytes by Percoll gradient	35
2.2.2 RNA extraction	36
2.2.2.1 Total RNA isolation Kit	36
2.2.2.2 Guanidium Isothiocyanate method	37
2.2.2.3 TRI reagent	38
2.2.2.4 Removal of DNA contamination from RNA samples	38
2.2.2.5 Measurement of RNA yield and purity	38
2.2.3 Differential Display	39
2.2.3.1 Reverse Transcription Polymerase Chain Reaction	39
2.2.3.2 Electrophoresis of RT-PCR products	40
2.2.3.3 Identification of differentially expressed cDNAs	41
2.2.3.4 Excision and elution of differentially expressed fragments	41
2.2.3.5 Reamplification of eluted products	41
2.2.4 Subcloning of PCR products	42
2.2.4.1 Preparation of PCR products and vector for subcloning	42
2.2.4.2 Ligation of vector and DNA	43
2.2.4.3 Transformation of competent cells	43

2.2.4.4 Purification of plasmid DNA	44
2.2.5 DNA sequencing	44
2.2.6 Reverse slot blot verification of differentially expressed cDNAs	45
2.2.7 Bioinformatic Analysis of Differentially Expressed Transcripts	47
2.2.7.1 Transcript Identification	47
2.2.7.2 Transcript Expression Profile	48
2.3 Results	49
2.3.1 Plasmodium falciparum cultures	49
2.3.1.1 Asexual parasites	49
2.3.1.2 Gametocytes	49
2.3.2 RNA Extractions	52
2.3.3 Differential Display	53
2.3.3.1 Identification and reamplification of differentially expressed cDNAs	53
2.3.3.2 Subcloning and sequencing of reamplified fragments	55
2.3.3.3 Identification and analysis of differentially expressed transcripts	56
2.3.3.4 Verification of differential display results	62
2.4 Discussion	69
2.4.1 Identification of differentially expressed genes	69
2.4.2 Analysis of differentially expressed gene transcripts	70
2.4.2.1 Differentially expressed genes encoding surface proteins	70
2.4.2.2 Differentially expressed genes involved in cell signalling/metabolism	71
2.4.2.3 Differentially expressed genes involved in RNA metabolism	72
2.4.2.4 Differentially expressed actin genes	74
2.4.2.5 Differentially expressed genes annotated as hypothetical	74
2.5 Conclusion	77
 CHAPTER 3- CHARACTERISATION OF PFMYB2	 78
 3.1 Introduction	 78
3.1.1 Common eukaryotic regulatory elements in <i>P. falciparum</i>	79

3.1.1.1 The TATA box	79
3.1.1.2 OCT1 transcriptional domains	80
3.1.1.3 The GC-rich region	80
3.1.1.4 The CAAT box	81
3.1.2 Regulatory sequences unique to <i>P. falciparum</i>	82
3.1.3 <i>P. falciparum</i> transcription factors	86
3.1.3.1 Basal transcription factors in <i>P. falciparum</i>	86
3.1.3.2 Specific transcription factors in <i>P. falciparum</i>	87
3.1.3.3 The cMyb family	88
3.2 Materials and Methods	90
3.2.1 Cloning of <i>PfMyb2</i>	90
3.2.1.1 Genomic DNA isolation from <i>P. falciparum</i>	90
3.2.1.2 Preparation of <i>PfMyb2</i> DNA for cloning	90
3.2.1.3 Preparation of pET-15b and pGEX-4T-2 expression vectors	91
3.2.1.4 Ligation of insert and vector	92
3.2.1.5 Confirmation of the presence of inserts by colony PCR	92
3.2.1.6 Confirmation of the presence of inserts by restriction enzyme analysis	92
3.2.1.7 Purification of plasmid DNA	93
3.2.1.8 Verification of insert sequence and orientation	93
3.2.1.9 Transformation of BL21-CodonPlus competent cells	93
3.2.2 Expression of 6x His- <i>PfMyb2</i> protein	94
3.2.2.1 Induction of target protein using IPTG	94
3.2.2.2 Induction of target protein using the Overnight Express™ Autoinduction System	95
3.2.2.3 Extraction of soluble proteins from BL21 cells	95
3.2.2.4 Immunoblot to confirm the expression of 6xHis- <i>PfMyb2</i>	96
3.2.2.5 Purification of 6xHis- <i>PfMyb2</i>	96
3.2.2.6 Purification of 6xHis- <i>PfMyb2</i> under denaturing conditions	96
3.2.2.7 Dialysis of denatured and purified 6xHis- <i>PfMyb2</i>	97
3.2.3 <i>In vitro</i> translation of <i>PfMyb2</i>	98

3.2.3.1 PCR primer design for <i>in vitro</i> translation	98
3.2.3.2 <i>In vitro</i> translation procedure	98
3.2.4 Electrophoretic mobility shift assays (EMSA)	99
3.2.4.1 Binding reactions using an EMSA kit	99
3.2.4.2 End-labelling of oligonucleotides	100
3.2.4.2 Binding reaction and electrophoresis	100
3.3 Results	102
3.3.1 Cloning <i>PfMyb2</i> into expression vectors	102
3.3.1.1 Amplification of <i>PfMyb2</i>	102
3.3.1.2 Sequence verification of <i>PfMyb2</i> insert	104
3.3.2 Expression and purification of 6xHis- <i>PfMyb2</i> protein	105
3.3.2.1 Purification of 6xHis- <i>PfMyb2</i> under native conditions	107
3.3.2.2 Purification of 6xHis- <i>PfMyb2</i> under denaturing conditions	108
3.3.2.3 <i>In vitro</i> translation of 6xHis- <i>PfMyb2</i>	110
3.3.2.4 Expression of 6xHis- <i>PfMyb2</i> using Overnight Express™	111
3.3.3 Binding studies with 6xHis- <i>PfMyb2</i>	112
3.3.3.1 6xHis- <i>PfMyb2</i> binds <i>in vitro</i> to Myb regulatory elements	113
3.3.4 Bioinformatic analysis of <i>PfMyb2</i>	117
3.4 Discussion	124
3.4.1 Recombinant <i>P. falciparum</i> protein expression	124
3.4.2 Functional analysis of 6xHis- <i>PfMyb2</i>	126
3.4.3 Structural analysis of the Myb protein family	128
3.4.4 Role of <i>PfMyb2</i> in parasite development	130
3.5 Conclusion	132
3.6 Concluding remarks	133
 APPENDIX	 135
 A-1 Reagents	 135
A-1.1 Parasite culture media	135

A-1.2 DNA analysis	136
A-1.3 Recombinant protein expression	136
A-1.4 Electrophoretic mobility shift assays (EMSAs)	138
A-2 Standard Laboratory Procedures	139
A-3 Plasmid maps and primer sequences	142
A-3.1 Primers for differential display	142
A-3.2 Nucleotide sequence of PF10_0327	143
A-3.3 pET-15b vector construct and cloning cassette	144
A-3.4 pGEX-4T-2 vector construct and cloning cassette	145
A-3.5 Primer design for <i>in vitro</i> translation	146
A3-6 Oligonucleotide sequences for EMSA assays	147
A-4 Differential Display Transcript Sequences	148
A-5 List of Suppliers of Chemicals and Equipment	152
REFERENCES.....	157

List of Figures

Figure 1- Life cycle of <i>Plasmodium falciparum</i>	3
Figure 2- Gametocyte morphology	5
Figure 3- Asexual parasites in culture	49
Figure 4- Gametocytes in culture	50
Figure 5- Percoll gradient purification of gametocytes	51
Figure 6- Purified gametocytes after Percoll gradient centrifugation	51
Figure 7- RNA analysis	53
Figure 8- Differential display analysis of <i>P. falciparum</i> gene expression	54
Figure 9- Reamplification of differential display fragments	55
Figure 10- Verification of plasmid inserts after transformation of DH5 α cells	56
Figure 11- Reverse slot blot for the verification of differential display data	63
Figure 12- Amino acid sequence of <i>PfMyb2</i>	102
Figure 13- Amplification of <i>PfMyb2</i> for cloning	103
Figure 14- Verification of the presence of <i>PfMyb2</i> inserts	103
Figure 15- Restriction enzyme analysis of vector constructs	104
Figure 16- Partial DNA sequence of 6xHis- <i>PfMyb2</i> in pGEX-4T-2	105
Figure 17- Expression of 6xHis- <i>PfMyb2</i>	106
Figure 18- Immunoblot to confirm the expression of 6xHis- <i>PfMyb2</i>	106
Figure 19- Assay to assess the solubility of 6xHis- <i>PfMyb2</i>	107
Figure 20- Effect of glycyl-glycine on the solubility of 6xHis- <i>PfMyb2</i>	108
Figure 21- Purification of 6xHis- <i>PfMyb2</i> under denaturing conditions	109
Figure 22- Precipitation of 6xHis- <i>PfMyb2</i> during dialysis	109
Figure 23- Primer for the <i>in vitro</i> translation of <i>PfMyb2</i>	110
Figure 24- SDS-PAGE analysis of <i>in vitro</i> translation products	111
Figure 25- Purification of 6xHis- <i>PfMyb2</i> from the Overnight Express TM system	112
Figure 26- Oligonucleotides used for EMSA analysis	113
Figure 27- 6xHis- <i>PfMyb2</i> binds to <i>Pfcrk1</i> and <i>Pfmap1</i>	114

Figure 28- Autoradiograph indicating an interaction between 6xHis- <i>PfMyb2</i> and <i>Pfmap1</i>	114
Figure 29- 6xHis- <i>PfMyb2</i> binds to the consensus cMyb oligonucleotide	115
Figure 30- 6xHis- <i>PfMyb2</i> binds to the mim-1 oligonucleotide	116
Figure 31- 6xHis- <i>PfMyb2</i> does not bind to the NfKB oligonucleotide	116
Figure 32- mRNA Expression profile of <i>PfMyb2</i>	117
Figure 33-Clustal-W alignment comparing the binding domains of <i>PfMyb2</i> and <i>PfMyb1</i>	118
Figure 34- Helix-turn-helix structure of <i>PfMyb2</i> binding domains	119
Figure 35- An alignment between <i>PfMyb2</i> and human CDC5	120
Figure 36- Model of putative tertiary structure of <i>PfMyb2</i> binding domains and sequence comparison with homologous proteins	122
Figure 37 -Model of the interaction between Myb and its DNA binding domain	123
Figure 38- Three dimensional perspective of the structure of <i>PfMyb2</i>	123

List of Tables

Table 1-Transcripts with nucleotide homology to the PlasmoDB database	58
Table 2-Criteria for the classification of transcripts using sequence similarity search outputs	59
Table 3-Classification of transcripts identified by differential display	61
Table 4-Reverse slot blot verification of differential display data	64
Table 5-Microarray verification of differential display results	66
Table 6-Comparison of expression data from all three methods of analysis	67

List of Abbreviations

A	adenine
A _{260/280/600}	absorbance at 260nm, 280nm and 600nm respectively
ACD	acid citrate/dextrose
APS	ammonium persulphate
ATP	adenosine 5'-triphosphate
BLASTN	basic local alignment search tool (nucleotide)
BLASTX	basic local alignment search tool (protein)
bp	base pair
BSA	bovine serum albumin
C	cytosine
cDNA	complementary deoxyribonucleic acid
C/EBP	CAAT enhancer binding protein
Ci	Curie
Cip	calf intestinal phosphatase
cm	centimetre
cpm	counts per minute
°C	degrees Celsius
dATP	deoxyadenosine triphosphate
ddNTP	dideoxynucleoside triphosphate
DD-RT-PCR	differential display reverse transcription polymerase chain
DEPC	diethylpyrocarbonate
DNA	deoxyribonucleic acid
dNTP	deoxynucleoside triphosphate
DTT	dithiothreitol
E value	expectation value
EDTA	ethylenediamine tetraacetic acid
EMSA	electrophoretic mobility shift assay
EPO	erythropoietin
EST	expressed sequence tag

g	gram
G	guanine
<i>g</i>	relative centrifugal force
GPI	glycosylphosphatidylinositol
HPRT	hypoxanthine phosphoribosyl transferase
<i>Pfhsp</i>	<i>P.falciparum</i> heat shock protein
IM	incomplete culture medium
IPTG	isopropyl-1-thio-β-D-galactopyranoside
kb	kilobase
kDa	kilodalton
l	litre
LB	luria broth
M	molar
mA	milliamps
mg	milligram
ml	millilitre
mm	millimetre
mM	millimolar
MRE	myb regulatory element
mRNA	messenger ribonucleic acid
MSP-1	merozoite surface protein 1
μCi	microCurie
μg	microgram
μl	microlitre
μm	micrometer
nm	nanometer
ORF	open reading frame
p	pico (10 ⁻¹²)

p	primer
PBS	phosphate buffered saline
PCR	polymerase chain reaction
<i>Pf</i> EMPI	<i>Plasmodium falciparum</i> erythrocyte membrane protein I
<i>Pfgig</i>	<i>Plasmodium falciparum</i> gene implicated in gametocytogenesis
PMSF	phenylmethylsulfonylfluoride
PNK	polynucleotide kinase
Poly(A)	polyadenylate
PVM	parasitophorous vacuolar membrane
rRNA	ribosomal ribonucleic acid
RNA	ribonucleic acid
RNase	ribonuclease
RT-PCR	reverse transcription polymerase chain reaction
SAGE	serial analysis of gene expression
SAP	shrimp alkaline phosphatase
SDS	sodium dodecyl sulphate
SDS-PAGE	sodium dodecyl sulphate- polyacrylamide gel electrophoresis
snRNP	small nuclear ribonucleoprotein
SSC	standard saline citrate
SSH	suppression subtractive hybridisation
SSCP	single strand conformation polymorphism
SSU	small subunit
T	thymine
TBE	tris/borate/EDTA
TBP	TATA binding protein
TE	tris-EDTA
TEA	tris/EDTA/acetate
TEMED	N,N,N,N-tetramethylethylenediamine
TIGR	The Institute for Genomic Research
Tris	hydroxymethyl methylamine
U	units
USE	upstream sequence element

UV	ultraviolet
V	volt
v/v	volume/volume
W	watts
w/v	weight/volume

CHAPTER 1- INTRODUCTION

1.1 The burden of malaria

Malaria is a blood disease caused by the protozoan parasite of the genus *Plasmodium*, and transmitted by the *Anopheles* mosquito. In 2002, 2.2 billion people were exposed to the threat of *P. falciparum* malaria, resulting in approximately 515 million clinical cases (Snow et al., 2005). Of these, 70% were children and infants in sub-Saharan Africa, 25% occurred in South East Asia and the remaining 5% in South America.

The frequency of malaria has been increasing at an alarming rate (Wahlgren and Chen, 2002; Carucci, 2004) due to various factors. Among these is the increasing resistance of the mosquito vector to insecticides, resistance of the parasite to antimalarial drugs, a climate that is more conducive to the survival of the mosquito, mass movement of people into areas of high transmission, and the continued inability of developing countries to afford the necessities for disease control.

Humans are susceptible to infection from four species of *Plasmodium*: *P. falciparum*, *P. vivax*, *P. ovale* and *P. malariae*. Two characteristics that distinguish *P. falciparum* from the other malarias are its ability to invade erythrocytes of all ages causing extremely high parasitaemias, and the capacity to avoid splenic clearance by adherence to host capillary endothelium (Heddini, 2002). These features all contribute to making *P. falciparum* the most pathogenic of the four species of *Plasmodium*. Some of the difficulties inherent in combating this pathogen are highlighted when examining the complexities of the parasite life cycle.

1.2 *P. falciparum* life cycle

The life cycle of *P. falciparum* (Figure 1) alternates between a mosquito vector and a human host. Sporozoites are inoculated into the human with the bite of a female *Anopheles* mosquito (1), upon which they migrate immediately to the liver and invade hepatic cells (2). A dormant period ensues lasting approximately two weeks, during which the sporozoites mature into schizonts and cause rupture of the hepatocytes (3). A hepatocyte can release up to 30 000 merozoites, each capable of invading an erythrocyte. Thus begins a period of asexual cycling in the erythrocytes (4), where a single merozoite invading a cell is capable of producing as many as 36 merozoites in a mature multinucleated schizont. On average in *P. falciparum*, 16 new merozoites are released from an erythrocyte every 48 hours, each of which is able to invade a new cell. The resultant exponential increase in parasitaemia is largely responsible for the onset of clinical symptoms (Suh et al., 2004).

At a point during the multiple rounds of asexual erythrocytic cycling, a proportion of parasites are stimulated to differentiate into gametocytes (5), the sexual form essential for transmission. This process is one of the most significant events in ensuring the survival and dissemination of the parasite. Unlike the asexual erythrocytic stages, the gametocyte is able to survive the environment of the mosquito midgut, and hence its presence in the infected host when the mosquito takes a blood meal is of paramount importance to ensure the continued survival of the parasite. Male and female gametocytes are ingested by the mosquito during a blood meal (6), and the male exflagellates under the influence of xanthurenic acid to form eight microgametes each of which is capable of fertilizing the female macrogamete to form a zygote (7). The zygote, or ookinete, passes through the epithelial cells of the midgut and lodges in the basal lamina where it grows into an oocyst that develops into numerous sporozoites (8). Upon rupture of the oocyst, the sporozoites invade the salivary glands (9) to complete the cycle.

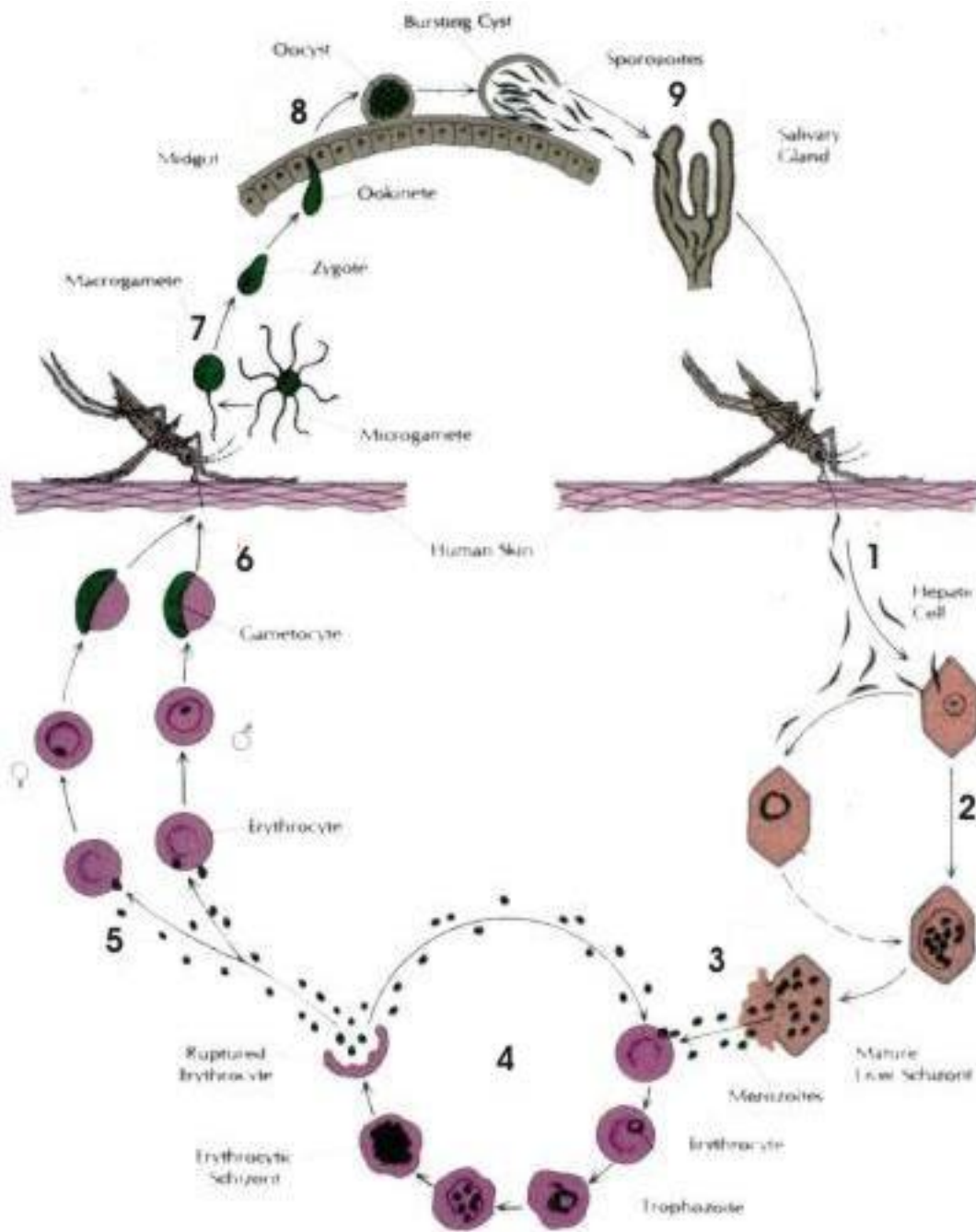


Figure 1- Life cycle of *Plasmodium falciparum*

A diagrammatic representation of the life cycle of *P. falciparum*, as it alternates between the *Anopheles* mosquito vector and human host. See text for details (<http://www.luc.edu/depts/biology/lifecyl.gif>).

1.3 Sexual development in *P. falciparum*

Upon rupture of an erythrocyte, each released merozoite has one of two developmental pathways it can follow. It can either embark on another round of schizogony, or it can convert to a single male or female gametocyte (Diebner et al., 2000). *Plasmodium* is one of the few Apicomplexa to exhibit asexual schizogony, and as such it is possible that this has evolved as a means of extending the period of transmissibility of the parasite, thereby increasing its chances of survival (Dyer and Day, 2000). An intricate balance must be maintained between the asexual and sexual cycles; prolonging the asexual cycle and thereby reducing the exposure of the antigenic gametocytes must be balanced with ensuring that gametocytes are available when a mosquito takes a blood meal (Dyer and Day, 2000).

Currently the mechanisms mediating the switch in the life cycle of the parasite have not been fully elucidated, but an understanding of sexual development is of prime importance in the development of transmission blocking vaccines and as a contribution to the general understanding of parasite biology.

1.3.1 Gametocyte morphology

Several highly distinctive morphological and genetic traits distinguish gametocytes from asexual parasites (Figure 2). The entire growth period of the gametocyte has been divided into five stages (I-V) spanning a period of 8-17 days following erythrocyte invasion of a sexually committed merozoite (Hawking et al., 1971). Stage I gametocytes appear within two days of red cell invasion, but are morphologically indistinguishable from asexual trophozoites. Stage II gametocytes are produced after two days of growth and begin to elongate within the erythrocyte through the formation of a subpellicular cytoskeleton. Cells at stage III can be distinguished into male and female forms by the varying arrangement of the cytoplasm, with the male cytoplasm having fewer ribosomes, mitochondria and

endoplasmic reticulum (ER) than the female. The erythrocyte is also becoming clearly distorted by this stage (2-8 days post-invasion). Sexually differentiated stage IV gametocytes are spindle-shaped with pointed poles, and have taken on a more symmetrical conformation corresponding to the complete enclosure of the parasite in a subpellicular cytoskeleton. There is a marked increase in the density of ribosomes, endoplasmic reticulum, Golgi vesicles and mitochondria in the female macrogametocyte relative to the male microgametocyte, which reflects the subsequent development of the macrogametocyte as the fertilized egg. Collapse of the pointed, spindle shape into a crescent shape with rounded extremities marks the final transformation of the stage IV parasite to the morphologically mature stage V gametocyte (Talman et al., 2004).

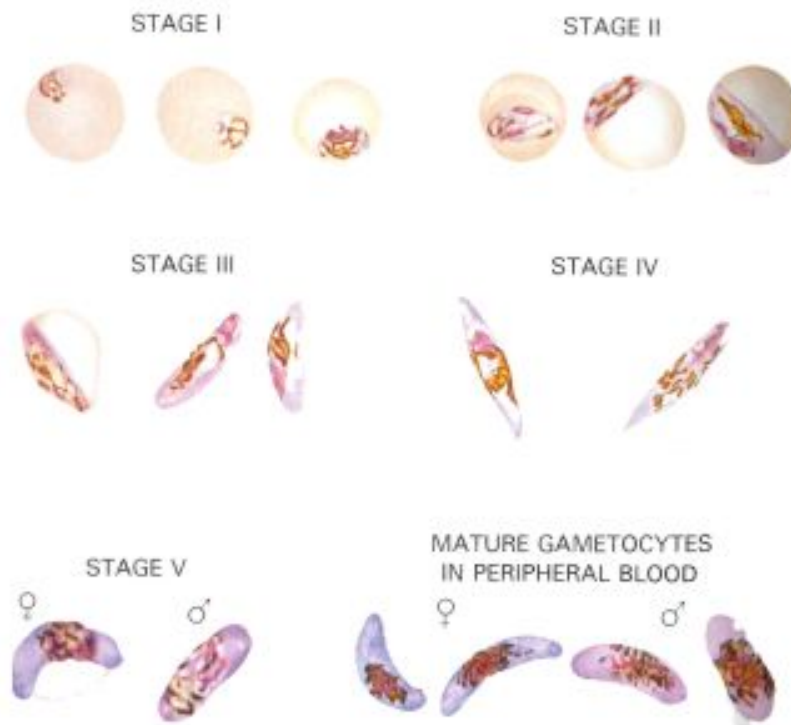


Figure 2- Gametocyte morphology

The five morphological stages of *P. falciparum* gametocytes are illustrated (Carter and Graves, 1988) See text for details.

1.3.2 Sexual stage-specific gene expression

Development of gametocytes is accompanied by a coordinated expression of sexual stage-specific genes, in addition to the many genes whose expression is up- or down-regulated (Janse and Waters, 2004). In their analysis of the genome using a high density microarray, Le Roch et al. (2003) identified 152 genes showing elevated levels of expression in gametocytes, and 218 genes expressed uniquely in sexual stages. Of these 370 genes, approximately 84% were identified only as hypothetical proteins, highlighting both the difficulties inherent in isolating sufficient quantities of gametocyte RNA for analysis, and the relative lack of research interest that has hitherto characterised this developmental stage.

Proteomic analysis of this stage has confirmed the existence of stage-specific gene expression, with numerous proteins found to be over-represented in the gametocyte (Florens et al., 2002). Some of these include proteins involved in cell cycle/DNA processing, which are required to respond immediately, upon uptake by the mosquito, to the stimuli to release the gametocyte from its arrested state in the G_0 stage of the cycle and initiate gametogenesis. Proteins involved in the mitochondrial tricarboxylic acid (TCA) cycle are also in evidence as this source of energy replaces the dependence on glycolysis seen in asexual parasites (Young et al., 2005). Proteins involved in protection against oxidative stress are upregulated in sexual stages to compensate for the elevated levels of oxidative agents associated with the erythrocytic stages (Lasonder et al., 2002). The number of proteins expressed uniquely in sexual stages, and the diversity of their functions, reflect the fundamental changes and adaptations that the parasite must make to ensure its survival in a new environment.

Because the expression of many of the sexual-stage specific genes, especially the surface antigens, is not essential for asexual proliferation, they provide a good target for gene disruption studies, and several have thus been quite extensively characterised.

Pfs16 and Pfg27/25

One of the earliest genes identified as being specific to sexual stages is *Pfs16*, which encodes a protein that localises to the parasitophorous vacuolar membrane (PVM) of the parasite. The synthesis of *Pfs16* protein is specific to gametocytes, but promoter activity is evident in ring stages and mRNA can be detected within 24 hours of erythrocyte invasion (Dechering et al., 1997). Transcriptional activity of *Pfs16* is therefore often used as a marker of sexually committed ring stages (Schneider et al., 2004). The early onset of expression and its abundance led to speculation that it was intricately involved in the formation of gametocytes. Knockout studies refuted this belief however, as *Pfs16* mutants were still capable of forming gametocytes with no apparent morphological changes. However, a significant reduction in the conversion rate to gametocytes was observed (Kongkasuriyachai et al., 2004) and male mutants were unable to exflagellate and were not infectious to mosquitoes. Thus it appears that *Pfs16*, although not essential for gametocyte growth, is required for optimal production of mature gametocytes.

In parasites committed to the sexual pathway, *Pfg27/25* is expressed approximately 30 hours after erythrocyte invasion. By the time these committed parasites have matured to stage I and II gametocytes, it represents approximately 5-10% of the cytoplasmic protein content of the cell (Lobo et al., 1999), coinciding with a time at which the transcription rate of this gene is beginning to decrease (Alano et al., 1996). Pologe (1994) ascribed defective gametocytogenesis to DNA rearrangements upstream of the *Pfg27/25* locus, but this was disputed by Alano et al. (1996). In an attempt to better understand the transcriptional regulation of the *Pfg27/25* gene, Alano et al (1996) studied the upstream genomic sequences of the gene. A highly conserved sequence in the 5' untranslated region of the gene was identified, consisting of five direct and one inverted repeat of a 90bp unit. Each unit contained a poly-dT tract, a 47bp conserved tract and a short d(AT) tract. Analysis of the upstream polymorphic region in gametocyteless strains showed that no correlation existed between the upstream regions and the ability to convert to gametocytes.

Based on the close proximity of the polymorphic region to the transcriptional control apparatus of the gene, Alano et al. (1996) proposed that the gene is under the transcriptional influence of the polymorphic region, which contributes to the irregular expression of this gene. Subsequent to this, complete disruption of the *Pfg27/25* locus through homologous recombination has resulted in a complete loss of the sexual phenotype (Lobo et al., 1999), again implicating this gene as a crucial element in gametocyte development. Despite not maturing to gametocytes, all transgenic mutants maintained expression of *Pfs16*, indicating that their commitment to sexual differentiation had not been affected. This shows that a clear genetic distinction exists between the commitment to sexual development and the physical development of this stage. This knockout further implicates a region upstream of the *Pfg27/25* locus in sexual development.

Pfs25

Whereas *Pfs16* is suitable as a marker of early stage gametocytes, identification of late stage gametocytes requires quantification of the *Pfs25* gene, whose mRNA is expressed in stage V gametocytes (Schneider et al., 2004) and in mosquito sexual stages (gametes/ookinetes) (Kongkasuriyachai and Kumar, 2002). *Pfs25* encodes a surface antigen containing epidermal growth factor-like domains, which may be involved in invasion of the midgut epithelium. Antibodies raised against *Pfs25* can block the development of oocysts in the midgut of the mosquito thereby significantly reducing transmission of the parasite; this makes it a leading transmission-blocking vaccine candidate (Barr et al., 1991; Kaslow et al., 1991).

Pfs230* and *Pfs48/45

A unique protein family containing several cysteine repeats has been identified in sexual stages of *Plasmodium*. *Pfs230* and *Pfs48/45* are two members of this family that have been well characterised and identified as potential transmission-blocking vaccine candidates (Moreira et al., 2004). Expression of both of these proteins begins in stage III gametocytes.

Pfs230 encodes a 360kDa protein that contains six tandemly repeated six-cysteine domains and is localised to the plasma membrane. Full length *Pfs230* (360kDa) is proteolytically cleaved to a 310kDa form as soon as the parasite emerges from the red cell in the mosquito midgut. It is postulated that this stage specific processing is an immune evasion strategy, where an N-terminal 50kDa span containing an immunodominant region is cleaved from the protein to attract the immune response, such that the 310kDa form is not vulnerable to antibodies upon release from the red cell (Williamson et al., 1996). The role of *Pfs230* in sexual stage parasites has yet to be determined, as mutants expressing truncated forms of this protein are still able to form gametocytes (Eksi et al., 2002).

Pfs48/45 encodes a 55kDa membrane protein containing two six-cysteine domains (Moreira et al., 2004). Knockout mutants lacking this gene were still able to form gametocytes that differentiated into gametes, but male gametes were unable to adhere to and penetrate female gametes. Fertilization and zygote formation were thus severely affected (van Dijk et al., 2001).

Exflagellation, and thus male gamete formation in the mosquito, is induced by xanthurenic acid (Billker et al., 1998) through a calcium-dependant pathway requiring calcium-dependant protein kinases (CDPK) (Billker et al., 2004). Recently knock-out of CDPK4 in *P. berghei* abrogated the formation of male gametes, and the existence of six to seven CPDKs in the *P. falciparum* genome has led to speculation that these could play a similarly important role in gamete formation in this organism (Billker et al., 2004).

1.3.3 Environmental factors influencing sexual development

The first to demonstrate a parasite sensitivity to the environment were Carter and Miller (1979) who showed that gametocyte conversion rate could be modulated by the addition of fresh erythrocytes to the culture. This was confirmed by Bruce et al.

(1990) who found that during rapid asexual growth, the proportion of schizonts forming gametocytes was very low immediately following the dilution of the culture with fresh erythrocytes. This proportion increased as the levels of parasitaemia rose and asexual growth slowed, but upon addition of fresh erythrocytes asexual growth was again favoured. Using a co-culture technique, where experimental and control cultures were separated by a semi-permeable membrane, Dyer and Day (2003) demonstrated that the conversion of asexual parasites to gametocytes occurred in a density dependant manner regulated through the action of diffusible molecules. Thus conversion to gametocytes is inhibited during conditions that favour asexual growth, but as the parasitaemia increases up to and beyond a critical level asexual growth is inhibited and the parasites are released from the inhibition of sexual development. In this model the formation of sexual stages is the default developmental pathway, whilst the asexual stages serve to increase the transmission potential of the parasite (Dyer and Day, 2003).

Irrespective of whether diffusible factors stimulate sexual growth, or override an inhibition on sexual development, the inference is that the parasites have a quantitative sensitivity to the environment, and are able to regulate their own growth and development (Dyer and Day, 2003). In the model proposed by Dyer and Day (2000), an intricate balance exists between stimulatory/inhibitory environmental signals, signal transduction pathways and transcription factors in the nucleus of the parasite. Variation in concentration of any of these components above or below a certain threshold can trigger a switch in the developmental cycle of the parasite. Some of these environmental influences are described below.

Nacher et al. (2002) described a linear relationship between the likelihood of observing gametocytes in peripheral blood smears and the degree of anaemia, which is a cause of severe morbidity in malaria (Chang and Stevenson, 2004). The onset of malarial anemia is precipitated by sub-optimal erythropoiesis following the increased production of erythropoietin (EPO) in the kidney. The production of EPO is

regulated in a feed-back loop with the degree of tissue oxygenation and Nacher et al. (2002) thus postulate that hypoxia could be one of the many signals for the onset of gametocytogenesis.

Host immunity is one of the most important factors affecting the rate of conversion to gametocytes. Immune response is associated with both increased and decreased gametocyte load, possibly because a decrease in gametocytaemia is accompanied by an increased rate of commitment to gametocytogenesis (Dyer and Day, 2000; Talman et al., 2004). The immune response to gametocytes has also been proposed as one of the reasons for the small number of asexual parasites that actually form gametocytes (Taylor and Read, 1997). If the transmission-blocking immunity that has been demonstrated for several surface antigens of *P. falciparum* is dependant on the gametocyte density, as has been proposed (Taylor and Read, 1997), it is advantageous to keep this number of gametocytes to a minimum.

In addition to the mechanisms postulated by Taylor and Read (1997), alternative explanations do exist for the surprisingly small number of asexual parasites that convert to gametocytes. One theory is that the parasites attempt to minimise the damage inflicted on the mosquito vector. Penetration of the stomach wall by the ookinete can be harmful to the mosquito; potentially fatal damage is reduced by minimizing the number of gametocytes as well as through the apoptotic potential of the ookinete (Talman et al., 2004). Another explanation, as described by Piper et al. (1999), is a naturally acquired age-dependant immunity to *PfEMP-1*. Trophozoites and gametocytes share the same repertoire of *var* genes, responsible for expression of the same combinations of *PfEMP-1* variants. Immunoepidemiological studies performed by Piper et al. (1999) showed that an antibody response is generated to an infected erythrocyte containing a *PfEMP1* variant irrespective of whether the erythrocyte contains a trophozoite or gametocyte, and that this immune response has the ability to regulate the density of both gametocytes and trophozoites.

Another host factor that increases the conversion to gametocytes is exposure of asexual parasites to host hormones, such as insulin, progesterone and testosterone. The same effect is seen upon exposure to host steroids and corticosteroids, but in both cases the mechanisms by which this occurs have yet to be ascertained (Dyer and Day, 2000; Talman et al., 2004).

Anti-malarial drugs have different effects on the development of gametocytes depending on their modes of action. Treatment with anti-folate drugs and chloroquine increases the number of gametocytes in malaria infected individuals (Barkakaty et al., 1988), potentially as a response to a declining asexual parasitaemia. Conversely a six-dose regimen of co-artemether administered to children in Gambia was found to reduce gametocyte prevalence, duration of gametocyte carriage and infectiousness to mosquitoes (Sutherland et al., 2005). *In vitro* studies showed that riboflavin has potent anti-malarial activities against both asexual and gametocyte stages of the disease, specifically when used in combination with other antimalarial drugs (Akompong et al., 2000), but the effects *in vivo* have yet to be ascertained.

1.3.4 Genetic factors regulating gametocytogenesis

To date very few genetic mechanisms have been implicated in the process of gametocytogenesis. It has been reported that a decrease in gametocyte development in parasites maintained in culture for a long time is associated with a deletion on the short arm of chromosome 9 (Alano et al., 1995), whilst defects in the production of male gametocytes have been associated with a region on chromosome 12 (Guinet and Wellems, 1997). Gardiner et al. (2005) were the first to report the identification of some of the genes on chromosome 9 thought to be associated with gametocyte formation, specifically a gene they have designated as *P. falciparum* gene implicated in gametocytogenesis (*pfgig*). This gene is expressed predominantly in schizonts, coinciding with the approximate time at which a decision to commit to a gametocyte is made. Up-regulation of this gene resulted in an increased expression of *Pfs16*, an

early marker of sexual development, whilst genetic silencing resulted in decreased gametocytogenesis. Significant expression of this gene has been recently confirmed using an early-stage gametocyte microarray, supporting the notion that it is involved in early sexual development (Gardiner et al., 2005; Young et al., 2005). Eksi et al (2005) have reported the identification of a sub-telomeric gene family that is expressed during the transition from asexual to sexual stages. A *P. falciparum*-specific sub-telomeric gene family consisting of 36 genes (designated DXF, for differentially expressed exported family) was found to contain six members that were differentially expressed in parasites committed to sexual development. The expression pattern of these genes is consistent with the commitment to sexual development, i.e. in schizonts of the cycle preceding gametocyte formation. This is similar to the expression pattern of the other markers of sexual development, *Pfs16* and *Pfs25*.

1.3.5 Signalling pathways implicated in gametocytogenesis

A prerequisite for a developmental cycle that is highly dependant on the external environment is the existence of signalling pathways that transmit environmental stimuli to the transcriptional apparatus of the parasite. Cyclin-dependant kinases are key regulators of the progression of the cell cycle, and numerous eukaryotic homologues of protein kinases and phosphatases have been cloned in *P. falciparum* (Doerig and Chakrabarti, 2004). Many of these are highly conserved and are either upregulated or exclusively expressed in gametocytes. Two mitogen activated protein kinases (MAPK) of the ERK1/ERK2 sub-family, *PfMAP-1* and *PfMAP-2*, have been identified as being expressed predominantly in gametocytes (Doerig, 1997; Dorin et al., 1999). Kinases belonging to this subfamily play a central role in regulation of cellular proliferation suggesting possible roles for these proteins in the process of sexual development, potentially either maintaining or releasing the gametocyte from its arrested state in the G₀ stage of the cell cycle (Dorin et al., 1999).

Evidence collected thus far points to the involvement of both a phorbol ester-induced pathway (Trager and Gill, 1989) and a cyclic AMP-dependant pathway during commitment to sexual development (Inselberg, 1983). In the latter pathway, protein kinase A is activated by the binding of cyclic AMP (cAMP), the levels of which are controlled by the activity of adenylyl cyclase. All three of these components have been identified in *P. falciparum* (Doerig, 1997), though the mechanism of action of the adenylyl cyclase differs to its mammalian counterpart in that it utilises Mn^{2+} ATP as a substrate rather than Mg^{2+} ATP, and its activity is unaffected by compounds that inhibit G-proteins (Doerig, 1997). Support for a role for cAMP pathways in sexual differentiation comes from stimulation of gametocytogenesis in cultures where cAMP levels were artificially elevated through the addition of caffeine (Brockelman, 1982) and 8-bromo-cAMP (Trager and Gill, 1989).

Xanthurenic acid, which stimulates gamete growth in the mosquito, also stimulates guanylate cyclase activity in purified gametocyte membranes. This may be through one of the two guanylate cyclase genes that have been identified in *P. falciparum*, both of which are specifically expressed in gametocytes (Doerig and Chakrabarti, 2004).

Dyer and Day (2000^a) proposed the existence of a G protein-dependant signal transduction pathway in the parasite based on a response to the cholera toxin. Heterotrimeric G proteins are conserved in eukaryotes and often control a cell's sensitivity to its environment by the coupling of membrane-bound signal receptors to downstream effector mechanisms. These proteins are activated through the exchange of GDP for GTP, and are able to regulate their own activity through an intrinsic GTPase activity. Dyer and Day (2000^a) induced gametocyte growth through addition of cholera toxin to cultures; this removes the GTPase activity of the protein causing uncontrolled signalling cascades. Despite these results, which indirectly implicate G-proteins in a signalling cascade, there is no evidence from the genome for the

existence of G-protein signalling in the parasite (Aravind et al., 2003; Doerig and Chakrabarti, 2004).

The decision to commit to sexual development is a multifactorial one, based on an intricate relationship between environmental stimuli and developmental pathways acting on stage specific genes. Although the mechanisms involved remain poorly understood, it is generally accepted that conversion to a gametocyte occurs when environmental conditions no longer favour asexual growth. This can be due to, amongst other as yet unidentified factors, a rapidly increasing asexual parasitaemia, drug pressure, host immune response and severe anaemia. Continued survival must then take precedence over multiplication, and the emphasis shifts to initiating mechanisms that will ensure the successful proliferation of the parasite. The complexities involved in controlling the life cycle to this extent can be better appreciated if one examines some of the gene regulatory mechanisms at play in the parasite.

1.4 *P. falciparum* gene regulation

Plasmodium is a small, haploid, genomically complicated organism that is able to adapt its gene expression to suit the myriad environments in which it finds itself (Bannister and Mitchell, 2003). During the course of natural infection, the malaria parasite invades and multiplies within several different cell types: hepatocytes and erythrocytes in humans, and gut, vasculature and salivary glands in mosquitoes. Upon entry into each cell immediate adaptations must be made to ensure survival and proliferation in the new environment, and this must require a highly coordinated pattern of gene expression. Investigations into the mechanisms of gene regulation in *Plasmodium* were for a long time hampered by the limited techniques that could be applied, as well as difficulties inherent in working with the different developmental stages of the parasite. Only with the advent of transfection technology in *P. falciparum* (Wu et al., 1995) has light been shed on some of the elements that control

the essential gene regulatory mechanisms in the parasite (Crabb and Cowman, 1996; Horrocks and Kilbey, 1996; Wickham et al., 2003). Homologues of all 12 subunits of RNA Polymerase II, TATA-box binding protein (TBP) and TFIIB have been identified in the genome using bioinformatic or experimental analysis (McAndrew et al., 1993; Coulson et al., 2004; Ruvalcaba-Salazar et al., 2005). Even though the specific sequences and some of the regulatory factors involved may differ, the overall mechanism of gene regulation in the parasite shows similarities to the classic eukaryotic principles of gene regulation. The correct timing and expression levels of genes are controlled by promoter and terminator sequences which flank the coding regions (Horrocks et al., 1998), and the rate of transcription is regulated by the binding of transcription factors (*trans*-acting elements) to activation sequences (*cis*-acting elements) within promoter regions. Binding of transcription factors to DNA alters the conformation of DNA in that region, which either allows access of other transcription factors to previously unreachable sequences, or alternatively prevents the binding of other transcription factors. Other features in common with eukaryotic gene regulation include monocistronically transcribed genes, 5' and 3' untranslated regions, the presence of introns, capped mRNA and polyA tails (Kumar et al., 2004).

Transcriptional regulation has been shown to play a major role in controlling gene expression in many of the *Plasmodium* species. Microarray experiments (Bozdech et al., 2003; Florent et al., 2004) have demonstrated fluctuating levels of RNA throughout the parasite's life cycle for many genes, implying control of gene regulation at the level of RNA synthesis or mRNA stability. Nuclear run-on experiments have demonstrated transcriptional regulation of genes required for pathogenesis (Scherf et al., 1998), sexual differentiation (Alano et al., 1996) and cytoadherence (Lanzer et al., 1992). There is also evidence for posttranscriptional regulation of gene expression in the parasite (Coulson et al., 2004), and the identification of the Puf family of RNA binding proteins in *P. falciparum* (Cui et al., 2002) suggests regulation of gene expression at the translational level. Thus the

control of gene expression is likely to occur at multiple levels, including epigenetic mechanisms.

1.4.1 Epigenetic gene regulation in *P. falciparum*

Regulation of gene expression that is mediated by chromatin with no involvement of transcription factors is referred to as “epigenetic.” In most instances this involves the silencing of individual genes through condensation of the surrounding chromatin fibre, rendering binding sites required for initiation of transcription inaccessible to transcription factors or other proteins (Horrocks et al., 1998). The conformation of the chromatin fibre is controlled by acetylation, methylation or phosphorylation of the histone proteins, themselves differentially expressed, which form part of the nucleosome (Sternier and Berger, 2000). The expression of these proteins is strictly regulated, but the processes by which this is mediated in *Plasmodium* have not been elucidated (Lobo and Kumar, 1999).

There is much evidence to suggest that mechanisms for the chromatin-mediated regulation of gene expression do exist in the parasite, such as the non-random distribution of gene clusters in the parasite genome (Le Roch et al., 2003) combined with the identification of proteins involved specifically in chromatin remodelling (Ji and Arnot, 1997; Fan et al., 2004). Furthermore, there is evidence to suggest that the extent of chromatin packaging in gametocytes is reduced compared to asexual parasites, possibly to accommodate the three rounds of rapid replication that male gametocytes must undergo as they enter the mosquito (Pace et al., 1998).

The chromatin environment defines specific genes as either transcriptionally active or silent, and transformation from one state to the other is accomplished primarily through the covalent modifications of histones. Alterations to chromatin structure could account for variation in patterns of gene expression, and hence morphological and metabolic disparities between the different developmental stages of the parasite.

Probably the best studied case of epigenetic gene regulation in *Plasmodium* is the *var* gene family. Malaria parasites rely on a mechanism of antigenic variation whereby they evade the antibody response of the host by altering the antigenic phenotype of the infected red blood cell. The *var* genes encoding the erythrocyte membrane protein family (*Pf*EMP1) are primarily responsible for this mechanism. *P. falciparum* possesses a family of 40-60 *var* genes whose expression must be tightly regulated. Expression of more copies than is necessary would result in premature expenditure of the antigenic repertoire, but the rate of switching between genes must be sufficiently rapid to stay ahead of the corresponding antigenic response of the host (Horrocks et al., 1998; Deitsch, 2004). Regulation of *var* gene switching appears to be controlled at the level of transcription initiation, as evidenced by the identification of only a single transcribed *var* gene in nuclear run-on experiments. Switching between *var* genes is not accompanied by any conformational changes in DNA as is the case in many species exhibiting clonal antigenic variation (Scherf et al., 1998). Therefore regulation of the expression of *var* genes most likely involves epigenetic mechanisms.

1.4.2 Transcriptional gene regulation

Transcriptional regulation implies the control of gene expression regulated by the interplay between promoters, regulatory sequences and transcription factors. This mechanism of gene regulation will be discussed in detail in Chapter 3.

1.4.3 Post-transcriptional gene regulation

Alterations in the protein-coding sequence of a pre-mRNA after its synthesis, RNA editing, regulation of the stability of mRNAs and regulation of the subcellular location of specific mRNAs are all post-transcriptional mechanisms by which a cell can modulate the levels of protein to suit its developmental requirements (Lodish et al., 1999). In *Plasmodium*, it has been postulated that there is a heavy reliance on

post-transcriptional mechanisms to control protein expression. This assertion is based on the paucity of recognisable transcription factors or their regulatory domains, an increased number of potential RNA binding proteins in the genome and the high prevalence of antisense transcripts in the genome (Patankar et al., 2001). Furthermore, the *Plasmodium* genome encodes approximately one third of the number of proteins involved in transcriptional processes when compared to the genomes of other eukaryotes (Coulson et al., 2004). Conversely, it encodes nearly twice the number of CCCH-type zinc finger proteins which function in regulating mRNA stability and localisation.

Given the regulatory sequences and transcription factors that have been identified in the parasite genome, it is likely that post-transcriptional mechanisms work in concert with transcriptional regulation to control gene expression. Post-transcriptional control allows a more rapid response to new environments such as is required during the transmission from mosquito to host, and also places little limitation on the nucleotide composition of intergenic sequences, which would explain the approximately 90% AT content seen in *P. falciparum* (Coulson et al., 2004).

1.4.4 Translational mechanisms of gene regulation

Translational control plays an essential role in the regulation of gene expression during the development of most eukaryotic cells. Identification of the Puf family of RNA binding proteins (Cui et al., 2002) provides some evidence for the existence of this regulatory mechanism in *Plasmodium*. These evolutionarily conserved proteins are an important family of translational regulators that control the expression of multiple genes by binding to the 3' untranslated region of the target mRNA and repressing translation (Wickens et al., 2002).

Examples of translational gene regulation in *P. falciparum* are the stage specific expression of *Pfs25*, and the *var* gene family. In both cases the mRNA is evident in

early ring stages but protein is only detected later. The mechanisms that allow nascent mRNA to remain untranscribed in the cell have yet to be ascertained.

Ribosomal RNAs (rRNAs) are central to defining the ribosome, the complex that is formed between RNA and proteins to facilitate translation (McCutchan et al., 1995). Regulation of the rate of rRNA synthesis is an essential mechanism by which a cell controls its growth and development, and numerous regulation strategies have been identified. The distinct types of rRNA units that exist between erythrocytic and mosquito stages of the parasite life cycle may be a unique response to the innate cellular need to optimise protein synthesis under different physiological conditions (Mercereau-Puijalon et al., 2002).

1.5 Aims and Objectives

The aim of this research was to gain some insight into the mechanism by which an asexual parasite develops into a gametocyte. The two-fold strategy entailed identifying genes differentially expressed between these two developmental stages, followed by analysis of a transcription factor, *PfMyb2*, which may play a role in development.

Chapter 2:

- To utilise differential display to identify some of the genes differentially expressed between the asexual and gametocyte parasite stages of *P. falciparum*
- To characterise and categorise the gene transcripts identified through differential display
- To confirm the expression profiles by comparison with available microarray data
- To ascertain the utility of differential display as a genome profiling method in an era of high-throughput techniques such as microarrays

Chapter 3:

- To express and purify 6xHis-*PfMyb2*, a *P. falciparum* transcription factor
 - To perform binding studies with recombinant 6xHis-*PfMyb2* to ascertain the *in vitro* binding potential of this protein with consensus Myb regulatory elements
-

CHAPTER 2- IDENTIFICATION OF *P. FALCIPARUM* STAGE SPECIFIC GENE EXPRESSION

2.1 Introduction

Consequent to the successful sequencing of the *P. falciparum* genome a new era has dawned in *Plasmodium* research. Identification of the approximately 5300 genes (Gardner et al., 2002) that are co-ordinately expressed during the life cycle of the parasite represents but the tip of the iceberg in the quest for a greater understanding of the biological complexities inherent in the parasite. By themselves, the millions of bases of DNA sequence that have been identified shed no light on gene function, cellular processes or targets for drug and vaccine development (Lockhart and Winzeler, 2000). The focus of research has thus shifted from the identification of genes to the determination of the protein products of these genes and their patterns of expression between different life cycle stages and under the influence of different environmental stimuli. This new field of functional genomics aims to achieve a greater understanding of the complex interplay that occurs between the constituent components of a biological system (Lockhart and Winzeler, 2000).

In recent years gene expression profiling has replaced traditional gene-by-gene analysis as the quintessential tool for studying gene function (Le Roch et al., 2003), and has been successfully employed to gain insight into many fundamental biological processes. Large scale genome screening methods are ideally suited to an organism such as *Plasmodium* where the application of routine genetic tools presents unique challenges. The complexity of the life cycle, the difficulties in maintaining most stages in routine culture and the predominantly haploid life cycle, which complicates any attempts at gene disruption (Wellems et al., 1999), are all motivating factors to exploit new high throughput methods to accumulate information about genes. To this end, numerous large-scale genome profiling methods have been employed in

Plasmodium, of which DNA microarrays (Hayward et al., 2000; Mamoun et al., 2001; Bozdech et al., 2003; Le Roch et al., 2003), serial analysis of gene expression (SAGE) (Munasinghe et al., 2001; Patankar et al., 2001), suppression subtractive hybridisation (SSH) (Dessens et al., 2000; Spielmann and Beck, 2000; Florent et al., 2004) and differential display (Cui et al., 2001) have proven to be particularly valuable.

2.1.1 Differential display

The technique of differential display, first described by Liang and Pardee (1992), is the method of choice in many laboratories for the direct comparison of two or more RNA populations. It is a simple, sensitive and reliable technique that was developed to overcome the limitations of earlier methods in the search for differentially expressed genes (Liang, 2002). The method relies on the amplification by polymerase chain reaction (PCR) of cDNA sequences derived from mRNA populations by reverse transcription. Poly(dT) primers, with two anchoring bases at their 3' end to anneal immediately upstream of the polyA tail, are used in combination with defined arbitrary 5' primers to generate a pool of differently sized cDNA fragments which can be resolved on an acrylamide gel and visualised by autoradiography. In this manner mRNA can be preferentially reverse transcribed from a pool of total RNA. The method provides a comparative snapshot of the gene expression between cells from different developmental stages or under the influence of different environmental perturbations.

The 5' primer is usually a short (10-15mer) arbitrary oligonucleotide that will anneal to multiple sites on the cDNA. Because the 5' primer will be used for reamplification following identification of differentially expressed fragments, the optimal length needs to be empirically determined: a shorter primer will anneal to the end of the cDNA molecule more frequently generating more fragments, whilst a longer primer is required for efficient reamplification of differentially expressed fragments. In their

original description of the method Liang and Pardee (1992) utilised a 10 base primer for both the initial RT-PCR reaction and the subsequent reamplification.

Following the random amplification of cDNA fragments, bands on the autoradiograph that are unique to one sample are considered to be differentially expressed, and are reamplified after excision and elution from the acrylamide gel. Sequencing and bioinformatic analysis of the reamplified fragments allows identification of genes differentially expressed between the biological samples being compared.

Cui et al. (2001) were the first to use differential display to compare gene expression between the different developmental stages of *P. falciparum*, with a view to identifying some of the genes that are differentially expressed during gametocytogenesis. Prior to this publication it was thought that the high AT content of the *Plasmodium* genome would preclude the use of differential display, as the oligo(dT) primers could anneal to any of the polyA stretches that are found throughout the genome. However, Cui et al. identified 96 genes as being stage specific thereby showing that the application of differential display to *Plasmodium* genomics can be highly valuable. Furthermore, they considered the annealing of oligo(dT) primers to internal A-rich regions of mRNAs as being advantageous because the majority of the differential display products were found to be within coding regions and could be rapidly identified using BLASTN and BLASTX analysis.

Subsequent to the pioneering work by Liang and Pardee there have been countless papers published, describing both the successful application of the technique and the high rate of false positives that seemed to be prevalent. Numerous modifications and improvements have been proposed in an attempt to reduce the rate of the false positives, as discussed below.

The first anomaly arises with the use of the short random 5' primer which can anneal degenerately during the reverse transcription step making the reproducibility of the technique difficult. This can be exacerbated if there are even slight temperature differences between tubes, or different batches of polymerase are used (Liang and Pardee, 1995). Inconsistent annealing between samples can erroneously identify a gene as being differentially expressed. Furthermore, subcloning of the fragments is required to get accurate sequence, as the short primer is not optimal for direct sequencing of the PCR products. This problem can be circumvented by the use of a longer 5' oligonucleotide (20mer). A low annealing temperature can be used for the first PCR cycle to achieve arbitrary priming, and the annealing temperature can then be increased to attain greater stringency, and hence reproducibility. This method has been successfully utilised in differential display to improve the reproducibility and facilitate subsequent band reamplification and sequencing (Zhao et al., 1995).

Another practical limitation to the technique is the presence of contaminating sequences, which can arise in one of two ways: identically sized cDNA fragments can co-migrate with the band of interest on the acrylamide gel, or alternatively a contaminant can be introduced during the reamplification procedure. Whilst the contaminating sequence may initially be present in tiny amounts relative to the fragment of interest, after reamplification by PCR the two DNA sequences may be amplified to almost equivalent levels (Miele et al., 1998). Aside from the obvious precautions that should be taken to avoid contamination in the reamplification or gel elution process, two further methods have been described for the reduction of false positives by contaminating sequences. The first describes the use of a modified single strand conformation polymorphism (SSCP) reaction following elution of the candidate fragments. SSCP allows separation of single stranded DNA molecules on the basis of conformation rather than size, and the "correct" fragment can then be eluted from the SSCP gel (Miele et al., 1998). In the second method, Callard et al. (1994) only subcloned half of the reamplified product, and used the balance as a hybridisation probe to identify the plasmid colonies containing the correct insert. In

both instances the authors have managed to isolate the true differentially expressed fragment from contaminating sequences.

A final aspect of the technique that has come in for some criticism has been its questioned ability to identify rare transcripts (Matz and Lukyanov, 1998; Lievens et al., 2001). It has been suggested (Bertioli et al., 1995) that the competition for substrates between the PCR components would be the limiting factor for amplification. For example, vital components such as dNTPs could be depleted before rare transcripts have had a chance to be amplified to perceptible levels. This is an important point as the difficulty in identifying genes responsible for a specific function is often a result of the gene being expressed at low levels. However, depending on the cell types being compared it may not present a problem, as the greater the number of differentially expressed transcripts, the higher the probability of detecting some of them (Matz and Lukyanov, 1998). Thus if the two samples being compared are expected to have large variation in gene expression, some differentially expressed transcripts will inevitably be identified.

Despite the intricacies of the technique, differential display is still the simplest and most economical method for mRNA comparison, particularly where a pronounced difference is expected between the populations being compared and identification of a few differentially expressed transcripts is sufficient. Used in combination with a reliable method of verification it can be a very powerful technique for the identification of stage specific gene expression.

2.1.2 Microarrays

Microarray technology is designed to measure fluctuations in mRNA expression levels to facilitate an understanding of biological processes (Yue et al., 2001) and is capable of screening entire genomes simultaneously. This means that the expression profile of every developmental stage can be elucidated, and functions ascribed on the

basis of the cell physiology at the time of expression. This has been confirmed in *Plasmodium* by the microarray analysis of Bozdech et al (2003), who used a high density oligonucleotide microarray to analyse the complete transcriptome of the intraerythrocytic developmental cycle of the parasite. The discovery in *Plasmodium* that the temporal expression of a gene is integrally linked to the cellular requirement for its protein product was confirmed by Le Roch et al. (2003), who organised groups of genes into clusters based on their temporal expression patterns, and concluded that all genes within a cluster are likely to be involved in similar cellular processes and hence the expression profile of a gene can give insight into its biological role. Further verification from a proteomics perspective was provided by Florens et al. (2002), who utilised a combination of high-resolution liquid chromatography and tandem mass spectrometry to compare the protein complement of the intraerythrocytic stages of the parasite.

All three research groups illustrated that the induction of genes in the erythrocytic stages of the parasite occurs only once per cycle, and only when there is a need for the protein product. Sporozoites for example, following inoculation by the female *Anopheles* mosquito, are primarily concerned with evasion of the host immune system and invasion of the hepatic parenchymal cells, and their protein complement predominantly mirrors these requirements. Merozoites require the machinery to facilitate invasion of erythrocytes following release from the liver, as well as the proteins that will migrate to the surface of the infected erythrocyte to direct the complex immune evasion strategy of the parasite. The maturation of the trophozoite sees concomitant expression of proteins required for metabolism, sequestration and immune evasion, and later the DNA replication machinery required for the development of multinucleated schizonts. Gametocytes contain the protein complement required for the resumption of the cell cycle upon entry into the mosquito as they react to the stimuli to initiate gametogenesis (Florens et al., 2002). Thus as the malaria parasite undergoes the distinct morphological and metabolic changes that characterise its life cycle, a tightly coordinated pattern of gene

expression is mandatory to ensure that the protein requirement of each stage is fulfilled. Evidence by Bozdech et al (2003) also shows that as much as 60% of the genome is active throughout the entire life cycle, with discrete peaks of expression at specific times. Thus what seems to be the induction of a gene at a time corresponding to the need for its protein product may in fact be up-regulation of a constitutively active gene.

Although operons do not exist in the parasite, as gene expression appears to be monocistronic (with the exception of the mitochondrial genes) (Carlton, 1999), it has been postulated that in addition to being temporally co-expressed, groups of functionally related genes are clustered in similar positions on chromosomes (Florens et al., 2002). For example, genes belonging to the *var* gene family that encode erythrocyte membrane proteins cluster subtelomerically (Fischer et al., 1997), and genes required in the initial stages of gametocytogenesis seem to be localised to a subtelomeric region of chromosome 9 (Alano et al., 1995). Genome profiling has produced conflicting data for whether functionally related genes are spatially clustered. Whereas proteomic analysis by Florens et al. (2002) and functional clustering by Le Roch et al. (2003) demonstrated that genes with similar expression patterns and functions are clustered on chromosomes, the microarray analysis of Bozdech et al. (2003) has shown that this rarely occurs. Perhaps the difference between the sets of data arises from the differing methods used to identify clusters as well as different criteria for what constitutes a gene cluster. Certainly there is agreement on the proposal by Le Roch et al. (2003), that telomeric genes are primarily concerned with erythrocyte remodelling whereas genes towards the centromeres are more likely to be involved in cell growth and maintenance.

In an organism such as *Plasmodium*, where approximately 60% of the identified open reading frames (ORFs) have no known sequence similarity to genes in other organisms (Gardner et al., 2002), the synchronicity between gene expression and developmental progression is particularly relevant, and can be used to glean insight

into the functions of the unique genes. Genes that are co-expressed at different developmental stages or under the effect of different environmental stimuli, or are present in clusters on chromosomes, can be assigned similar functions in a “guilt by association” model. For example, based on their presence in clusters of co-expressed genes, Florens et al. (2002) proposed putative functions for 24 proteins annotated as hypothetical in the PlasmoDB database.

Taken together, the microarray data that have been published thus far go a long way toward facilitating an understanding of the complex patterns of gene expression and regulation in the parasite. However, the costs of performing these experiments are substantial and not feasible for the majority of laboratories. The method also requires a high degree of technical expertise to standardise and interpret results. Furthermore the method has been questioned for its accuracy and sensitivity, specifically where cDNA probes are used (Liang, 2002). As with any new technique, refinements and modifications will doubtless develop to increase the accuracy of the technique, but problems do exist with the reproducibility of results, the sensitivity and cross-hybridisation of the probes, and interpretation and management of data (Liang, 2002).

In light of the urgency in identifying novel drug and vaccine targets numerous laboratories have performed alternative genome profiling methods and have successfully contributed to increasing the knowledge-base surrounding the parasite.

2.1.3 Serial Analysis of Gene Expression

SAGE is a highly sensitive and quantitative study of the gene expression profile of a given cell. The technique is based on two principles: first, a short (10mer) oligonucleotide from a defined position within a transcript can uniquely identify a gene; second, ligation of several short tags into a single molecule allows for successful sequencing of these molecules. The expression levels of a gene can then be quantitated by the relative abundance of its representative tag (Velculescu et al.,

1995; Munasinghe et al., 2001). SAGE has been successfully used to screen the *Plasmodium* genome for variations in gene expression despite its high AT content (Munasinghe et al., 2001). Whereas microarray studies are limited to the analysis of known genes in a so-called “closed platform,” SAGE analysis is not dependant on the prior availability of transcript information (Velculescu et al., 1995) and can thus assist in the identification of uncharacterised ORFs. It has been particularly useful in its identification of the large number of antisense transcripts that characterise the *Plasmodium* genome (Patankar et al., 2001). This discovery is important for its implications in the mechanisms of post-transcriptional regulation in the parasite. The biggest drawback with the SAGE technique is that, being of the “open platform” variety, numerous sequencing reactions are required to gauge accurate expression information. This is both costly and time consuming.

2.1.4 Suppression Subtractive Hybridisation

Suppression subtractive hybridisation (SSH) is a profiling technique that has been used to screen the *Plasmodium* genome for differentially expressed genes (Dessens et al., 2000; Spielmann and Beck, 2000; Florent et al., 2004). SSH is a modification of the standard subtractive hybridisation (SH) technique that compares two populations of mRNA and produces clones of genes that are unique to one population. The principle of the subtraction step is that following hybridisation of the two populations the remaining unhybridised cDNAs represent genes unique to one population. The modification of SSH is that these differentially expressed sequences are isolated by selective PCR amplification rather than by physical separation (Diatchenko et al., 1996). Dessens et al. (2000) first proved the utility of SSH in *Plasmodium* parasites when they compared gene expression between mosquito stages (gametes, zygotes and ookinetes) and asexual parasites from a non-gametocyte producing strain of *P. berghei*. Florent et al. (2004) set out to analyse the expression profiles of the late erythrocytic stages of *P. falciparum* and in their analysis of 50 gene transcripts differentially expressed during merozoite morphogenesis, they recognised that SSH

was able to identify alternative mRNA splice sites, and hence refined the description of two genes previously annotated in the PlasmoDB database. The differences in splice site organisation may be due to strain-dependant factors (SSH was performed on the FcB1 parasite strain whereas PlasmoDB is based on 3D7 parasites) or may have more substantial relevance with regard to alternative splicing mechanisms in the parasite. Like all genome profiling methods SSH has certain drawbacks. The procedure favours highly differentially expressed genes and thus the primary application should be to detect dramatic changes in gene expression (Ji et al., 2002). The technique is also technically demanding and requires a large investment of time and money.

In light of the importance of understanding gene expression in the malaria parasite, differential display was used in this study to compare the expression profiles of asexual and gametocyte stages of *P. falciparum*. Despite the immense advances that have been made in *Plasmodium* genomics subsequent to the initiation of this research, small-scale genome profiling efforts such as this can still make a contribution to the knowledge-base relating to the complex biology of the malaria parasite.

2.2 Materials and Methods

2.2.1 *Plasmodium falciparum* culture methods

The 3D7 strain of *Plasmodium falciparum* parasites was used throughout this research. Initial stocks were a gift from the Department of Pharmacy and Pharmacology of the University of the Witwatersrand.

All parasite culture work was performed in a laminar flow hood, and standard sterile techniques used throughout. Equipment was sterilised by autoclaving (120°C, 15lb/in²) prior to use, and reagents and media were sterilised either by autoclaving or by filtration through 0.22µm filter unit with the aid of a peristaltic pump for larger volumes.

MilliQ water was used for the preparation of all solutions and media.

All centrifugation steps were performed in a Jouan BR 3.11 centrifuge at room temperature.

2.2.1.1 Preparation of parasite culture from frozen stock

Frozen parasite cultures were removed from liquid nitrogen and placed in a preheated water bath at 37°C. The thawed suspension (1ml) was transferred to calibrated Falcon tubes, and 0.1ml 12% NaCl was added whilst swirling the tubes. The cultures were incubated for five minutes at room temperature, after which nine volumes of 1.6% NaCl were added in a drop wise manner. The tubes were centrifuged at 600g for five minutes, the supernatant was removed, and the pellet resuspended in nine volumes of a 0.9% NaCl/ 0.2% glucose solution. The tubes were centrifuged as before and fresh erythrocytes were added to the pellet to a volume of 750µl. This was transferred to a sterile Nunc 24cm² tissue culture flask and 5ml complete RPMI culture medium (A-1) supplemented with 20% human AB plasma was added. The flask was gassed with 93% N₂, 5% CO₂, and 2% O₂ and incubated at 37°C for 48

hours before microscopic analysis was performed to ascertain the viability of the parasites.

2.2.1.2 Freezing of asexual parasite cultures

Asexual parasites mainly in the ring stage were centrifuged at 800g for five minutes and the supernatant was removed. The packed cells were resuspended in a 1:1 ratio with freezing solution (A-1). 1ml of this suspension was aliquoted into sterile 1.8ml cryotubes, incubated at room temperature for 5 minutes and stored in liquid nitrogen.

2.2.1.3 Preparation of erythrocytes for malaria cultures

Whole blood was collected in Acid Citrate Dextrose (ACD) tubes from volunteers. Tubes were centrifuged at 900g for ten minutes and the plasma and white cell layers were removed. The remaining red cells were resuspended in two volumes sterile PBS and centrifuged as before. This procedure was repeated three times to ensure complete removal of contaminating white cells. An equal volume of incomplete RPMI medium (A-1) was added to obtain a haematocrit of 50%. Washed red blood cells were stored at 4°C for up to two weeks. Whole blood could be stored at 4°C in ACD tubes for up to one month and washed prior to use.

2.2.1.4 Continuous culture method

Parasite cultures were incubated at 37°C in an atmosphere consisting of 93% N₂, 5% CO₂, and 2% O₂ according to a modification of the method of Trager and Jensen (1976). Cultures were maintained at a haematocrit of 5% in 80cm² or 175cm² Nunc tissue culture flasks containing 20ml or 30ml complete RPMI medium (A-1) respectively. Parasites were monitored daily by microscopic analysis of stained blood smears (A-2), and parasitaemia was calculated by counting the number of parasitized erythrocytes as a percentage of total red cells in the microscope field. This was kept within a constant range of 4-10% either through the aspiration of

culture or by dividing cultures into two flasks. Culture medium was removed and pre-warmed complete culture medium was added daily. The culture flasks were kept out of the incubator for as little time as possible.

2.2.1.5 Sorbitol synchronisation of cultures

D-sorbitol provokes osmotic lysis of all of the parasite stages except rings (Lambros and Vanderberg, 1979). Cultures containing predominantly early ring stage parasites were transferred to 50ml Nunc tubes and centrifuged at 600g for five minutes. The supernatant was aspirated and ten volumes (approximately 20ml) of 5% D-sorbitol in PBS were added. The solution was incubated at 37°C for 40 minutes, followed by centrifugation at 600g for five minutes. The supernatant was removed and the pellet resuspended in complete culture medium, and incubated as described.

2.2.1.6 Preparation of gametocytes

Exposure to certain stressful conditions induces the formation of gametocytes in the 3D7 strain of *P. falciparum*. An increase in parasitaemia, a diminished surface area in which to grow and no addition of fresh erythrocytes are factors that stress the parasites resulting in their conversion to sexual stages (Carter et al., 1993). On average, eight flasks of synchronous ring-stage parasites were prepared at a time. The method used to culture gametocytes was a modification of the method of Carter et al. (1993). Cultures were initiated at a parasitaemia of 2% and a haematocrit of 5% in an 80cm² tissue culture flask, with 20ml complete culture medium supplemented with 10% human serum. Dilutions were prepared from stock cultures that had been recently thawed, as the gametocyte conversion rate decreases after prolonged time in culture. Flasks were left in an upright position to reduce the surface area available to the parasites. The medium was replaced daily, but only 15ml medium was removed each day to avoid aspirating the gametocytes which don't adhere to the flask surface. Culture flasks were gassed daily with 2% O₂, 5% CO₂ and 93% N₂ for 30 seconds.

Microscopic analysis was performed every alternate day. Gametocyte cultures were harvested after 12-16 days.

One of two methods was used to purify the gametocytes from residual asexual parasites. The cultures were either purified through a Percoll density gradient (see below) or were subjected to two rounds of D-sorbitol treatment 24 hours apart (Petmitr et al., 1997) starting on day 12, and the gametocytes collected by centrifugation at 900g for 10 minutes.

2.2.1.7 Isolation of gametocytes by Percoll gradient

After 12-16 days medium was removed; cultures were combined into two 50ml Nunc tubes, centrifuged at 900g for five minutes, and the pellet resuspended in 6ml 1X incomplete culture medium (A-1).

10X incomplete culture medium (IM) was used to make 90% Percoll, and this was diluted with 1X IM to make 35%, 50%, 65% and 80% Percoll (in both instances, incomplete culture medium could be replaced with PBS). 2.5ml of each dilution were carefully layered into a 15ml Falcon tube to form a step-wise gradient. 3ml resuspended gametocyte cultures were added to the top and the tubes were centrifuged at 1100g for 40 minutes at 37°C.

Layers were washed with 15ml incomplete RPMI culture medium by centrifugation at 900g for ten minutes. Pellets were resuspended in 1ml PBS, and 5µl were used to prepare a smear to verify the presence of gametocytes. Relevant layers were pooled and washed twice in PBS. The supernatant was discarded and the pellet was either frozen in liquid nitrogen or used immediately for an RNA extraction.

2.2.2 RNA extraction

The primary concern when working with RNA is the control of RNases, tiny amounts of which are sufficient to destroy RNA. RNases may be released from cells immediately after harvesting, or may arise through contamination from skin or clothing. Thus all RNA work was performed in a dedicated area of the laboratory which was wiped thoroughly with RNaseZap prior to use. In addition, the following basic precautions were taken:

Samples were kept on ice, which slows down the activity of RNases.

Latex gloves were worn and changed regularly.

Disposable sterile polypropylene Falcon tubes were used at all times.

A dedicated set of pipettes was used for all RNA work, and filter pipette tips used at all times.

All tubes and pipette tips used were certified RNase/DNase free by the manufacturers.

Only water treated with diethylpyrocarbonate (DEPC), or provided as nuclease-free, was used. 0.1% DEPC was prepared in water, mixed overnight, and then autoclaved to destroy the DEPC.

2.2.2.1 Total RNA isolation Kit

The Genelute™ Mammalian Total RNA Miniprep kit was used for gametocyte extractions and asexual parasite extractions from fewer than two 80cm² tissue culture flasks. The kit was used as per manufacturer's instructions. Pelleted cells were resuspended in saponin at a final concentration of 0.05%. Cells were left at room temperature for approximately five minutes to allow for cell lysis and the parasites were collected by centrifugation at 1100g for five minutes at 4°C and washed three times with PBS. Gametocytes were lysed and homogenized in guanidine thiocyanate and β-mercaptoethanol to release RNA and inactivate RNases. Lysates were centrifuged through a filtration column to remove debris and shear DNA. The filtrate

was then applied to a high capacity silica column to bind total RNA, followed by washing and elution in sterile distilled water. Centrifugation steps were performed in a Sorvall RMC14 microcentrifuge at 4°C.

2.2.2.2 Guanidium Isothiocyanate method

This method (Chomczynski and Sacchi, 1987) was used with larger quantities of starting material, such as RNA extractions from asexual parasite stages where up to four 175cm² culture flasks were combined.

Four 175cm² sterile tissue culture flasks, each containing 30ml culture medium at 10-15% parasitaemia, were used. Parasites were transferred to 50ml sterile Nunc tubes and centrifuged for five minutes at 1100g at 4°C in a Jouan BR3.11 centrifuge. The supernatant was removed and saponin was added to the pellet at a final concentration of 0.05% to lyse the erythrocyte pellet. Tubes were left to stand at room temperature for five minutes and were then centrifuged for ten minutes at 1100g. The parasite pellet was washed three times in PBS. 5ml ice-cold D* solution was added (4M guanidine thiocyanate, 0.75M sodium citrate pH 7.0, 10% N-laurylsarcosine and 0.072% β-mercaptoethanol) and the pellet was resuspended completely using a 21 gauge needle. 0.5ml 2M sodium acetate (pH 4.0 with glacial acetic acid) was added and mixed by gentle inversion. RNA was extracted twice with buffered phenol (pH 4.2) and chloroform at room temperature and precipitated with isopropanol overnight at -20°C (A-2). After a wash with 75% ethanol, the pellet was resuspended in 20μl nuclease free water and heated to 65°C for ten minutes to increase solubility. Concentration was determined by measuring the absorbance of the sample at a wavelength of 260nm (2.2.2.5) on a Beckman DU-65 spectrophotometer, and RNA samples were diluted to a concentration of 1μg/μl and stored at -70°C.

2.2.2.3 TRI reagent

This method became the method of choice for all RNA extractions. 20ml parasite culture were transferred to 50ml sterile tubes (Nunc) and centrifuged for five minutes at 1100g in a Jouan BR3.11 centrifuge. The pellet was washed once in sterile PBS, and 20ml TRI reagent pre-warmed to 37°C were added to the pellet. The reagent is a solution of phenol and guanidine isothiocyanate which disrupts cells and dissolves cellular components whilst maintaining the integrity of the RNA. Addition of chloroform followed by centrifugation separates the solution into an organic phase and an aqueous phase. The RNA can be purified from the aqueous phase by precipitation with isopropanol (A-2).

2.2.2.4 Removal of DNA contamination from RNA samples

To minimise false positives in the differential display experiments the template RNA had to be devoid of any contaminating DNA. Digestion of residual DNA with DNaseI ensured that all products amplified in the RT-PCR reactions originated from RNA.

40 units (1µl) RNase-free DNaseI in 1X Buffer L (1mM Tris-HCl pH 7.5, 1mM MgCl₂) (Roche, Germany) were added for every 100µl of reaction, and the tubes were incubated at 37°C for 30 minutes. The enzyme was inactivated at 99°C for five minutes.

2.2.2.5 Measurement of RNA yield and purity

An RNA solution with a concentration of 40µg/ml has an absorbance of 1.0 at a fixed wavelength of 260nm. The absorbance of 1µl of RNA in 49µl TE buffer (A-1) was measured at 260nm (A₂₆₀) and 280nm (A₂₈₀) UV light in a Beckman DU65 spectrophotometer calibrated with TE buffer. The purity of an RNA preparation is ascertained by calculating the A₂₆₀:A₂₈₀ ratio. This value should be between 1.8 and 2.0. A smaller value is indicative of protein contamination, and if necessary the RNA

can be purified a second time by phenol: chloroform extraction.

Further verification of the quality of the RNA came from analysing an aliquot on a 1% agarose gel containing 10µg/ml ethidium bromide, electrophoresed at 90V for 30 minutes in 1X TAE buffer (A-1). The 28S and 18S ribosomal RNA bands are used as an indication of the quality of the RNA. These two subunits should ideally be visible as distinct bands on an agarose gel with the intensity of the 28S subunit approximately twice that of the 18S subunit.

2.2.3 Differential Display

2.2.3.1 Reverse Transcription Polymerase Chain Reaction

RNA was reverse transcribed using Superscript III reverse transcriptase. The quantity of RNA ranged from 0.5-1µg in a 20µl reaction. The quality of RNA was an important factor in minimising false positives and a high concentration of template RNA helped to ensure that genes expressed at low levels were still amplified. Asexual and gametocyte stage parasite preparations were kept separate to minimize the chances of cross-stage contamination. Two-base anchored primers were designed in such a way as to selectively anneal to the 5' end of the poly-A tail, thereby preferentially transcribing messenger RNA (mRNA). Primers were designated T₁₂VN, where V represents G, A and C and N represents any of the four nucleotides.

A separate reaction tube was set up for each anchored primer, containing 0.5µM primer, 0.5-1µg RNA, 0.5mM dNTP in a 13µl final volume. The reaction was incubated at 65°C for five minutes, and chilled quickly on ice. The volume was increased to 20µl with the addition of 1X First Strand Buffer (50mM Tris-HCl pH 8.3, 75mM KCl, 3mM MgCl₂), 5mM DTT, 40U RNaseOUT and 200U SuperscriptIII Reverse Transcriptase. Reaction tubes were incubated at 50°C for 55 minutes, after

which the enzyme was inactivated at 70°C for 15 minutes. cDNA was used immediately for a PCR reaction or stored at -20°C until required.

Differential display relies on RT-PCR amplification of mRNA utilizing as many combinations of random primers as possible. Reactions involving asexual and gametocyte stage parasite RNA were performed separately, and tubes kept apart at all times to prevent contamination of samples. Three random 10-mers and eight random 13-mers were used with each of the four cDNA samples for PCR reactions. Although these were arbitrary primers, they were designed in such a manner as to accommodate the high AT content of the *P. falciparum* genome. The 2-base anchored primer corresponding was used as the downstream primer. A 25µl PCR included 0.2mM primers, 0.5µl cDNA, 2µM dNTPs, 1.5mM MgCl₂, 1X PCR buffer (50mM KCl and 10mM Tris-HCl pH 8.3), 2µCi α-³²P(dATP) (800Ci/mmol, 10µCi/µl) and 0.625U Takara *Taq* Polymerase. All primer sequences are listed in A-3.

Reactions were assembled on ice and placed immediately into an Eppendorf Gradient thermal cycler preheated to 94°C. The PCR program for the 10 base primers comprised one cycle of 94°C for one minute, 30°C for five minutes, and 72°C for five minutes, followed by 40 cycles of denaturation at 94°C for one minute, annealing at 35°C for 30 seconds and extension at 72°C for one minute. The first low-temperature cycle allows the arbitrary 5' primer to anneal as often as it can, and stringency is then increased to reduce non-specific annealing and hence limit false positives on the differential display gel. For the 13-base primers the annealing temperatures of the two stages were increased to 35°C and 40°C respectively.

2.2.3.2 Electrophoresis of RT-PCR products

3µl PCR reaction were mixed with 2µl stop solution and denatured at 90°C for five minutes prior to loading onto a 6% acrylamide sequencing gel (A-2) pre-heated to approximately 46°C by electrophoresis for 45 minutes. A 6% acrylamide gel was

used to facilitate elution of PCR products. cDNA was electrophoresed for four to six hours at 60W with current and voltage set to be non-limiting, and the gel was exposed to X-ray film overnight.

2.2.3.3 Identification of differentially expressed cDNAs

The X-ray film was developed in a Kodak X-Omat film processor in the casualty ward of the Johannesburg General Hospital, South Africa. The incorporation of α -³²P d(ATP) into the PCR reaction allowed the visualisation of PCR products, and thus the comparison of gene expression between the two parasite stages was possible. Asexual parasite and gametocyte expression profiles were compared visually and fragments that were present in only one stage were identified as potential differentially expressed genes.

2.2.3.4 Excision and elution of differentially expressed fragments

A critical step in the differential display process is the accurate excision of the identified cDNA fragments, which requires precise realignment of the autoradiograph with the acrylamide gel. A second exposure of the gel was made and the bands representing genes of interest were excised from the autoradiograph. This second autoradiograph was realigned with the gel, and PCR products were excised with a sterile scalpel by cutting the acrylamide gel through the holes that had been created in the autoradiograph. Excised fragments were sliced into small pieces, placed in 30 μ l water and centrifuged through a pre-soaked 200 μ l pipette filter tip at 15000*g* for ten minutes in a Sorvall RMC-14 centrifuge. A further exposure of the gel was then made to verify that the fragments had been accurately excised.

2.2.3.5 Reamplification of eluted products

Reamplification was performed with the same primer combinations as had been used for the original PCR reaction. The PCR reaction was set up as described (2.2.3.1) but

the dNTP concentration was increased to 70 μ M and the radio-label was omitted. 10 μ l of eluate was used for an initial 100 μ l PCR (designated PCR1), 1 μ l of which was then used as a template for a second round of PCR. PCR2 was set up in triplicate, and the reactions were pooled and precipitated with ethanol, and the pellet resuspended in final volume of 30 μ l water. 2 μ l of this was assessed on a 1% agarose gel in the presence of 10 μ g/ml ethidium bromide.

2.2.4 Subcloning of PCR products

2.2.4.1 Preparation of PCR products and vector for subcloning

Reamplified products were subcloned into the pGEM-3Z plasmid vector. Phosphorylation and blunt-ending reactions were performed in a single tube. For ligation with the dephosphorylated vector, a 5' phosphate group was added to the PCR product by combining PCR product, 1mM ATP, 200 μ M dNTPs, 1X T4 polynucleotide kinase buffer (70mM Tris-HCl pH 7.6, 10mM MgCl₂, 5mM DTT) and 10U T4 polynucleotide kinase in a 50 μ l reaction volume. This was incubated at 37°C for 60 minutes. Thereafter, the additional adenine that *Taq* polymerase tends to add at the 3' end during PCR amplification was removed by the addition of 2 μ l (2U) T4 DNA polymerase to the reaction tubes. The reaction was allowed to continue at 37°C for five minutes after which the enzyme was inactivated at 75°C for 15 minutes.

The DNA was purified by phenol: chloroform extraction and ethanol precipitation (A-2) and the pellet was resuspended in 10 μ l 1X DNA dilution buffer supplied with the Rapid DNA Ligation Kit.

The pGEM3Z vector was prepared for subcloning by linearising it with the restriction enzyme *HincII* (Roche, Germany). 1 μ g vector, 2 μ l 1X Buffer M (10mM Tris-HCl pH 7.5, 10mM MgCl₂, 50mM NaCl, 1mM DTT) (Roche, Germany) and 2 μ l (6-20U) *HincII* were combined with water in a 20 μ l reaction volume. The reaction was incubated at 37°C for four hours. The enzyme created a blunt ended linear vector

making it compatible for ligation with the PCR product. The vector was dephosphorylated to prevent self-ligation by adding 1U Calf Intestinal Phosphatase to the above reaction and incubating it at 50°C for 60 minutes. The sample was extracted with phenol/chloroform and purified by ethanol precipitation. The dephosphorylated vector was resuspended in 10µl 1 X DNA dilution buffer from the Rapid DNA ligation kit and stored at -20°C until required.

2.2.4.2 Ligation of vector and DNA

Aliquots of the dephosphorylated, linearised vector and the phosphorylated blunt-ended PCR product were electrophoresed on a 1% agarose gel alongside 2µl of DNA mass ladder to estimate the yield of DNA. The molar ratio of vector DNA to insert DNA was calculated such that a molar ratio of 1:2 was used for the ligation reaction. In addition the total amount of DNA in the reaction was kept below 200ng.

Ligation of the vector and insert DNA was carried out using the Rapid DNA Ligation Kit. 9µl DNA (vector + insert) in 1X DNA dilution buffer, 10µl 2X T4 DNA ligation buffer and 1µl T4 DNA ligase were combined in a 1.5ml Eppendorf tube, and the reaction incubated at room temperature for 30 minutes. A mock ligation reaction, containing all the reagents except the DNA, was prepared to check for self-ligation of the vector. Linearised vector is not taken up by competent cells during the transformation and thus no colonies should be observed on these plates.

2.2.4.3 Transformation of competent cells

Subcloning efficiency DH5α competent cells (stored at -70°C in 50µl aliquots) were thawed on ice and transformed according to the manufacturer's instructions. In brief, 5µl of ligation mix was added to 50µl of ice-cold cells in 1.5ml Eppendorf tubes and gently mixed. The tubes were left on ice for 30 minutes, and then heat-shocked at 37°C for 20 seconds. The tubes were incubated on ice for a further two minutes.

500µl of Luria Broth (LB) medium (A-1) were added, and the cells incubated at 37°C for 60 minutes in an environmental shaker at 250rpm. 50µl and 450µl aliquots of these transformed cells were spread onto LB agar plates (A-1) containing 100µg/ml ampicillin, and incubated at 37°C overnight.

The following day the plates were removed from the incubator and stored at 4°C. Four to six discrete colonies were transferred from the agar plates to 10ml growth tubes containing 2ml LB medium with 100µg/ml ampicillin, using a sterile loop or pipette tip. These tubes were incubated overnight at 37°C in an environmental shaker at 250rpm.

2.2.4.4 Purification of plasmid DNA

Glycerol stocks of the individual colonies were prepared for permanent storage at -70°C by adding 500µl of cells from the overnight cultures to 500µl 60% sterile glycerol. Plasmid DNA was liberated from the *E. coli* host cells by SDS/alkaline lysis followed by precipitation with ethanol (A-2).

The presence of plasmid inserts was verified by digesting 10µl of the plasmid DNA with 10U *Eco*R1 and 10U *Hind*III (Roche, Germany) which cut the vector adjacent to the cloning site. The reaction was performed in 1X buffer B (10mM Tris-HCl pH 8.0, 5mM MgCl₂, 100mM NaCl, 1mM β-mercaptoethanol) at 37°C for four hours. The digested products were electrophoresed on a 1% agarose gel in 1X TAE at 90V for 40 minutes (A-1).

2.2.5 DNA sequencing

DNA sequencing was performed using the Sequenase[®] 2.0 sequencing kit as per manufacturer's instructions. All reactions were performed in 1.5ml Eppendorf tubes, which were kept closed to prevent evaporation of the small volumes.

Two methods were used to prepare the DNA for sequencing depending on whether the template was plasmid DNA or PCR product.

Plasmids containing inserts were denatured in preparation for sequencing by the addition of 2µl 2M NaOH, 2mM EDTA to 20µl plasmid DNA followed by incubation at room temperature for five minutes. 3µl 3M sodium acetate (pH 5.0) and 7µl water were added and mixed thoroughly to neutralise the reaction. 75µl ice cold 100% ethanol was added to precipitate the DNA at -70°C for ten minutes, after which it was centrifuged for ten minutes. The precipitate was washed with 200µl 70% ethanol for five minutes, allowed to air dry, and then resuspended in 7µl water. Plasmid DNA and the T7 or Sp6 RNA polymerase sequencing primers were annealed by heating 1µM primer, 1X Sequenase® buffer and 7µl plasmid DNA at 65°C for two minutes and then cooling slowly to room temperature. The tubes were stored on ice and used for the sequencing reaction within one hour.

The PCR products were prepared for sequencing using a PCR product pre-sequencing kit. The kit contained two enzymes: Shrimp Alkaline Phosphatase (SAP), which dephosphorylates any residual dNTPs from the PCR reaction and exonuclease I (exo), which removes remaining single strand primers. 7µl PCR product was incubated at 37°C for 15 minutes with 1U SAP and 1U exo, and then heated to 80°C for 15 minutes to inactivate the enzymes. 5-10pmol of the same primer that had been used as the upstream primer in the original PCR reaction was added to the pre-treated PCR product, boiled for five minutes, and then snap-cooled in an ice water bath which promotes primer annealing and prevents renaturation of PCR template.

2.2.6 Reverse slot blot verification of differentially expressed cDNAs

One of the accepted vagaries of the differential display technique is the number of false positives that can be generated. This propensity for false positives requires

verification of the expression profile of the eluted fragments. A novel and efficient method for this purpose is a reverse slot blot (Meyersfeld and Coetzer, 2003).

RNA from new asexual parasite and gametocyte cultures was reverse transcribed as described (2.2.3.1). cDNA was amplified with the same arbitrary primers as had been used previously and α -³²P-labelled RT-PCR products were pooled to generate asexual and gametocyte stage probes.

For the slot blot, 2 μ l of either plasmid DNA or reamplified differentially expressed cDNA fragments from the differential display gel (2.2.3.5) were added to 198 μ l denaturing solution (25mM EDTA, 0.4M NaOH). This was heated at 95°C for two minutes and stored on ice until required.

A Biorad slot blot apparatus was assembled using ten pieces of damp Whatman no. 1 filter paper overlaid with pre-soaked Hybond™-N nylon membrane. A vacuum was applied until the membrane was almost dry, the denatured DNA was added to the slot apparatus and the vacuum was again applied. The slot blot was washed with 400 μ l neutralisation solution (1.5M NaCl, 0.5M Tris-HCl pH 7.2, 1mM EDTA) and the vacuum applied to draw all solutions through the membrane. The membrane was then dried in an 80°C oven for ten minutes, wrapped in Gladwrap and placed DNA-side down on a UV transilluminator at 302nm for four minutes to crosslink the DNA. The blot was prepared in duplicate so that the two membranes could be probed separately with asexual and gametocyte stage RNA. The pGEM-3Z vector was spotted onto the membrane to be used as a negative control.

Membranes were inserted into 100ml Hybaid hybridisation tubes with approximately 5ml hybridisation solution (5X SSC, 5X Denhardts solution, 0.5% SDS) and incubated at 42°C for 40 minutes with rotation in a Hybaid incubator. 50X Denhardts solution consisted of 1% (w/v) bovine serum albumin, 1% (w/v) Ficoll™ and 1% (w/v) polyvinylpyrrolidone; 5X SSC comprised 0.75M NaCl, 75mM Na-citrate pH 7.0.

α -³²P –labelled probes were denatured at 95°C for five minutes and added to the hybridisation tubes. These were incubated overnight at 42°C, following which the blots were washed twice with 2X SSC at room temperature for 15 minutes, shaking vigorously. Blots were exposed to X-ray film for 40 minutes with intensifying screens. If further washes were required to reduce background radioactivity, stringency was increased by performing washes at 65°C, and using 0.5X SSC/ 0.1% SDS as a washing agent. This was repeated until background had been eliminated. The intensity of the bands on the autoradiographs was quantitated using Syngene GeneTools version 2.20.02. An initial overnight exposure was performed and this was increased as required.

2.2.7 Bioinformatic Analysis of Differentially Expressed Transcripts

2.2.7.1 Transcript Identification

Transcript sequences were analysed by comparison with the PlasmoDB (<http://plasmodb.org>) and TIGR (www.tigr.org/tdb/tgi/pfgi/) databases. Where discrepancies existed between the two databases the match with the better Expectation (E) value was selected. A combination of nucleotide-nucleotide (BLASTN) and nucleotide-protein (BLASTX) similarity searches (Altschul et al., 1997) were performed to increase the confidence with which a gene identity was assigned. Transcripts were classified into categories on the basis of the E values of the first and second BLASTN and BLASTX matches. Transcripts that were members of multi-gene families could also be identified in this manner. For a transcript identity to be assigned, a threshold E value was arbitrarily set at 10⁻⁵ after inspection of the E values of all matches. Protein analysis of the identified transcripts was obtained from the ExPASy Proteomics Server (<http://us.expasy.org/>) as well as geneDB (www.genedb.org/). The clustering analysis of Le Roch et al. (2003) was downloaded from <http://carrier.gnf.org/publications/CellCycle/>.

2.2.7.2 Transcript Expression Profile

PlasmoDB contains expression profiles for almost every gene in the *Plasmodium* genome, generated using a high-density oligonucleotide array (Le Roch et al., 2003). mRNA expression values of *P. falciparum* genes for nine different life cycle stages from 17 microarray experiments were obtained from the RNA Abundance Database (<http://www.cbil.upenn.edu/RAD/>) (Manduchi et al., 2004) and compared with data obtained from differential display and reverse slot blot.

2.3 Results

2.3.1 Plasmodium falciparum cultures

2.3.1.1 Asexual parasites

Asexual parasites were maintained in culture as described (2.2.1.4). Figure 3 shows a thin blood smear of mixed asexual parasites, representative of the stages the culture flask contained when parasites were harvested for RNA extractions.

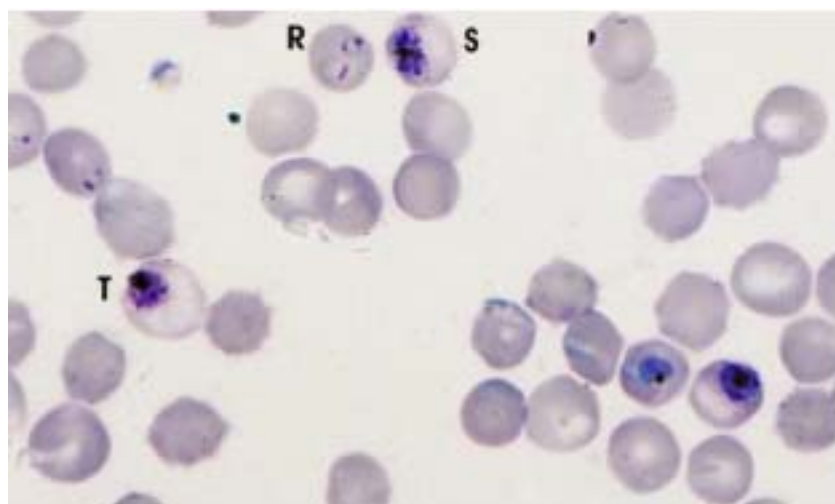


Figure 3-Asexual parasites in culture

Thin blood smear stained with Giemsa, showing mixed asexual *P. falciparum* erythrocytic stage parasites at 1000x magnification. Arrows indicate late trophozoite (T), ring (R) and schizont (S) stages.

2.3.1.2 Gametocytes

The 3D7 strain of parasite was used in this study because of its ability to form gametocytes in culture when exposed to stressful conditions (2.2.1.6). Early stage gametocytes were visible within eight days of the cultures being initiated, but cultures were continued until at least day 14, by which stage thin blood smears indicated a

gametocytaemia of between 1-2%. Figure 4 shows thin blood smears of gametocytes at different times of culture.

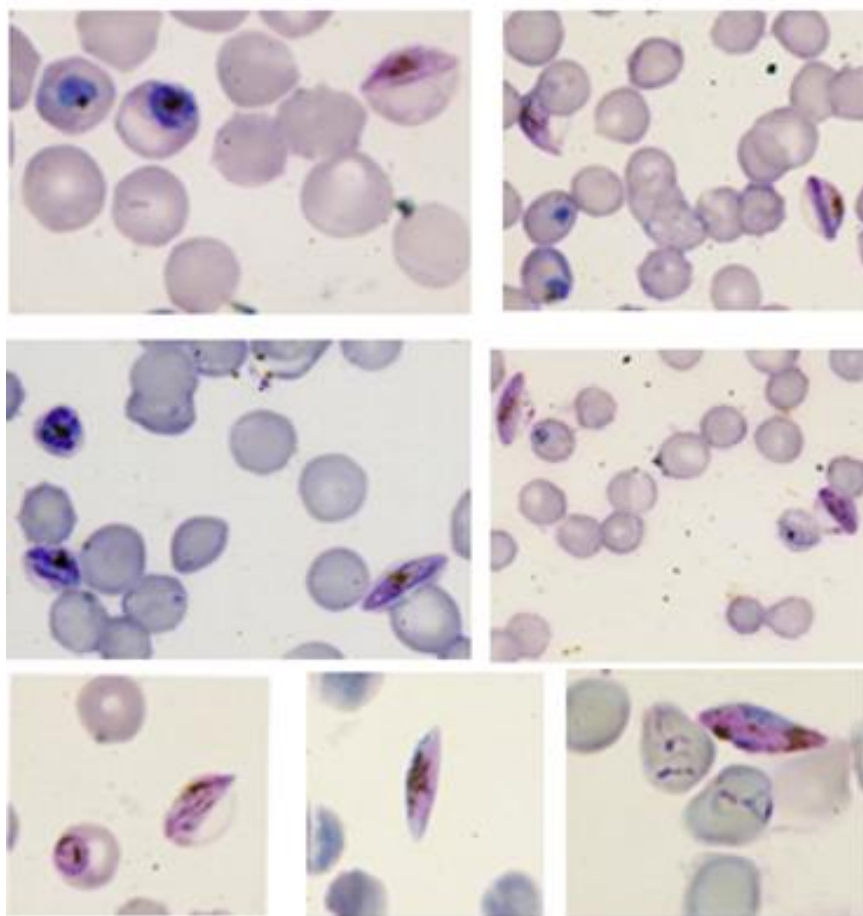


Figure 4- Gametocytes in culture

Thin blood smears made between 10 and 12 days after initiation of gametocyte cultures, showing stage III and IV gametocytes under 1000x magnification.

Gametocytes were purified from residual asexual parasites by centrifugation through a Percoll density gradient (Figure 5).

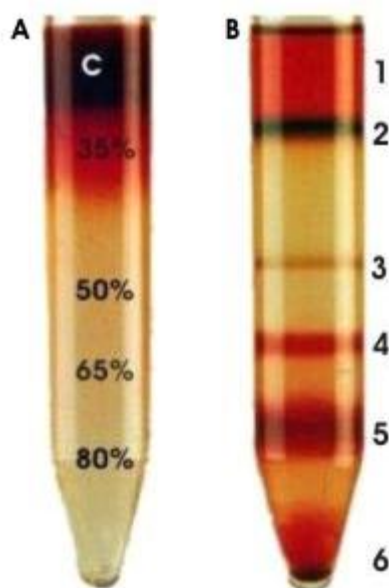


Figure 5- Percoll gradient purification of gametocytes

Gametocyte cultures before (A) and after (B) centrifugation through a Percoll density gradient. The fractions contained predominantly 1) lysed erythrocytes and extra-erythrocytic asexual parasites, 2) and 3) debris; 4) debris and gametocytes, 5) gametocytes and unparasitised erythrocytes and 6) residual asexual parasites and unparasitised erythrocytes. “Debris” refers to haemozoin and lysed erythrocyte fragments. C: Gametocyte culture layered on top of the Percoll gradient.

Layers 4) and 5) were collected and pooled as described (2.2.1.7). Centrifugation of the parasite cultures through the Percoll density gradient resulted in pure gametocytes, devoid of any contaminating asexual parasites.

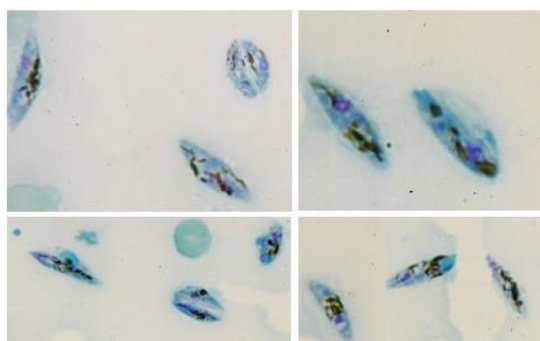


Figure 6- Purified gametocytes after Percoll gradient centrifugation

These images, taken from the same slide, show predominantly stage III and IV gametocytes after purification through a Percoll density gradient.

Despite repeated attempts with numerous combinations of Percoll densities, the yield of gametocytes purified by this method remained low. An alternate method, as described in 2.2.1.6, involved lysing erythrocytes that contained asexual stage parasites from day 12 with two rounds of D-sorbitol treatment 24 hours apart. D-sorbitol has no effect on ring stages or gametocytes. Thus the initial treatment would eliminate trophozoites but leave the early rings intact. These would be removed by the second treatment by which time they would have matured to trophozoites. Erythrocytes containing gametocytes were then collected by centrifugation, whilst dead asexual parasites remained in the supernatant. The yield of gametocytes increased considerably utilising this method.

2.3.2 RNA Extractions

Maximum RNA yield from asexual parasites was obtained by harvesting cultures containing predominantly late stages. Yield varied according to the method used. Typically, 80-120µg RNA was obtained from a 40ml culture at a haematocrit of 5% and a parasitaemia of approximately 12% using TRI Reagent (2.2.2.3). The integrity and purity of the RNA were assessed by agarose gel electrophoresis (Figure 7) and spectrophotometry respectively. The 28S (± 4000 b) and 18S (± 2100 b) ribosomal RNA subunits migrate at approximately 1100bp and 800bp relative to the double-stranded DNA size marker (Figure 7). Messenger RNA (mRNA) is sometimes visible as a faint smear between these two ribosomal subunits.

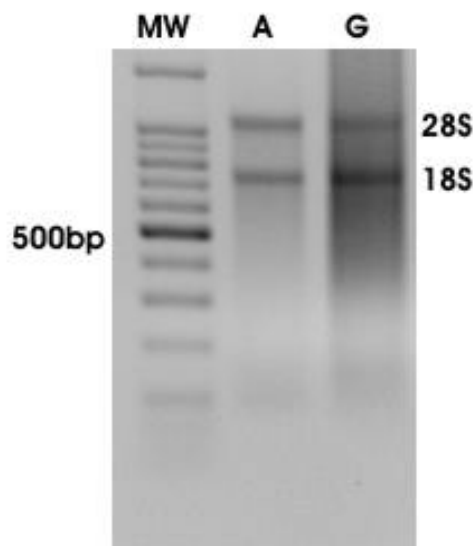


Figure 7- RNA analysis

RNA from mixed asexual stage parasites (**A**) and gametocytes (**G**) was electrophoresed on a 1% agarose gel. The two bands in the RNA lanes represent the 28S and 18S ribosomal RNA subunits.

Spectrophotometric analysis of isolated RNA gave $A_{260}:A_{280}$ ratios that ranged between 1.8 and 2.0 which is within the prescribed range of a pure RNA preparation. An indication of good quality RNA is that the approximate intensity of the 28S ribosomal subunit should be twice that of the 18S subunit. In the gametocyte RNA shown in Figure 7 the 18S subunit is more intense indicating that some degradation has occurred. Using TRI reagent to extract RNA from 160ml of gametocyte culture at a gametocytaemia of approximately 2% yielded 100-150 μ g RNA.

2.3.3 Differential Display

2.3.3.1 Identification and reamplification of differentially expressed cDNAs

Differential display was used to compare gene expression between predominantly late stage asexual parasites and gametocytes of *P. falciparum*. In each experiment, RT-PCR products from gametocyte cultures were compared with a minimum of two

different asexual stage RNA extractions, due to the higher degree of variation expected between the asexual stage parasite cultures. To reduce the number of false positives in a technique known to be error-prone, only gene fragments unique to a developmental stage were chosen for further analysis (Figure 8).

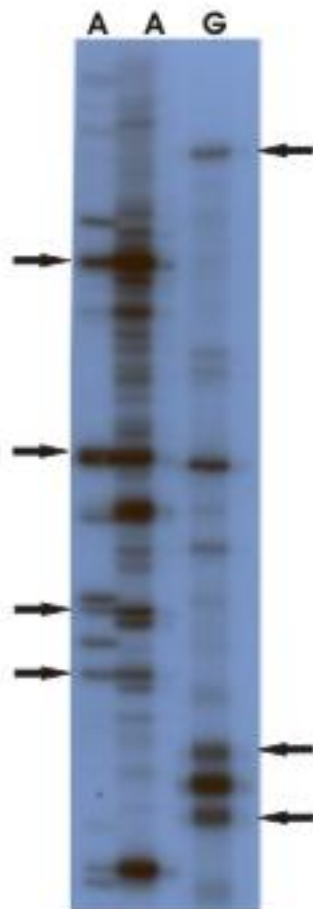


Figure 8- Differential display analysis of *P. falciparum* gene expression

Autoradiograph of α - ^{32} P-labelled RT-PCR products separated on a 6% acrylamide gel. Arrows indicate some of the bands that would have been identified as differentially expressed. A: asexual; G: gametocyte.

A total of 55 differentially expressed fragments were excised from five differential display gels encompassing 44 arbitrary primer combinations. Twenty-one of these transcripts were expressed uniquely in gametocytes, whilst the remaining 34 were expressed in asexual stages. 37 transcripts were successfully reamplified (2.2.3.5) using two rounds of PCR, a success rate of 67%. Elution of small quantities of DNA

from the acrylamide gel probably accounted for the low amplification rate. Reamplified eluates were analysed by agarose gel electrophoresis (Figure 9).

Agarose gel analysis occasionally showed more than one product in the reamplification (eluates 48 and 51 in Figure 9), arising from either contamination or co-migration of similarly sized bands in the original differential display gel. Transcripts such as these were discarded as candidates for sequencing directly from the PCR product. Even when the amplification appeared to be specific, often the direct sequencing was of a poor quality due to contaminating bands that were not resolved on the agarose gel, or the fact that short (13mer) sequencing primers were used. In total, six transcripts were directly sequenced whilst 31 were subcloned into the pGEM3Z vector to facilitate sequencing (2.2.4, 2.2.5).

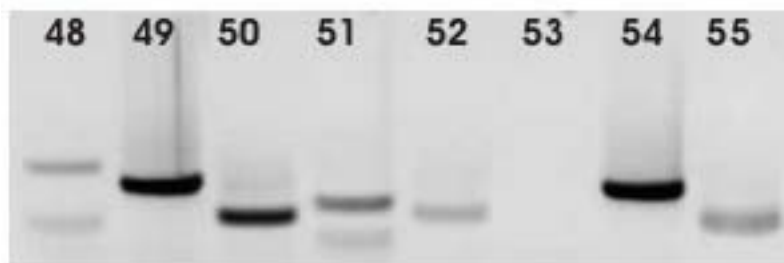


Figure 9- Reamplification of differential display fragments

1% agarose gel showing some of the fragments reamplified from a differential display gel. Two bands are evident in eluates **48** and **51**. The lack of product in eluate **53** was either failed PCR or insufficient PCR template due to inefficient elution of the band from the acrylamide gel. Amplified products ranged in size from 200-600bp.

2.3.3.2 Subcloning and sequencing of reamplified fragments

Multiple colonies were selected from overnight incubations of transformed DH5 α cells, and the presence of insert verified by restriction digest analysis of plasmid DNA (Figure 10). All colonies from a plate would have contained the same size

insert if only one fragment had been excised from the acrylamide gel. In some instances, however, either through the introduction of a contaminant or because similarly sized bands had co-migrated on the acrylamide gel, inserts of different sizes were identified on the agarose gel (e.g. E23). Either of these could have represented a differentially expressed fragment, and so plasmids from both colonies were sequenced.

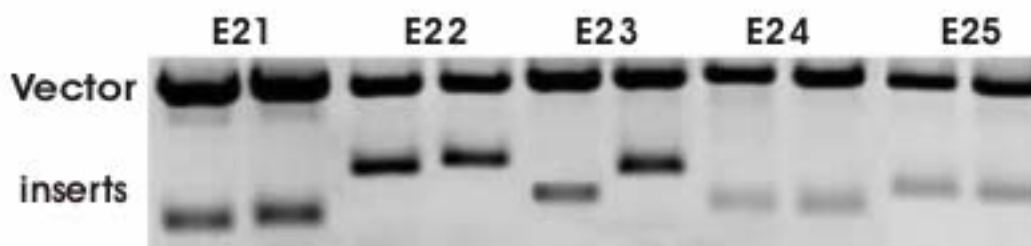


Figure 10- Verification of plasmid inserts after transformation of DH5 α cells

The presence of plasmid inserts was verified on a 1% agarose gel following digestion of plasmid DNA with *Eco*R1 and *Hind*III. Multiple colonies from each plate were analysed. The lanes representing E23 contain inserts of different sizes.

Inserts were sequenced using the Sp6 and T7 sequencing primers, which anneal adjacent to the multiple cloning site of the vector. The subcloning of 31 reamplified fragments generated 35 unique sequences since four clones contained inserts of different sizes. These, combined with six transcripts that were sequenced directly from PCR products, generated a total of 41 unique sequences.

2.3.3.3 Identification and analysis of differentially expressed transcripts

Sequences were submitted to the *P. falciparum* genome databases (PlasmoDB and TIGR) which use the BLAST algorithm to search for homologous sequences within the genome. The length of the sequences submitted to the database ranged from 66-232 nucleotides. Sequence identity with the database was not always 100%, due to errors that could have been introduced by *Taq* polymerase during the multiple rounds

of PCR, as well as sequencing ambiguities. Where a discrete band on the sequencing gel could not be discerned an “N” was inserted into the sequence to represent all of the four nucleotides.

The Expectation (E) value is a reference value which indicates the probability of the query sequence being present in the genome by chance and thus apportions a statistical significance to any sequence match. 27 sequences (A-4) had a match to the *P. falciparum* genome with an E value below the set threshold of 10^{-5} . The threshold value was arbitrarily assigned after inspection of the E values of all matches. The gene source and the stage in which the genes were expressed are presented in Table 1. Of the remaining 14 sequences, 12 had nucleotide homology to sequences within the human genome and two matched the *Plasmodium* database but had E values above the set threshold.

Table 1-Transcripts with nucleotide homology to the PlasmoDB database

Eluate	Stage	Gene Locus	Gene Product	Nucleotide Identity	E Value
E2-3	Asexual	PFF0435w	Ornithine aminotransferase	87%	6e-09
E2-4	Asexual	PFD0625c	<i>Pf</i> EMP1	83%	2e-06
E4-2	Asexual	MAL7P1.166	Regulator of nonsense transcripts	91%	5e-25
E5-1	Asexual	chr5.rRNA-1-28s-A	rRNA	96%	1e-40
E 6-8	Asexual	PF10_0121	HPRT	92%	8e-18
E9-1	Asexual	PFL2665c	<i>Pf</i> EMP1	85% ^T	2e-19
E 10	Gametocyte	PF11_0361	Hypothetical protein	98%	3e-09
E12-2	Gametocyte	PF14_0338	Hypothetical protein	93%	5e-26
E16-3	Gametocyte	MAL13P1.226	Hypothetical protein	92%	1e-07
E17-1	Asexual	PF14_0124	Actin II	97%	3e-38
E18-1 ^T	Asexual	NP660290	<i>Pf</i> EMP1	93%	5e-11
E19-1	Asexual	chr7.rRNA-1-28s	rRNA	88%	2e-25
E23-1	Asexual	MAL13P1.70	Hypothetical protein	100%	2e-09
E25-2	Asexual	chr7.rRNA-1-28s	rRNA	96%	2e-25
E26-2	Asexual	PF07_0050	<i>Pf</i> EMP1	86%	2e-21
E27-2	Gametocyte	PFD0020c	<i>Pf</i> EMP1	96%	8e-17
E28-1 ^T	Asexual	NP660290	<i>Pf</i> EMP1	84%	3e-22
E37-6	Asexual	PFI0210c	Hypothetical protein	93%	3e-13
E39-1	Asexual	PFI1475w	MSP-1	100%	1e-24
E39-2	Asexual	PFD1120c	Integral membrane protein	92%	8e-16
E39-3	Asexual	PFF0745c	Putative ribonuclease	93%	4e-12
E42	Asexual	chr5.rRNA-1-28s-A	rRNA	98%	3e-27
E43-1	Asexual	PF14_0076	Plasmepsin 1 precursor	97%	1e-25
E43-2	Asexual	PFE0925c	snRNP	94%	2e-17
E51-3	Asexual	PF10_0025	PF70 protein	87%	2e-14
E52-1	Asexual	chr7.rRNA-1-28s	rRNA	100%	3e-13
E54-4	Gametocyte	PF10_0244	Hypothetical protein	100%	4e-10

^T TIGR database

Differentially expressed transcripts were identified using a BLASTN sequence similarity search on PlasmoDB. The gene identity is that of the best (first) match using only a BLASTN search. Data were obtained from the PlasmoDB version 4.3 accessed on 2 February 2005.

To increase the confidence with which the gene source could be allocated to a transcript it is necessary to consider both nucleotide and protein sequence similarity searches. Gene annotations from Table 1 were classified into categories based on the best matches from BLASTN and BLASTX similarity searches, as well as the E-values of the first and second best matches. These criteria enabled a more stringent appraisal of the value of the gene annotation when compared to identifying a significant best match from a single similarity search. The criteria for the categories are presented in Table 2 and the categorised transcripts are shown in Table 3.

Table 2-Criteria for the classification of transcripts using sequence similarity search outputs

Transcript Category	1 st Match	First Match E Value		Second Match E value	
	BLASTN and BLASTX	BLASTN	BLASTX	BLASTN	BLASTX
A	Identical	$\leq T$	$\leq T$	$> T$	$> T$
B	Identical	$\leq T$	$\leq T$	$\leq T$	$\leq T$
C	Identical	$\leq T$	$> T$		
D	Non-identical rRNA gene	$\leq T$			

Threshold (T) was set at 10e-05. Category A contains transcripts whose gene source could be confidently assigned. Members of gene families were placed into category B, where BLASTN and BLASTX first and second matches were both below threshold. Category C genes were similar to those in category A but for the fact that their BLASTX matches were above threshold. Category D contained those transcripts identified as rRNA genes.

Transcripts whose gene source could be confidently ascertained were placed into category A (Table 3). Genes of known or predicted function in category A are PFI1475w (merozoite surface protein 1, precursor), PFD1120c (integral membrane protein), PFF0745C (putative ribonuclease), PF10_0121 (hypoxanthine phosphoribosyltransferase), PF10_0025 (PF70 protein) and PFE0925c (small nuclear

ribonucleoprotein). The remaining transcripts were identified as hypothetical proteins.

Transcripts in category B (Table 3) had both BLASTN and BLASTX first and second match E-values below the 10^{-5} threshold, making it difficult to ascertain the specific gene source. Gene families in category B are *PfEMP1* (PFD0020c, PFL2665c, PF07_0050), actin (PF14_0124) and plasmepsin (PF14_0076).

Five gene transcripts in this category were members of the *PfEMP1* family, four of which were found to be expressed maximally in asexual stages. Because the specific gene source cannot be defined, it is difficult to assign an expression profile to these transcripts. Thus, even though a tentative gene assignment has been made on the basis of the best BLASTN and BLASTX match, the transcript may actually be representing the second best match. Despite having a BLASTX second match greater than threshold, transcript E-18 was placed into category B after the remaining matches all identified it as a member of the *PfEMP1* family.

Inspection of the BLASTN alignment of transcript E4-2 to the genome showed that the plus strand encoded an erythrocyte membrane-associated antigen (MAL7P1.12) whilst the minus stranded coded for a regulator of nonsense transcripts (MAL7P1.166). Thus the two DNA strands were coding for two unrelated genes, either of which could have represented the differentially expressed transcript. These two genes are not necessarily members of gene families, but category B was still the most appropriate classification.

Table 3-Classification of transcripts identified by differential display

Eluate	First Match				Second Match ^a			
	BLASTN		BLASTX		BLASTN		BLASTX	
	Gene Locus	E-value	Gene locus	E-value	Gene Locus	E-value	Gene Locus	E-value
A – Single Genes								
E39-1	PFI1475w	1e-24	PFI1475w	2e-18	PF11_0317	0.035	PF11_0178	0.21
E39-2	PFD1120c	8e-16	PFD1120c	1e-11	MAL13P1.19	0.997	NM	
E23-1	MAL13P1.70	2e-09	MAL13P1.70	3e-05	NM		NM	
E39-3	PFF0745c	4e-12	PFF0745c	5e-06	PF13_0073	0.70	NM	
E37-6	PFI0210c	3e-13	PFI0210c	7e-08	PFL2300w	0.67	NM	NM
E6-8	PF10_0121	8e-18	PF10_0121	6e-19	PF14_0192	0.008	PF07_0104	1.0
E51-3	PF10_0025	2e-14	PF10_0025	9e-09	NM		NM	
E43-2	PFE0925c	2e-17	PFE0925c	2e-11	NM		PF08_0096	0.002
E10	PF11_0361	3e-09	PF11_0361	9e-06	PF10_0032	0.52	NM	
E12-2	PF14_0338	5e-26	PF14_0338	2e-14	NM		NM	
E54-4	PF10_0244	4e-10	PF10_0244	9e-07	NM		NM	
B – Gene families								
E27-2	PFD0020c	8e-17	PFD0020c	7e-11	MAL7P1.50	2e-05	PFL1950w	1e-06
E9-1	PFL2665c	2e-19	PFL2665c	5e-17	PFA0765c	4e-11	PFD0005w	1e-15
E26-2	PF07_0050	2e-21	PF07_0050	2e-16	MAL7P1.1	9e-11	PF08_0141	1e-11
E17-1	PF14_0124	2e-38	PF14_0124	4e-29	PFL2215w	5e-28	PFL2215w	3e-25
E18-1 [†]	NP660290	5e-11	NP660290	10e-07	PFC0005w	4e-05	NP660287	3e-03
E43-1	PF14_0076	1e-25	PF14_0076	3e-14	PF14_0078	2e-15	PF14_0077	4e-10
E4-2	MAL7P1.12	1e-24	MAL7P1.12	6e-08	MAL7P1.166	1e-24	MAL7P1.166	7e-07
C – Single genes								
E16-3	MAL13P1.226	1e-07	MAL13P1.226	0.004	NM		NM	
E2-3	PFF0435w	6e-09	NM	0.01	NM		NM	
E2-4	PFD0625c	2e-06	PFD0625c	0.011	NM		NM	
E28-1 [†]	NP660290	3e-22	NP660290	7e-03	TC12934	2e-05	NM	
D – Ribosomal RNA gene families								
E5-1	chr5.rRNA-1-28s-A	1e-40	PF11_0501	5e-23	chr1.rRNA-1-28s	3e-34	MAL13P1.455	5e-23
E25-2	chr5.rRNA-1-28s-A	2e-25	PF11_0501	0.045	chr1.rRNA-1-28s	1e-18	MAL13P1.455	0.045
E42	chr7.rRNA-1-28s	3e-27	PF11_0500	2e-06	chr5.rRNA-1-28s-A	4e-27	MAL13P1.450	2e-06
E19-1	chr7.rRNA-1-28s	3e-13	NM		chr5.rRNA-1-28s-A	6e-13	NM	
E52	chr7.rRNA-1-28s	2e-25	NM		chr5.rRNA-1-28s-A	3e-25	NM	

^a NM; no match to the database. [†] TIGR database

The 27 differentially expressed transcripts with sequence homology to the PlasmoDB database were categorised according to the criteria presented in Table 2. **w**: Watson (antisense) DNA strand; **c**: Crick (sense) DNA strand

The transcripts assigned to category C (Table 3) had BLASTN and BLASTX first match E-values below and above the threshold respectively. For transcript E28-1 this could be explained by the fact that the sequence contained a high number of unreadable nucleotides thus reducing the E value of the BLASTX match, whilst for E16-3 the BLASTX value is reduced because the sequence overlaps with an intronic region. Intronic sequence can originate either from DNA contamination or RNA that has not yet been processed (pre-mRNA). The latter explanation is more likely in this instance considering that RNA was treated with DNase prior to use. This category highlights the manner in which the dual sequence similarity search can prevent erroneous identification of gene transcripts.

Category D has five transcripts identified as rRNA. However, because of the high nucleotide overlap among the ribosomal RNA genes as evidenced by the E-values of BLASTN first and second matches, it was not possible to confidently assign the transcript to a gene locus. The transcripts in category D would be expected to have BLASTX E values above threshold, as they do not code for proteins. Despite this there is some homology to genes coding for hypothetical proteins as evidenced by the BLASTX E values of transcripts E5-1 and E42, indicating that these transcripts could in fact represent hypothetical proteins.

2.3.3.4 Verification of differential display results

The stage of maximum expression of all transcripts identified by differential display was verified using a reverse slot blot (Meyersfeld and Coetzer, 2003). Either reamplified transcripts or plasmid DNA of subcloned transcripts were spotted onto a nylon membrane for hybridisation with ³²P-labelled asexual and gametocyte RT-PCR reactions from new RNA extractions (Figure 11). The intensities of the bands on the autoradiograph were scanned using Syngene GeneTools software.

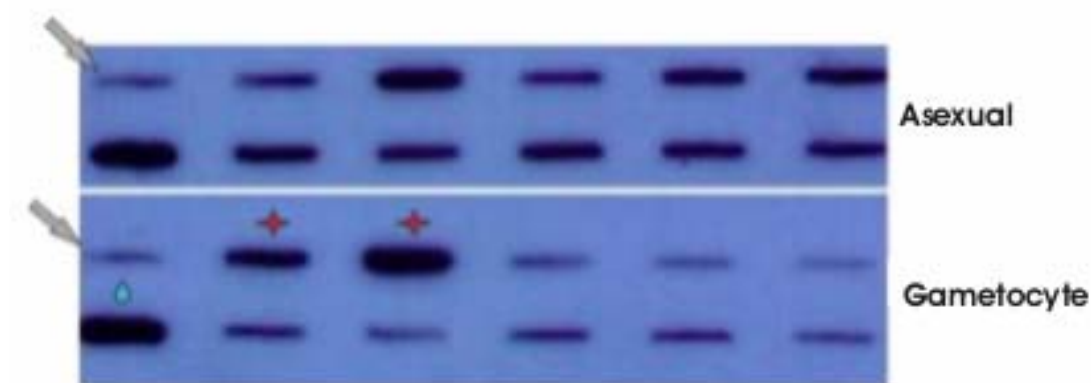


Figure 11- Reverse slot blot for the verification of differential display data

The top of the image represents the membrane probed with ^{32}P -labelled RT-PCR products from asexual stage parasites, whilst the bottom membrane contains the identical fragments, probed with gametocyte ^{32}P -labelled RT-PCR products. The two arrows indicate the pGEM3Z vector which was used as a negative control; the intensity of these bands was taken to be background, and subtracted from the values of all other bands. Bands over-expressed in gametocytes are indicated with a red star; the band marked with a blue circle required a shorter exposure time.

Table 4 illustrates the comparison between differential display data and reverse slot blot results for the confidently identified category A transcripts. Seven of the ten transcripts showed a correlation between the two methods, with discrepancies probably being attributable to variations between parasite cultures at the time of RNA extraction. When the two methods of analysis were in agreement the stage of maximum expression of the transcript could be confidently assigned.

Analysis of the ratio of expression between asexual stage parasites and gametocytes shows that the majority are close to 1, indicating some level of expression in both developmental stages. This points to a gene expression pattern in the parasite whereby a large portion of genes are constitutively active, with up-regulation at specific points in the developmental cycle. Thus more convincing ratios would probably have been obtained had highly synchronous parasite cultures been used.

Table 4-Reverse slot blot verification of differential display data

Eluate	Gene Locus	Description of gene product	Differential Display	Reverse Slot blot	Ratio A:G
E39-1	PFI1475w	MSP-1	Asexual	Asexual	1.2
E39-2	PFD1120c	Integral membrane protein	Asexual	Asexual	1.2
E23-1	MAL13P1.70	Hypothetical protein	Asexual	Asexual	1.3
E39-3	PFF0745c	Ribonuclease, putative	Asexual	Asexual	1.2
E6-8	PF10_0121	HPRT	Asexual	Asexual	2.7
E54-4	PF10_0244	Hypothetical protein	Gametocyte	Gametocyte	0.8
E37-6	PFI0210c	Hypothetical protein	Asexual	Asexual	1.3
E51-3	PF10_0025	PF70 protein	Asexual	Gametocyte	0.7
E43-2	PFE0925c	snRNP	Asexual	Gametocyte	0.8
E12	PF14_0338	Hypothetical protein	Gametocyte	Asexual	3.7

The intensities of the signals on the asexual (A) and gametocyte (G) autoradiographs are expressed as an A:G ratio. A value of one indicates that expression was equivalent; anything less than one represents an over-expression in gametocytes, whilst any value greater than one indicates that the transcript is over-expressed in asexual stage parasites.

A second method of verification was subsequently used to confirm the stage of maximum expression of the transcripts identified by differential display. With the progression of the PlasmoDB database and the availability of microarray data for almost all annotated and predicted genes, the results of the differential display for category A genes were compared with the available microarray data (Table 5). Absolute expression values of the transcripts were obtained from the RNA abundance database (<http://www.cbil.upenn.edu/RAD/>) (Manduchi et al., 2004).

Of the 11 genes in category A whose expression profiles were analysed by microarray, nine were maximally expressed in the stage in which they had been identified by differential display, leaving two discrepancies. The first transcript for which there was disagreement was PFE0925c (E43-2), the small nuclear ribonucleoprotein. PFE0925c was identified in asexual stages in differential display, but reverse slot blot and microarray both showed elevated levels of expression in gametocytes, although the microarray values are not significantly different. Mass spectrometry data available for this protein on PlasmoDB showed evidence for PFE0925c peptide fragments in both asexual parasite and gametocyte stages.

E37-6 (PFI0210c) was the second transcript for which there was a discrepancy in results. It represents the only instance where differential display and reverse slot blot correlated (both showing increased expression in asexual stages), but contradicted microarray data. However, the available microarray data were below the threshold of confidence as there were either too few probes per slide or the expression was too low. Furthermore, mass spectrometry has evidenced peptide fragments of this protein in merozoites confirming that the differential display and reverse slot data are probably correct.

Table 5-Microarray verification of differential display results

Gene Locus		Description of gene product	Differential Display	Microarray	
				Asexual	Gametocyt
E39-1	PFI1475w	MSP-1	Asexual	1677	100
E39-2	PFD1120c	Integral membrane	Asexual	2017	163
E23-1	MAL13P1.70	Hypothetical protein	Asexual	70	3
E39-3	PFF0745c	Ribonuclease,	Asexual	144	42
E6-8	PF10_0121	putative HPRT	Asexual	4583	1199
E51-3	PF10_0025	PF70 protein	Asexual	736	22
E43-2	PFE0925c	snRNP	Asexual	178	217
E10	PF11_0361	Hypothetical protein	Gametocyte	87	458
E37-6	PFI0210c	Hypothetical protein	Asexual	5	56
E12	PF14_0338	Hypothetical protein	Gametocyte	418	454
E54-4	PF10_0244	Hypothetical protein	Gametocyte	337	416

mRNA expression values of *P. falciparum* genes (Le Roch et al., 2003) were obtained from the RNA Abundance Database (<http://www.cbil.upenn.edu/RAD/>). The peak expression between all asexual life stages was compared with the differential display data, and the stage of maximum expression is indicated in bold italics.

Table 6 shows a comparison of the stages of maximum expression as determined by differential display, reverse slot blot and microarrays.

Table 6-Comparison of expression data from all three methods of analysis

Eluate	Gene Locus	Differential Display	Reverse slot blot	Microarray
E39-1	PFI1475w	Asexual	Asexual	Asexual
E39-2	PFD1120c	Asexual	Asexual	Asexual
E23-1	MAL13P1.70	Asexual	Asexual	Asexual
E39-3	PFF0745c	Asexual	Asexual	Asexual
E6-8	PF10_0121	Asexual	Asexual	Asexual
E54-4	PF10_0244	Gametocyte	Gametocyte	Gametocyte
E10	PF11_0361	Gametocyte	ND	Gametocyte
E37-6	PFI0210c	Asexual	Asexual	Gametocyte
E51-3	PF10_0025	Asexual	Gametocyte	Asexual
E43-2	PFE0925c	Asexual	Gametocyte	Gametocyte
E12	PF14_0338	Gametocyte	Asexual	Gametocyte

ND No data

Ultimately the selection criteria that were employed resulted in the identification of only one false positive (E43-2), where differential display did not correlate with either reverse slot blot or microarrays. When reverse slot blot results did not correlate with microarrays or differential display (E12 and E51-3), the discrepancy was attributed to variations between parasite stages of the various cultures from which RNA was extracted, although again the microarray values for E12 are not significantly different in asexual and gametocyte stages. The final permutation, where reverse slot blot and differential display contradicted microarrays, occurred only once (E37-6) and mass spectrometry showed the microarray data to be erroneous. Despite the small sample number, the low proportion of false positives

(9%) was satisfactory considering the reported error rates of the differential display technique (Callard et al., 1994; Luce and Burrows, 1998; Miele et al., 1998).

The data presented here indicate that differential display is still a viable alternative to microarrays for small laboratories wishing to perform genome profiling experiments. The data analysis has been enhanced by the microarray data that have become available but which had not yet been published at the time this research was initiated. Microarray data can serve as an excellent guide as to the accuracy of the differential display results despite the fact that microarrays themselves are not infallible. Thus any discrepancies in results should still be verified using an independent method.

2.4 Discussion

Differential display has been used to compare gene expression between predominantly late stage asexual parasites and predominantly late stage gametocytes of *P. falciparum*. The positive identification of 27 genes that are differentially expressed between these two developmental stages confirms the utility of differential display, both as a genome profiling method in *Plasmodium* and as a viable alternative to microarrays in small laboratories. In this research differential display was preferred to SSH or SAGE due to the rapidity with which results could be generated and the fact that fewer downstream manipulations are required. This represents a small-scale effort to contribute to knowledge of the molecular biology of the malaria parasite.

2.4.1 Identification of differentially expressed genes

A total of 41 transcripts were identified, which is quite a small number given the vast difference in gene expression of parasites only a few hours apart in their development (Bozdech et al., 2003). This number would probably have been larger had highly synchronous asexual parasite cultures been used. The aim of this study, however, was not to identify the maximum number of genes, but rather to compare gene expression between mixed asexual and gametocyte parasite cultures to identify candidate differentially expressed genes that would provide opportunity for further investigation. Numerous genes were identified that play pivotal roles in the structural, metabolic and immune defence properties of the parasite.

The majority of identified genes were from asexual stage parasites. This is similar to what was found by Cui et al. (2001) where almost two thirds of the transcripts came from asexual parasites. Potentially this is due to the greater transcriptional activity of the asexual parasites, mirroring the increased metabolic requirements of this stage.

2.4.2 Analysis of differentially expressed gene transcripts

Sequences were submitted to the PlasmoDB database where the WashU BLAST algorithm was used to search for homologous sequences. Two sequence similarity searches, BLASTN (transcript nucleotide to genome) and BLASTX (transcript nucleotide to proteome), were used to confidently assign identities to the reamplified transcripts. Nucleotide sequence identities ranged from 83% to 100%, with the majority being in excess of 90%. The less than perfect match is attributable to both problematic sequencing where individual nucleotides could not be discerned, and errors incorporated by *Taq* polymerase during the two rounds of PCR that were required for reamplification. *Taq* polymerase has a reported error rate of 10^{-4} errors per base pair (Barnes, 1992). Ultimately the stringency of band selection from the differential display gel experiments, combined with the dual similarity searches, facilitated the reduction of false positives. Of the 11 confidently identified transcripts whose expression was compared with microarray data, only two transcripts showed a discrepancy.

A proportion of transcripts matched to the human genome database rather than *Plasmodium*. Contamination during PCR reactions is not uncommon, and by the nature of the reaction a small contamination could be amplified to detectable levels in subsequent analysis. Investigation of the transcripts originating from a human contaminant showed that the majority of these were excised from the same differential display acrylamide gel with a possible source being residual white blood cells in the blood used for culturing.

2.4.2.1 Differentially expressed genes encoding surface proteins

Differentially expressed surface proteins would most likely play integral roles in the immune evasion strategies of the parasite. Transcript E39-1 (PFI1475w), over-expressed in asexual parasite stages, was identified in the PlasmoDB database as

merozoite surface protein 1 (MSP-1 precursor). This gene encodes a protein involved in adhesion of the merozoite to the surface of the erythrocyte through a glycosylphosphatidylinositol (GPI) anchor. During schizogony the protein is cleaved into fragments which remain associated with the parasite surface and elicit a protective immune response in the host, making MSP-1 a leading vaccine candidate (Pan et al., 1999).

E51-3 (PF10_0025), also detected at elevated levels in asexual parasites, was identified as Pf70, a 70kDa exoantigen consisting of two transmembrane domains. Several peptides of Pf70 have been synthesized (James et al., 1993) and have been found to be highly immunogenic in rabbits. Furthermore, studies in Zaire (Tshefu and James, 1995) have shown that infected individuals with antibodies to Pf70 had lower parasitaemias than those who did not develop antibodies against this antigen.

PfEMP1 is a surface protein encoded by a family of approximately 60 *var* genes. Infected erythrocytes express only a single *var* gene at any given time, with switching between genes occurring at a rate of approximately 2% per generation (Roberts et al., 1992). This mechanism of gene switching contributing to antigenic variation is fundamental to the parasite's ability to successfully evade host immune responses. *PfEMP1* is also an adhesive molecule that binds to a host of human receptors to mediate the process of sequestration, whereby parasitized erythrocytes adhere to the capillary endothelium to avoid splenic clearance. The high number of *PfEMP1* genes identified in this study reflects the high copy number and diversity of this gene family, two factors integral to the immune evasion strategy of the parasite.

2.4.2.2 Differentially expressed genes involved in cell signalling/metabolism

Transcript E39-2 (PFD1120c) is a gene encoding a 14.8kDa integral membrane protein that is predicted to act as a signalling peptide (<http://www.genedb.org/>). The protein consists primarily of transmembrane domains indicating a potential function

in cell signalling, erythrocyte invasion (given its elevated expression in asexual parasites), transport or immune evasion.

Transcript E6-8 (PF10_0121) was identified as hypoxanthine phosphoribosyltransferase (HPRT), which is involved in the purine salvage pathway of the parasite. *P. falciparum* is capable of *de novo* synthesis of pyrimidines (cytosine, thymine and uracil) but utilises a salvage pathway to obtain purines. Hypoxanthine is the primary purine salvaged from the host, and by a series of enzymatic reactions can be converted to ATP (dATP) or GTP (dGTP) to be incorporated into RNA and DNA respectively. HPRT is one of numerous enzymes involved in this process, specifically responsible for the conversion of hypoxanthine to inosine monophosphate (Keough et al., 1999). Despite being found to be over-expressed in asexual stage parasites, gametocytes by necessity would also require purines as they are still transcriptionally active, as shown in the microarray data. The synthesis of HPRT mRNA may therefore occur in the early gametocyte stages which were not included in this study, or the gene may be expressed at a level not detectable by this method.

E43-1 (PF14_0076) is a precursor to plasmepsin I, which belongs to a family of ten aspartyl proteases involved in haemoglobin degradation. Intraerythrocytic stages of *Plasmodium* rely on red blood cell haemoglobin as a vital source of nutrients, degrading vast quantities of it. Haemoglobin is endocytosed from the erythrocyte cytosol and trafficked to an acidic food vacuole, where it is degraded by parasite proteases. Plasmepsins I and II, which are highly homologous, initiate this process (Le Bonniec et al., 1999).

2.4.2.3 Differentially expressed genes involved in RNA metabolism

E39-3 (PFF0745C) is annotated as a putative ribonuclease involved in RNA degradation. Ribonucleases act at the crossroads between transcription and

translation, and as such the elevated levels of expression of this gene in the asexual parasites probably represents the greater transcriptional activity of this developmental stage.

E43-2 (PFE0925c) was identified as a small nuclear ribonucleoprotein (snRNP), which contains a DEAD/DEAH box helicase domain at its carboxy terminal. These proteins are a component of the spliceosome, involved in numerous aspects of RNA metabolism including nuclear transcription, pre-mRNA splicing, translation and organellar gene expression.

A significant proportion (11%) of the transcripts identified by differential display corresponded to various ribosomal RNA (rRNA) genes. These molecules are responsible for defining the nature of the ribosomes, the protein complexes fundamental to the process of translation. In contrast to other species, these genes in *Plasmodium* are structurally distinct and developmentally controlled (McCutchan et al., 1995), and differ from other eukaryotes in that their copy number within the genome is unusually low (Langsley et al., 1983). The high proportion of these gene transcripts identified in this study indicates that despite their low copy number they are transcriptionally very active. The small subunit (SSU) rRNA genes exist in two distinct forms, designated A and S, that are expressed during the intraerythrocytic and mosquito stages, respectively. No explanation is currently available for this anomaly but it raises the possibility that other components of the transcription and translation machinery may also be stage specific. All RNA genes identified in this study were of the A-type, and were identified in asexual erythrocytic stages. The advantage for the parasite in having distinct ribosomes at different developmental stages has not been determined.

Another aspect where *Plasmodium* species are unique is in the chromosomal arrangement of the ribosomal DNA units. In eukaryotes, multiple repeats are commonly clustered together in a head to tail arrangement, whereas in *Plasmodium*

the five copies are each located on a different chromosome (chromosomes 1, 5, 7, 11, and 13). The significance of these ribosomal characteristics to the translational mechanisms of the parasite is not known.

2.4.2.4 Differentially expressed actin genes

PF14_0124 was identified as Actin II, a motility associated protein that assists in the formation of the axoneme, a specialised structure consisting of two parallel microtubules responsible for motility of male gametes. Actin II was identified in asexual stage parasites in differential display whilst according to microarray analysis there is a large peak of expression of PF14_0124 in gametocytes. Despite the apparent divergence between members of the *P. falciparum* actin gene family (Wesseling et al., 1988) it is possible that this transcript actually represents actin (PFL2215w), which is expressed in both developmental stages of the parasite (Wesseling et al., 1989). Indeed BLASTN and BLASTX similarity searches identified actin as the second best match, both with E values below threshold. This again highlights the difficulty in performing expression analyses for transcripts that belong to gene families, where the specific gene source is not identified.

2.4.2.5 Differentially expressed genes annotated as hypothetical

Five of the differentially expressed transcripts were found to encode proteins identified as hypothetical, i.e. the ORFs encoding these proteins have no known sequence similarity to genes in other organisms. Assigning putative functions to these hypothetical proteins in the absence of sequence similarity is difficult but the fact that they are unique does classify them as potential drug or vaccine targets. Some insight into the biological roles that these hypothetical proteins may play can be gained by examining the protein domains they contain. In each instance only the domain with the best score, representing the most confident assignment, is discussed.

Transcript E54-4 (PF10_0244) is a gametocyte stage-specific hypothetical protein that contains two EF-hand motifs. These motifs are involved in binding intracellular calcium, inducing a conformational change that is transmitted to target proteins often catalyzing enzymatic reactions.

E37-6 (PFI0210c) encodes a hypothetical protein expressed in asexual stages. This protein contains the motif of an ion transport protein, a family of proteins comprising sodium, calcium or potassium ion channels. Ion transport is an essential mechanism for intraerythrocytic survival and growth of the parasite. A small stretch of 13 amino acids at the carboxy terminal of the protein corresponds to a Myb-like DNA binding domain.

E23-1 (MAL13P1.70) and E10 (PF11_0361) are hypothetical proteins containing multiple transmembrane domains. They are likely to be members of one of the large families of surface proteins, playing a role in transport, immune evasion, sequestration, or in the case of the asexual stage-specific MAL13P1.70, erythrocyte invasion.

E12 (PF14_0338) contains a domain implicating it as a member of the group of Sm proteins which are central to RNA metabolism and are involved in diverse processes such as pre-mRNA splicing and telomere formation.

Analysing individual protein domains in isolation provides only partial information as far as gaining an understanding of the molecular function of these uncharacterised proteins is concerned. Le Roch et al (2003) organised genes into clusters based on the premise that genes with similar temporal patterns of expression are involved in comparable functions. The dataset containing these clusters was downloaded (<http://carrier.gnf.org/publications/CellCycle/>) and searched for the hypothetical proteins identified in this study. Only PF14_0338 (E12) and PFI0210c (E37-6) were present in this dataset.

PF14_0338 was found in cluster 8, along with genes mainly involved in cell cycle regulation and progression. The genes in this cluster are expressed predominantly in trophozoites but with a broad variation of expression (Le Roch et al., 2003). This does not really correlate with the function of RNA metabolism ascribed on the basis of being part of the Sm family of proteins; however, PF14_0338 may represent one of the exceptions to the cluster's characteristics.

PFI0210c was found in cluster 6 along with genes expressed in asexual stages, a finding that correlates with results from this study. The genes in this cluster code for proteins involved in protein synthesis, such as translation initiation factors and ribosomal proteins. It is unclear how this correlates with the presence of an ion transport motif.

Further analysis of these hypothetical proteins is required before any functions can be confidently assigned. As more functional information becomes available it will be interesting to ascertain the validity of expression-based clustering of genes.

2.5 Conclusion

Differential display is a technique that can be prone to false positives, and independent verification of results is thus required. In this research, reverse slot blot and available microarray data were used for verification of the stage of maximum expression of differentially expressed transcripts identified in asexual and gametocyte parasites of *P. falciparum*. A stage of expression could be confidently assigned to the gene transcript when there was correlation between all three methods of analysis.

The majority of the genes that were identified in this study reflect the diverse molecular and biochemical requirements of the two developmental stages, as opposed to being involved directly in the conversion of the parasite from an asexual form to a gametocyte. The exception to this was transcript E37-6, the hypothetical protein containing a Myb-like DNA binding domain. This particular protein was not characterised further due to the small number of homologous amino acids involved, and the divergence of this sequence from the two identified Myb proteins in *Plasmodium*, PF10_0327 and PF13_0088.

The inability to identify transcripts that could play a role in sexual development could be due to the questioned ability of differential display to identify rare transcripts. Transcription factors would be expected to play a vital role in the regulation of genes implicated in gametocytogenesis, and these have been shown by microarrays to be expressed at very low levels in *Plasmodium*. Also, the fact that predominantly late-stage gametocytes were used for the comparison means that some of the transcribed messages may already have been translated, and potentially degraded. Conversion to a sexual transmissible form requires a highly coordinated pattern of gene expression, and characterisation of one of the few annotated transcription factors in *Plasmodium*, *PfMyb2*, could help in gaining some insight into this process.

CHAPTER 3- CHARACTERISATION OF *PFMYB2*

3.1 Introduction

Despite the prevalence of post-transcriptional mechanisms of gene regulation in *P. falciparum*, microarray analysis has also highlighted the stringent transcriptional regulation that occurs throughout the intraerythrocytic developmental cycle of the parasite. There is evidence that the basic structure of the transcriptional apparatus in *Plasmodium* mirrors that seen in eukaryotes (Horrocks et al., 1998). Thus the transcriptional machinery consists of promoter and enhancer elements under the control of RNA Polymerase II, which is directed to transcription initiation sites by a set of transcription factors to form a transcription initiation complex. A number of promoters and upstream regulatory regions have been characterised from several *Plasmodium* species, but many of these are divergent both within the *Plasmodium* genus and from other eukaryotic regulatory domains (Coulson et al., 2004; Deitsch, 2004). The only similarity these sequences have to common eukaryotic regulatory elements is the presence of homopolymeric dA:dT tracts which have been shown in some eukaryotes to play an integral role in the regulation of transcription (Horrocks et al., 1998)

Identification of the *cis*-acting elements has been particularly difficult due to the high AT content of the intergenic regions of the parasite. For example, regions resembling the classic TATA box motif (consensus TATAAA/TA) are abundant upstream of most genes purely as a function of the AT content of the genome and not in a regulatory capacity. In addition, regulatory domains in the parasite appear to be located much further upstream of the transcription initiation site than is generally seen (Kumar et al., 2004). The identification of only a handful of recognisable eukaryotic regulatory sequences has led to speculation that in addition to the basal transcriptional apparatus, *Plasmodium* has developed its own unique set of specific

transcription factors that interact with regulatory sequences. Thus the identification of gene regulatory elements and the transcription factors that bind to them is still an important goal in unravelling the complexities of gene regulation in this organism.

3.1.1 Common eukaryotic regulatory elements in *P. falciparum*

3.1.1.1 The TATA box

A primary event that initiates transcription is the binding of TATA binding protein (TBP) to the TATA region of a promoter. The TATA box is a region consisting of a consensus sequence (TATAAA/TA) that has been extensively characterised in eukaryotes as a core promoter element, invariably located 25-35bp upstream of the transcription initiation site. *In vitro* mutagenesis studies have shown that even a single-base change in this nucleotide sequence drastically decreases transcription of genes adjacent to a TATA box, highlighting the importance of this regulatory domain for efficient transcription (Lodish et al., 1999).

Despite the difficulty in recognising these motifs within the AT-rich *P. falciparum* genome, several TATA-like motifs that are essential for efficient gene expression have been identified. The *P. falciparum* knob-associated histidine-rich protein (*kahrp*) gene promoter contains a consensus TATA box (TATAA) 81bp upstream of the transcription initiation site and the *gbp130* gene contains a TATA-like motif (TGTA) 186bp upstream of the start site. Chromatin immunoprecipitation assays have shown that *Pf*TBP interacts with both of these divergent TATA-like motifs *ex vivo* (Ruvalcaba-Salazar et al., 2005) suggesting that structural distortion of the sequences may be more important for protein-DNA binding than is base pair complementarity. In both instances these TATA-like motifs appear unrelated to sequences previously identified as essential for efficient promoter activity: 160bp upstream of the RNA initiation site in *Pfkahrp* (Lanzer et al., 1992) and 507bp upstream of the *gbp130* RNA initiation site (Horrocks and Lanzer, 1999). Whether

these sequences are mutually exclusive or work in tandem to ensure efficient promoter activity remains to be ascertained.

TATA boxes have also been identified up to 60bp upstream of transcriptional start sites of the *P. falciparum* *hsp86* gene (Su and Wellems, 1994) and upstream of the calmodulin gene (Robson and Jennings, 1991), *P. falciparum* proliferating cell nuclear antigen gene (*PfPCNA*) (Horrocks and Kilbey, 1996) and *Pfs25* (Dechering et al., 1999).

3.1.1.2 OCT1 transcriptional domains

A major transcriptional start site has been mapped 960bp upstream of the translational start site of the *PfPCNA* gene (Horrocks and Kilbey, 1996). Two sequences sharing homology to the consensus sequence of the OCT1 transcription factor binding site (ATGCAAAT) were found approximately 270bp (ATTCAAAT) and 230bp (ATGCATAT) upstream of the transcriptional start site. Consensus OCT1 binding sites have been identified in other *Plasmodia*, making it one of the more conserved eukaryotic regulatory sequences identified thus far.

3.1.1.3 The GC-rich region

In eukaryotes, transcription of genes with promoters containing a TATA box begins at a specific initiation site. In contrast, transcription of many protein-coding genes has been shown to begin at any one of multiple potential sites over an extended region spanning 20-200bp. As a result, such genes give rise to mRNAs with multiple alternative 5' ends. These genes often contain a GC-rich region (consensus GGGCGG) within approximately 100 base pairs of the start-site region. These regions are particularly prevalent in constitutively expressed housekeeping genes (Lodish et al., 1999).

Multiple transcriptional start sites have been identified for numerous *P. falciparum* genes, such as calmodulin (Robson and Jennings, 1991), *PfPCNA* (Horrocks and Kilbey, 1996) and MSP-2 (Wickham et al., 2003). GC patches have been identified between 370 and 500bp upstream of the transcriptional start site of the proliferating cell nuclear antigen (*PfPCNA*) gene (Horrocks and Kilbey, 1996) and also upstream of the *hsp86* transcription initiation site (Wu et al., 1995) but no homologue of SP1, the 100kDa protein that interacts with this region, has been identified in *P. falciparum*. If these GC-rich domains do indeed function as transcriptional regulators, it highlights again the distinguishing feature of *Plasmodium* transcriptional machinery whereby the regulatory sequences are much further upstream of the transcription initiation site than is seen in other eukaryotes.

3.1.1.4 The CAAT box

The CAAT box (consensus GGCCAATCT) is another common promoter element that is usually found approximately 80bp upstream of the transcription initiation site. In eukaryotes the transcription factor NF1, a 60kDa protein, binds to the CAAT box (Berg et al., 2005). NF1 is also known as CBF (CCAAT box-binding factor) and CTF (CCAAT binding transcription factor).

Three sequences have been identified in *Plasmodium* that have homology with CBF (Coulson et al., 2004) and putative CAAT boxes have been identified upstream of the *Pfs25* (Dechering et al., 1999) and calmodulin genes (Robson and Jennings, 1991). In addition, an inverted CAAT box was identified upstream of the *hsp86* transcription initiation site (Wu et al., 1995).

In their intensive analysis of regulatory sequences in the *Plasmodium* genome, using the *hsp* gene family as a model, Militello et al. (2004) report the absence of any distinguishable CAAT boxes. This, combined with the findings of Dechering et al. (1999), who show that the CAAT box upstream of the *Pfs25* gene is not essential for

efficient transcriptional activity, indicates that the functional significance of these regulatory regions in the *Plasmodium* genome has yet to be determined, and underscores the postulate that *P. falciparum* utilises its own unique set of regulatory elements.

3.1.2 Regulatory sequences unique to *P. falciparum*

Two distinct approaches have been used to identify unique regulatory sequences within the *Plasmodium* genome. In the classic gene-by-gene approach the effects of serial deletions on the upstream regions of transcription initiation sites are analysed using reporter gene expression. Although this has been successful in *Plasmodium*, as will be discussed below, it is more tedious than bioinformatic approaches, which exploit the availability of the complete genome sequence to identify over-represented sequence elements upstream of co-ordinately regulated genes. Some of the identified regulatory sequences are discussed below.

A *cis*-acting element identified in *P. falciparum* that is essential for efficient promoter activity is a 24-bp region upstream of the CDP-diacylglycerol synthase transcription initiation site (Osta et al., 2002). In comparison to other eukaryotes the distance from the transcriptional start site is vast, with the regulatory domains occurring between 1640 and 1569bp upstream of the translation initiation codon. Mapping of this region identified a 44bp region that bound specifically to trophozoite-stage nuclear extracts, and a 24bp region contained within the 44bp domain that mediated this specific interaction (Osta et al., 2002). The 24bp region that is essential for transcription factor binding shows no homology to any known eukaryotic binding motifs.

gbp130 is a developmentally regulated gene expressed uniquely in trophozoites, encoding a glycophorin binding protein (Horrocks and Lanzer, 1999). Horrocks and Lanzer (1999) identified a five bp element between 544 and 507bp upstream of the

transcriptional start site that binds nuclear factors in a specific manner and is responsible for efficient activity of the *gbp130* promoter. This region is encompassed within a 300bp repeat element that is similar to those seen upstream of the *pfg27/25* gene (Alano et al., 1996), but appear to bear no relation to the putative TATA-like motif identified 186bp upstream of the transcription initiation site (Ruvalcaba-Salazar et al., 2005).

pfs16 and *pfs25* are two sexual stage-specific genes that have been identified as markers of sexual differentiation. Dechering et al. (1999) analysed the promoter regions of these genes and found sequence elements that are essential for efficient transcriptional activity. The upstream regions of both genes show extreme AT-bias, and contain long homopolymeric dA:dT tracts, a common feature of eukaryotic intergenic regions. Studies conducted in *Dictyostelium discoideum* and yeast have shown the ability of similar homopolymeric tracts to stimulate transcription either by the introduction of bends into the DNA which enhances the affinity of transcription factors for their binding sites, or by the direct binding of certain transcription factors (Horrocks et al., 1998). Despite the instability of these tracts, which would ordinarily render them susceptible to elimination during replication and repair, bioinformatic analysis shows that they are over-represented in the *P. falciparum* genome, suggesting that they confer some selective advantage. The fact that they primarily localise to regions flanking genes gives rise to the speculation that they play a similar role in *P. falciparum* gene regulation as is seen in other eukaryotes.

In addition to these AT tracts, two elements with homology to the yeast MAT α 2 transcription factor binding site were identified upstream of the *Pfs16* transcription initiation site. An eight bp binding domain (AAGGAATA) was identified 409bp upstream of the transcriptional start site of *Pfs25* that is recognised by the protein PAF-1, a protein not found in any other eukaryotic databases (Dechering et al., 1999)

Another *P. falciparum* gene containing novel regulatory sequences is the knob-associated histidine rich protein (*kahrp*). The *kahrp* gene, developmentally regulated during the intraerythrocytic cycle, is under the transcriptional control of a palindromic sequence located 160bp upstream of the RNA initiation site (Lanzer et al., 1992). Utilising mobility shift assays, Lanzer et al. showed that this site is recognised in a specific manner by parasite nuclear extracts, but the interactions varied depending on the stage from which the extracts were derived, in accordance with the expression profile of the gene. The mechanism of regulation through this sequence element has not been elucidated.

Merozoite surface protein-2 (MSP-2) is an important molecule involved in parasite invasion of erythrocytes. Wickham et al. (2003) have characterised the promoter of this gene and defined certain regions that are required for expression on the merozoite surface. A single putative transcriptional start site has been identified 256bp upstream of the transcriptional start site, and transcriptional activity of the promoter was increased by a deletion of a region 170bp upstream of the initiation site, identifying this region as a potential negative regulator of the gene.

Bioinformatic approaches have also had success in identifying regulatory regions in *P. falciparum*, although the functionality of these regions is more hypothetical and has yet to be determined. Based on the premise that over-represented sequence elements upstream of co-ordinately regulated genes are often regulatory elements, Militello et al. (2004) devised a new bioinformatic strategy for identification of regulatory elements in *P. falciparum*. The result of their analysis was the G-box, which was detected upstream of seven out of 18 heat shock protein (*hsp*) genes and appears to be a *bona fide* regulatory element despite not interacting specifically with any nuclear factors. The regulatory element is a palindromic sequence consisting of two G-boxes located 195bp upstream of the transcription initiation site; the sequence is not homologous to any known transcription factor binding site, but its palindromic nature suggests that it may be recognised by a transcription factor unique to

Plasmodium. Transcription factors often interact with palindromic sequences, however Militello et al. (2004) were unable to experimentally demonstrate any interaction between the G box and DNA-binding proteins, leading to speculation on mechanisms by which the G box could function. Some possibilities include regulation of transcription through its intrinsic DNA structure, a mechanism which has been shown to occur in *Plasmodium* (Porter, 2002). Alternatively it may function as a stem-loop RNA element, structures which are known to play a role in many aspects of gene regulation including RNA stability and translation (Mignone et al., 2002). Despite not being present in mature *hsp86* RNA, it is possible that the G box is present in the nascent RNA and is lost upon processing.

Despite the *hsp86* gene requiring dual, palindromic G-boxes, single G boxes were also over-represented in the *hsp* gene family 5' upstream sequences, indicating that two G boxes are not always required for activation of gene expression (Militello et al., 2004).

Transfection experiments indicated that, as with many of the genes described above, an upstream sequence element, 5' UTR, G-boxes, 3' UTR and the region containing the transcription start site are all required for efficient reporter gene expression. Genome-wide analysis revealed palindromic G-boxes upstream of five more *Plasmodium* genes, indicating that they are not unique to the *hsp* gene family. The 'GC' patches identified by Horrocks and Kilbey (1996) upstream of the *PfPCNA* gene are not included amongst these five genes. Whether or not these patches represent G-boxes, or indeed if they have a regulatory role in the transcription of the *PfPCNA* gene, remains to be determined.

Peterson et al. (2004) utilised Bayesian Decomposition to analyse microarray data in a search for regulatory motifs upstream of co-regulated sets of genes. This resulted in a handful of motifs being classified as transcription factor binding sites, based on the expectation that these sites would be enriched amongst co-regulated sets of genes (Militello et al., 2004; Peterson et al., 2004). It remains to be determined whether the

motifs identified in this manner have any binding capacity when incubated with nuclear extracts or whether, like the G boxes of Militello et al. (2004), they show no binding.

3.1.3 *P. falciparum* transcription factors

The eukaryotic gene regulatory machinery makes use of two distinct types of transcription factors. Basal transcription factors are required for all regulatory events, and assist RNA polymerase in the recognition of promoter sequences and unravelling the DNA double helix. Modulatory or specific transcription factors regulate the expression of specific genes at a specific time, and allow differential expression at different developmental stages of a life cycle, such as are seen in *Plasmodium*.

3.1.3.1 Basal transcription factors in *P. falciparum*

Despite a great deal of research, not much is known thus far about the *trans*-acting elements directing transcriptional regulation in *P. falciparum*. Genome mining and proteome analysis, in addition to traditional molecular biology techniques, have indicated a scarcity of recognisable eukaryotic regulatory transcription factors, although some components of the basal transcriptional apparatus have been identified. These include a highly divergent TATA binding protein (McAndrew et al., 1993; Ruvalcaba-Salazar et al., 2005) and homologues of all 12 subunits of RNA polymerase II and TFIIB, TFIID, TFIIE α and TFIIH (Coulson et al., 2004). The identification of TATA boxes and the *P. falciparum* TATA binding protein (PfTBP) heralded the first instance where *cis*-acting elements were identified in concert with their corresponding *trans*-acting factor (Ruvalcaba-Salazar et al., 2005). TBP is part of transcription factor IID (TFIID), which is a sequence specific basal transcription factor that forms the scaffold upon which the rest of the transcriptional apparatus assembles prior to transcription initiation (Orphanides et al., 1996). TBP binds directly to the TATA box element, after which a preinitiation complex is formed

consisting of RNA Polymerase II and basal transcription factors TFIIA, B, E, F and H. *Pf*TBP has been implicated in the transcriptional regulation of *kahrp* and *gbp-130*, and given the identification of TATA-like motifs in the calmodulin and *hsp86* genes, is likely to be involved in the transcriptional regulation of these genes in addition to many more genes in the parasite genome.

3.1.3.2 Specific transcription factors in *P. falciparum*

Despite the findings that the regulatory mechanisms in *P. falciparum* are similar to those seen in eukaryotes, and many components of the basal transcriptional machinery have been recognized, very few modulatory transcription factors have been identified.

An interesting aspect of *Plasmodium* protein biology is that the proteins appear to be much larger than their eukaryotic counterparts, with a high degree of variation outside of the functional domains. Thus simple overall sequence similarity searches are possibly too stringent to accurately identify specific classes of proteins in the database. Kumar et al. (2004) searched for conserved transcription factors in the *Plasmodium* genome based on the ability of nuclear extracts to interact with sequences representing highly conserved transcription factor binding motifs localised on a DNA array. A total of 20 positive hits were returned, of which *cis*-acting elements had been identified for six genes. The specificity of the interaction of eight of the most prominent positive hits, cMyb, CREB, EGR, MEF-1, NFkB, E2F1, Smad 3/4, and HSE, was ascertained using mobility shift assays. Highly specific binding was only observed for c-Myb and MEF-1 indicating that binding on the array may have been non-specific or that the binding proteins may have a broader binding specificity than previously described.

To identify differences between gametocyte-producing and gametocyteless strains of parasite, Gissot et al. (2004) used a gene-specific microarray printed with 153 PCR

products representing genes potentially involved in cell cycle. Their analysis included mapping of the expression profiles of five putative transcription factors; *PfMyb3*, *pfhmg2*, *pfphD2*, *pfphDB* and *pfkrox*. The annotation of these proteins as *bona fide* transcription factors in *P. falciparum* has not yet been validated.

3.1.3.3 The cMyb family

The Myb family of DNA-binding proteins is amongst the most recognisable of the eukaryotic transcription factors to have been identified in *P. falciparum*. Vertebrate *c-myb* is known to encode sequence-specific DNA-binding proteins that are involved in the proliferation and differentiation of haematopoietic cells, and have been implicated in regulating the G₁/S transition in the cell cycle (Thompson et al., 1986). The Myb proteins have been well conserved throughout evolution specifically in their N-terminal DNA binding domains. These generally contain three repeats (R1, R2 and R3) of approximately 50 amino acids with three regularly spaced tryptophan molecules (Lipsick, 1996). The three tandem repeats each contain a helix-turn-helix (HTH) motif which is critical for DNA binding. Despite the structural conservation, there is a large degree of variation in transactivation by the Myb proteins, depending on cell type and promoter structure (Oh and Premkumar Reddy, 1999). This implies that Myb is reliant on other cell-type specific co-factors to mediate transactivation, a fact confirmed by its association with a CAAT enhancer-binding protein (C/EBP) (Ogata et al., 2003).

PfMyb1 was the first specific transcription factor to be characterised in *P. falciparum* (Boschet et al., 2004). Nuclear extracts containing this protein were shown by mobility shift assay to interact in a specific manner with consensus Myb DNA binding domains. Subsequent analysis demonstrated that knock-down of this gene using dsRNA severely inhibited growth of *P. falciparum* cultures (Gissot et al., 2005), suggesting that this transcription factor plays an essential role in the intraerythrocytic growth of the parasite. Chromatin immunoprecipitation assays

identified six genes that may be directly regulated by *Pf*Myb1, and several more that may be regulated by transcription factors under the control of *Pf*Myb1.

The identification of the Myb family of DNA binding proteins in *P. falciparum*, and evidence for the vital role played by one member of this family in transcriptional regulation, suggests that further investigation of this gene family is warranted to gain more insight into the mechanisms of transcriptional regulation in the parasite.

To this end PF10_0327, identified on the PlasmoDB database as *Pf*Myb2, was investigated. Bioinformatic analysis was used to identify the DNA binding domains of the protein which were cloned into an expression vector. Purified recombinant protein was incubated with end-labelled consensus cMyb DNA binding domains and analysed by EMSA.

3.2 Materials and Methods

3.2.1 Cloning of *PfMyb2*

3.2.1.1 Genomic DNA isolation from *P. falciparum*

P. falciparum strain 3D7 was maintained in culture as described (2.2.1) and late stage parasites were harvested for DNA extraction when parasitaemia reached 10-15%. A 20ml culture was transferred to a 50ml Nunc tube, and cells were pelleted by centrifugation at 1100g for 15 minutes. The cell pellet was washed once in PBS by centrifugation at 1100g for five minutes, and the packed erythrocytes lysed in 0.05% saponin. The parasite pellet was resuspended by gentle pipetting in 250µl lysis buffer (10mM Tris-HCl pH 8.0, 20mM EDTA, 0.5% SDS) supplemented with 0.025mg/ml Proteinase K to remove any contaminating proteins. The solution was incubated at 37°C for three hours with occasional mixing. 1ml MilliQ water was added after three hours, and the DNA was purified by phenol: chloroform extraction and ethanol precipitation (A-2).

3.2.1.2 Preparation of *PfMyb2* DNA for cloning

The N terminal region of the *PfMyb2* transcription factor contains two DNA binding domains. From the first methionine, a region spanning 196 amino acids encompassing both DNA binding domains was selected for amplification and cloning into the pET15-b and pGEX-4T-2 expression vectors. Protein expression was under the control of the bacteriophage T7 promoter (pET-15b) or the *tac* promoter (pGEX-4T-2). *Nde*I (5') and *Bam*HI (3') restriction sites were added to the PCR primers for ligation with the pET-15b expression vector, and *Bam*HI (5') and *Xho*I (3') restriction sites were appended for ligation with the pGEX-4T-2 expression vector (A-3). 0.5-1.0µg DNA, forward and reverse primers at a final concentration of 0.4µM and 2X High Fidelity PCR Master Mix were combined in a 50µl PCR reaction. A two step PCR was performed in an Eppendorf Gradient thermal cycler.

Five cycles of denaturation at 94°C for 30 seconds, annealing at 50°C for 30 seconds and extension at 72°C for 30 seconds to allow annealing of the gene-specific region of the primer, were followed by 25 cycles of denaturation at 94°C for 30 seconds, annealing at 65°C for 30 seconds and extension at 72°C for 30 seconds for amplification of the entire region including the appended restriction sites. PCR products were purified by phenol: chloroform extraction and ethanol precipitation (A-2) and the pellet resuspended in 50µl nuclease-free water. Amplification was confirmed by electrophoresis of 10µl PCR product on a 1% agarose gel at 60mA for 30 minutes.

The PCR products were subsequently digested with the appropriate restriction enzymes for cloning into pET-15b and pGEX-4T-2 expression vectors. 40µl PCR product, 5µl 10X restriction enzyme buffer, 0.5µg acetylated BSA and 1U appropriate enzyme were combined in a microcentrifuge tube and the volume adjusted to 50µl with MilliQ water. *Nde*I (Roche Buffer H: 50mM Tris-HCl pH 7.5, 10mM MgCl₂, 100mM NaCl, 1mM DTT) required overnight digestion at 37°C whilst two hours incubation was sufficient for complete *Xho*I digestion (Roche Buffer H) of the PCR product. Products were purified by phenol:chloroform extraction and ethanol precipitation and subsequently digested with *Bam*HI (Roche Buffer B: 10mM Tris-HCl pH 8.0, 5mM MgCl₂, 100mM NaCl, 1mM β-mercaptoethanol) at 37°C for two hours. DNA was again purified and was resuspended in 10µl 1X DNA dilution buffer from the Rapid DNA Ligation Kit.

3.2.1.3 Preparation of pET-15b and pGEX-4T-2 expression vectors

1µg of each vector was digested with either *Nde*I (pET-15b) or *Xho*I (pGEX-4T-2) to create compatible ends for ligation with *PfMyb*2. Reactions were incubated at 37°C for two hours, after which 1U Calf Intestinal Phosphatase (CIP) (Roche, Germany) was added and the incubation continued for a further 30 minutes. CIP removes the 5' phosphate from the linearised vector thereby preventing religation.

The vector was purified by extraction with phenol and chloroform followed by ethanol precipitation. The pellet was resuspended in a final volume of 40µl nuclease-free water, and was digested with *Bam*H1 followed by treatment with CIP. The vector was purified and resuspended in 10µl 1X DNA dilution buffer from the Rapid DNA Ligation kit.

3.2.1.4 Ligation of insert and vector

Aliquots of prepared *Pf*Myb2 and linearised dephosphorylated vector were loaded onto a 1% agarose gel alongside a DNA mass ladder to establish concentrations for ligation. A molar ratio of 3:1 - 5:1 of insert: vector was used for the ligation reaction but the total amount of DNA in the reaction did not exceed 200ng. The ligation reaction was set up using the Rapid DNA Ligation Kit as described (2.2.4.2), but the ligation temperature was decreased to 16°C. 5µl of ligated products were used to transform competent DH5α cells as described (2.2.4.3).

3.2.1.5 Confirmation of the presence of inserts by colony PCR

Individual transformed colonies were picked from agar plates with a sterile pipette tip, resuspended in 7µl nuclease-free water, and lysed by boiling for 5 minutes. 5µl was combined with a final concentration of 0.1µM forward and reverse primers and 2X Roche PCR Master Mix in a 25µl PCR reaction. pGEX-4T-2 vector primers (A-3) were used for amplification, which append an additional 147bp to the size of the PCR product, whereas gene-specific primers were used for the insert in the pET-15b vector. The PCR comprised 30 30-second cycles of denaturation at 94°C, annealing at 60°C and elongation at 72°C.

3.2.1.6 Confirmation of the presence of inserts by restriction enzyme analysis

Further verification of the presence of inserts was obtained by restriction analysis of purified plasmid DNA. Following the purification of plasmid DNA from transformed

colonies (3.2.1.7), pET-15b was digested with *Hind*III (Promega, USA) and *Xba*I (Promega, USA) in 1X Promega Buffer B (6mM Tris-HCl pH 7.5, 6mM MgCl₂, 50mM NaCl, 1mM DTT), whilst pGEX-4T-2 was digested with *Bam*HI and *Xho*I in 1X Promega Buffer C (10mM Tris-HCl pH 7.9, 10mM MgCl₂, 50mM NaCl, 1mM DTT) in 20µl reaction volumes. The digestion of the respective constructs with these enzymes increased the size of the *Pf*Myb2 inserts by 399bp (pET-15b) and 25bp (pGEX-4T-2) respectively. Reactions were incubated at 37°C for two hours, and the presence of an insert was verified by agarose gel electrophoresis.

3.2.1.7 Purification of plasmid DNA

The Eppendorf FastPlasmid™ Mini Kit was used for the rapid isolation and purification of plasmid DNA. The method allows DNA to be captured on a solid phase support and then eluted in a low-salt buffer. Plasmid DNA was purified as per the instructions in the kit and stored at 4°C until required.

3.2.1.8 Verification of insert sequence and orientation

To verify the fidelity of the *Pf*Myb2 amplification, single colonies were spread onto an agar plate and sent for sequencing (Inqaba Biotech, South Africa). Analysis of the sequence data using ChromasPro Version 1.2 identified an error in the *Pf*Myb2/pGEX-4T-2 construct which precluded further work with this particular construct. All further investigation was thus performed only with *Pf*Myb2/pET-15b.

3.2.1.9 Transformation of BL21-CodonPlus competent cells

BL21-CodonPlus® (DE3)-RIL competent cells were chosen as the host cells for expression of recombinant protein. The CodonPlus-RIL cells contain additional copies of the tRNAs for the arginine, isoleucine and leucine codons prevalent in AT-rich genomes, in addition to a pACYC-based plasmid that confers chloramphenicol resistance. A 100µl aliquot of cells was thawed on ice and mixed with 2µl 10%

XL10-Gold β -mercaptoethanol, which is included with the kit to increase transformation efficiency. The cells were incubated on ice for ten minutes with occasional mixing. Approximately 50ng plasmid DNA containing the *PfMyb2* construct (quantitated by agarose gel electrophoresis) were used for transformation, which was performed as per manufacturer's instructions.

3.2.2 Expression of 6x His-*PfMyb2* protein

3.2.2.1 Induction of target protein using IPTG

Conditions for the optimal expression of 6xHis-*PfMyb2* had to be determined empirically. Variables such as isopropyl-1-thio- β -D-galactopyranoside (IPTG) concentration, length and temperature of induction and optical density of cultures at the time of induction all had an influence on the efficiency of protein expression. In addition, the induction was performed in the presence of 0.2M-1M glycyl-glycine in an attempt to increase the solubility of the recombinant protein (Gosh et al., 2004). The range of values for these variables is indicated in parentheses below.

1ml aliquots of LB broth, containing 50 μ g/ml chloramphenicol and 100 μ g/ml ampicillin, were inoculated with single transformed colonies (3.2.1.9) in 14ml round bottomed Falcon tubes. These were incubated at 37°C with vigorous rotation for 16 hours. A 1:10 dilution was made into fresh LB medium (1-50ml) and the cells were incubated (24°C-37°C) until the absorbance at 600nm was between 0.5-1.8. IPTG was added at a final concentration of 0.4-1.0mM to and incubated at 37°C (1-20 hours). 100 μ l aliquots of induced cultures and uninduced control cultures (no IPTG) were solubilised in 1X SDS buffer (A-1) and 20 μ l were analysed by SDS-PAGE on a 10% mini gel (Laemmli, 1970) prepared in a Hoefer Mighty Small Dual Casting Tray. Gels were electrophoresed for 90 minutes at 35mA with cooling to 4°C and stained overnight with 0.5% Coomassie Brilliant Blue in 10% acetic acid/25%

methanol. Gels were destained in 10% acetic acid/ 10% methanol followed by 10% acetic acid. The expected size of the 6xHis-P/Myb2 fusion protein was 24.8kDa.

3.2.2.2 Induction of target protein using the Overnight Express™ Autoinduction System

The Overnight Express™ Autoinduction systems are designed for high-level protein expression with all IPTG-inducible expression systems. The system comprises three solutions, which control growth of the cultures, use lactose to induce expression at optimal cell density and control the pH of the cultures.

LB medium containing 50µg/ml chloramphenicol and 100µg/ml ampicillin was inoculated from glycerol stocks and grown overnight at 37°C with shaking. A 500µl aliquot of this was added to 200ml Overnight Express medium and grown at room temperature for 24 hours with shaking. The soluble fraction was purified as described (3.2.2.3).

3.2.2.3 Extraction of soluble proteins from BL21 cells

Following IPTG induction, cells were pelleted by centrifugation at 15000g for one minute. The pellet was frozen at -70°C for five minutes, and allowed to thaw slowly to room temperature. The pellet was resuspended in 200µl BugBuster HT Protein Extraction Reagent containing 2.5µl Protease Inhibitor Cocktail Set III per 1ml induced cells. This specific cocktail of protease inhibitors targets serine, cysteine, trypsin-like and aspartic proteases, as well as aminopeptidase B and leucine aminopeptidase. The cell suspension was incubated at room temperature for 20 minutes with gentle inversion. Insoluble proteins were removed by centrifugation at 15000g for 15 minutes at 4°C, and a 20µl aliquot of supernatant was solubilised in 1X SDS solubilisation buffer for SDS-PAGE analysis. The insoluble pellet was resuspended in 200µl 0.1M phosphate buffer pH 8.0, and a 20µl aliquot solubilised as above for electrophoresis.

3.2.2.4 Immunoblot to confirm the expression of 6xHis-*PfMyb2*

Proteins were transferred to nitrocellulose membrane at 50V for four hours in a TE series Transphor Electrophoresis Unit system at 4°C (A-2). The presence of transferred proteins was confirmed by immersing the membrane for ten minutes with slow agitation in 1% Ponceau S diluted in 7% acetic acid. The membrane was destained in water to visualise the proteins. The presence of 6xHis-*PfMyb2* was confirmed with mouse anti-His antibody detected by goat anti-mouse peroxidase conjugated IgG using 4-chloro-1-naphthol as a substrate (A-2).

3.2.2.5 Purification of 6xHis-*PfMyb2*

pET-15b (5708bp) encodes an N-terminal His-tag comprising six histidine residues that allows purification of the resultant recombinant protein by the affinity of histidine for Ni²⁺- charged magnetic agarose beads. Beads from either Novagen (USA) or Sigma (USA) were used. In both instances, 10-30µl packed beads, equilibrated in binding buffer, were incubated with the soluble extract from induced cells for 1 hour at room temperature with slow rotation. Beads were collected with a magnet, and washed twice with wash buffer (A-1). 6xHis-*PfMyb2* was eluted from the beads in two to five bead volumes elution buffer (A-1), and immediately dialysed into 1X electrophoretic mobility shift assay (EMSA) binding buffer containing 2mM DTT to reduce the conserved cysteines. Aliquots representing 10% of all fractions were electrophoresed on a 10% acrylamide gel (35mA for 90 minutes at 4°C) and visualised with 0.5% Coomassie Brilliant Blue to assess the quality of the protein purification.

3.2.2.6 Purification of 6xHis-*PfMyb2* under denaturing conditions

The insoluble pellet produced after lysis of the bacterial pellet was retained for purification of inclusion bodies. The pellet was resuspended by vortexing in the same volume of BugBuster HT that was used for the original pellet, and lysozyme was

added to a final concentration of 1KU/ml. This was mixed and allowed to incubate at room temperature for five minutes. The suspension was transferred to a 15ml Nunc tube and six volumes of 10% BugBuster HT were added and vortexed for one minute. The suspension was centrifuged at 5000g for 15 minutes at 4°C and the supernatant was removed. The inclusion bodies were resuspended in 0.5 times the original culture volume of 10% BugBuster HT, mixed by vortexing and centrifuged as before. This was repeated twice, and the pellet was then resuspended a final time and centrifuged at 16000g for 15 minutes at 4°C. The final pellet was resuspended in 1-2ml binding buffer containing 8M urea, and left on ice for one hour to ensure complete denaturation of the target protein. Protein was purified as above (3.2.2.4) with the inclusion of 8M urea in all solutions.

An alternative protocol to purify protein under denaturing conditions involved denaturation of the entire insoluble pellet, as opposed to isolation of inclusion bodies. The insoluble pellet from 3.2.2.2 was resuspended in binding buffer containing 8M urea, and allowed to incubate on ice for one hour to ensure complete denaturation of the protein. Purification was then performed as described (3.2.2.4).

3.2.2.7 Dialysis of denatured and purified 6xHis-*PfMyb2*

Purified *PfMyb2* was subjected to step-wise dialysis to remove all traces of urea, and allow the protein to refold and revert to its native conformation. Purified denatured protein in 8M urea was aliquoted into ¼ inch dialysis tubing and placed into 100ml PBS containing 7M urea. The urea concentration was decreased in a step-wise manner by the addition of urea-free PBS every 45 minutes for five hours, after which the dialysed products were transferred to a fresh beaker containing 100ml urea-free PBS for 30 minutes. A 100µl aliquot was removed after each dilution and centrifuged at maximum speed for 60 seconds. The pellet was diluted into 1x SDS solubilisation buffer and electrophoresed alongside the supernatant of the same fraction on a 10% acrylamide gel.

3.2.3 *In vitro* translation of *PfMyb2*

The *in vitro* system is based on transcription with a T7 bacteriophage polymerase followed by translation in an optimised rabbit reticulocyte lysate. The PROTEINscript[®] II kit is designed for linked *in vitro* transcription and translation and was used for translation of *PfMyb2* using either PCR product or plasmid DNA from transformed DH5 α cells as a template.

3.2.3.1 PCR primer design for *in vitro* translation

A 5' PCR primer was designed to incorporate the T7 promoter, ribosome binding site and a histidine tag which, in addition to the *PfMyb2* gene-specific region, resulted in a primer 84 nucleotides in length (A-3). *P.falciparum* total RNA was reverse transcribed as described (2.2.3.1), but with one of the two *PfMyb2* gene-specific primers instead of the 2-base anchored primer (A-3). A two-step PCR was used, which entailed five cycles of denaturation at 94°C, annealing at 45°C and elongation at 72°C, all for 30 seconds, followed by 25 cycles of 30 seconds denaturation at 94°C, 30 seconds annealing at 60°C and extension for 30 seconds at 72°C. The first step in the PCR allowed annealing of the gene-specific portion of the primer, after which the stringency could be increased for annealing of the full length *in vitro* translation primer.

3.2.3.2 *In vitro* translation procedure

A 10 μ l transcription reaction was prepared incorporating 2 μ l 5X transcription mix, 2 μ l enzyme mix (supplied with the kit) and 0.5 μ g plasmid DNA or 2-5 μ l PCR product. Components were mixed by pipetting and incubated at 30°C for 60 minutes. Any transcription reaction not used immediately in the translation reaction was stored at -20°C for later use. The RNA produced during the transcription reaction was translated by combining 2.5 μ l 20X translation mix, 2.5 μ l 500 μ M methionine, 35 μ l

reticulocyte lysate, 8µl nuclease-free water and 2µl transcription reaction. These reagents were combined in a 500µl PCR tube and incubated at 30°C for 60 minutes. A negative control reaction contained all components except the RNA template. This was to confirm the successful translation of *PfMyb2*, as well as assist in the identification of this protein amongst the 200mg/ml endogenous proteins present within the reticulocyte lysate. In general, four translation reactions were initiated, and these were then pooled and diluted to 500µl in 0.1M Na₂PO₄ pH 8.0 for purification of 6xHis-*PfMyb2* (3.2.2.4). An immunoblot (3.2.2.3) was used to confirm the identity of the protein.

3.2.4 Electrophoretic mobility shift assays (EMSA)

Mobility shift assays were performed to assess the binding potential of 6xHis-*PfMyb2* to oligonucleotides representing consensus Myb DNA binding sequences (A-3). *P. falciparum* putative Myb regulatory elements (MREs) (*pfcrk1* and *pmap1*), as well as the chicken *mim-1* gene element, were used as potential binding sequences, in addition to the consensus sequence of the Myb binding domain (Boschet et al., 2004). The sequence representing the binding domain of NfKB was used as a negative control.

3.2.4.1 Binding reactions using an EMSA kit

Nucleic acids can be detected in the acrylamide gel under UV light using the Molecular Probes' fluorescence-based EMSA kit. The kit contains SYBR Green for the detection and staining of a minimum of 1ng DNA. 40ng oligo was incubated for 20 minutes at room temperature with 8µl purified 6xHis-*PfMyb2* in 1x Binding buffer (75mM KCl, 0.1M DTT, 0.1M EDTA, 10mM Tris pH 7.4). Glycerol was added to a final concentration of 5% immediately prior to loading onto a 6% acrylamide gel which had been pre-electrophoresed at 160V for 2 hours. Samples

were electrophoresed at 130V for 90 minutes and visualised as per the instructions in the kit

To increase the sensitivity of detection, as well as ascertain the specificity of the interaction with competition experiments, the oligonucleotides were end-labelled for detection with autoradiography.

3.2.4.2 End-labelling of oligonucleotides

Complementary single stranded oligonucleotides were annealed by combining forward and reverse sequences in a 500µl PCR tube, and heating to 90°C for five minutes. Tubes were cooled slowly to room temperature over a period of 30 minutes. 500ng double stranded oligonucleotides were incubated at 37°C for 60 minutes in a 30µl reaction containing 20U T4 polynucleotide kinase (PNK), T4 PNK buffer (70mM Tris-HCl pH 7.6, 10mM MgCl₂, 5mM DTT) and 5µl [γ -³²P]ATP (6000Ci/mmol; 10µCi/µl). The reaction was stopped by the addition of 1µl 0.5M EDTA. End-labelled products were then diluted to 50µl with water, purified by centrifugation through Sephadex G-25 columns to remove unincorporated ATP, and the incorporation of ³²P measured using a scintillation counter. Oligonucleotides were used immediately or stored at -70°C and used within a week.

3.2.4.2 Binding reaction and electrophoresis

Binding reactions were prepared by combining 8µl 6xHis-PfMyb2, EMSA binding buffer (A-1) and 3x10⁶ cpm ³²P-labelled oligonucleotides (40ng) in a 10µl reaction. Competition reactions contained 10- and 100-fold excess of unlabelled homologous or non-homologous competitor oligonucleotide. Tubes were incubated at room temperature for 20 minutes, after which glycerol was added to a final concentration of 5%. Samples were loaded without loading dye onto a 4% non-denaturing acrylamide

gel and electrophoresed at 100V (4mA) for one hour at 4°C in 0.25X TBE buffer (A-1).

Gels were dried for 70 minutes at 80°C and exposed to X-ray film overnight at -70°C in a cassette containing intensifying screens.

3.3 Results

3.3.1 Cloning *PfMyb2* into expression vectors

PfMyb2 (PF10_0327) is described in the PlasmoDB database as a 2748bp gene encoding a 108kDa protein containing two putative Myb-like DNA binding domains. Consistent with the majority of proteins in the Myb family, these binding domains are found at the amino-terminal, specifically spanning amino acids 14-53 and 59-103 (Figure 12).

```

MRIQIKGGIWKNCEDEV LKAAVMKYGLNNWSRVASLLVRKSAKQCKARWYEWLDP SVRKTEWNKEEEEKL
LHLAKLFPTQWRTIAPIVGRTAQQC LEHYEYLLDEAEGKVYDKNKNPRHLRPG EIDPAPESKPARADPVD
MDEDEKEMLA EAKARLANTKGKKA KRKAREKQLEQARRLALLQKKRELKAAGITSLNYKRKDKNKIDHSK
EILFHRKPLKGFYDVKDEQNINDDIYENNKTNQKKS IKSMDEVENINDAMEYNKNKGKRQHQHNNNEEANL
LSTIENYDKQFNELSHLRKRVRNLNLP EPI LNENEIDEIIQINKEASAFNDI IKDQNDKLPINNILPSIES
SSFILNNKDKFSNLETDFYNNNNNKSIAFSSKLDLSIQQA AKNIISRKMNIPIGMNNDYNEEEFERKNN
IFQKSNKNVPDDLEYDDTN NYHNNNNNNNNNTSFNSIKNSTTLYNHLDVEKNIKDVQEEIEKQKKLNEED
QQKNQKEKLTIKNKIISDIKSFKKNISLYAHSIISYKNLKNDSMMDNQTIQTTEYYDDNYEEKIDRAKL
HIKASLANLPQETNLI ELQLNEEHPECDTDNIEKDEIEKD IQDIENEKRKNEERKEKEKFNKQNKIIRWN
LPRPYFLDKINLFMNYMHNEYEDVHNL IQREMLLLIKNDMFNYPLRNSTP VQNKVHVEDLENYVMNMAMK
SINEEFEDMYGEASLNNMNTKDDSNIDGCDEKSDNIDGCDEKSDTTNKGDDTSQCSIEHSSYNHIDVWEEI
NKNIIFCPSKNAYRFIEDVNENDKKENYKYKCEKLNKILNDMEHYKKLENKYDIYTKGYQLKIKSYKKS
YDTLFNSYINCINEKEALNVLHENEKIYALTRIKEEKKENKKEIEYHKS LQKFYQDLLETNHQLKETCKQ
TLKVP

```

Figure 12- Amino acid sequence of *PfMyb2*

PfMyb2 is a protein comprising 916 amino acids. The two Myb-like DNA binding domains are highlighted in red, and the region of protein that was cloned into the pET-15b and pGEX-4T-2 expression vectors is underlined.

3.3.1.1 Amplification of *PfMyb2*

The 5' region of PF10_0327 encompassing the two Myb-like DNA binding domains was amplified from *P. falciparum* 3D7 genomic DNA. Using gene specific primers with the appropriate restriction sites appended for subsequent cloning (A-3), a product of 616bp (pET-15b) or 611bp (pGEX-4T-2) was obtained. The amplified products are illustrated in Figure 13.

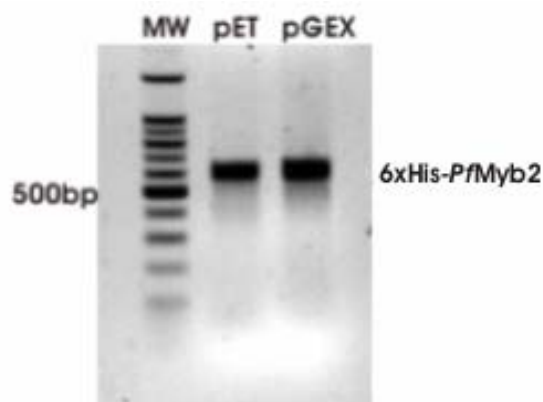


Figure 13- Amplification of *PfMyb2* for cloning

10µl of PCR product was electrophoresed alongside a 100bp DNA size marker (MW). **pET** and **pGEX** denote *PfMyb2* PCR product amplified with primers for the subsequent cloning into pET-15b and pGEX-4T-2 expression vectors.

After the transformation of DH5α cells (3.2.1.4), single colonies were picked from plates and screened for the presence of inserts via colony PCR (Figure 14).

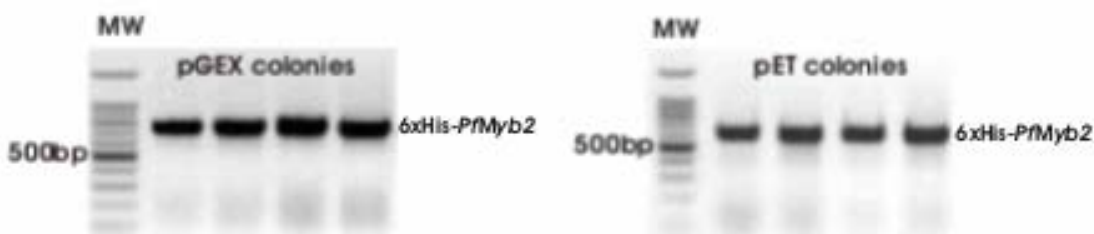


Figure 14-Verification of the presence of *PfMyb2* inserts

Colonies of transformed DH5α cells were used as a template for a PCR reaction with either vector primers flanking the cloning site (pGEX-4T-2, **A**) or the *PfMyb2* primers used for the original amplification (pET-15b, **B**). These PCR reactions generated products of 735bp and 616bp respectively. **MW**: 100bp DNA size marker; **pET**: pET-15b expression vector; **pGEX**: pGEX-4T-2 expression vector.

Further verification of the presence of inserts was obtained by restriction enzyme analysis of purified plasmid (Figure 15).

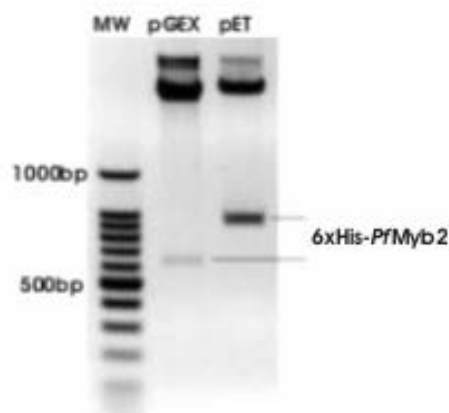


Figure 15-Restriction enzyme analysis of vector constructs

pGEX-4T-2 was digested with *Bam*H1 and *Xho*I, which increased the size of the product by 25bp to 613bp, whilst pET-15b was digested with *Hind*III and *Xba*I which created a 987bp product. The low intensity of the pGEX-4T-2 insert is probably a result of incomplete digestion.

3.3.1.2 Sequence verification of *PfMyb2* insert

After verifying the presence of inserts, plasmid DNA was sequenced to ensure that no errors had been introduced during the amplification and that the insert was in-frame. The sequencing results were analysed using ChromasPro version 1.2, which indicated that whereas the sequence of *PfMyb2* in pET-15b was correct, there was an error in the pGEX-4T-2/*PfMyb2* sequence (Figure 16). A T-C point mutation was introduced into the sequence at position 391 of *PfMyb2* during PCR amplification, which would result in a switch from a serine (TCC) to a proline (CCC). Whereas serine is often found in protein functional centres and is relatively common within tight turns on the protein surface, proline is a helix breaker, being unable to adopt a normal helical conformation. This mutation occurred outside of the DNA binding domain and did not result in a frame-shift, but despite this the different structures of these amino acids precluded further work with the *PfMyb2*/pGEX-4T-2 construct.

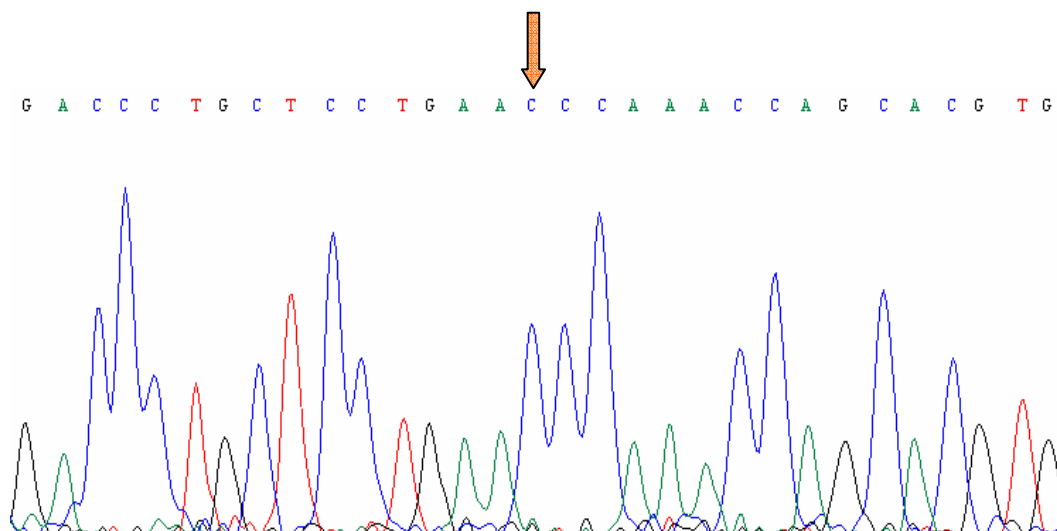


Figure 16-Partial DNA sequence of 6xHis-*PfMyb2* in pGEX-4T-2

The above sequence represents the portion of *PfMyb2* sequence in pGEX-4T-2 vector that had a mutation introduced during amplification. The arrow indicates a cytosine residue that should be a thymine, which would result in an amino acid change from a serine to a proline. It is clear from the quality of the sequence that this is a genuine mutation introduced during amplification, and is not a reading error.

3.3.2 Expression and purification of 6xHis-*PfMyb2* protein

Induced cells were lysed and proteins separated by SDS-PAGE (Figure 17). Induction times and IPTG concentrations were varied as described (3.2.2.1). As the levels of expression of *PfMyb2* were not significantly affected by these variables, subsequent inductions were performed with 0.4mM IPTG for four hours at 37°C. An immunoblot was performed to verify that the protein represented recombinant *PfMyb2*, and not a bacterial protein of a similar size.

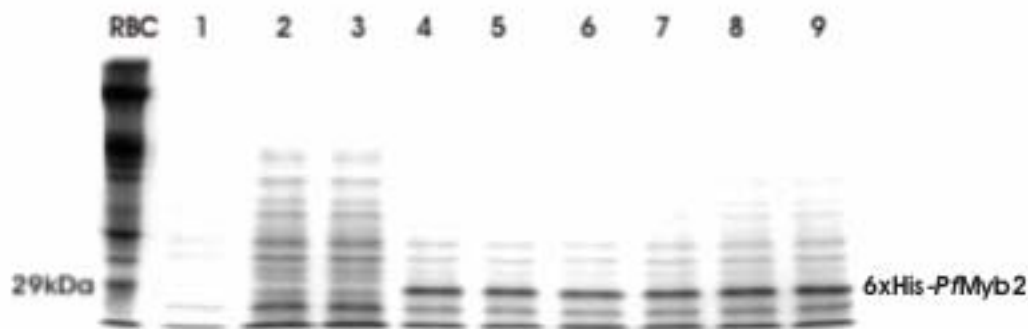


Figure 17-Expression of 6xHis-*PfMyb2*

Whole bacterial cell extracts from uninduced vector control (1), induced vector controls (2, 3), and induced vector constructs (4-10) were electrophoresed alongside a red blood cell membrane marker (RBC) on a 10% acrylamide gel. 4,5,6 represent samples induced with 0.4mM IPTG for 3,4 or 5 hours respectively, whilst 7,8,9 are samples induced with 1mM IPTG, also for 3, 4, or 5 hours respectively. All samples were induced at 37°C. Band 7 of the RBC membrane has a molecular weight of 29kDa. 6xHis-*PfMyb2*, deduced from amino acid sequence to be 24.8kDa, is evident only in lanes 4-9.

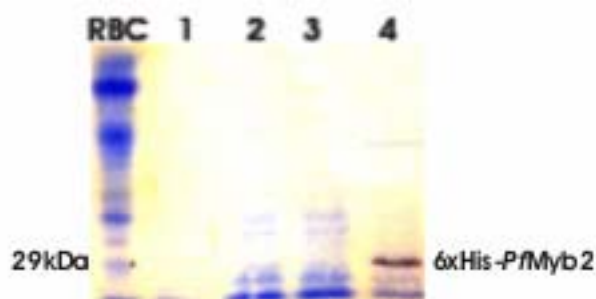


Figure 18- Immunoblot to confirm the expression of 6xHis-*PfMyb2*

Lanes 1-4 from the Coomassie Blue-stained gel in Figure 17 were transferred to nitrocellulose membrane and incubated overnight with anti-His antibody raised in mouse. 6xHis-*PfMyb2* was visualised with 1,3, diaminobenzidine which produces a brown colour. RBC: red blood cell membrane marker; 1: uninduced vector control; 2,3: induced vector controls; 4: induced 6xHis-*PfMyb2* construct.

3.3.2.1 Purification of 6xHis-PfMyb2 under native conditions

Separation of whole bacterial cell extracts into soluble and insoluble fractions indicated that most of the expressed protein was insoluble, as evidenced by SDS_PAGE analysis (Figure 19). This correlated with bioinformatic analysis of 6xHis-PfMyb2, which indicated that over-expression in *E. coli* would yield protein that had a 58% probability of being insoluble (www.biotech.ou.edu).

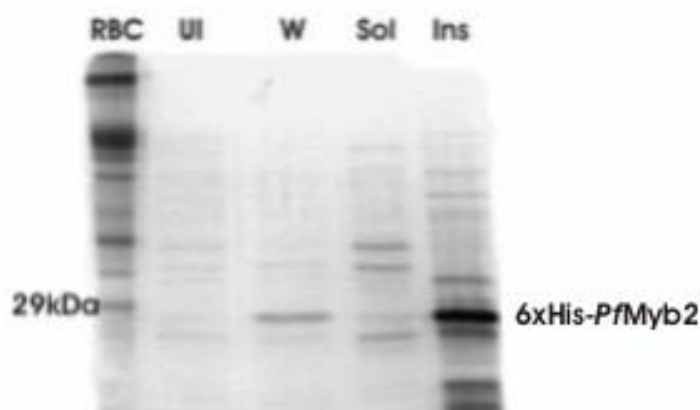


Figure 19-Assay to assess the solubility of 6xHis-PfMyb2

Cultures were induced with 0.4mM IPTG at 37°C for four hours. Soluble and insoluble protein fractions were analysed by SDS-PAGE. Most of the expressed protein is evident in the insoluble fraction. **RBC**: red blood cell membrane marker; **UI**: uninduced control; **W**: induced whole extract; **Sol**: soluble fraction; **Ins**: insoluble fraction.

Conditions that were varied in an attempt to obtain a significant yield of soluble protein included induction of the cultures at room temperature, in the presence of glycyl-glycine (Gosh et al., 2004) as well as at varying optical densities. These all decrease the rate of recombinant protein production, thereby allowing the cells sufficient time to correctly fold the newly synthesized protein rather than sequester it. It shows how the presence of glycyl-glycine had no marked effect on the expression profile as most of the expressed protein remained in the insoluble fraction.

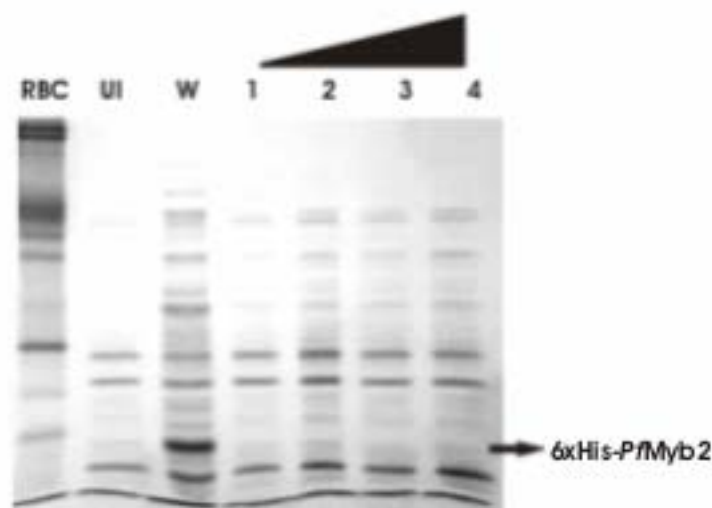


Figure 20- Effect of glycyl-glycine on the solubility of 6xHis-PfMyb2

6xHis-PfMyb2 was induced in the presence of increasing concentrations of glycyl-glycine, and the soluble fractions analysed by SDS-PAGE alongside an erythrocyte membrane marker (**RBC**). **UI**: uninduced control; **W**: induced whole extract; **1**: no glycyl-glycine added; **2**: 0.1M glycyl-glycine; **3**: 0.2M glycyl-glycine; **4**: 0.5M glycyl-glycine. Increasing the glycyl-glycine concentration to 1M also had no significant effect on solubility (not shown).

Despite some purified soluble protein being obtained, the concentration was always too low for use in the mobility shift reactions. From a 200ml culture, a 100µl elution would contain approximately 500ng purified protein, at a concentration of only 5ng/µl. As the maximum volume that could be included in the EMSA binding reaction was 8µl protein, this would equate to only 40ng protein in the reaction which may influence the detection sensitivity.

3.3.2.2 Purification of 6xHis-PfMyb2 under denaturing conditions

Attempts to purify the protein from inclusion bodies (3.2.2.5) were more successful as a pure pull-down with nickel beads was obtained (Figure 21). However, to obtain functional protein it had to be renatured through step-wise dialysis to remove any residual urea (3.2.2.6). Figure 22 shows the precipitate that was formed in 4M urea, which could have been due either the inability of the denatured protein to regain its

native conformation , or possibly the dialysis was too rapid to allow correct refolding of the protein.

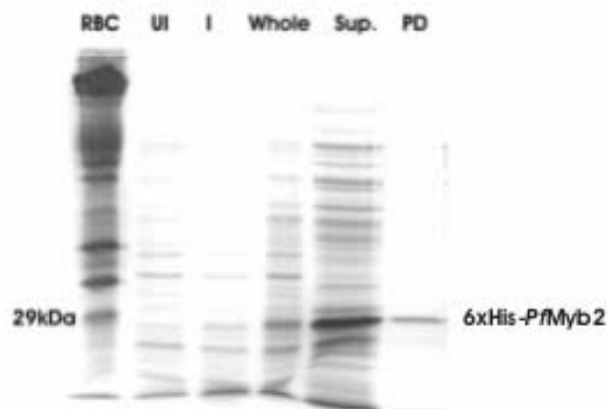


Figure 21- Purification of 6xHis-PfMyb2 under denaturing conditions

Inclusion bodies containing *PfMyb2* were denatured in 8M urea, purified by pull-down on Ni^{2+} magnetic beads, and electrophoresed on a 10% SDS gel alongside an erythrocyte membrane marker (**RBC**). **UI**, uninduced control; **I**, induced control; **whole**, induced whole extract; **Sup**, supernatant following pull-down of 6xHis-PfMyb2 and **PD**, pull-down containing pure denatured *PfMyb2*.



Figure 22- Precipitation of 6xHis-PfMyb2 during dialysis

P1, P2, P3: Pellets from 4, 2 and 0M urea respectively indicating precipitated 6xHis-PfMyb2. **S1, S2, S3**: supernatants from the same fractions; **RBC**: erythrocyte membrane proteins.

Thus despite initially obtaining a good yield of protein, the protein was lost through precipitation during the dialysis procedure. The dialysis protocol was not optimised

as there was no guarantee that the protein would regain its native conformation. Rather, an *in vitro* translation kit was used in an effort to obtain sufficient yield of soluble protein.

3.3.2.3 *In vitro* translation of 6xHis-PfMyb2

The PROTEINscript II *in vitro* translation kit uses either plasmid DNA or PCR product as template for the transcription and subsequent translation of the protein of interest in an optimised rabbit reticulocyte lysate. When using PCR product as a template, the components that would ordinarily be provided in the vector sequence had to be added to the PCR primer. These components included a histidine tag, a ribosome binding site and T7 promoter regions. The design of this primer is illustrated below (Figure 23).

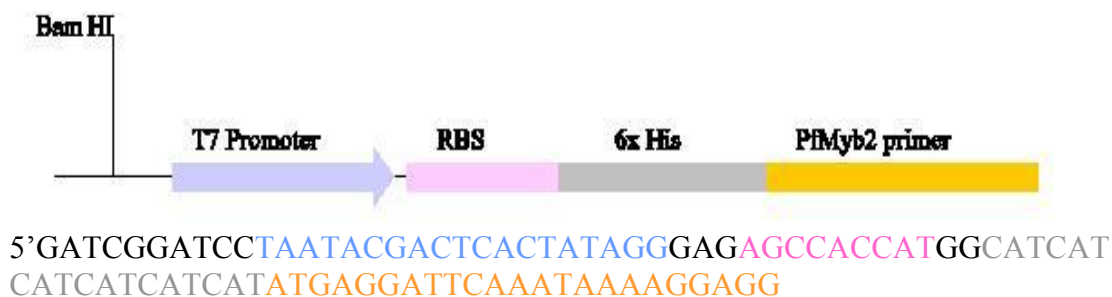


Figure 23-Primer for the *in vitro* translation of PfMyb2

The T7 promoter, Ribosome binding site (RBS) and histidine tag are indicated, as well as the gene-specific primer. Colour-coded nucleotides are shown below the image, with text colours corresponding to the various components of the primer.

This 84 nucleotide primer was used in combination with a gene-specific downstream primer to amplify PfMyb2 from *P. falciparum* genomic DNA, and the product of this reaction was then transcribed and used as a template in a translation reaction. Successful translation was confirmed by SDS-PAGE analysis in the presence of a negative control containing no RNA template (Figure 24).

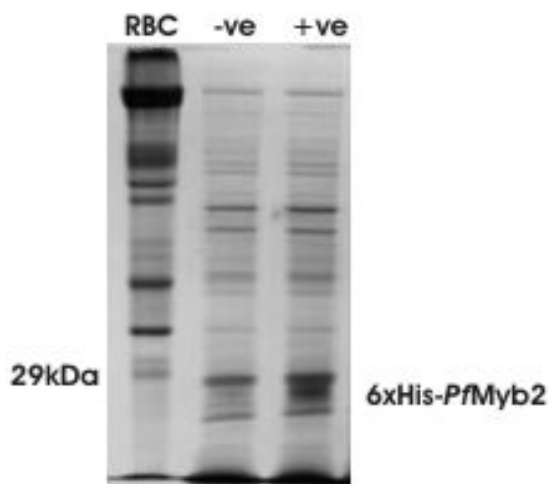


Figure 24- SDS-PAGE analysis of *in vitro* translation products

A negative control was electrophoresed alongside the reaction expressing 6xHis-PfMyb2.

Protein obtained in this manner was soluble, but again the yield was low, with only approximately 100ng protein obtained from a 50 μ l reaction. Setting up multiple reactions and pooling them increased the yield to approximately 500ng, and the concentration was increased to 10ng/ μ l by eluting into 50 μ l, but this was still too low for the EMSA reactions. Therefore the Overnight Express™ Autoinduction system was used in an attempt to obtain a high yield of concentrated protein.

3.3.2.4 Expression of 6xHis-PfMyb2 using Overnight Express™

The use of the Overnight Express™ Autoinduction system vastly improved the yield of soluble protein, as shown in Figure 25. 200ml cultures were grown for 24 hours at room temperature, the cells lysed in 3ml BugBuster® and protein was purified from the soluble fraction. Approximately 4 μ g protein was obtained in a 100 μ l elution volume, a concentration of 40ng/ μ l. Thus approximately 320ng protein could be used in the EMSA reaction thereby ensuring that the protein concentration was not a limiting factor.

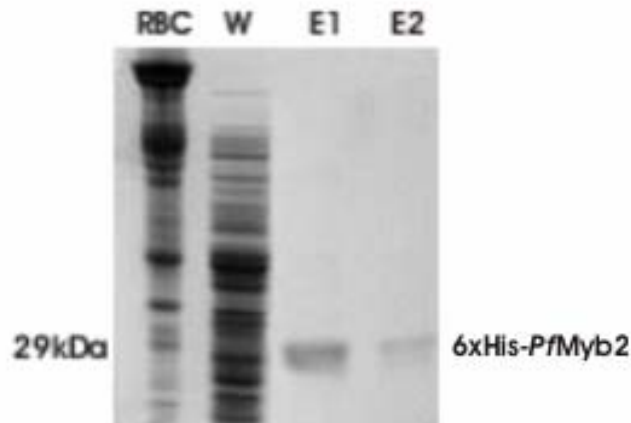


Figure 25- Purification of 6xHis-*PfMyb2* from the Overnight Express™ system

Purified 6xHis-*PfMyb2* was electrophoresed on a 10% acrylamide gel. **RBC**: red blood cell membrane marker; **W**: induced whole extract; **E1**: elution in 50μl imidazole; **E2**: a second elution in 100mM imidazole. 5μl of 80μl elutions are shown.

3.3.3 Binding studies with 6xHis-*PfMyb2*

Purified recombinant protein was incubated with double stranded oligonucleotides representing consensus Myb regulatory elements (A-3). *Pfmap1* and *Pfcrk1* represent portions of the promoter region of the parasite mitogen activated protein kinase (map1) and cyclin related kinase (crk1) genes respectively. c-Myb oligonucleotide represents the consensus c-Myb binding sequence as described by Kumar et al (2004), whilst the oligonucleotide denoted mim-1 represents a portion of the promoter region of the chicken mim-1 gene. NfKB was used as a non-homologous competitor and negative control. The oligonucleotide alignment with the consensus binding domain is shown in Figure 26. Whereas the oligonucleotides from the two vertebrate sequences conform exactly to the consensus sequence, the two *P. falciparum* sequences are slightly divergent, highlighting the unique gene sequences that are prevalent in this organism.

consensus	-----YAACKG-----	6
cMyb	-TACAGGCATAACGGTTCCGTAGTGA	25
mim-1	ACACAT-TATAACGGTTTT-TAGC--	22
Pfcrk1	-TAAA--AATAACCGACAC-CAAAAA	22
Pfmap1	-TATTTGAGAACTGGATGAACAG---	22
NfKB	-AGTTGAGGGGACTTTCCC-AGGC--	22

Figure 26- Oligonucleotides used for EMSA analysis

The oligonucleotides used for mobility shift assays with 6xHis-*PfMyb2* were compared using ClustalW software. The consensus nucleotides that are recognised by Myb proteins are highlighted green, whilst the divergent domains in the two oligonucleotides derived from *P. falciparum* genes are in orange. Y: T or C; K: G or T

Each experiment was performed with a different sample of purified protein which led to inconsistent binding; the gels shown below represent a selection of experiments where binding was observed.

3.3.3.1 6xHis-*PfMyb2* binds *in vitro* to Myb regulatory elements

The first evidence that recombinant *PfMyb2* was able to bind *in vitro* came from the fluorescence-based method of analysis (Figure 27), where a band-shift was evident when *Pfcrk1* and *Pfmap1* oligonucleotides were incubated with protein. Analysis with this method does not allow competition experiments, hence incubation with 100x excess of homologous competitor does not remove the signal, as in lane 5 below. Oligonucleotides were thus end-labelled with $\gamma^{32}\text{P}(\text{ATP})$ so that they could also be detected via autoradiography.

Pfmap1 in Figure 27 was end-labeled, so that the same gel could be dried and exposed to an autoradiograph. Here, the decrease in intensity of the signal in the presence of excess unlabelled oligonucleotide can be clearly seen in lane 3 (Figure 28) which points to a specific interaction between oligonucleotide and protein. The autoradiograph also indicates that 6xHis-*PfMyb2* forms two complexes when bound to oligonucleotide, something not evident in the less sensitive fluorescence-based assay.

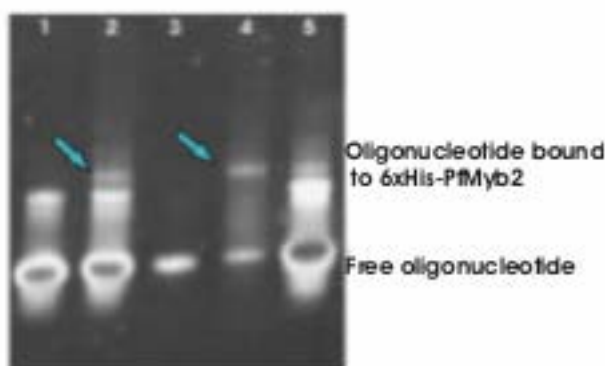


Figure 27-6xHis-PfMyb2 binds to *Pfcrk1* and *Pfmap1*

EMSA gel stained in SybrGreen and visualized under UV light. Arrows indicate the complex formed by 6xHis-PfMyb2 bound to oligonucleotide. 1: *Pfcrk1* oligonucleotide 2: *Pfcrk1* oligonucleotide with 6xHis-PfMyb2; 3: *Pfmap1* oligonucleotide; 4 *Pfmap1* oligonucleotide with 6xHis-PfMyb2; 5: 100x excess of homologous (*Pfmap1*) unlabelled competitor.

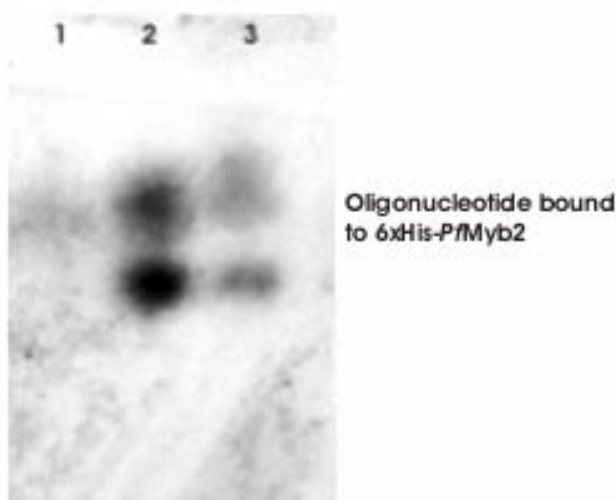


Figure 28- Autoradiograph indicating an interaction between 6xHis-PfMyb2 and *Pfmap1*

Overnight exposure of lanes 3, 4 and 5 of the gel shown in Figure 28. 1: *Pfmap1* oligonucleotide; 2: *Pfmap1* oligonucleotide incubated with 6xHis-PfMyb2; *Pfmap1* oligonucleotide incubated with 6xHis-PfMyb2 in the presence of 100 fold excess of unlabeled homologous competitor (*Pfmap1*).

Following confirmation that 6xHis-PfMyb2 was able to bind to the two slightly divergent *P. falciparum* regulatory elements, its ability to interact with the vertebrate

consensus sequences, cMyb and mim1, was assayed. A composite of three different mobility shift assays performed with cMyb oligonucleotide, using 6xHis-*PfMyb2* from different purifications, is shown in Figure 29.

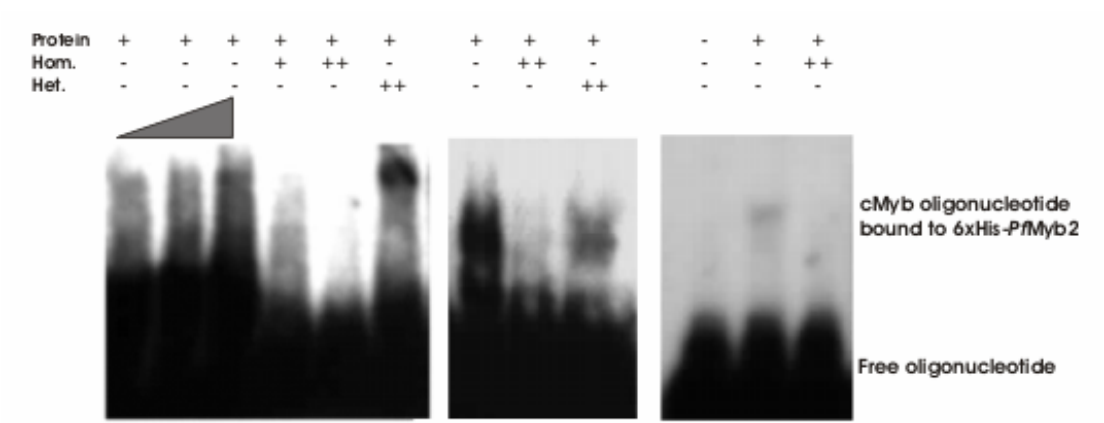


Figure 29- 6xHis-*PfMyb2* binds to the consensus cMyb oligonucleotide

Overnight exposure of three independent mobility shift assays using 6xHis-*PfMyb2* from different purifications, and the consensus cMyb oligonucleotide. The block triangle represents increasing concentrations of protein (100ng, 150ng, 200ng), and the components of each reaction are indicated as being included (+) or excluded (-). **Hom:** 10x (+) or 100x (++) excess of unlabelled cMyb oligonucleotide; **Het:** 100-fold excess of unlabelled NfKB oligonucleotide; all reactions contained 40ng oligonucleotide and 200-300ng protein.

The increasing intensity of the band representing oligonucleotide bound to 6xHis-*PfMyb2*, the loss of signal in the presence of unlabelled homologous competitor and the reappearance of the signal where the unlabelled competitor is non-homologous all indicate a specific interaction between the cMyb oligonucleotide and 6xHis-*PfMyb2*. As opposed to the two complexes formed between 6xHis-*PfMyb2* and *Pfmap1*, only one complex is formed from the interaction with cMyb.

The final assay was to determine the interaction between 6xHis-*PfMyb2* and the oligonucleotide representing a portion of the chicken mim-1 gene promoter (Figure 30). The smearing of the mim-1 oligonucleotide makes it difficult to ascertain the

specificity of the interaction, but binding can be clearly seen in lane 2. As with *Pfmap1*, two complexes are formed when 6xHis-*PfMyb2* binds to *mim-1*.

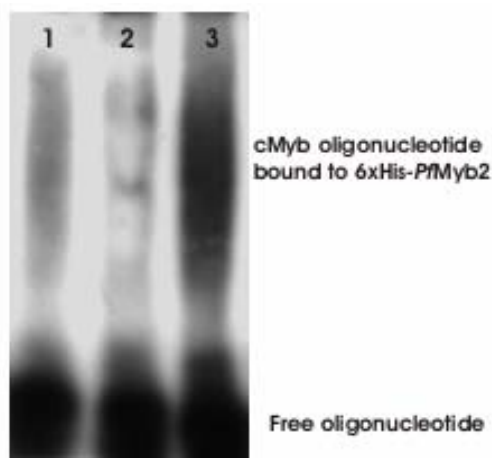


Figure 30- 6xHis-*PfMyb2* binds to the *mim-1* oligonucleotide

Overnight exposures of a mobility shift assay performed with *mim-1* oligonucleotide. **1:** *mim-1*; **2:** *mim-1* with 6xHis-*PfMyb2*; **3:** *mim-1* and 6xHis-*PfMyb2* in the presence of 100 fold excess of unlabelled homologous (*mim-1*) competitor.

No binding was observed when 6xHis-*PfMyb2* was incubated in the presence of NfKB oligonucleotide, which did not contain the consensus Myb recognition sequence, as indicated by the example in Figure 31.



Figure 31- 6xHis-*PfMyb2* does not bind to the NfKB oligonucleotide

This overnight exposure of a mobility shift assay performed with the negative control NfKB shows that no interaction occurs. This oligonucleotide consistently left a remnant in the well when electrophoresed on an acrylamide gel. **1:** NfKB oligonucleotide; **2:** NfKB incubated with 6xHis-*PfMyb2*.

3.3.4 Bioinformatic analysis of *PfMyb2*

PfMyb2 (PF10_0327) is described in the PlasmoDB database as a 2748bp gene encoding Myb2 protein (A-3). This single exon gene, located on chromosome 10, is expressed predominantly in the asexual parasite stages, with a decline in expression in gametocytes (Figure 32).

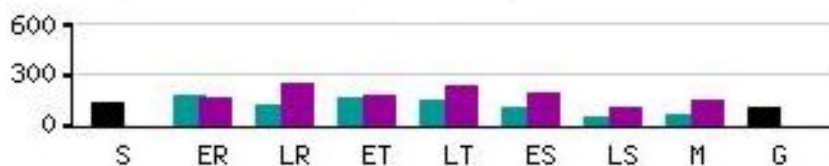


Figure 32- mRNA Expression profile of *PfMyb2*

Absolute expression data for *PfMyb2* mRNA were obtained from the PlasmoDB database (www.plasmoDB.org/). Expression is maximal during the late ring and late trophozoite stage, but is evident at low levels throughout the intraerythrocytic developmental cycle of the parasite. **S**: schizonts; **ER**: early rings; **LR**: late rings; **ET**: early trophozoites; **LT**: late trophozoites; **ES**: early schizonts; **LS**: late schizonts; **M**: merozoites; **G**: gametocytes. *Plasmodium* developmental stages were synchronised by sorbitol (green) or temperature (purple) (Le Roch et al., 2003).

The binding domains of Myb proteins are highly conserved at the amino terminal, generally consisting of three tandem repeats (R1, R2 and R3) of approximately 50 amino acids, with regularly spaced tryptophan residues which have been shown to be critical for the DNA binding activity of the protein. These binding domains specifically recognise the consensus sequence YAACT/GG (Biedenkapp et al., 1988), where Y represents T or C.

To date, *PfMyb1* (PF13_0088), a 50kDa transcription factor, is the only member of the Myb family to be characterised in *P. falciparum* (Boschet et al., 2004). ClustalW software (<http://www.ebi.ac.uk/clustalw/index>) was used to compare the amplified portion of *PfMyb2* binding domains with the binding domains of *PfMyb1* (Figure 33). Despite the fact that the binding domains are at the carboxy-terminus of *PfMyb1* and the amino terminus of *PfMyb2* there is still a good correlation (25%) between the

PfMyb2 MRIQIKGGIWKNCDEVLKAAVMKYGLNNWSRVASLLVR-KSAQCCKARWYEW-LDPSVR 58
PfMyb1BD IFSYTK-----NDKNIEHNFLYFSETFWNEVSEKLSNNQNAKECQKMQLYGCFFEDDK 53
: * : :: : . : . . *.*: . * : : ***: *: : :

PfMyb2 KTEWNKEEEEEKLHLAKLFP-TQWRTIAPIVG--RTAQCLEHYEYLLDEAEGKVYDKNK 115
PfMyb1BD QKKTKDEVDKLLCLSKKYEQRNWKTARELNTNRSPLSCFEQYIKIN----KLYE-NK 107
.:*.**: *** **: : :*: ** :. *: . *:*** : *:*: **

PfMyb2 NPRHLRPGEIDPAPESKPARADPVMDDEDEKEMLAEAKARLANTKGKKAKRKAREKQLEQ 175
PfMyb1BD EKVKLRIAFNVLEDIQLOILVSIIGDKN---WAEVKKHMESLNSNTSRIKKRKTNLNF 163
: :*. : : : :*: **.* : : . :.::: * *:..*

PfMyb2 ARRALLQKKRELKAAGITSL 196
PfMyb1BD FE----- 165

The conserved tryptophan (w) residues are coloured blue, whilst the two conserved cysteine (c) residues are indicated with an arrow. A conserved tyrosine residue (y), also characteristic of Myb binding domains, is coloured green. The three repeats that comprise the binding domains of *PfMyb1* (R1, R2 and R3) are underlined, and the *PfMyb2* binding domains are overlined. Identical (*), conservative (:) and semi-conservative (.) amino acid substitutions are indicated.

Each repeat encodes three helices which form a helix-turn-helix structure. The role of these tryptophans is to mediate DNA binding through preservation of the helical structure of the binding domains, rather than interacting directly with the DNA. Confirmation that *PfMyb2* conforms to this characteristic structure was obtained using SSPro software available from the Institute for Genomics and Bioinformatics (www.igb.uci.edu/servers/psss.html) (Figure 34).

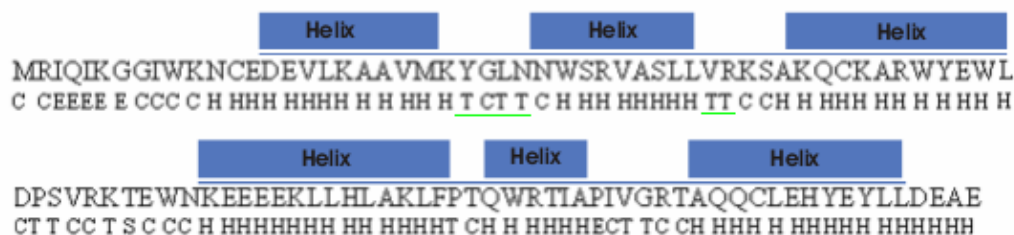


Figure 34- Helix-turn-helix structure of *PfMyb2* binding domains

SSPro was used for predicting secondary structure of the binding domains of *PfMyb2*, which verified that they conformed to the characteristic Myb structure. Helical regions (H) are indicated, as well as coils (C), extended strands (E), turns (T) and bends (S). The turn motifs are underlined. The *PfMyb2* binding domains are overlined.

Comparison between *PfMyb2* and its nearest human homologue, CDC5, reveals two characteristics common to many *P. falciparum* proteins (Figure 35). The first of these is that homology exists primarily in the functional (binding) domains, whilst the remainder of the protein shows minimal homology. An exception to this is the region indicated in parenthesis in Figure 35, which is highly homologous despite not being annotated as a functional domain. The significance of this region is not known. Secondly, many *P. falciparum* proteins contain long stretches of homopolymeric residues, most commonly asparagines (N), which in *Plasmodium* can vary from as few as 10 to more than a thousand amino acids.

It is also evident that the degree of homology between *PfMyb2* and human CDC5 is far greater than that between *PfMyb2* and *PfMyb1*. The divergence between these two members of the Myb family in the parasite may have arisen due to the dearth of specific transcription factors, and the need to perform unique functions.

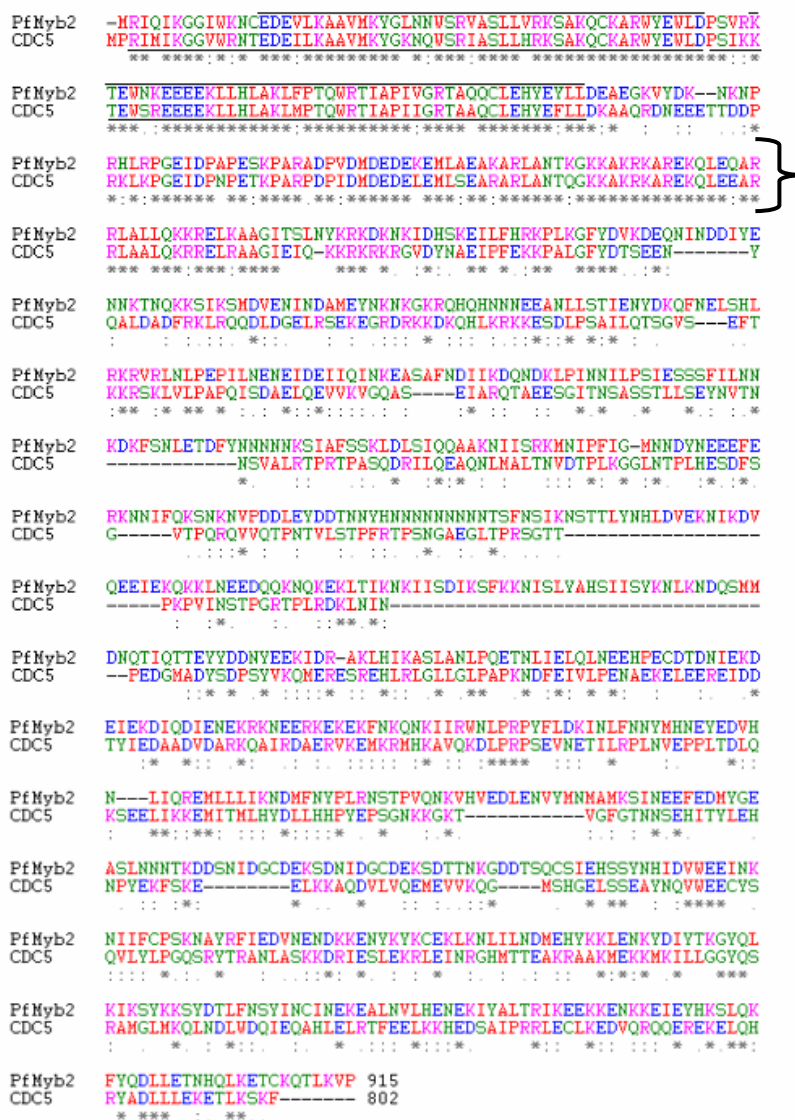


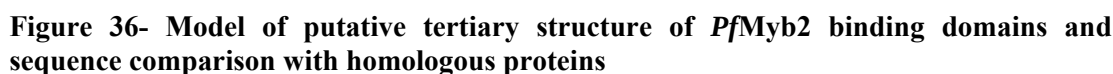
Figure 35- An alignment between *PfMyb2* and human CDC5

The protein sequences of *PfMyb2* and CDC-5 have been compared with ClustalW. The binding domains of *PfMyb2* are overlined, whilst the Myb binding domains of CDC-5 are underlined. The bracket indicates a region of the proteins that has no annotated function despite the high sequence similarity. Identical (*), conservative (:) and semi-conservative (.) substitutions are indicated.

A putative three dimensional structure of *Pf*Myb2 was obtained from Swiss-Model v3.5 (<http://swissmodel.expasy.org/>). This program compares a submitted sequence with all protein structures available in the database, and returns an alignment between these sequences and the query (target) sequence. In Figure 36, the proposed tertiary

structure of *PfMyb2* is shown. Despite entering all 196 amino acids from the region amplified, the returned structure contained only the first 93 amino acids, implying divergence of the remaining sequence from homologous sequences. Thus the last 10 amino acids of R2 are not shown. Only the conserved cysteine in R1 is shown, as the R2 cysteine is outside of the first 93 amino acids. Another cysteine residue is present immediately adjacent to R1, and this residue is highlighted for its potential to form disulfide bonds. Figure 36 (B) shows the alignment of the target sequence (*PfMyb2*) with its closest homologues.

The interaction between Myb and its cognate DNA is mediated by the helix-turn-helix structure of the binding domains. The general structure of the helix-turn helix (HTH) motif is shown in Figure 37 using the mouse Myb protein as a model. Of the three helices comprising the binding domain, the recognition helix fits into the major groove of the DNA where it mediates binding, whilst the stabilising helix supports the structure. It has not been determined which helices play these respective roles in *PfMyb2*. The role of the tryptophan residues in stabilising the structure, as opposed to interacting directly to the DNA, can be clearly seen.



B: Sequence comparison between the binding domains of *PfMyb2* (**target**) and closest homologous proteins. **1h8aC:** Myb Transforming Protein, Chain C (Homo sapiens); **1mseC:** Myb DNA binding domain, (Mus musculus); **1h89c:** Myb proto-oncogene protein (Homo sapiens); **1h88c:** Myb proto-oncogene protein (Homo sapiens).



Figure 37 -Model of the interaction between Myb and its DNA binding domain

The binding domains of mouse cMyb (Morikawa et al., 1995) are shown bound to the double helix of DNA (green). The conserved tryptophan residues are in red.

The protein has been rotated in Figure 38 to give an idea of its 3-dimensional structure.

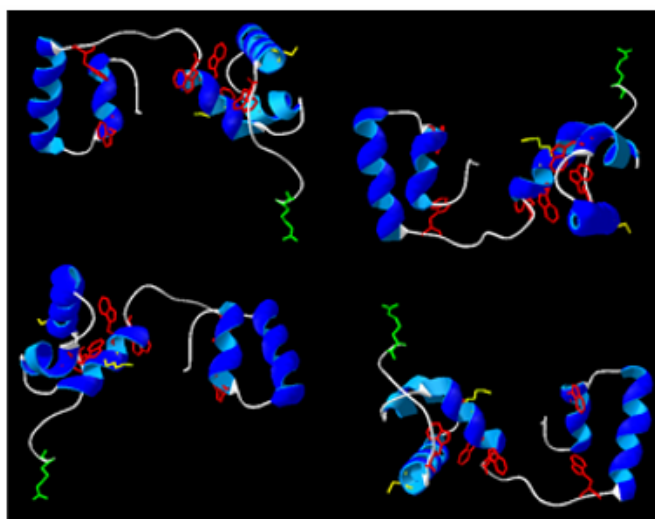


Figure 38- Three dimensional perspective of the structure of *PfMyb2*

PfMyb2 has been rotated around its central axis to get a different perspective of the three dimensional structure. The first arginine is indicated in green; the conserved tryptophans are coloured red and cysteines are shown in yellow. The protein was reduced with DTT prior to binding to prevent the formation of disulfide bonds.

3.4 Discussion

The identification and characterization of transcriptional regulators in *Plasmodium falciparum* is a critical step in enhancing our understanding of the gene regulatory mechanisms of the parasite. Despite the identification of numerous basal transcription factors, many of which mirror the eukaryotic gene regulatory apparatus, the only specific transcription factor to be characterized to date is the DNA binding protein *PfMyb1* (Boschet et al., 2004). Nuclear extracts containing *PfMyb1* bound in a sequence specific manner to oligonucleotides representing consensus *c-myb* DNA binding domains, as determined by polyclonal antibodies directed against *PfMyb1*. Purified recombinant *PfMyb1* was unable to bind, potentially due to incorrect folding or processing of the recombinant protein (Boschet et al., 2004). Though the importance of *PfMyb1* to the intraerythrocytic cycle of the parasite has been demonstrated through the use of RNAi (Gissot et al., 2005) its mechanisms of action and general significance remain unknown.

The work presented here represents the first example of functional studies performed with a recombinant *P. falciparum* specific transcription factor, *PfMyb2*.

3.4.1 Recombinant *P. falciparum* protein expression

An integral requirement for the functional analysis of proteins is the ability to express and purify the protein in a state that closely mimics its native conformation. One of the most common systems for the expression of recombinant proteins is the bacterium *Escherichia coli*. However, the expression of *P. falciparum* proteins in *E.coli* has often proven problematic due to the high AT content of the parasite genome (in excess of 80%) and the codon bias exhibited by this organism. Certain codons frequently used by *P. falciparum* in highly expressed genes are found in insufficient quantities in the bacterial host resulting in translation stalling, whereby the bacterium produces low yields or truncated forms of the protein (Baca and Hol, 2000). *PfMyb2* was therefore expressed in BL21 competent cells containing the RIL plasmid, which

encodes tRNAs commonly found in AT-rich organisms but which are particularly rare in *E. coli*, specifically arginine (R), isoleucine (I) and leucine (L).

Another criterion, not unique to *Plasmodium*, is that the bacterial host is highly sensitive to the production of foreign proteins, and over-expressed proteins often aggregate into insoluble inclusion bodies. Inclusion bodies form when the translation rate of the protein exceeds the capacity of the cell to correctly fold the protein. Purification of functional protein from inclusion bodies requires the use of a strong denaturant, followed by gradual removal of this denaturant under conditions optimal for protein refolding. There is no guarantee that the renatured protein will regain its native conformation, and functional studies may therefore be compromised, thus this approach was not pursued after initial attempts resulted in precipitation of the recombinant protein.

Slowing down the rate of translation is a primary consideration when inducing expression of recombinant protein. Despite experimenting with many of the parameters known to do this (induction was performed at different temperatures, with varying concentrations of IPTG and over a range of time periods), an insufficient yield was obtained to perform functional studies. None of the varied conditions of time or temperature had a marked effect on protein solubility, and the majority of expressed protein remained sequestered in inclusion bodies. The addition of a dipeptide such as glycylglycine to induced cultures has also been shown to increase the solubility of the expressed protein (Gosh et al., 2004). The mechanism by which this occurs is not well understood but three options have been postulated. The first is that the increased osmotic stress caused by the high concentration of dipeptide induces the expression of heat-shock proteins with chaperone-like activity, which enhances correct folding. Another option is that the dipeptide interacts directly with the expressed protein and acts as a chemical chaperone. Finally, the considerable energy spent by the bacterium in glycylglycine transport slows down the overall metabolic rate, including protein synthesis, thereby allowing sufficient time for

correct folding of the recombinant protein. Despite inducing cultures in up to 1M glycylglycine there was no increase in the solubility of expressed protein.

In an attempt to circumvent some of the problems associated with the bacterial expression of recombinant proteins, *PfMyb2* was expressed using a mammalian *in vitro* translation system, which uses optimised rabbit reticulocytes for the expression of recombinant protein. *In vivo*, reticulocytes are highly specialised cells primarily responsible for the synthesis of globin, which represents more than 90% of the protein made in the reticulocyte. These cells have lost their nuclei but contain the complete translation machinery for extensive protein synthesis. Soluble protein was obtained using this method, but the concentration remained too low for use in the mobility shift assays.

A high yield of soluble protein was ultimately obtained using the Overnight Express™ Autoinduction System, which allows bacterial growth to proceed to high cell density before induction by lactose. The return to a bacterial system enabled purification of sufficiently high yield of concentrated soluble protein, suitable for use in the mobility shift assays.

3.4.2 Functional analysis of 6xHis-*PfMyb2*

Two methods are commonly used for the assessment of protein/DNA interactions. If purified recombinant protein is available, the protein can be incubated with specific end-labelled DNA sequences and an interaction verified by autoradiography following gel electrophoresis. Alternatively, the same labelled oligonucleotides can be incubated with whole nuclear extracts and the interaction verified using an antibody directed against the protein of interest. In both cases the specificity of the interaction can be confirmed with competition experiments using unlabeled homologous or non-homologous competitor oligonucleotides. With nuclear extracts the *in vivo* state of the protein can be approximated, as any chaperones or co-factors

that may facilitate binding are present. The former method is more direct and the absence of contaminating proteins means the results are often more conclusive. However, obstacles to overcome include the solubility of the expressed protein and the manner in which the expressed protein folds. Nonetheless this was the preferred method in this research as an antibody to *PfMyb2* was not available.

Mobility shift assays were performed with purified 6xHis-*PfMyb2* and oligonucleotides representing a consensus Myb regulatory element (MRE) (Kumar et al., 2004), putative MREs identified in *Pfcrk1* and *Pfmap1* gene promoters (Boschet et al., 2004) and a MRE identified in the chicken *mim-1* gene promoter. Initial experiments to deduce the binding potential of 6xHis-*PfMyb2* yielded inconsistent results. Every time the protein is expressed, be it in a mammalian or bacterial system, the manner in which the protein folds may differ and this could influence its binding capability. Furthermore, the amount of protein in the reaction seemed to be a determining factor in the outcome of the experiment, implying that the results were dependant on the limits of detection of the method used for EMSA experiments as well as the concentration of protein. Increasing the sensitivity of detection by using autoradiography still required prolonged exposure of the autoradiograph to visualise the shifted bands. Smearing of some of the oligonucleotides during electrophoresis also made verification of the results more complicated, particularly in assessing the specificity of the reaction. However, upon increasing the yield and concentration of purified protein with the Overnight Express™ system results were more conclusive. Data indicated that 6xHis-*PfMyb2* bound to all oligonucleotides containing the consensus MRE sequence, in addition to the partially divergent *P. falciparum*-derived sequences, and the specificity of the interactions were confirmed with competition experiments. No binding was observed with NfKB oligonucleotide which again indicated that the observed interactions were specific, and reliant upon the presence of the Myb recognition sequence.

3.4.3 Structural analysis of the Myb protein family

The Myb family of oncogenes has been highly conserved during evolution, being present in all vertebrates studied to date as well as some invertebrates (Lipsick et al., 2001). They play an essential role in the cell, participating in the regulation of genes involved in cell cycle progression and differentiation. In addition to their function, the tripartite structure of these proteins has also been conserved, consisting of an N-terminal binding domain, central transactivation domain and a C-terminal negative regulatory domain (Kanei-Ishii et al., 1996). The highest degree of conservation is within the binding domains, which consist of three repeats (R1, R2 and R3) of approximately 50 amino acids containing regularly spaced tryptophan residues (Kanei-Ishii et al., 1996). Each repeat is predicted to form a helix-turn-helix structure which is critical for DNA binding (Pinson et al., 2001). Mutation analysis of c-Myb constructs (Saikumar et al., 1990) has determined that the role of these tryptophans is to mediate DNA binding through preservation of the helical structure of the binding domains, rather than interacting directly with the DNA. The creation of a hydrophobic platform that allows adjacent basic amino acids to bind DNA is essential for this process. In addition to the conserved tryptophan residues, the Myb proteins contain a critical cysteine residue in their second repeat (R2), which needs to be in a reduced state for DNA binding to occur *in vitro* (Guehmann et al., 1992). It is speculated that Myb function may therefore be regulated by a reduction-oxidation system (Guehmann et al., 1992).

The transactivation and negative regulatory domains of the conserved Myb proteins have not been identified in *PfMyb2*. This could be due to the incomplete annotation of the protein in the PlasmoDB database, or due to a degree of sequence divergence of these domains from homologous proteins.

Many proteins in *Plasmodium* have been shown to be up to 50% larger than orthologous proteins and are characterised by long stretches of homopolymeric

residues, most commonly asparagines (Aravind et al., 2003). These low complexity regions, when seen in other eukaryotes, generally occur in transcription factors and nuclear proteins, whereas they are prevalent in the vast majority of *Plasmodium* proteins (Pizzi and Frontali, 2001). The biological significance of these regions remains unknown, but the resultant sequence variation from host proteins makes them potential candidates for drug targets.

Structurally, *PfMyb2* conforms to the characteristics of a Myb transcription factor, with all conserved cysteines and tryptophans evident in the protein. In addition the binding domains are conserved at the amino terminal end of the protein, and form the characteristic helix-turn-helix structure when modelled with SwissPDB viewer. The general structure of the helix-turn helix (HTH) motif is composed of two α -helices connected by a short stretch of amino acids, the “turn”. The recognition helix fits into the major groove of the DNA where its amino acids play a major role in recognising the specific DNA binding region. This interaction is a versatile one, as it is thought that the manner in which the recognition helix is presented to the binding domain varies between proteins, increasing the number of proteins that can interact with the same motif (Alberts et al., 2002).

PfMyb1 and *PfMyb2* both contain conserved cysteine residues in R1 and R2 (cys⁴⁶, cys⁹⁵), but it is unclear whether the reduced state of one or both of these residues is required for DNA binding. These appear to be quite distant from each other in terms of amino acid location, but depending on the manner in which the protein folds they could be sufficiently close to form disulfide bridges. Although there was no evidence in *Plasmodium* that these would prevent DNA binding through the formation of disulfide bonds, this possibility was removed by reduction of the protein in 2mM DTT prior to binding. Preliminary data indicated that this was important for binding since attempts to replicate an experiment were unsuccessful when using protein that had been stored at 4°C for a few days, potentially because in that time the DTT had oxidized allowing disulfide bridges to reform (Getz et al., 1999).

In contrast to the majority of Myb proteins, *PfMyb1* has C-terminal binding domains (Boschet et al., 2004), as is the case in *D. discoideum*. Analysis of the two binding domains of *PfMyb2* indicated that they were most similar to R1 and R2 of *PfMyb1*. In vertebrates, R2 and R3 constitute the minimum requirements for sequence-specific DNA binding whilst R1 appears to have no direct interaction with DNA (Ogata et al., 1994). Despite this, deletion of a portion of R1 in c-myb abrogated DNA binding (Sala et al., 1995), and these authors thus speculate that R1 may activate a mechanism of c-myb function that involves interaction with other cellular factors. This could have important implications for functional studies using recombinant protein in an environment devoid of cellular factors.

It is not surprising that *PfMyb2* recognises and binds to a range of sequences. All of the oligonucleotides contained elements of the consensus binding domain, despite the slight divergence of the *P. falciparum*-derived sequences. A similar binding pattern was observed with *PfMyb1* (Boschet et al., 2004), leading to speculation that in an organism where very few specific transcription factors have been identified, one transcription factor may be responsible for the activation of numerous genes, and hence recognise different regulatory elements. Furthermore, studies have shown that *PfTBP* also interacts with highly divergent TATA-like motifs suggesting that the structure of the protein may be more important for binding than base-pair complementarity (Ruvalcaba-Salazar et al., 2005).

3.4.4 Role of *PfMyb2* in parasite development

Having ascertained that recombinant *PfMyb2* protein binds to consensus MRE sequences, and therefore established that it is a transcription factor in *P. falciparum*, one can speculate on the role that this specific transcription factor plays in the parasite. There is evidence for the critical role played by *PfMyb1* in the control of genes involved in cell cycle (Gissot et al., 2005) which is consistent with the conserved function of the Myb protein family. There is every reason to believe that

the closely related *PfMyb2* has a similarly important role in the cell. Expression of both genes is evident throughout the intraerythrocytic life of the parasite and both could therefore function during asexual or gametocyte stages. Two of the Myb regulatory elements to which *PfMyb2* bound are found in the promoters of genes that are involved in signalling pathways that control progression of the cell cycle. Despite binding to similar regulatory elements, the affinity or kinetics with which *PfMyb1* and *PfMyb2* bind could be different, and this redundancy could further denote the importance of these two proteins.

Pfmap-1 and *Pfcrk-1* are two genes expressed predominantly in asexual and sexual stages respectively, whose promoters contain putative Myb regulatory elements (Doerig et al., 1995; Doerig et al., 1996). *Pfcrk-1* encodes a novel cdc2-related protein kinase and is most closely related to a family of negative regulators of cell growth (Doerig et al., 1995). *Pfmap-1* encodes a novel protein kinase that contains a conserved cdc-2 region. Comparison with closely related proteins suggests that *Pfmap-1* plays an important role in adaptive response and signal transduction (Doerig et al., 1996). *PfMyb2* binds to oligonucleotides representing promoter regions of both of these genes, which implies that the protein functions in both the asexual parasite and gametocyte stages of the life cycle.

Some vertebrate Myb proteins participate in regulation of the G₀/G₁ transition of the cell cycle (Kolchanov et al., 2002). If a similar role for this protein exists in the parasite, *PfMyb2* could be an important component in the arrest of gametocytes in the G₀ stage of the cell cycle, where they remain until they reach the mosquito midgut. Although a correlation has not been conclusively shown between the classical cell cycle progression and the *P. falciparum* life cycle it is feasible that such a regulatory role may exist for this transcription factor.

3.5 Conclusion

The work presented here has focussed on gaining insight into one of the few annotated specific transcription factors in *P. falciparum*, *PfMyb2*. Recombinant 6xHis-*PfMyb2* was expressed, purified and shown to bind to consensus Myb regulatory sequences derived from *Plasmodium* and chicken gene promoters. Based on homology to conserved proteins, *PfMyb2* probably plays a role in regulating genes involved in growth and differentiation of the parasite. The identification and characterisation of transcription factors and elucidation of the roles they play in regulating the complex life cycle of the malaria parasite will contribute to the knowledge base surrounding the gene regulatory mechanisms of the malaria parasite and is fundamental to efforts to ultimately eradicate this disease.

3.6 Concluding remarks

In an era of high-throughput genomic profiling, smaller laboratories can still make valuable contributions to knowledge through the use of accessible techniques such as differential display. This research began by comparing the gene expression between asexual parasite and gametocyte stages of *P. falciparum*, with a view to identifying novel differentially expressed genes as well as genes that may be implicated in the gametocytogenesis process. The identification of such differentially expressed transcripts is important in garnering a better understanding of the complex gene regulatory mechanisms at play in the parasite that control all aspects of its intricate life cycle. The adaptability of the parasite to vastly different host environments, combined with sophisticated immune evasion strategies, has contributed to the parasite's ability to evade all attempts at eradication over the last century.

Numerous genes were identified whose differential expression was a consequence of the diverse requirements of the two developmental stages rather than playing a causative role in sexual development. The pursuit of a transcription factor belonging to a family known to regulate cell cycle progression and differentiation was therefore a natural progression of this work.

Having overcome the difficulties inherent in purifying recombinant *P. falciparum* proteins, binding experiments were performed indicating that 6xHis-PfMyb2 bound to cMyb, Pfcrk1, Pfmap1 and mim-1 oligonucleotides containing the Myb consensus binding sequence. These findings are consistent with results obtained in analysis of PfMyb1 (Boschet et al., 2004), the only other member of this protein family to be characterised in *P. falciparum*. That these proteins are able to bind promiscuously could be a function of the paucity of specific transcription factors in *Plasmodium* requiring that they act on numerous genes. The significance of having two closely related transcription factors in the cell at similar times, which recognise and bind to the same consensus sequence, alludes to the importance of these transcription factors

in the parasite. Comparative analysis of the kinetics of these two proteins, and their respective binding affinities, will shed light on this.

This research has opened the door for a multitude of future investigations. Obtaining an antibody to *PfMyb2* and performing binding studies with nuclear extracts will allow analysis of *PfMyb2* in the presence of any chaperone proteins or co-factors that may facilitate binding specificity and fidelity. In addition, chromatin immunoprecipitation assays can be performed to analyse *ex vivo* protein/DNA interactions at the different intraerythrocytic time points of the parasite. These will facilitate a better understanding of the transcriptional apparatus of the parasite.

APPENDIX

A-1 Reagents

A-1.1 Parasite culture media

Freezing solution:

72ml of PBS (for freezing) were combined with 28ml glycerol, and sterilised by filtration through a 0.22µm filter. This was stored at -20°C until required. For freezing, pelleted cultures were resuspended in a 1:1 ratio with freezing solution, and stored in liquid nitrogen.

Phosphate Buffered Saline (PBS) (pH 7.4):

10mM Na₂HPO₄, 1.4mM KH₂PO₄, 142mM NaCl and 2.7mM KCl.

PBS for freezing:

80mM Na₂HPO₄; 30mM KH₂PO₄; 120mM NaCl

RPMI culture medium (incomplete):

RPMI culture medium was supplemented with 25mM HEPES, 2mM L-Glutamine, 50µg/ml gentamycin, 44µg/ml hypoxanthine and 20mM D-glucose. Stored at 4°C until required.

RPMI complete culture medium:

90ml incomplete culture medium was supplemented with 10ml inactivated human type AB plasma and 2.8ml 5% sodium bicarbonate. This was stored at 4°C and was used within 3-4 days. Plasma was inactivated at 56°C for two hours prior to use, centrifuged at 1100g for 15 minutes, and the supernatant collected and stored at -70°C in 10ml aliquots.

A-1.2 DNA analysis

Agarose gel:

Agarose was diluted to 1% in 1X TAE buffer and 1 μ l ethidium bromide (10 μ g/ μ l) was added per 10ml.

50x TAE buffer (pH 8.0):

2M Tris-acetate, 100 ml 0.5M EDTA pH 8.0 in 1000ml

10x TBE buffer:

890mM Tris base, 890mM Boric Acid, 20mM EDTA

TE buffer (pH 8.0):

10 mM Tris-HCl pH 8.0, 1mM EDTA

A-1.3 Recombinant protein expression

Coomassie Brilliant Blue stain:

0.5g Coomassie Blue R-250 was mixed with 25% isopropanol and 10% acetic acid in two litres.

Laemlli gel:

This consists of a lower (resolving) gel and an upper (stacking) gel. The resolving gel comprised 10% acrylamide, 0.1% bis-acrylamide, 0.375M Tris-HCl pH 8.8, 1% SDS, 50 μ l 10% APS and 20 μ l TEMED. The stacking gel comprised 4% acrylamide, 0.15% bis-acrylamide, 0.125M Tris-HCl pH 6.8, 0.15% SDS, 50 μ l APS and 10 μ l TEMED. Gels were electrophoresed at 35mA for 90 minutes with cooling to 4°C.

LB medium:

10g bactotryptone, 5g yeast extract, 10g NaCl and 10ml of 1M Tris-HCl pH 7.5 were combined in 1 litre water and sterilized by autoclaving (120°C, 15lb/in²).

LB agar:

LB medium was supplemented with 15g agar per litre. The solution was allowed to cool to the touch before the addition of sterile ampicillin to a final concentration of 100µg/ml.

Novagen BugBuster[®] Ni-NTA His Bind Purification kit**Binding buffer:**

0.5M NaCl, 20mM Tris-HCl pH 7.9, 5mM imidazole

Wash buffer:

0.5M NaCl, 20mM Tris-HCl pH 7.9, 20-60mM imidazole

Elution buffer:

0.25M NaCl, 10mM Tris-HCl pH7.9, 30-100mM imidazole

5X SDS Solubilisation buffer (pH 8.0):

50mM Tris-HCl, 5mM EDTA, 5% SDS, 0.5% bromophenol blue, 50% glycerol and 5% β-mercaptoethanol

SDS electrophoresis buffer (pH 8.3):

6.06g Tris base, 28.8g glycine, 2.0g SDS. H₂O to 1000ml

Sigma His-Select[™] HC Nickel Magnetic Beads**Binding buffer:**

50mM NaPO₄ pH8.0, 0.3M NaCl

Wash buffer:

50mM NaPO₄ pH 8.0, 0.3M NaCl, 20mM imidazole

Elution buffer:

50mM NaPO₄ pH8.0, 0.3M NaCl, 50-200mM imidazole

SOB medium:

20g tryptone, 5g yeast extract and 0.5g NaCl were combined in a final volume of 1000ml water and sterilised by autoclaving. 10ml filter-sterilised 1M MgSO₄ and 10ml filter-sterilised 1M MgCl₂ were added immediately prior to use.

SOC medium:

2ml sterile 20% glucose was made up to 100ml with SOB medium.

For denaturing conditions, urea was added to all buffers to a final concentration of 8M.

A-1.4 Electrophoretic mobility shift assays (EMSAs)**TBE acrylamide gel:**

5% glycerol, 0.25X TBE, 4% acrylamide, 0.2% bis-acrylamide, 75μl 10% APS and 20μl TEMED were combined. Gels were allowed to polymerise for two hours before running, and were pre-run at 160v (10mA) for 90 minutes before samples were loaded. Samples were electrophoresed at 4°C for 90 minutes at 100v (4mA) in a 0.25X TBE running buffer.

5X EMSA binding buffer (pH 7.4):

750mM KCl, 50mM Tris-HCl, 0.5mM EDTA.

1X EMSA binding buffer (pH 7.4) for dialysis:

0.3g DTT in 100ml 1X EMSA binding buffer.

A-2 Standard Laboratory Procedures

Microscopic analysis of stained blood smears:

Smears were fixed in methanol and were stained using either the Rapid Haematology Staining Kit as per manufacturer's instructions, or by immersion for 15 minutes in 10% Giemsa solution diluted in 6.7mM phosphate buffer (pH 7.1). Slides were washed briefly under running water and were analysed using oil immersion microscopy at 1000x magnification.

Preparation and electrophoresis of DNA sequencing gels:

Glass sequencing plates (30cm x 40cm) were thoroughly washed with Extran detergent and assembled with 0.4mm spacers and two 19 well combs in an EZCast gel casting boot. The gel was assembled in a BRL model S2 electrophoresis apparatus. The gel was pre-electrophoresed at 60W for 45 minutes so that it had reached approximately 47°C by the time the samples were ready to be loaded. Samples were electrophoresed at 60W constant power, with voltage and amperage set to be non-limiting in 1X TBE buffer for between two and four hours. When electrophoresis of more than six hours was required, power was reduced to 55W. The gel was exposed to X-ray film at -70°C overnight, and the film then developed using a Kodak automatic developer in the emergency ward of the Johannesburg General Hospital (Johannesburg, South Africa).

DNA sequencing gel:

A 60ml gel contained 6M urea, 1X TBE, 12µl TEMED, 0.3% bis-acrylamide, 6% acrylamide and 480µl 10% ammonium persulphate (APS).

Stop solution:

95% deionised formamide, 20mM EDTA pH 7.5, 0.1% bromophenol blue 0.1% xylene cyanol FF

Phenol: chloroform extraction and ethanol precipitation of DNA and RNA:

DNA was extracted with Phenol pH 8.0, whilst Phenol pH 4.2 was used for RNA extractions.

For extraction of DNA or RNA, 500µl 1:1 phenol: chloroform was added to the resuspended pellets, and centrifuged briefly. The aqueous supernatant was removed, and combined with 250µl phenol. Samples were mixed thoroughly, centrifuged briefly, and the supernatant again removed and precipitated with 100% ice cold ethanol.

2½ volumes ice cold 100% ethanol and 10% 3M sodium acetate were combined with the aqueous supernatants and left at -70°C for 30 minutes. Samples were electrophoresed at maximum speed at 4°C for 30 minutes, and the pellet washed in 70% ice cold ethanol. The pellet was air dried and resuspended in an appropriate volume of water. For RNA extractions 100% ethanol could be replaced with isopropanol.

Alkaline lysis plasmid DNA mini-preparation:

All centrifugation steps were performed at 15000g in a Sorvall RMC-14 microfuge at 4°C.

Bacterial cultures were collected by centrifugation and the cells resuspended in ice-cold lysis buffer (10mM EDTA, 50mM glucose, 25mM Tris-HCl pH 8.0) and incubated at room temperature for five minutes. 0.2M NaOH, 1% SDS solution was added to the lysed cells and mixed by gentle inversion. The mixture was neutralized by the addition of ice cold 5M potassium acetate (pH 4.8) and mixed by gentle inversion. The tubes were placed on ice for 15 minutes, allowing the chromosomal DNA and bacterial proteins to precipitate, and these could then be removed by centrifugation. The supernatant was transferred to a fresh 1.5ml Eppendorf tube and the DNA purified by extraction with phenol/chloroform (1:1) followed by ethanol precipitation.

Removal of RNA contamination from plasmid DNA:

Following precipitation with ethanol the DNA was pelleted by centrifugation, air-dried, and resuspended in 100µl TE buffer containing 1µl DNase-free RNase (500µg/ml). The reaction was incubated at 37°C for 30 minutes following which 40µl 5M potassium acetate (pH 4.8) and 260µl water were added to each tube. The DNA was extracted first with phenol/chloroform and then with chloroform, and precipitated with 2½ volumes of ice-cold 100% ethanol at -70°C for 30 minutes.

Protein transfer and immunoblot:

Following electrophoresis the SDS-acrylamide gel was soaked in transblot buffer (25mM Tris-HCl, 192mM glycine, 20% methanol) for ten minutes. A similarly sized piece of nitrocellulose membrane was cut and equilibrated in transblot buffer along with two pieces of filter paper. The gel was placed on one piece of filter paper in a wet blot cassette, and the membrane and second piece of filter paper were layered on top of the gel. The cassette was closed and placed into a Transblot system at 4°C with pre-cooled transfer buffer. The system was connected to a power supply and proteins transferred for four hours at 35V.

For protein transfer from SDS gels that had been stained with Coomassie Brilliant Blue, gels were thoroughly destained, rinsed in numerous changes of distilled water to remove the last traces of acetic acid (including an overnight wash), and proteins were recharged for one hour at 4°C in SDS incubation buffer (25mM Tris-HCl, 192mM glycine, 1% SDS) prior to protein transfer.

Nitrocellulose membrane containing the transferred proteins was incubated in 3% BSA in TBS (50mM Tris-HCl pH 7.5, 150mM NaCl) for one hour at room temperature, and washed in TBS. The blot was incubated overnight with primary antibody (mouse anti-His antibody) diluted 1:2000 in 1% BSA in TBS. The membrane was washed four times in TBS/0.05% Tween 20 for five minutes with rapid shaking. The secondary antibody was goat anti-mouse IgG conjugated to horseradish peroxidase Type IV. A 1:2000 dilution in 1% BSA in TBS was added to

the membrane and left to incubate for one hour. The blot was washed four times with TBS/0.05% Tween 20 for five minutes. To initiate the colour reaction the blot was covered in 4-chloro-1-naphthol.

4-chloro-1-naphthol:

15mg dissolved in 5ml methanol, and added to 25ml TBS. 12.5µl 30% hydrogen peroxide was added immediately prior to use.

When performing the colour reaction for stained gels, 4-chloro-1-naphthol was replaced with 1,3, diaminobenzidine.

1, 3, diaminobenzidine:

15mg was dissolved for 30 minutes in 25ml TBS in the dark with rapid stirring. 12.5µl 30% hydrogen peroxide was added immediately prior to use.

A-3 Plasmid maps and primer sequences

A-3.1 Primers for differential display

Arbitrary 13mers:

60-2: 5' AAACCTCCTTGATG 3'
60-1: 5' TTACACAAGCACA 3'
55-6: 5' CTCCTACTCTATG 3'
55-5: 5' CTCACTATCCATC 3'
55-4: 5' TCCCTTTAGCATC 3'
55-3: 5' AGGATAGGTGATG 3'
55-2: 5' CAATGCGTCTTCC 3'
55-1: 5' GTCCAATTAGATG 3'

Arbitrary 10mers:

AT70: 5' TGAAACTAGT 3'
AT60: 5' TATGACGTCT 3'
AT50: 5' GTGCAATGAG 3'

Anchored primers:

P129: 5' TTTTTTTTTTTTVA 3'
P130: 5' TTTTTTTTTTTTVG 3'

P131: 5' TTTTTTTTTTTTVC 3'

P132: 5' TTTTTTTTTTTT VT 3'

V= A/G/C

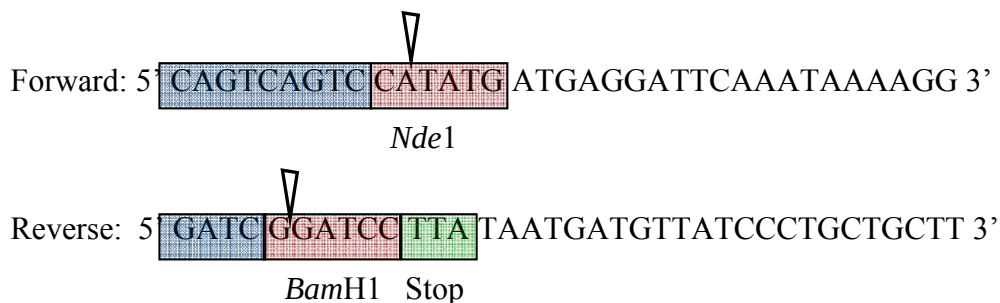
A-3.2 Nucleotide sequence of PF10_0327

ATGAGGATTCAAATAAAAGGAGGTATATGGAAAAATTGTGAGGATGAAGTTCTTAAAGCAGCTGTTAT
 GAAATATGGTTTAAATAATTGGTCAAGGGTCGCATCTTTATTAGTTCGTAAATCAGCTAAACAGTGTA
 AGGCGAGGTGGTATGAATGGCTTGATCCATCAGTTAGAAAAACCGAATGGAATAAAGAAGAAGAAGAA
 AAACTTTTACATCTAGCCAAATTGTTTCCAACACAATGGAGAACCATAGCACCTATAGTAGGAAGAAC
 AGCACAACAATGTTTAGAACATTATGAATATCTACTTGATGAAGCGGAAGGAAAGGTTTATGATAAAA
 ATAAGAACCCACGACATCTAAGGCCAGGTGAAATAGACCCTGCTCCTGAATCCAAACCAGCACGTGCA
 GATCCTGTAGATATGGATGAAGATGAAAAAGAAATGCTTGCAGAAGCTAAGGCAAGACTTGCAAATAC
 GAAAGGTAAAAAGCAAAAAGAAAAGCTAGAGAAAAACAACCTGGAACAAGCTAGGAGATTAGCTTTAT
 TACAAAAAAGAGAGAATTAA**AGCAGCAGGGATAACATCATT**AAATTATAAAAGAAAAGATAAAAAAT
 AAAATTGATCATTGAAAGAAATATTATTTTCATAGGAAACCATTAAAGGTTTTTATGATGTTAAAGA
 TGAACAAAATATTAATGATGATATATGAAAATAATAAGACAAATCAGAAGAAAAGTATAAAATCAA
 TGGATGTTGAAAATATTAATGATGCTATGGAATATAATAAAAAATAAGGTAAACGACAACATCAACAT
 AATAATAATGAAGAAGCAAATTTGTTATCTACCATAGAAAATTATGATAACAATTTAACGAATTAAG
 TCATTTAAGAAAAAGGGTTCGATTGAATTTACCAGAACCTATATTAATGAAAATGAAATAGATGAAA
 TAATACAAATAAATAAAGAAGCATCAGCATTTAACGATATTATAAAAGATCAAAACGATAAATTACCA
 ATAAATAATATTTTACCAAGTATTGAAAGCTCCTCATTTATATTAACAATAAAGATAAATTTTCAAA
 CTTGGAACAGATTTTTATAATAATAATAATAAATCAATTGCTTTTTCTAGTAACTAGATCTAA
 GTATACAACAAGCAGCAAAAAATATCATTTCTCGAAAAATGAATATTCCATTTATAGGAATGAACAAT
 GATTATAATGAAGAAGAATTGAAAGAAAAAATAATATATTTCAAAGAGTAATAAGAATGTTCCCTGA
 TGATTTAGAATATGATGACACGAACAATTATCATAATAATAATAATAATAATAATAATAACTTCTT
 TTAATTCGATCAAAAATTCACAACCCCTTTATAATCATTTGGATGTAGAAAAAATATAAAAGACGTA
 CAAGAAGAAATTGAAAAACAAAAGAAATTAATGAAGAAGATCAACAAAAGAATCAAAAAGAAAAGTT
 AACTATTAATAAATAAATTTATTAGTGATATTAAGTTTAAAAAATATTAGTTTATATGCTCATT
 CTATTATTTCTTATAAATAAATTTAAAAAATGATCAATCTATGATGGATAATCAAACATACAAACAAC
 GAATATTATGATGATAATTATGAAGAGAAAAATCGATCGAGCAAAATTACATATTAAGCATCGTTAGC
 CAATTTACCCCAAGAAACGAATCTTATAGAACCTCAATTAATGAAGAACATCCAGAATGTGATACGG
 ACAATATAGAAAAAGATGAAATAGAAAAAGATATACAAGATATTGAAAATGAAAAAGAAAAACGAA
 GAAAGAAAAGAAAAAGAAAAATTTAATAAGCAAAATAAAATTATTAGATGGAATTTACCTAGACCATA
 TTTCTTAGATAAAATTAATCTCTTTAATAATTATATGCACAATGAATATGAAGATGTCCATAATTTAA
 TACAAAGAGAGATGCTCCTCTTAATAAAAAATGATATGTTTAATTATCCCCTCAGAAATTCAACACCA
 GTACAAAATAAAGTTCATGTGGAAGATCTCGAAAACGTATACATGAATATGGCTATGAAGAGTATAAA
 TGAGGAGTTCGAAGATATGTATGGCGAAGCATCATTAATAACAACACAAAGGATGATTCTAATATTG
 ATGGGTGTGATGAAAAGAGTGATAATATTGATGGTTGTGATGAAAAAGTGATACAATAAAGGT
 GATGATACGTCCCAATGTAGTATCGAGCATAGTTCATATAATCATATCGATGTGTGGGAAGAAATTAA
 CAAAAATATAATTTTTGCCCATCAAAAAATGCATATCGCTTTATTGAAGATGTTAATGAAAACGATA
 AAAAAAGAAATTACAAATACAAATGTGAAAAATTAATAATTTAATTTAATGATATGGAACATTAT
 AAAAAATTAGAAAATAAATATGATATTTATACAAAAGGATATCAACTAAAAATTAAAGTTATAAAAA
 ATCATATGATACTCTATTTAATTCATATATTAATTGTATTAACGAAAAAGAAGCTTTAATGTTCTAC
 ATGAAAATGAAAAATATATGCATTAACAAGAATTAAGAAGAAAAAAGGAAAATAAAAAAGAAATC
 GAATATCATAAATCATTACAAAAGTTTTATCAAGACCTTTTAGAAACAAATCATCAATTAAGAAGAAC
 ATGTAAACAAACGTTAAAGGTGCCATAA

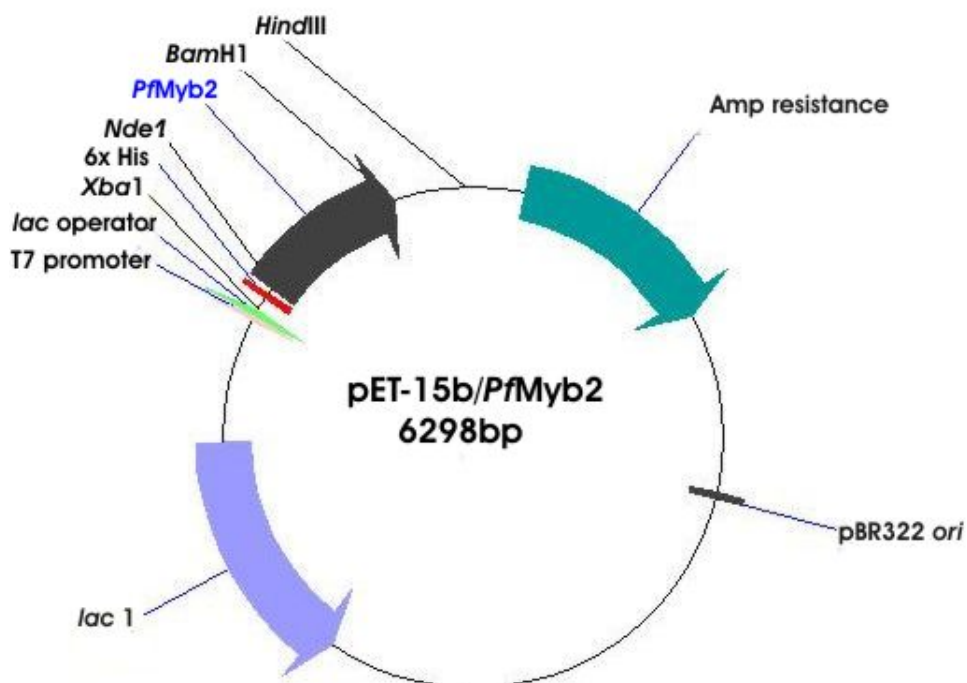
Forward (yellow) and reverse (grey) primers for amplification and cloning.

A-3.3 pET-15b vector construct and cloning cassette

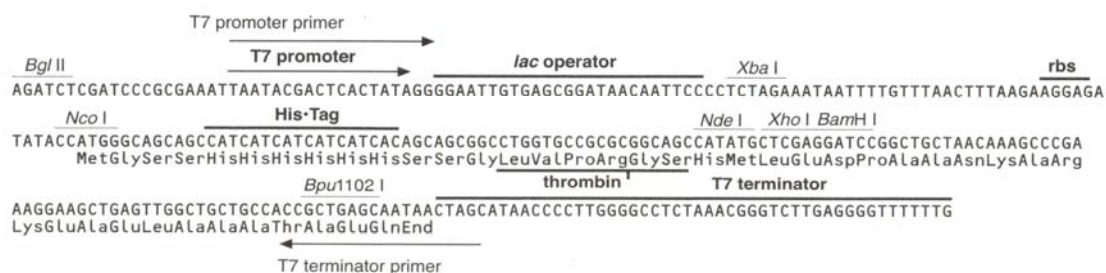
Primers for cloning of *PfMyb2* into pET-15b expression vector:



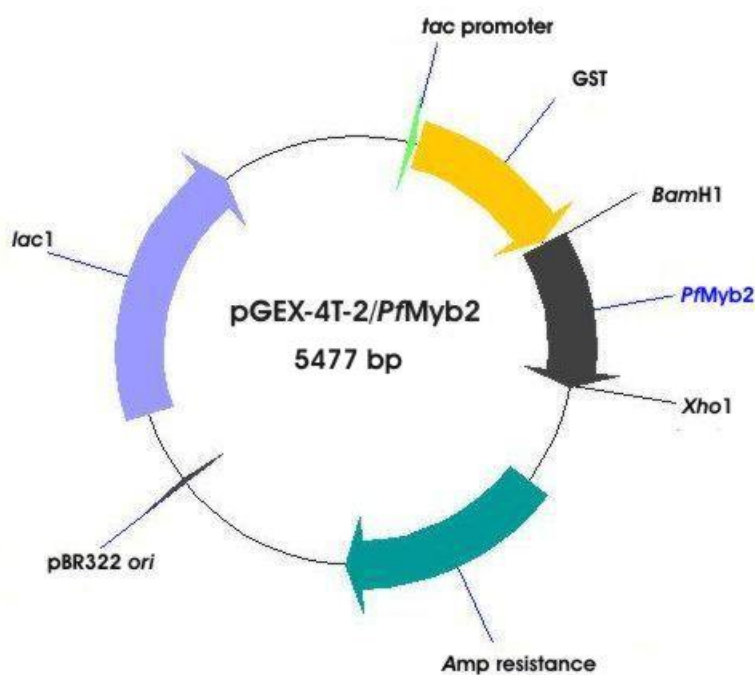
Nucleotides highlighted in blue are the additional bases required for restriction enzyme digestion. Recognition sequences are highlighted in purple, restriction sites are indicated with a triangle and stop codons are highlighted in green.



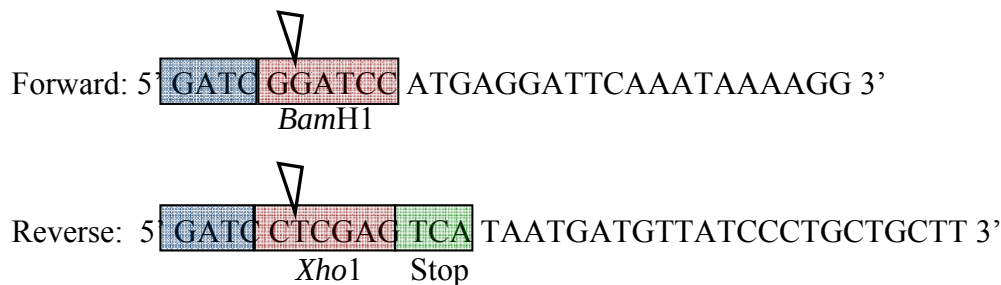
PfMyb2 (600bp with appended restriction sites) was cloned into the pET-15b expression vector (5698bp after digestion) to generate a 6298bp construct.



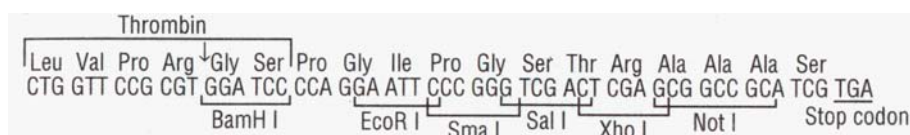
A-3.4 pGEX-4T-2 vector construct and cloning cassette



Primers used for the cloning of *PfMyb2* into the pGEX-4T-2 expression vector:



Nucleotides highlighted in blue are additional sequences flanking the restriction enzyme recognition sequences (highlighted in purple) required for enzyme binding. Restriction sites are indicated with a triangle, and stop codons are highlighted in green.



The above cloning cassette spans nucleotides 930-967.

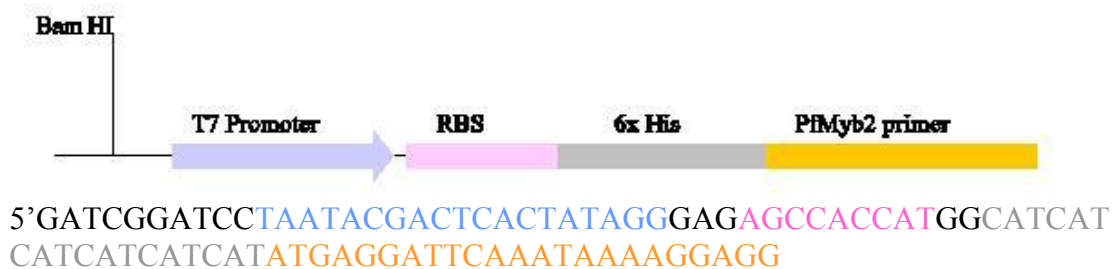
pGEX-4T-2 vector primers:

Forward: 5' GGGCTGGCAAGCCACGTTTGGTG 3'

Reverse: 5' CCGGGAGCTGCATGTGTCAGAGG 3'

The forward primer spans nucleotides 869-891, whilst the reverse primer spans 1040-1018. Using these primers for amplification appends 147 nucleotides to the insert size, creating a PCR product of 735bp when used for amplification of *PfMyb2*.

A-3.5 Primer design for *in vitro* translation



A3-6 Oligonucleotide sequences for EMSA assays

NfKBf: AGTTGAGGGGACTTTCCCAGGC

NfKBr: GCCTGGGAAAGTCCCCTCAACT

cMybf: TACAGGCATAACGGTTCGTTAGTGA

cMybr: TCACTACGGAACCGTTATGCCTGTA

Pfmap1f: TATTTGAGAACTGGATGAACAG

Pfmap1r: CTGTTTCATCCAGTTCTCAAATA

Pfcrk1f: TAAAAATAACCGACACCAAAAA

Pfcrk1r: TTTTGGTGTCGGTTATTTTA

mim-1f: ACACATTATAACGGTTTTTAGC

mim-1r: GCTAAAAACCGTTATAATGTGT

A-4 Differential Display Transcript Sequences

The sequences below represent the raw data as read off the sequencing gels. Where ambiguity existed in the sequence, the nucleotide was substituted with an “N.”

E2-3

CCCTTTNGTAAAACTCATTTATAAATTNTCATCAAAAATTTAANNNTTTT
TACAATAATTTTCATGCACATTCATCTCATTT

E2-4

CAAGGATACACAAGGTGTAGCGACAGGNCTACAAACCATCAACCGTACC
CTTGAAAATAGTGGTGAACAAAACCCGAACAAGAGTTAC

E4-2

GACATAATTATTATCGGTAAGATTTTTATTTTGTATATAGGGATTAAAGAA
TTGGAGCTATTATTATTATTGTTATCGTTGTTTTCGTCAACCTTGTTGTTAT
GAAACATGGTCTGATGCAGTGTTAATAGATTCGTCATTCTG

E5-1

TGTATCTAACCAGTGACGCGCATGAATGGATTAAACGAGATTCCCCTGTC
CCTACTTGCTATCTAGCGAAACCACAGCCAAGGGAACGGGCTTGGCAAAA
TCAGCGGGGAAAGAAGACCCTGTTGAGCTTTACTCTAGTCTGGCTTGTGA
AACGACTTAAGAGGTGTAGCATAAGTGGGAGTAGAACTGAATATGTTTA
CCGATCAGTGAAATACCACTA

E6-8

ATATCTTAACTTACTTTAGAAAATGACCCAATACCAAATAATCCAGNNNC
TGGTGAAAATNCCTTTGACNCCGTTTTTCGTAAAGGATGACGATGGTTATG
ACCTTGATTCTTTTATGATCCCTGCACATTATAAAAAATATCTTACCAAGG
TCTTAGTTCAAANGTCATAAAANNGATTGAGAATNCTTATGATATTAAAG
GTGTCACATGCGAGTTTCATAT

E9-1

GTCTTTTTTTTTTTTTGGGATTGGGTATATCATATGTTACACGATTCTTTAGA
GTGGAAGACAGAACTTAGTAAGTGTATAAATAATAACACTAATGGCAAC
ACATGTAGAAACAATAATAAATGTAAAACAGATTGTGGTTGTTTTGCAAA
ATGGGTTGTTAAAAAAAAGACGATGGGACAAAATAAAGCCATTTACA

E10

GCTACGTCATTGATAGTCTTTTCTATAGATGAAAAGATAAACTTTATATTA
TATGTTGATGTTATTT

E12-2

CATCAAGTTTTGGCAAGGCGGTTCCAAATTTAACGCGTTAATACCGAAAA
TTCCGAATTTTTNCTCACTGATTTTTTTGAAGGAATTTATTTTTTTTATTT
GCGCAAATATTTTCTTCAATGGATGAAGATACTGTAAAATTTATTTGCTTC
TGCTTATTTGACAGTACTATAATTATTACTTGGATATTTTTTTNGTTTCCTT
TGACCTAGTTTCAGACTGAGGACTGCAGTCGAG

E16-3

TTTTTTTTTTTTNNNCTTCATTTNTTTTTACTTCCTTTGTGTAAATAACTTTG
TCCATTTTTTCCAGAATTTTTTTNTTCTTTTTTCGTCACCAGATTC

E17-1

GAACAAAAAGAATGCTATGTAGGTGATGAAGCTCAAATAAAAGAGGTA
TTTTAACTTTAAAATATCCAATAGAACATGGTATTGTTACAAATTGGGATG
ATATGGAAAAAATATGGCATCATACGTTTCTATAATGAATTGAGAGTTTC
ACCTGAAGAACATCCAGTCTTAGTTAATCA
GAAGCTCCATTAATCTAAACAA

E18-1

CCATCACTACCACTTNCCTCTTTATATGTTAGTNCACATATCATTCCTCTCC
ATATATNTTGACCGTTTTTCGCATCCCA

E19-1

AAACTTTNGTTTCTTCTGATTCCATTTNNTTTGCTCATACTCTTAATACTTG
ATATGATCGTACCAACAACCGCATCAGGNCTCAAGGTAAACAGCCGTTCT
GGTTCCC

E23-1

GCATAATATTTTTCTCTAGACTTACTTACACATTTTTCTTTTATTAACGTAT
TTATTATCAAATCA

E25-2

CCTATAAATGGATAAGTAAAACAACCTTTAAAAGTAGTGGTATTTCACTGA
CGTAAAAACATATTTTCAGTTTCTACTCCCACTTATGCTGCACCTCTTAAGT
CGTTTCACAAAGCCAGACTAGAGTAAACCAACAGGGCTCTTCTT

E26-2

TAAATCACCATCTGGAGTTTTTGCAGATATGTCTGTATCCAAACATATATC
TCTATAATCTCCAAACGTATAGAACATGGATCGCAAAAATTCGGGAGGGA
TGATCCCACTATCTAATA

E27-2

CCATTATCTTTGTATGTTAGGGCGTAAATCATACCTTTCCAGATATGTTCA
CCGTTTGCTTCCCACCAGGTTTTACGTTTTTCCGTCACTAGGTTTGGATTCA
GGTG

E28-1

CCATCACTACCACTTCCCTCTTTATATGTTAGTCGCACATATCATTCCTCTC
CATATATCGTTTCGACCGTTTTTCGCATCCACNNNNNNNGGGGTTTTTAACC
ACTATTTNNGGGGACAGTAGGNNCCAGGTTCGTTCCGTAAATTTTCTTT
TATTTATCATTAA

E37-6

CCTCCTACNCTATGTTTCATCATAACTGCATAAAACAAAATAATACACACAA
TTTAATTATTTCTATATATATCAAATCGTATAAAATATAAAAA

E39-1

ACACACATTATCCCAATCAGGAGAAACAGAAGTAACAGAAGAAACAGAA
GAAACAGAAGAAACAGTAGGACACACAACAACGGTAACAATAACATTAC
CACCAACACAACCATCACCACCAAAGAAGTAAAA

E39-2

CTGATACTTCTGGTTTTTTATTAGATTTACTAAGATAACCAAACCTAAAG
CACTACCTAAGAGTAAAGCAAGACCAGTAACAACAGTAGATAT

E39-3

ACGTAAAGTGATTCTTAATAAATCGAACCTTCTTTCTTTTAAAAATGATTA
CTTAAATATAACAAATGAATAACGAAAGACGGTGTTGTGGAGCTATTGCT
CTTGAAAAATTTAAATGTCGCAAAAAAANNNAATTACACAAGCACA
CACACACACACACAGC

E42

TTCTTTCACTTTACCTTTATAACAAACCTTGGAATCAATTTACTTGGAGAA
GAGGTTTCGTTGAACTCAATTCAAAAAATTTATTGAACGGGGTTTTTCGGAT
TCAGTTCATTTTATTTTTTTGTTTTGTAGCAATAGTAATTCGTTTT

E43-1

TGTCGAATTAAAAAATCAAAATAAAATCAAACAAGTCCGTTTCTACCTTT
TATTTACCATTTGATGATAAACACAAAGGTATTTAACCATAGGAGGTAT
TGAAGATAGATTTTATGAAGGTCAATTAAGTTATGAAAATT

E43-2

ATCGATACCTCTACCAGCAACATCTGTAGCTACTAGTATATCAAATTCTGC
ATTTTTGAAAGCACTTAATGTTTGTTCTCGTATTTCTTGAGCTTTTCC

E51-3

ACCATAGTNNNNCTTCATCACATCAANNNGGAAAAGAATAATGAAAATGA
AACTGAAAAGAAANCTGATCAAAATGAAACTGGAACAAAAANNTTCTAA
ATATGATCC

E52-1

GATGGTAACATATATATAAATGAACTCCTTTACATAGGCTTTACACTCGG
GGTGCGTTTTCTTTGCACTTTACCTTTATAACAAACCTTGGAATCAATT
CTTGGAGAAGAGGTTTCGTTGAACTCAATTCAA

E54-4

CCGAAGGGAAAGTGCTTTCCTTATTTTGAACAGATTGTAACAACTCCATG
GGTACAGAAAGAGGGATATTAATTATCCTCTCCTTTCTTACTTCTACAATT
TTATTCATAATAAAAAATTCA

A-5 List of Suppliers of Chemicals and Equipment

Acrylamide	BDH, UK
ACD vacutainer tubes	Beckton Dickinson, UK
Agar	Oxoid, UK
Agarose	FMC, USA
Ammonium Chloride	Sigma, USA
Ammonium Persulphate	Stratagene, USA
Ampicillin	Sigma, USA
β -mercaptoethanol	Merck, USA
Bactotryptone	Oxoid, UK
Beckman SU65 spectrophotometer	Beckman, USA
Beckman, J2-21 centrifuge and JA-17 rotor	Beckman, USA
BL21-CodonPlus [®] (DE3)-RIL competent cells	Stratagene, USA
Blue/orange loading dye	Promega, USA
Boric acid	USA, USA
Bovine serum albumin	Pierce, USA
Bromophenol blue	BDH, UK
BRL model S2 electrophoresis apparatus	BRL, USA
BugBuster HT protein extraction reagent	Novagen, USA
4-chloro-1-naphthol	Sigma, USA
Calf intestinal Phosphatase	Roche, Germany
Centrifuge tubes	Nunc, Denmark
Chloroform	SMM Chemicals, RSA
Coomassie brilliant blue R-250	BDH, UK
Cryotubes	Nunc, Denmark
DC power supply, PS 500X	Hoefer, USA
DEPC	Sigma, USA
DH5 α competent cells	GibcoBRL, USA
Dialysis tubing	Sigma, USA
1,3, diaminobenzidine	Sigma, USA
100bp DNA ladder	Promega, USA

DNA mass ladder	GibcoBRL, USA
DNase-free RNase	Roche, Germany
EDTA	Roche, Germany
EMSA kit	Invitrogen, USA
Eppendorf Mastercycler gradient thermal	Eppendorf, Germany
Eppendorf tubes	Eppendorf, Germany
Ethanol	BDH, UK
Ethidium bromide	Roche, Germany
Extran	Merck, Germany
EZCast Gel Casting Boot	GibcoBRL, USA
FastPlasmid™ Mini Kit	Eppendorf, Germany
Falcon tubes	Becton Dickinson, USA
Filter tips	QSP, USA
Filter paper	Whatman, USA
0.22µm filters	Millipore, USA
Gel 3-place casting tray	CBS Scientific, USA
Geldoc Scanning System	Synoptics, UK
Gel Casting Tray	Hoefer, USA
Genelute Mammalian Total RNA Miniprep Kit	Sigma, USA
Gentamycin solution	Sigma, USA
Goat anti-mouse peroxidase conjugated IgG	Sigma, USA
Glacial acetic acid	Saarchem, RSA
d-Glucose	Saarchem, RSA
Glycerol	Sigma, USA
Glycine	BDH, UK
Glycyl-glycine	Sigma, USA
Guanidium thiocyanate	Fluka, Switzerland
Heating block	Hägar, RSA
High Fidelity PCR Master Mix	Roche, Germany
Hybond-N nylon membrane	Amersham,
Hybridisation tubes	Hybaid, UK
Hydrochloric acid	Saarchem, RSA

Hypoxanthine solution	Sigma, USA
IPTG	Invitrogen, USA
Isopropanol	Saarchem, RSA
Jouan BR3.11 centrifuge	Jouan, France
Laminar flow hood	Labotec, RSA
Lysozyme	Sigma, USA
Methanol	Saarchem, RSA
Mighty Small Dual Casting Tray	Hoefer USA
Mineral oil	Sigma, USA
Mouse anti-His antibody	Amersham
N-laurylsarcosine	Sigma, USA
Oligonucleotide primers	Inqaba Biotech, RSA; IDT USA
Overnight Express Autoinduction system	Novagen, USA
Parafilm	Whatman, UK
PCR Master Mix Kit	Roche, Germany
PCR product pre-sequencing kit	Amersham, UK
PCR tubes	QSP, USA
Peristaltic pump SJ-1211	SIS, USA
PET-15b expression vector	Novagen, USA
pGEM-3Z vector	Promega, USA
PGEX-4T-2 expression vector	Novagen, USA
Phenol	ICN Biochemicals, USA
PMSF	Roche, Germany
Polyvinylpyrrolidone	Sigma, USA
Ponceau S	Sigma USA
Potassium acetate	Saarchem, RSA
Potassium chloride	Saarchem, RSA
Potassium di-hydrogen orthophosphate	Saarchem, RSA

Prepared dialysis tubing	Gibco BRL, USA
Protease Inhibitor Cocktail Set III	Calbiochem, USA
PROTEINscript® II in vitro translation kit	Ambion, USA
$\alpha^{32}\text{P}(\text{dATP})$	Amersham, UK
$\gamma^{32}\text{P}$ ATP	Amersham, UK
Rapid DNA Ligation Kit	Roche, Germany
Rapid Haematology Staining Kit	SAIMR, SA
RNaseZap	Ambion, USA
RPMI culture medium	Invitrogen, USA
Saponin	USB, UK
SDS	BDH, UK
Sequenase® PCR Product sequencing Kit	Amersham, UK
Sequenase® version 2.0 DNA sequencing kit	Amersham, UK
d-Sorbitol	Sigma, USA
di-Sodium hydrogen orthophosphate	Saarchem, SA
Sodium acetate	BDH, UK
Sodium chloride	Saarchem, RSA
Sodium citrate	Saarchem, RSA
Sodium di-hydrogen orthophosphate	Saarchem, RSA
Sodium hydroxide	Saarchem, RSA
Sorvall RMC-14 centrifuge	Sorvall, USA
Sp6 RNA Polymerase promoter	Promega, USA
Superscript III Rnase H ⁻ Reverse Transcriptase	Invitrogen, USA
T4 DNA polymerase	Roche, Germany
T4 polynucleotide kinase	Promega, USA
T7 RNA polymerase promoter	Promega, USA
TE Transphor Electrophoresis Unit	Hoefer, USA
TEMED	Promega, USA
Thermohybrid IP22 Shake'n'Stack	Hybaid, UK
TRI reagent	Sigma, USA
Tris	BDH, UK
Triton X-100	Sigma, USA

Tryptone	Sigma, USA
Tween20	Merck, Germany
Urea	GibcoBRL, USA
UV transilluminator	UVP, USA
X-ray film	Agfa, Germany
Yeast extract	Oxoid, UK
Zeiss Axiostar microscope	Zeiss, Germany

REFERENCES

- Akompong, T., Eksi, S., Williamson, K. and Haldari, K. (2000). "Gametocytocidal activity and synergistic interactions of riboflavin with standard antimalarial drugs against growth of *Plasmodium falciparum* in vitro." Antimicrobial Agents and Chemotherapy **44**(11): 3107-3111.
- Alano, P., Roca L., Smith D., Read D., Carter R. and Day K. (1995). "*Plasmodium falciparum*: Parasites Defective in Early Stages of Gametocytogenesis." Experimental Parasitology **81**(2): 227-235.
- Alano, P., Silvestrini, F. and Roca, L. (1996). "Structure and polymorphism of the upstream region of the pfg27/25 gene, transcriptionally regulated in gametocytogenesis of *Plasmodium falciparum*." Molecular and Biochemical Parasitology **79**(2): 207-217.
- Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. and Walter, P. (2002). Basic Genetic Mechanisms in Molecular Biology of the Cell Garland Science. New York.
- Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D. J. (1997). "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs." Nucleic Acids Research **25**(17): 3389-3402.
- Aravind, L., Iyer, L. M., Wellems, T. E. and Miller, L. H. (2003). "*Plasmodium* biology: genomic gleanings." Cell **115**(7): 771-785.
- Baca, A. M. and Hol, W. G. J. (2000). "Overcoming codon bias: A method for high level overexpression of *Plasmodium* and other AT-rich parasite genes in *Escherichia coli*." International Journal for Parasitology **30**: 113-118.
- Bannister, L. and Mitchell, G. (2003). "The ins, outs and roundabouts of malaria." Trends in Parasitology **19**(5): 209-213.
- Barkakaty, B. N., Sharma, G. K. and Chakravorty, N. K. (1988). "Studies on efficacy of treatment with sulfamethoxazole + trimethoprim and sulfalene + pyrimethamine combinations in *Plasmodium falciparum* malaria of known and unknown resistant status." Journal of Communicable Diseases **20**: 165-174.
- Barnes, W. M. (1992). "The fidelity of *Taq* polymerase catalyzing PCR is improved by an N-terminal deletion." Gene **112**(1): 29-35.
-

-
- Barr, P. J., Green, K. M., Gibson, H. L., Bathurst, I. C., Quakyi, I. A. and Kaslow, D. C. (1991). "Recombinant Pfs25 protein of *Plasmodium falciparum* elicits malaria transmission-blocking immunity in experimental animals." Journal of Experimental Medicine **174**(5): 1203-1208.
- Berg, J. M., Tymoczko, J. L. and Stryer, L. (2005). Synthesizing the Molecules of Life in Biochemistry W H Freeman and Company. New York.
- Bertioli, D. J., Schlichter, U. H. A., Adams, M. J., Burrows, P. R., Steinbib, H.-H. and Antoniow, J. F. (1995). "An analysis of differential display shows a strong bias towards high copy number mRNAs." Nucleic Acids Research **23**: 4520-4523.
- Biedenkapp, H., Borgmeyer, U., Sippel, A. E. and Klempnauer, K. H. (1988). "Viral myb oncogene encodes a sequence-specific DNA-binding activity." Nature **335**(6193): 835-837.
- Billker, O., Dechamps, S., Tewari, R., Wenig, G., Franke-Fayard, B. and Brinkmann, V. (2004). "Calcium and a calcium-dependant protein kinase regulate gamete formation and mosquito transmission in a malaria parasite." Cell **117**: 503-514.
- Billker, O., Lindo, V., Panico, M., Etienne, A. E., Paxton, T., Dell, A., Rogers, M. J., Sinden, R. E. and Morris, H. R. (1998). "Identification of xanthurenic acid as the putative inducer of malaria development in the mosquito." Nature **392**: 289-292.
- Boschet, C., Gissot, M., Briquet, S., Hamid, Z., Claudel-Renard, C. and Vaquero, C. (2004). "Characterization of PfMyb1 transcription factor during erythrocytic development of 3D7 and F12 *Plasmodium falciparum* clones." Molecular and Biochemical Parasitology **138**(1): 159-163.
- Bozdech, Z., Llinas, M., Pulliam, B. L., Wong, E. D., Zhu, J. and DeRisi, J. L. (2003). "The transcriptome of the intraerythrocytic developmental cycle of *Plasmodium falciparum*." Public Library of Science, Biology **1**(1): E5.
- Brockelman, C. R. (1982). "Conditions favouring gametocytogenesis in the continuous culture of *Plasmodium falciparum*." Journal of Protozoology **29**: 454-458.
- Bruce, M. C., Alano, P., Duthie, S. and Carter, R. (1990). "Commitment of the malaria parasite *Plasmodium falciparum* to sexual and asexual development." Parasitology **100 Pt 2**: 191-200.
- Callard, D., Lescure, B. and Mazzolini, L. (1994). "A method for the elimination of false positives generated by the mRNA differential display technique." Biotechniques **16**(6): 1096-1103.
-

-
- Carlton, J. M.-R. (1999). "Gene Synteny across *Plasmodium* spp: Could 'Operon-like' Structures Exist?" Parasitology Today **15**(5): 178-179.
- Carter, R. and Graves, P. M., Eds. (1988). Gametocytes in Malaria: Principles and practice of malariology. London, Churchill Livingstone.
- Carter, R. and Miller, H. (1979). "Evidence for environmental modulation of gametocytogenesis in *Plasmodium falciparum* in continuous culture." Bulletin of the World Health Organisation **1**: 37-52.
- Carter, R., Ranford-Cartwright, L. and Alano, P. (1993). "The culture and preparation of gametocytes of *Plasmodium falciparum* for immunochemical, molecular, and mosquito infectivity studies." Methods in Molecular Biology **21**: 67-88.
- Carucci, D. (2004). "Know thine enemy." Nature **430**: 944-945.
- Chang, K.-H. and Stevenson, M. M. (2004). "Malarial anaemia: mechanisms and implications of insufficient erythropoiesis during blood-stage malaria." International Journal for Parasitology **34**: 1501-1016.
- Chomczynski, P. and Sacchi, N. (1987). "Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction." Analytical Biochemistry **162**(1): 156-159.
- Coulson, R. M. R., Hall, N. and Ouzounis, C. A. (2004). "Comparative genomics of transcriptional control in the human malaria parasite *Plasmodium falciparum*." Genome Research **14**: 1548-1554.
- Crabb, B. S. and Cowman, A. F. (1996). "Characterization of promoters and stable transfection by homologous and nonhomologous recombination in *Plasmodium falciparum*." Proceedings of the National Academy of Sciences of the U S A **93**: 7289-7294.
- Cui, L., Fan, Q. and Li, J. (2002). "The malaria parasite *Plasmodium falciparum* encodes members of the Puf RNA-binding protein family with conserved RNA binding activity." Nucleic Acids Research **30**: 4607-4617.
- Cui, L., Rzomp, K. A., Fan, Q., Martin, S. K. and Williams, J. (2001). "*Plasmodium falciparum*: differential display analysis of gene expression during gametocytogenesis." Experimental Parasitology **99**: 244-254.
- Dechering, K., Kaan, A. M., Mbacham, W., Wirth, D. F., Eling, W., Konings, R. N. and Stunnenberg, H. G. (1999). "Isolation and functional characterisation of two distinct sexual-stage-specific promoters of the human malaria parasite *Plasmodium falciparum*." Molecular and Cellular Biology **19**(2): 967-978.
-

-
- Dechering, K. J., Kaan, A. M., Mbacham, W., Wirth, D. F., Eling, W., Konings, R. N. and Stunnenberg, H. G. (1999). "Isolation and functional characterisation of two distinct sexual-stage-specific promoters of the human malaria parasite *Plasmodium falciparum*." Molecular and Cellular Biology **19**(2): 967-978.
- Dechering, K. J., Thompson, J., Dodemont, H. J., Eling, W. and Konings, R. N. H. (1997). "Developmentally regulated expression of pfs16, a marker for sexual differentiation of the human malaria parasite *Plasmodium falciparum*." Molecular and Biochemical Parasitology **89**(2): 235-244.
- Deitsch, K. W. (2004). Gene Expression in Malaria Parasites: Genomes and Molecular Biology. Waters, A. P. and Janse, C. J. Caister Academic Press. Wymondham: 205-227.
- Dessens, J. T., Margos, G., Rodriguez, M. C. and Sinden, R. E. (2000). "Identification of Differentially Regulated Genes of *Plasmodium* by Suppression Subtractive Hybridization." Parasitology Today **16**(8): 354-356.
- Diatchenko, L., Lau, Y. F., Campbell, A. P., Chenchik, A., Moqadam, F., Huang, B., Lukyanov, S., Lukyanov, K., Gurskaya, N., Sverdlov, E. D. and Siebert, P. D. (1996). "Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries." Proceedings of the National Academy of Sciences of the U S A **93**(12): 6025-6030.
- Diebner, H. H., Eichner, M., Molineaux, L., Collins, W. E., Jeffery, G. M. and Dietz, K. (2000). "Modelling the transition of asexual blood stages of *Plasmodium falciparum* to gametocytes." Journal of Theoretical Biology **202**: 113-127.
- Doerig, C. D. (1997). "Signal transduction in malaria parasites." Parasitology Today **13**(8): 307-313.
- Doerig, C. D. and Chakrabarti, D. (2004). Cell cycle control in *Plasmodium falciparum*: A genomics perspective in Malaria Parasites: Genomes and Molecular Biology. Waters, A. P. and Janse, C. J. Caister Academic Press. Wymondham: 249-288.
- Doerig, C. D., Horrocks, P., Coyle, J., Carlton, J., Sultan, A., Arnot, D. and R., C. (1995). "Pfck-1, a developmentally regulated cdc2-related protein kinase of *Plasmodium falciparum*." Molecular and Biochemical Parasitology **70**: 167-174.
- Doerig, C. M., Parzy, D., Langsley, G., Horrocks, P., R., C. and Doerig, C. D. (1996). "A MAP kinase homolog from the human malaria parasite *Plasmodium falciparum*." Gene **177**: 1-6.
-

-
- Dorin, D., Alano, P., Boccaccio, I., Ciceron, L., Doerig, C., Sulpice, R., Parzy, D. and Doerig, C. (1999). "An Atypical Mitogen-activated Protein Kinase (MAPK) Homologue Expressed in Gametocytes of the Human Malaria Parasite *Plasmodium falciparum*. Identification of a MAPK signature." Journal of Biological Chemistry **274**(42): 29912-29920.
- Dyer, M. and Day, K. P. (2000). "Commitment to gametocytogenesis in *Plasmodium falciparum*." Parasitology Today **16**(3): 102-107.
- Dyer, M. and Day, K. P. (2000^a). "Expression of *Plasmodium falciparum* trimeric G proteins and their involvement in switching to sexual development." Molecular and Biochemical Parasitology **110**: 437-448.
- Dyer, M. and Day, K. P. (2003). "Regulation of the rate of asexual growth and commitment to sexual development by diffusible factors from in vitro cultures of *Plasmodium falciparum*." American Journal of Tropical Medicine and Hygiene **68**(4): 403-409.
- Eksi, S., Haile, Y., Furuya, T., Ma, L., Su, X. and Williamson, K. C. (2005). "Identification of a subtelomeric gene family expressed during the asexual-sexual stage transition in *Plasmodium falciparum*." Molecular and Biochemical Parasitology **143**(1): 90-99.
- Eksi, S., Stump, A., Fanning, S. L., Shenouda, M. I., Fujioka, H. and Williamson, K. C. (2002). "Targeting and sequestration of truncated Pfs230 in an intraerythrocytic compartment during *Plasmodium falciparum* gametocytogenesis." Molecular Microbiology **44**: 1507-1516.
- Fan, Q., An, L. and Cui, L. (2004). "PfADA2, a *Plasmodium falciparum* homologue of the transcriptional coactivator ADA2 and its in vivo association with the histone acetyltransferase PfGCN5." Gene **336**(2): 251-261.
- Fischer, K., Horrocks, P., Preuss, M., Wiesner, J., Wunsch, S., Camargo, A. A. and Lanzer, M. (1997). "Expression of var genes located within polymorphic subtelomeric domains of *Plasmodium falciparum* chromosomes." Molecular and Cellular Biology **17**(7): 3679-3686.
- Florens, L., Washburn, M. P., Raine, J. D., Anthony, R. M., Grainger, M., Haynes, J. D., Moch, J. K., Muster, N., Sacci, J. B., Tabb, D. L., Witney, A. A., Wolters, D., Wu, Y., Gardner, M. J., Holder, A. A., Sinden, R. E., Yates, J. R. and Carucci, D. J. (2002). "A proteomic view of the *Plasmodium falciparum* life cycle." Nature **419**(6906): 520-526.
- Florent, I., Charneau, S. and Grellier, P. (2004). "*Plasmodium falciparum* genes differentially expressed during merozoite morphogenesis." Molecular and Biochemical Parasitology **135**(1): 143-148.
-

-
- Gardiner, D. L., Dixon, M. W. A., Spielmann, T., Skinner-Adams, T. S., Hawthorne, P. L., Ortega, M. R., Kemp, D. J. and Trenholme, K. R. (2005). "Implication of a *Plasmodium falciparum* gene in the switch between asexual reproduction and gametocytogenesis." Molecular and Biochemical Parasitology **140**(2): 153-160.
- Gardner, M. J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R. W., Carlton, J. M., Pain, A., Nelson, K. E., Bowman, S., Paulsen, I. T., James, K., Eisen, J. A., Rutherford, K., Salzberg, S. L., Craig, A., Kyes, S., Chan, M. S., Nene, V., Shallom, S. J., Suh, B., Peterson, J., Angiuoli, S., Pertea, M., Allen, J., Selengut, J., Haft, D., Mather, M. W., Vaidya, A. B., Martin, D. M., Fairlamb, A. H., Fraunholz, M. J., Roos, D. S., Ralph, S. A., McFadden, G. I., Cummings, L. M., Subramanian, G. M., Mungall, C., Venter, J. C., Carucci, D. J., Hoffman, S. L., Newbold, C., Davis, R. W., Fraser, C. M. and Barrell, B. (2002). "Genome sequence of the human malaria parasite *Plasmodium falciparum*." Nature **419**: 498-511.
- Getz, E. B., Xiao, M., Chakrabarty, T., Cooke, R. and Selvin, P. R. (1999). "A comparison between the sulfhydryl reductants Tris(2-carboxyethyl)phosphine and Dithiothreitol for use in protein biochemistry." Analytical Biochemistry **273**: 73-80.
- Gissot, M., Briquet, S., Refour, P., Boschet, C. and Vaquero, C. (2005). "PfMyb1, a *Plasmodium falciparum* transcription factor, is required for intra-erythrocytic growth and controls key genes for cell cycle regulation." Journal of Molecular Biology **346**(1): 29-42.
- Gissot, M., Refour, P., Briquet, S., Boschet, C., Coupe, S., Mazier, D. and Vaquero, C. (2004). "Transcriptome of 3D7 and its gametocyte-less derivative F12 *Plasmodium falciparum* clones during erythrocytic development using a gene-specific microarray assigned to gene regulation, cell cycle and transcription factors." Gene **341**: 267-277.
- Gosh, S., Rasheedi, S., Rahim, S. S., Banerjee, S., Choudhary, R. K., Chakhaiyar, P., Ehtesham, N. Z., Mukhopadhyay, S. and Hasnain, S. E. (2004). "Method for enhancing solubility of the expressed recombinant proteins in *Escherichia coli*." Biotechniques **37**(3): 418-423.
- Guehmann, S., Vorbrueggen, G., Kalkbrenner, F. and Moelling, K. (1992). "Reduction of a conserved cys is essential for Myb-DNA binding." Nucleic Acids Research **20**: 2279-2286.
- Guinet, F. and Wellems, T. E. (1997). "Physical mapping of a defect in *Plasmodium falciparum* male gametocytogenesis to an 800kb segment of chromosome 12." Molecular and Biochemical Parasitology **90**: 343-346.
-

-
- Hawking, F., Wilson, M. E. and Gammage, K. (1971). "Evidence for cyclic development and short-lived maturity in the gametocytes of *Plasmodium falciparum*." Transactions of the Royal Society of Tropical Medicine and Hygiene **65**(5): 549-559.
- Hayward, R. E., Derisi, J. L., Alfadhli, S., Kaslow, D. C., Brown, P. O. and Rathod, P. K. (2000). "Shotgun DNA microarrays and stage-specific gene expression in *Plasmodium falciparum* malaria." Molecular Microbiology **35**(1): 6-14.
- Heddi, A. (2002). "Malaria Pathogenesis: a jigsaw with an increasing number of pieces." International Journal for Parasitology **32**: 1587-1598.
- Horrocks, P., Dechering, K. and Lanzer, M. (1998). "Control of gene expression in *Plasmodium falciparum*." Molecular and Biochemical Parasitology **95**(2): 171-181.
- Horrocks, P. and Kilbey, B. J. (1996). "Physical and functional mapping of the transcriptional start sites of *Plasmodium falciparum* proliferating cell nuclear antigen." Molecular and Biochemical Parasitology **82**: 207-215.
- Horrocks, P. and Lanzer, M. (1999). "Mutational analysis identifies a five base pair cis-acting sequence essential for *gpb130* promoter activity in *Plasmodium falciparum*." Molecular and Biochemical Parasitology **99**(1): 77-87.
- Inselberg, J. (1983). "Stage-specific inhibitory effect of cyclic AMP on asexual maturation and gametocyte formation of *Plasmodium falciparum*." Journal of Parasitology **69**: 592-597.
- James, M. A., Montenegro-James, S. and Fajfar-Whetstone, C. (1993). "Immunogenicity and antigenic reactivity of a carrier-free synthetic peptide complex derived from a 70-kDa *Plasmodium falciparum* exoantigen." Parasitology Research **79**(6): 501-507.
- Janse, C. J. and Waters, A. P. (2004). Sexual development of malaria parasites in Malaria Parasites: Genomes and Molecular Biology. Janse, C. J. and Waters, A. P. Caister Academic Press. Wymondham: 445-474.
- Ji, D.-D. and Arnot, D. E. (1997). "A *Plasmodium falciparum* homologue of the ATPase subunit of a multi-protein complex involved in chromatin remodelling for transcription." Molecular and Biochemical Parasitology **88**(151-162).
- Ji, W., Wright, M. B., Cai, L., Flament, A. and Lindpaintner, K. (2002). "Efficacy of SSH PCR in isolating differentially expressed genes." BMC Genomics **3**(1): 12.
-

-
- Kanei-Ishii, C., Nomura, T., Ogata, K., Sarai, A., Yasukawa, T. and Tashiro, S. (1996). "Structure and function of the proteins encoded by the myb gene family." Current Topics in Microbiology and Immunology **211**: 89-98.
- Kaslow, D. C., Isaacs, S. N., Quakyi, I. A., Gwadz, R. W., Moss, B. and Keister, D. B. (1991). "Induction of *Plasmodium falciparum* transmission-blocking antibodies by recombinant vaccinia virus." Science **252**: 1310-1313.
- Keough, D. T., Ng, A.-L., Winzor, D. J., Emmerson, B. T. and de Jersey, J. (1999). "Purification and characterization of *Plasmodium falciparum* hypoxanthine-guanine-xanthine phosphoribosyltransferase and comparison with the human enzyme." Molecular and Biochemical Parasitology **98**(1): 29-41.
- Kolchanov, N. A., Ignatieva, E. V., Ananko, E. A., Podkolodnaya, O. A., Stepanenko, I. L., Merkulova, T. I., Pozdnyakov, M. A., Podkolodny, N. L., Naumochkin, A. N. and Romashchenko, A. G. (2002). "Transcription Regulatory Regions Database (TRRD): its status in 2002." Nucleic Acids Research **30**(1): 312-317.
- Kongkasuriyachai, D., Fujioka, H. and Kumar, N. (2004). "Functional analysis of *Plasmodium falciparum* parasitophorous vacuole membrane protein (Pfs16) during gametocytogenesis and gametogenesis by targeted gene disruption." Molecular and Biochemical Parasitology **133**: 275-285.
- Kongkasuriyachai, D. and Kumar, N. (2002). "Functional characterisation of sexual stage specific proteins in *Plasmodium falciparum*." International Journal for Parasitology **32**(13): 1559-1566.
- Kumar, N., Cha, G., Pineda, F., Maciel, J., Haddad, D., Bhattacharyya, M. and Nagayasu, E. (2004). "Molecular complexity of sexual development and gene regulation in *Plasmodium falciparum*." International Journal for Parasitology **34**(13-14): 1451-1458.
- Laemmli, U. K. (1970). "Cleavage of structural proteins during the assembly of the head of bacteriophage T4." Nature **227**: 680-685.
- Lambros, C. and Vanderberg, J. P. (1979). "Synchronization of *Plasmodium falciparum* erythrocytic stages in culture." Journal of Parasitology **65**(3): 418-420.
- Langsley, G., Hyde, J. E., Goman, M. and Scaife, J. G. (1983). "Cloning and characterisation of the rRNA genes from the human malaria parasite *Plasmodium falciparum*." Nucleic Acids Research **11**(24): 8703-8717.
- Lanzer, M., de Bruin, D. and Ravetch, J. V. (1992). "A sequence element associated with the *Plasmodium falciparum* KAHRP gene is the site of developmentally
-

- regulated protein-DNA interactions." Nucleic Acids Research **20**(12): 3051-3056.
- Lasonder, E., Ishihama, Y., Andersen, J. S., Vermunt, A. M., Pain, A., Sauerwein, R. W., Eling, W. M., Hall, N., Waters, A. P., Stunnenberg, H. G. and Mann, M. (2002). "Analysis of the *Plasmodium falciparum* proteome by high-accuracy mass spectrometry." Nature **419**(6906): 537-542.
- Le Bonniec, S., Deregnaucourt, C., Redeker, V., Banerjee, R., Grellier, P., Goldberg, D. E. and Schrevel, J. (1999). "Plasmeprin II, an acidic hemoglobinase from the *Plasmodium falciparum* food vacuole, is active at neutral pH on the host erythrocyte membrane skeleton." Journal of Biological Chemistry **274**(20): 14218-14223.
- Le Roch, K. G., Zhou, Y., Blair, P. L., Grainger, M., Moch, J. K., Haynes, J. D., De La Vega, P., Holder, A. A., Batalov, S., Carucci, D. J. and Winzeler, E. A. (2003). "Discovery of Gene Function by Expression Profiling of the Malaria Parasite Life Cycle." Science **301**: 1503-1508.
- Liang, P. (2002). "A decade of differential display." Biotechniques **33**: 338-346.
- Liang, P. and Pardee, A. B. (1992). "Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction." Science **257**(5072): 967-971.
- Liang, P. and Pardee, A. B. (1995). "Recent advances in differential display." Current Opinion in Immunology **7**(2): 274-280.
- Lievens, S., Goormachtig, S. and Holsters, M. (2001). "A critical evaluation of differential display as a tool to identify genes involved in legume nodulation: looking back and looking forward." Nucleic Acids Research **29**(17): 3459-3468.
- Lipsick, J. S. (1996). "One billion years of Myb." Oncogene **13**: 223-235.
- Lipsick, J. S., Manak, J., Mitiku, N., Chen, C.-K., Fogarty, P. and Guthrie, E. (2001). "Functional evolution of the Myb oncogene family." Blood Cells, Molecules and Diseases **27**(2): 456-458.
- Lobo, C.-A., Fujioka, H., Aikawa, M. and Kumar, N. (1999). "Disruption of the Pfg27 Locus by Homologous Recombination Leads to Loss of the Sexual Phenotype in *P. falciparum*." Molecular Cell **3**(6): 793-798.
- Lobo, C. A. and Kumar, N. (1999). "Differential transcription of histone genes in asexual and sexual stages of *Plasmodium falciparum*." International Journal for Parasitology **29**(9): 1447-1449.
-

-
- Lockhart, D. J. and Winzeler, E. A. (2000). "Genomics, gene expression and DNA arrays." Nature **405**(6788): 827-836.
- Lodish, H., Berk, A., Zipursky, L., Matsudaira, P., Baltimore, D. and Darnell, J., Eds. (1999). RNA Processing, Nuclear Transport, and Post-Transcriptional Control in Molecular Cell Biology. New York, W H Freeman and Company.
- Luce, M. J. and Burrows, P. D. (1998). "Minimizing false positives in differential display." Biotechniques **24**(5): 766-768.
- Mamoun, B. C., Gluzman, I. Y., Hott, C., MacMillan, S. K., Amarakone, A. S., Anderson, D. L., Carlton, J. M., Dame, J. B., Chakrabarti, D., Martin, R. K., Brownstein, B. H. and Goldberg, D. E. (2001). "Co-ordinated programme of gene expression during asexual intraerythrocytic development of the human malaria parasite *Plasmodium falciparum* revealed by microarray analysis." Molecular Microbiology **39**(1): 26-36.
- Manduchi, E., Grant, G. R., He, H., Liu, J., Mailman, M. D., Pizarro, A. D., Whetzel, P. L. and Stoeckert, C. J., Jr. (2004). "RAD and the RAD Study-Annotator: an approach to collection, organization and exchange of all relevant information for high-throughput gene expression studies." Bioinformatics **20**(4): 452-459.
- Matz, M. V. and Lukyanov, S. A. (1998). "Different strategies of differential display: areas of application." Nucleic Acids Research **26**(24): 5537-5543.
- McAndrew, M. B., Read, M., Sims, P. F. and Hyde, J. E. (1993). "Characterisation of the gene encoding an unusually divergent TATA-binding protein (TBP) from the extremely A+T-rich human malaria parasite *Plasmodium falciparum*." Gene **124**: 165-171.
- McCutchan, T. F., Li, J., McConkey, G. A., Rogers, M. J. and Waters, A. P. (1995). "The cytoplasmic ribosomal RNAs of *Plasmodium* spp." Parasitology Today **11**(4): 134-138.
- Mercereau-Puijalon, O., Barale, J. C. and Bischoff, E. (2002). "Three multigene families in *Plasmodium* parasites: facts and questions." International Journal for Parasitology **32**(11): 1323-1344.
- Meyersfeld, D. R. and Coetzer, T. L. (2003). "Reverse slot blot for the verification of cDNAs identified through differential display." Biotechniques **34**(2): 270-272.
- Miele, G., MacRae, L., McBride, D., Manson, J. and Clinton, M. (1998). "Elimination of false positives generated through PCR re-amplification of differential display cDNA." Biotechniques **25**(1): 138-144.
-

-
- Mignone, F., Gissi, C., Liuni, S. and Pesole, G. (2002). "Untranslated regions of mRNAs." Genome Biology **3**: 4.1-4.10.
- Militello, K. T., Dodge, M., Bethke, L. and Wirth, D. F. (2004). "Identification of regulatory elements in the *Plasmodium falciparum* genome." Molecular and Biochemical Parasitology **134**(1): 75-88.
- Moreira, C. K., Marrelli, M. T. and Jacobs-Lorena, M. (2004). "Gene expression in *Plasmodium*: from gametocytes to sporozoites." International Journal for Parasitology **34**(13-14): 1431-1440.
- Morikawa, S., Ogata, K., Sekikawa, A., Sarai, A., Ishii, S., Nishimura, Y. and Nakamura, H. (1995). "Determination of the NMR solution structure of a specific DNA complex of the Myb DNA-binding domain." Journal of Biomolecular NMR **6**(3): 294-305.
- Munasinghe, A., Patankar, S., Cook, B. P., Madden, S. L., Martin, R. K., Kyle, D. E., Shoaibi, A., Cummings, L. M. and Wirth, D. F. (2001). "Serial analysis of gene expression (SAGE) in *Plasmodium falciparum*: application of the technique to A-T rich genomes." Molecular and Biochemical Parasitology **113**(1): 23-34.
- Nacher, M., Singhasivanon, P., Silachamroon, U., Treeprasertsuk, S., Tosukhowong, T., Vannaphan, S., Gay, F., Mazier, D. and Looareesuwan, S. (2002). "Decreased hemoglobin concentrations, hyperparasitemia, and severe malaria are associated with increased *Plasmodium falciparum* gametocyte carriage." Journal of Parasitology **88**(1): 97-101.
- Ogata, K., Morikawa, S., Nakamura, H., Sekikawa, A., Inoue, T., Kanai, H., Sarai, A., Ishii, S. and Nishimura, Y. (1994). "Solution structure of a specific DNA complex of the Myb DNA-binding domain with cooperative recognition helices." Cell **79**: 639-648.
- Ogata, K., Sato, K. and Tahirov, T. H. (2003). "Eukaryotic transcriptional regulatory complexes: cooperativity from near and afar." Current Opinion in Structural Biology **13**(1): 40-48.
- Oh, I.-H. and Premkumar Reddy, E. (1999). "The *myb* gene family in cell growth, differentiation and apoptosis." Oncogene **18**(19): 3017-3033.
- Orphanides, G., Lagrange, T. and Reinberg, D. (1996). "The general transcription factors of RNA polymerase II." Genes and Development **10**: 2657-2683.
- Osta, M., Gannoun-Zaki, L., Bonnefoy, S., Roy, C. and Vial, H. J. (2002). "A 24 bp cis-acting element essential for the transcriptional activity of *Plasmodium*
-

- falciparum* CDP-diacylglycerol synthase gene promoter." Molecular and Biochemical Parasitology **121**(1): 87-98.
- Pace, T., Birago, C., Janse, C. J., Picci, L. and Ponzi, M. (1998). "Developmental regulation of a *Plasmodium* gene involves the generation of stage-specific 5' untranslated sequences." Molecular and Biochemical Parasitology **97**(1-2): 45-53.
- Pan, W., Ravot, E., Tolle, R., Frank, R., Mosbach, R., Turbachova, I. and Bujard, H. (1999). "Vaccine candidate MSP-1 from *Plasmodium falciparum*: a redesigned 4917 bp polynucleotide enables synthesis and isolation of full-length protein from *Escherichia coli* and mammalian cells." Nucleic Acids Research **27**(4): 1094-1103.
- Patankar, S., Munasinghe, A., Shoaibi, A., Cummings, L. M. and Wirth, D. F. (2001). "Serial analysis of gene expression in *Plasmodium falciparum* reveals the global expression profile of erythrocytic stages and the presence of anti-sense transcripts in the malarial parasite." Molecular Biology of the Cell **12**(10): 3114-3125.
- Peterson, A. J., Kossenkov, A. V. and Ochs, M. F. (2004). Linking gene expression patterns and transcriptional regulation in *Plasmodium falciparum*. The fifth International Conference for the Critical Assessment of Microarray Data Analysis, North Carolina, USA.
- Petmitr, P., Pongvilairat, G. and Wilairat, P. (1997). "Large scale culture technique for pure *Plasmodium falciparum* gametocytes." South East Asian Journal for Tropical Medicine and Public Health **28**(1): 18-21.
- Pinson, B., Brendeford, E. M., Gabrielsen, O. S. and Daignan-Fornier, B. (2001). "Highly conserved features of DNA binding between two divergent members of the *Myb* family of transcription factors." Nucleic Acids Research **29**(2): 527-535.
- Piper, K. P., Hayward, R. E., Cox, M. J. and Day, K. P. (1999). "Malaria transmission and naturally acquired immunity to *PfEMP-1*." Infection and Immunity **67**(12): 6369-6374.
- Pizzi, E. and Frontali, C. (2001). "Low-complexity regions in *Plasmodium falciparum* proteins." Genome Research **11**(2): 218-229.
- Pologe, L. G. (1994). "Aberrant transcription and the failure of *Plasmodium falciparum* to differentiate into gametocytes." Molecular and Biochemical Parasitology **68**: 35-43.
-

-
- Porter, M. E. (2002). "Positive and negative effects of deletions and mutations within the 5' flanking region of *Plasmodium falciparum* DNA polymerase delta." Molecular and Biochemical Parasitology **122**: 9-19.
- Roberts, D. J., Craig, A. G., Berendt, A. R., Pinches, R., Nash, G., Marsh, K. and Newbold, C. I. (1992). "Rapid switching to multiple antigenic and adhesive phenotypes in malaria." Nature **357**(6380): 689-692.
- Robson, K. J. H. and Jennings, M. W. (1991). "The structure of the calmodulin gene of *Plasmodium falciparum*." Molecular and Biochemical Parasitology **46**: 19-34.
- Ruvalcaba-Salazar, O. K., Ramirez-Estudillo, M. d. C., Montiel-Condado, D., Recillas-Targa, F., Vargas, M. and Hernandez-Rivas, R. (2005). "Recombinant and native *Plasmodium falciparum* TATA-binding-protein binds to a specific TATA box element in promoter regions." Molecular and Biochemical Parasitology **140**(2): 183-196.
- Saikumar, P., Murali, R. and Premkumar Reddy, E. (1990). "Role of tryptophan repeats and flanking amino acids in Myb-DNA interactions." Proceedings of the National Academy of Sciences of the U S A **87**: 8452-8456.
- Sala, A., Bellon, T., Melotti, P., Peschle, C. and Calabretta, B. (1995). "Inhibition of erythro-myeloid differentiation by constitutive expression of a DNA binding deficient c-myb mutant: implication for c-myb function." Blood **86**(9): 3404-3412.
- Scherf, A., Hernandez-Rivas, R., Buffet, P., Bottius, E., Benatar, C., Pouvelle, B., Gysin, J. and Lanzer, M. (1998). "Antigenic variation in malaria: in situ switching, relaxed and mutually exclusive transcription of var genes during intra-erythrocytic development in *Plasmodium falciparum*." EMBO Journal **17**(18): 5418-5426.
- Schneider, P., Schoone, G., Schallig, H., Verhage, D., Telgt, D., Eling, W. and Sauerwein, R. (2004). "Quantification of *Plasmodium falciparum* gametocytes in differential stages of development by quantitative nucleic acid sequence-based amplification." Molecular and Biochemical Parasitology **137**(1): 35-41.
- Snow, R. W., Guerra, C. A., Noor, A. M., Myint, H. Y. and Hay, S. I. (2005). "The global distribution of clinical episodes of *Plasmodium falciparum* malaria." Nature **434**: 214-217.
- Spielmann, T. and Beck, H. P. (2000). "Analysis of stage-specific transcription in *Plasmodium falciparum* reveals a set of genes exclusively transcribed in ring stage parasites." Molecular and Biochemical Parasitology **111**: 453-458.
-

-
- Sterner, D. E. and Berger, S. L. (2000). "Acetylation of histones and transcription-related factors." Microbiology and Molecular Biology Reviews **64**: 435-459.
- Su, X. Z. and Wellems, T. E. (1994). "Sequence, transcript characterization and polymorphisms of a *Plasmodium falciparum* gene belonging to the heat-shock protein (HSP) 90 family." Gene **151**(1-2): 225-230.
- Suh, K. N., Kain, K. C. and Keystone, J. S. (2004). "Malaria." Canadian Medical Association Journal **170**(11): 1693-1702.
- Sutherland, C. J., Ord, R., Dunyo, S., Jawara, M. and Drakeley, C. J. (2005). "Reduction of malaria transmission to *Anopheles* mosquitoes with a six dose regimen of co-artemether." PloS Medicine **2**(4).
- Talman, A. M., Domarle, O., McKenzie, F. E., Arie, F. and Robert, V. (2004). "Gametocytogenesis: the puberty of *Plasmodium falciparum*." Malaria Journal **3**(1).
- Taylor, L. H. and Read, A. F. (1997). "Why so few transmission stages? Reproductive restraint by malaria parasites." Parasitology Today **13**: 135-140.
- Thompson, C. B., Challoner, P. B., Neiman, P. E. and Groudine, M. (1986). "Expression of the *c-myc* proto-oncogene during cellular proliferation." Nature **319**(6052): 374-380.
- Trager, W. and Gill, G. S. (1989). "*Plasmodium falciparum* gametocyte formation in vitro: its stimulation by phorbol diesters and by 8-bromo cyclic adenosine monophosphate." Journal of Protozoology **36**: 451-454.
- Trager, W. and Jensen, J. B. (1976). "Human malaria parasites in continuous culture." Science **193**(4254): 673-675.
- Tshefu, K. and James, M. A. (1995). "Relationship of antibodies to soluble *Plasmodium falciparum* antigen (Pf70) and protection against malaria in a human population living under intense transmission in Kinshasa, Zaire." Tropical Medicine and Parasitology **46**(2): 72-76.
- van Dijk, M. M., Janse, C. J., Thompson, J., Waters, A. P., Braks, J. A. M., Dodemant, H. J., Stunnenberg, H. G., van Gemert, G., Sauerwein, R. W. and Eling, W. (2001). "A central role for Pfs48/45 in malaria parasite male gamete fertility." Cell **104**: 153-164.
- Velculescu, V. E., Zhang, L., Vogelstein, B. and Kinzler, K. W. (1995). "Serial analysis of gene expression." Science **270**(5235): 484-487.
-

-
- Wahlgren, M. and Chen, Q. (2002). "*Plasmodium falciparum*: A painful malaria parasite is being unveiled." www.nature.com/nm/special_focus/malaria/introduction
- Wellems, T. E., Su, X. Z., Ferdig, M. and Fidock, D. A. (1999). "Genome projects, genetic analysis, and the changing landscape of malaria research." Current Opinion in Microbiology **2**(4): 415-419.
- Wesseling, J. G., Smits, M. A. and Schoenmakers, J. G. G. (1988). "Extremely diverged actin proteins in *Plasmodium falciparum*." Molecular and Biochemical Parasitology **30**(2): 143-153.
- Wesseling, J. G., Snijders, P. J., van Someren, P., Jansen, J., Smits, M. A. and Schoenmakers, J. G. (1989). "Stage-specific expression and genomic organization of the actin genes of the malaria parasite *Plasmodium falciparum*." Mol Biochem Parasitol **35**(2): 167-176.
- Wickens, M., Bernstein, D. S., Kimble, J. and Parker, R. (2002). "A PUF family portrait: 3'UTR regulation as a way of life." Trends in Genetics **18**: 150-157.
- Wickham, M. E., Thompson, J. K. and Cowman, A. F. (2003). "Characterisation of the merozoite surface protein-2 promoter using stable and transient transfection in *Plasmodium falciparum*." Molecular and Biochemical Parasitology **129**(2): 147-156.
- Williams, C. E. and Grotewald, E. (1997). "Differences between plant and animal Myb domains are fundamental for DNA binding activity, and chimeric Myb domains have novel DNA binding specificities." The Journal of Biological Chemistry **272**: 563-571.
- Williamson, K. C., Fujioka, H., Aikawa, M. and Kaslow, D. C. (1996). "Stage-specific processing of Pfs230, a *Plasmodium falciparum* transmission-blocking vaccine candidate." Molecular and Biochemical Parasitology **78**(1-2): 161-169.
- Wu, Y., Sifri, C. D., Lei, H. H., Su, X. Z. and Wellems, T. (1995). "Transfection of *Plasmodium falciparum* within human red blood cells." Proceedings of the National Academy of Sciences of the U S A **92**: 973-977.
- Young, J. A., Fivelman, Q. L., Blair, P. L., de la Vega, P., Le Roch, K. G., Zhou, Y., Carucci, D. J., Baker, D. A. and Winzeler, E. A. (2005). "The *Plasmodium falciparum* sexual development transcriptome: A microarray analysis using ontology-based pattern identification." Molecular and Biochemical Parasitology **143**(1): 67-79.
-

-
- Yue, H., Eastman, P. S., Wang, B. B., Minor, J., Doctolero, M. H., Nuttall, R. L., Stack, R., Becker, J. W., Montgomery, J. R., Vainer, M. and Johnston, R. (2001). "An evaluation of the performance of cDNA microarrays for detecting changes in global mRNA expression." Nucleic Acids Research **29**(8): E41-41.
- Zhao, S., Ooi, S. L. and Pardee, A. B. (1995). "New primer strategy improves precision of differential display." Biotechniques **18**(5): 842-850.
-