

https://dspace.org/

Cloning and expressing PfMyb2, a Plasmodium falciparum DNA-binding protein

Baker, Gillian Lee

2009-04-09T11:50:35Z
http://hdl.handle.net/10539/6886

Downloaded from a DSpace repository. https://dspace.org/

Cloning and expressing *Pf*Myb2, a *Plasmodium falciparum* DNA-binding protein

Gillian Lee Baker

A dissertation submitted to the Faculty of Science, University of the Witwatersrand, Johannesburg, in fulfilment of the requirements for the degree of Master of Science.

March 2008

DECLARATION

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

_____

G. L. Baker

_____ day of _____ 2008

ABSTRACT

The malaria parasite, *Plasmodium falciparum*, is responsible for more than 3000 deaths in Africa per day. The complex lifecycle of the parasite in the human and mosquito hosts requires a tight control of gene regulation. This area of the parasite's biology is not well understood but recent advances in the application of molecular techniques to *P. falciparum* and the publication of the parasite genome sequence have provided the means to gain insight into mechanisms of gene regulation in the parasite.

*Pf*Myb2 is annotated as a putative transcription factor. It contains two Myb-like DNA-binding domains which have been cloned and demonstrated to bind Myb regulatory elements in the promoters of two *P. falciparum* genes, therefore validating it as a DNA-binding protein (Meyersfeld, 2005). In this study the full length *Pf*Myb2 gene was cloned into the pGEX-4T-2 expression vector and induced to express recombinant *Pf*Myb2 protein fused to a glutathione-S-transferase (GST) tag in *E. coli*. This prokaryotic expression system produced mainly insoluble recombinant protein. The protein was extracted from the bacterial inclusion bodies, solubilised, refolded and purified using affinity chromatography. A 500ml culture yielded ~20-40ug purified *Pf*Myb2-GST in total.

Bioinformatic analysis revealed homologues of *Pf*Myb2 in other *Plasmodium* species and an unusually high homology to the human CDC5 protein, a multifunctional protein involved in transcriptional regulation of the cell cycle as well as in pre-mRNA splicing. *Pf*Myb2 and CDC5 both contain a REB1 domain, part of a yeast protein. This domain binds to DNA to enhance transcription factor binding and it may also remodel chromatin resulting in activation or silencing of promoters. Based on the high homology to CDC5, including the REB1 domain, it is speculated that *Pf*Myb2 is a multifunctional protein that plays a role in transcription, chromatin remodelling and RNA processing in the parasite.

The full length recombinant *Pf*Myb2 produced in this study lays the foundation for future studies to validate these speculative functions.

PRESENTATIONS ARISING FROM THIS WORK

Poster Presentation: Bioinformatic analysis of *Pf*Myb2, a *Plasmodium falciparum* DNA-binding protein (Gillian L. Baker and Theresa L. Coetzer)

Presented at:

SASBMB xxth Conference, 2-5 July 2006

Faculty of Health Sciences Research Day, 23 August 2006

The Molecular and Biology Group Symposium, 5 October 2006

Publication: Manuscript in preparation (Meyersfeld, D., Baker, GL., Vaquero, C. and Coetzer, TL.)

To be published in Open Parasitology.

To Mom and Dad

ACKNOWLEDGEMENTS

Firstly I would like to thank my Supervisor, Professor Theresa L. Coetzer for her constant guidance and encouragement. Her expertise and support have made this study and dissertation possible.

I would also like to thank the following people:

- Bridge, Cath and Sheena for all the laughs, coffees and sushi. You guys have been a great source of support and companionship throughout.
- Kuben and Danny, for your guidance in the laboratory.
- The rest of the Red Blood Cell Membrane Unit
- My family near and far, for their constant support at all hours of the day and night (thanks to modern technology).
- Department of Pharmacy and Pharmacology (University of the Witwatersrand) for the stocks of *P. falciparum* 3D7 strain.
- Finally to Colin, for your love, support and guidance in the last year -without which I'm not sure what I would have done. Thank you also for the help with MS word and making this dissertation possible.

Thanks to the following research agencies:

- National Research Foundation
- Belgian Embassy for the BTC Fellowship
- South African Malaria Initiative

ETHICS CLEARANCE

Ethical clearance was obtained in order to culture the malaria parasites in human blood from healthy volunteers: University of Witwatersrand, protocol number M03-11-06, committee for research on human subjects (medical).

TABLE OF CONTENTS

DECLARATION ..... **ii**
ABSTRACT ..... **iii**
PRESENTATIONS ARISING FROM THIS WORK ..... **iv**
ACKNOWLEDGEMENTS ..... **vi**
ETHICS CLEARANCE ..... **vi**
TABLE OF CONTENTS ..... **vii**
LIST OF FIGURES ..... **xi**
LIST OF TABLES ..... **xiii**
ABBREVIATIONS ..... **xiv**

1. INTRODUCTION **1**
1.1 The impact of malaria 1
1.2 The *P. falciparum* lifecycle 2
1.3 Eukaryotic gene regulation 4
1.4 The *P. falciparum* genome study and parasite biology 5
1.4.1 *P. falciparum* genome 5
1.4.2 *P. falciparum* proteome 6
1.4.3 *P. falciparum* transcriptome 6
1.4.4 The study of *P. falciparum* gene regulation 6
1.5 Aspects of gene regulation in *P. falciparum* 7
1.5.1 Epigenetic gene regulation 8
1.5.2 Post-transcriptional control of gene regulation 9
1.5.3 Transcriptional gene regulation 9
1.5.3.1 Regulatory sequences in *P. falciparum* 10
1.5.3.2 Transcriptional control of *P. falciparum* gene expression 11
1.5.3.3 Specific transcription factors in *P. falciparum* 12
1.6 The *Myb*-gene family and proteins 13
1.7 A specific *P. falciparum* transcription factor, *Pf*Myb1 15
1.8 A putative transcription factor, *Pf*Myb2 16
1.9 Objectives of the study 17

2. MATERIALS AND METHODS 18
2.1 *P. falciparum* parasite culture 18
2.1.1 Preparation for parasite culture 18
2.1.1.1 Culture media 18
2.1.1.2 Heat inactivation of plasma 18
2.1.1.3 Erythrocytes 18
2.1.2 Preparation of culture from frozen stocks 19
2.1.3 Culture maintenance 19
2.1.3.1 Viewing parasite cultures 19
2.1.3.2 Changing medium of cultures 20
2.1.3.3 Parasite culture synchronisation 20
2.2 DNA extraction and analysis 21
2.2.1 DNA extraction protocol 21
2.2.2 Determining DNA concentration and purity 22
2.3 PCR amplification of *Pf*Myb2 24
2.3.1 PCR conditions 24
2.3.2 Phenol/chloroform purification of PCR product 26
2.4 Cloning *Pf*Myb2 26
2.4.1 Large-scale preparations of expression vectors 26
2.4.2 Enzyme digestion of *Pf*Myb2 PCR product and vectors 28
2.4.3 Ligation and subcloning of *Pf*Myb2 into pET-41a and pGEX-4T-2 29
2.4.4 Transformation of DH5α *E.coli* cells with pET-41a and pGEX-4T-2 constructs 30
2.4.5 Sequencing of *Pf*Myb2 insert 32
2.4.6 Ligation and subcloning of *Pf*Myb2 into pTriEx-3 32
2.4.7 Transformation of Rosetta 2(DE3) *E. coli* cells with expression vector constructs 33
2.5 Expression of Recombinant *Pf*Myb2 33
2.5.1 Induction of protein expression 33
2.5.1.1 Induction with Overnight Express™ Instant TB Medium 33
2.5.1.2 Induction with IPTG 35
2.5.2 SDS-PAGE analysis 35
2.5.3 Western blot analysis 38

2.6. Purification of recombinant *Pf*Myb2 39
2.6.1 Purification of inclusion bodies and refolding of insoluble protein 39
2.6.2 Affinity purification of the GST-linked recombinant *Pf*Myb2 41
2.6.2.1 Affinity purification of *Pf*Myb2-GST from the soluble protein fraction 41
2.6.2.2 Affinity purification of refolded *Pf*Myb2 42
2.7 Bioinformatic analysis of *Pf*Myb2 43

3. RESULTS 44
3.1 DNA Extraction 44
3.2. PCR amplification of *Pf*Myb2 45
3.2.1 Optimisation of annealing temperatures for PCR segments 45
3.2.2 Optimisation of amount of DNA used for PCR 46
3.3 Cloning of *Pf*Myb2 into pET-41a 48
3.4. Cloning and expression of pGEX-4T-2-*Pf*Myb2 constructs 50
3.4.1 Cloning of *Pf*Myb2 into pGEX–4T-2 50
3.4.2 Induction of *Pf*Myb2-GST fusion protein in Rosetta *E. coli* 53
3.5. Cloning of *Pf*Myb2 into pTriEx-3 61
3.5.1 Cloning into pTriEx-3 62
3.5.2 Induction of pTriEx-*Pf*Myb2 constructs 62
3.6 Bioinformatic analysis of *Pf*Myb2 64
3.6.1 General features of *Pf*Myb2 64
3.6.2 *Plasmodium* homologues of *Pf*Myb2 65
3.6.3 Comparison between *Pf*Myb2 and *Pf*Myb1 67
3.6.4 Homology between *Pf*Myb2 and human CDC5 68

4. DISCUSSION 70
4.1 Recombinant *Pf*Myb2-GST protein expression in *E. coli* 70
4.1.1 Expression of *P. falciparum* proteins in *E. coli* 70
4.1.2 Strategies to improve expression of soluble *Pf*Myb2-GST 71
4.1.3 Purification of *Pf*Myb2-GST from inclusion bodies 71
4.1.4 Additional strategies to enhance soluble expression of *Pf*Myb2 73
4.2 Possible cellular functions of *Pf*Myb2 74

4.2.1 Role of *Pf*Myb2 as a transcription factor 74
4.2.2 Possible role of *Pf*Myb2 in RNA processing 77
4.2.3 Possible role of *Pf*Myb2 in chromatin remodelling 78
4.3 Future studies 79
4.4 Conclusion 80

5. LIST OF SUPPLIERS 81

6. APPENDIX 84
A-1: pET-41a 84
A-2: pGEX-4T-2 85
A-3: *Pf*Myb2-GST gene sequence in pGEX-4T-2 86
A-4: *Pf*Myb2-GST fusion protein sequence 87
A-5: pTriEx-3 88
A-6: *Pf*Myb2-his gene sequence in pTriEx-3 89
A-7: *Pf*Myb2-his fusion protein sequence 89
A-8: Sequence analysis of the *Pf*Myb2 gene inserted into pGEX-4T-2 90

7. REFERENCES 94

LIST OF FIGURES

1. INTRODUCTION

Figure 1.2.1: The lifecycle of *P. falciparum* 2
Figure 1.3.1: Transcriptional activation of genes 5
Figure 1.6.1 The general structure of Myb proteins. 15

3. RESULTS

Figure 3.1: Mixed parasite culture 44
Figure 3.2: *P. falciparum* DNA extraction 45
Figure 3.2.1: Optimisation of annealing temperatures for *Pf*Myb2 PCR 47
Figure 3.2.2: Optimisation of DNA concentration per PCR reaction 48
Figure 3.3.1: The pET-41a vector and *Pf*Myb2 gene insert digested with *Bam*HI and *Xho*I for cloning 49
Figure 3.3.2: Digested pET-41a plasmid constructs 50
Figure 3.4.1.1: The pGEX-4T-2 vector and *Pf*Myb2 gene insert digested with *Bam*HI and *Xho*I for cloning 51
Figure 3.4.1.2: PCR amplification of pGEX-4T-2 transformed DH5α *E. coli* colonies 52
Figure 3.4.1.3: The pGEX-4T-2-*Pf*Myb2 constructs, digested with *Bam*HI and *Xho*I, from transformed DH5α *E. coli* cells 53
Figure 3.4.2.1: Western blot of *Pf*Myb2-GST protein expressed from three of the pGEX-4T-2-*Pf*Myb2 constructs 55
Figure 3.4.2.2: Laemmli gel of induced Rosetta *E. coli* containing either pGEX2:2 or pGEX-4T-2 56
Figure 3.4.2.3: Western blot of soluble and insoluble fractions of Rosetta *E. coli* expressing the pGEX2:2 construct 57
Figure 3.4.2.4: Laemmli gel of refolded *Pf*Myb2-GST 58
Figure 3.4.2.5: Western blot of refolded *Pf*Myb2-GST 59
Figure 3.4.2.6: Laemmli gel of refolded *Pf*Myb2-GST using varying elution conditions 60

Figure 3.4.2.7: Western blot of refolded *Pf*Myb2-GST using varying elution conditions 61
Figure 3.5.1.1: PCR amplification of pTriEx-3 transformed DH5α *E. coli* colonies 63
Figure 3.5.1.2: Restriction enzyme digest of pTriEx-3-*Pf*Myb2 constructs 63
Figure 3.6.1: Bioinformatic predictions of *Pf*Myb2 64
Figure 3.6.2: Expression levels of *Pf*Myb2 mRNA in the intraerythrocytic life stages of *P. falciparum* 65
Figure 3.6.3: Alignment of *Pf*Myb2, *P. berghei* Myb2 and *P. vivax* Myb2 66
Figure 3.6.4: The conserved DNA-binding domains in *Pf*Myb2 and *Pf*Myb1 68
Figure 3.6.5: Alignment of *Pf*Myb2 and human CDC5 protein 69

4. DISCUSSION

Figure 4.2 The general structure of *Pf*Myb2, *Pf*Myb1 and human CDC5 proteins 76

LIST OF TABLES

1. INTRODUCTION

Table 1.5.1: Putative transcription factors in *P. falciparum.* 14

2. MATERAILS AND METHODS

Table 2.3.1: Components of the PCR reaction using the Promega Master Mix 25
Table 2.3.2: Components of the PCR reaction using the Eppendorf TripleMaster PCR system 25
Table 2.3.3: Optimised PCR conditions for *Pf*My2 PCR 25
Table 2.4.2.1: The components of the restriction enzyme digestion 29
Table 2.4.3.1: Components of the ligation reaction 30
Table 2.5.1.1: Protease inhibitors contained in the Calbiochem Protease Inhibitor Cocktail Set III. 34
Table 2.5.2.1: Components of the Laemmli separating and stacking gels 36

3. RESULTS

Table 3.6: Comparison of *Pf*Myb2 whole protein and DNA-binding domains with other homologues proteins 67

ABBREVIATIONS

A	adenine
ACD	acid citrate/dextrose
BLASTP	basic local alignment search tool (protein)
bp	base pair
BSA	bovine serum albumin
C	cytosine
CDC5	cell division cycle 5
cm	centimetre
cm^2	square centimetre
°C	degrees Celsius
DNA	deoxyribonucleic acid
DTT	dithiothreitol
E.	*Escherichia*
EDTA	ethylenediamine tetraacetic acid
EMSA	electrophoretic mobility shift assays
g	gram
G	guanine
GBP130	glycophorin binding protein 130
GST	glutathione-S-transferase
HRP	horseradish peroxidase
HTH	helix-turn-helix
hsp	heat shock proteins
IPTG	Isopropyl-1-thio-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
MCS	multiple cloning site
ng	nanogram
nm	nanometre
P.	*Plasmodium*
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PMSF	phenylmethylsulfonylfluoride
rRNA	ribosomal ribonucleic acid
RNA	ribonucleic acid
RNase	ribonuclease
REB1	RNA polymerase I enhancer binding
rpm	revolutions per minute
RPMI	Roswell Park Memorial Institute (medium)
PAGE	poly acrylamide electrophoresis
PBS	phosphate buffered saline
*Pf*PCNA	proliferating cell nuclear antigen
pI	isoelectric point
PMSF	phenylmethylsulfonylfluoride
Pol II	RNA polymerase II
Pu	purine
Py	pyrimidine
SDS	sodium dodecyl sulphate
*Pf*Sir2	silent information regulator 2 protein
SV40	simian virus 40
T	thymine
TAE	Tris-acetate-EDTA
TE	Tris-EDTA
TEMED	N,N,N,N-tetramethylethylenediamine
TF	Transcription factor
Tris	hydroxymethyl methylamine
TBS	Tris Buffered Saline
U	units
ul	microlitre
uM	micromolar
W	Watts

1. INTRODUCTION

1.1 The impact of malaria

Malaria is a disease caused by four species of protozoan parasites belonging to the genus *Plasmodium*. It affects between 300-600 million people worldwide and is responsible for more than 1 million deaths annually (Snow et al., 2005, Greenwood et al., 2005). The *Plasmodium* parasites are transmitted to humans by the female *Anopheles* mosquito and the geographical distribution of this insect vector therefore determines which regions will be affected by malaria. The disease is endemic to Africa, Asia, Central and South America and hence is a problem mainly in developing countries (Snow et al., 2005). The disease not only causes ~3000 deaths per day (mostly children in sub-Saharan Africa) having great social implications, but it also affects productivity and negatively impacts on the economy of the endemic areas. In Africa alone, the total estimated cost of malaria is US$ 12 billion. *Plasmodium falciparum* (*P. falciparum*) is the major cause of malaria in Africa (Greenwood et al., 2005). In South Africa this disease affects only certain areas of the Limpopo province, Mpumalanga and KwaZulu Natal (www.doh.gov.za/facts/stats-notes/2004/malaria.htm).

Efforts to eradicate malaria were implemented after the end of World War II through the use of insecticide to kill the mosquito vector and chloroquine to treat the disease. In many areas this was successful and these efforts have impacted greatly on the distribution of the disease worldwide. Certain areas however still remain affected and in the last decade there has been a resurgence of malaria, in Africa particularly. A number of factors have contributed to this. Civil unrest and war in malaria areas has lead to a lack of medical facilities and treatment for the disease. Global warming may play a role by increasing mosquito vector survival in previously cooler areas, resulting in the larger distribution of the vector and spread of malaria. However, the most relevant factor for the resurgence of malaria is the resistance that the parasite is developing to drugs commonly used to fight the disease, as well as resistance in mosquitoes to insecticides used to control the vector (Greenwood et al., 2005). With the increase in drug-resistant *Plasmodium* species, there is a greater urgency to find

new drug targets and interventions to control the disease. By studying the parasite's biology and pathogenesis, novel means for intervention may be uncovered.

1.2 The *P. falciparum* lifecycle

Plasmodium falciparum has evolved a complex lifecycle to survive within the human host and the insect vector. During its lifecycle this single celled parasite will invade human hepatic and erythrocytic cells, while avoiding the host immune response. Once taken up by the mosquito it will need to survive in the gut and salivary glands of the vector. Throughout this complex lifecycle the parasite undergoes many morphological changes as seen in figure 1.2.1.

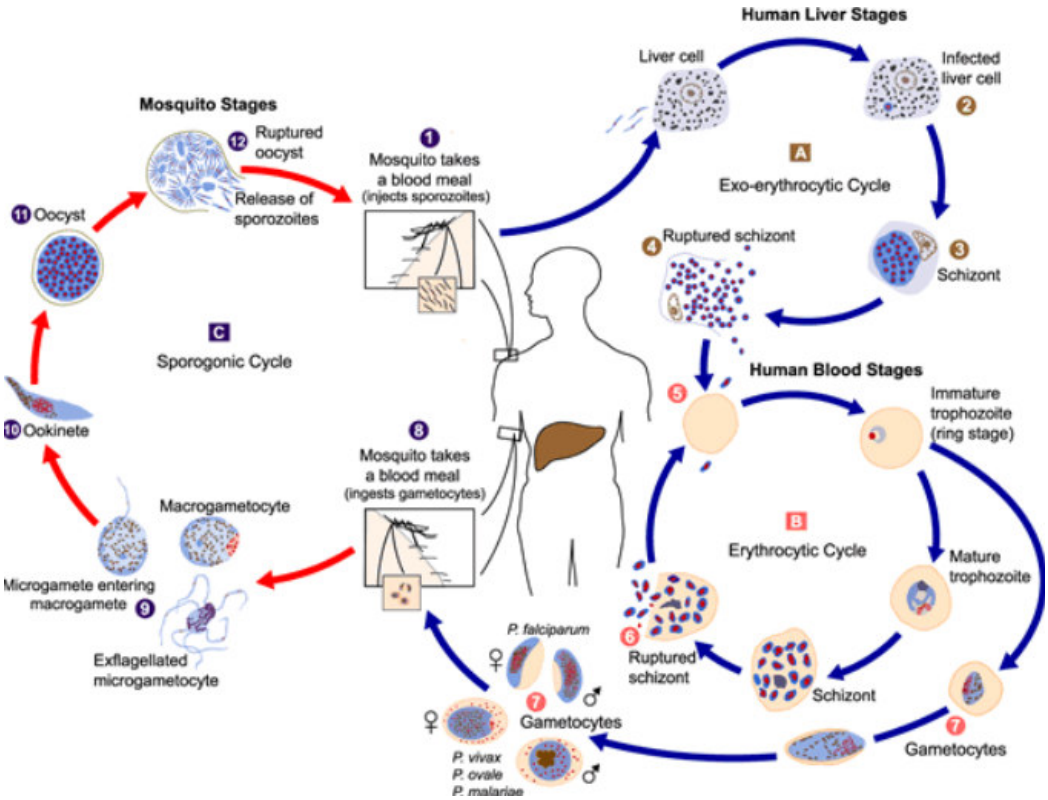


Figure 1.2.1: The lifecycle of *P. falciparum*
This diagram of the *P. falciparum* lifecycle is adapted from the Center for Disease Control and Prevention web page (http://www.dpd.cdc.gov/dpdx/Default.htm).
Explanation of the lifecycle stages is in the text.

The process of human invasion begins when a *P. falciparum*-infected mosquito takes a blood meal, injecting sporozoite forms of the parasite beneath the skin or into the bloodstream. The sporozoites move on to invade hepatic cells of the liver, where they

undergo a number of asexual division cycles, over a period of two weeks, to form thousands of merozoites (stages 1-4 of A in figure 1.2.1).

The merozoite forms of the parasite enter into the bloodstream where each one is able to invade an erythrocyte. Once in the erythrocyte the parasite will undergo different morphological changes over a 48 hour period to produce more merozoites. These stages are known as rings, trophozoites, schizonts and merozoites (stages 5-6 of B in figure 1.2.1). This intraerythrocytic cycle of the parasite is responsible for the pathogenesis of malaria. Each merozoite that infects an erythrocyte is able to produce up to 20 new merozoites, which burst from the infected cell. This process destroys the infected cells, increases parasite number rapidly due to successive infections of erythrocytes and results in the onset of disease symptoms. Destruction of the red blood cells leads to chills, fevers and anaemia. Small blood vessels become blocked with parasitised erythrocytes adhering to their lining, causing a decreased blood and oxygen supply to tissues. The inflammatory response of the host is also affected. All these factors contribute to cause severe malaria (Miller et al., 2002).

It is also during the intraerythrocytic stage that sexual differentiation of *P. falciparum* takes place. This is important for dissemination of the disease as the *Anopheles* mosquito takes up gametocytes during a blood meal. Upon infection of the erythrocyte some parasites are induced to produce either macro- or microgametocytes (stage 7 of B in figure 1.2.1) instead of undergoing asexual division. Once ingested by the mosquito the gametocytes undergo gametogenesis. During this phase a single female and male gametocyte produce one female gamete and 8 male gametes respectively. The gametes fuse to form a zygote that undergoes meiotic divisions within the oocyst and produce haploid sporozoites. These then move from the mosquito gut into the salivary glands, where they will eventually be injected into another human host during a blood meal (stages 8-12 of C in figure 1.2.1) (Lobo and Kumar, 1998).

This complex lifecycle that *P. falciparum* has developed has been well studied. The biology of the parasite such as mechanisms of erythrocyte invasion, metabolism, host immune evasion, cell cycle control and differentiation, as well as induction of asexual

division and sexual differentiation are slowly being elucidated. The application of molecular techniques to *P. falciparum* and the sequencing of the parasite genome have resulted in major advances in our knowledge. However, not much is known about how the parasite regulates the different stages of its lifecycle but it is clear that a tight control of gene expression is required (Deitsch, 2004). Therefore the study of gene regulation is of great importance and may lead to novel means of intervention and control of this important human pathogen.

1.3 Eukaryotic gene regulation

In eukaryotes the initiation of transcription is the most important step in gene expression. The RNA polymerase II (pol II) enzyme complex is responsible for the transcription of protein coding genes into pre-mRNA. General transcription factors such as TFIIA, TFIIB, TFIIC and TFIID, including the TATA binding protein (TBP), guide pol II to the transcription-initiation site (figure 1.3.1) to begin transcription of the gene. Promoters in eukaryotes have a bipartite structure containing a basal promoter that is regulated by an upstream enhancer element. Promoters are generally close to transcription start sites, whereas enhancer sequences can be located 50kb or more from the coding region. Specific transcription factors bind to promoters and only activate certain genes, while the general transcription factors are required by pol II to transcribe all activated genes. Initiation of transcription is influenced by several parameters including chromatin structure and concentration of specific transcription factors. Post-transcriptional mechanisms such as mRNA stability and processing provide further control of gene expression (Lodish et al., 2000).

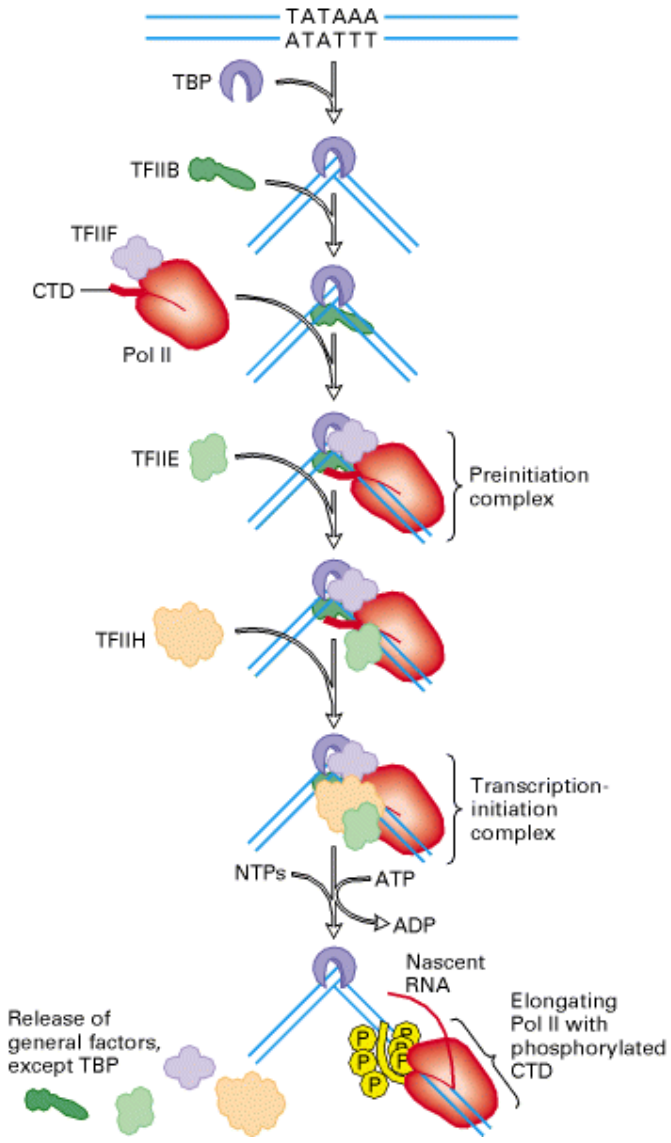


Figure 1.3.1: Transcriptional activation of genes
This figure was adapted from Lodish et al. (2000) and depicts transactivation of genes in eukaryotes by general transcription factors and the initiation of transcription by RNA polymerase II.
(Transcription factors: TFIID, TFIIE, TIIF and TFIIH; TATA binding protein: TBP;
C-terminal repeat domain: CTD; Adenosine triphosphate: ATP; Adenosine diphosphate: ADP; Ribonucleoside triphosphates: NTPs; RNA polymerase II: pol II)

1.4 The *P. falciparum* genome study and parasite biology

1.4.1 *P. falciparum* genome

The full genome sequence of *P. falciparum* (Gardner et al., 2002) has revealed an extreme AT bias, which distinguishes it from other genomes. Over 80% of its genome

is made up of A-T nucleotides and intergenic regions can contain up to 90%. Genes are monoscistronically transcribed and the resulting mRNA contains 3' and 5' untranslated regions, a poly-A-tail and a 5'cap. Genes are under the control of bipartite promoters, which consist of a basal promoter and an upstream enhancer element. Some of these promoters contain homopolymeric (dA:dT) tracts. Unlike eukaryotes, these *cis*-acting elements are located much further upstream from the transcriptional start site (Horrocks et al., 1998, Deitsch, 2004). The promoters discovered in *P. falciparum* are discussed in more detail in section 1.5.3.1.

1.4.2 *P. falciparum* proteome

The *P. falciparum* genome encodes ~5300 proteins and 60% of these are hypothetical with no known function. Parasite proteins that have homology to those in eukaryotes are often very large and can be up to 50% bigger than their eukaryote homologues. Many contain low complexity regions made up of stretches of repeating amino acids, usually asparagines (Aravind et al., 2003, Gardner et al., 2002). The function of these regions is not certain but they may be involved in protein-protein interactions as they loop out around the protein. These regions are thought to change frequently and may therefore play a role in host immune evasion as antibodies will no longer recognise the exposed and changed epitope (Brocchieri, 2001).

1.4.3 *P. falciparum* transcriptome

Microarray analysis of the intraerythrocytic lifecycle showed that genes with related functions are expressed at specific stages, even if they are not clustered in the genome and each lifecycle stage expresses different levels of mRNA. Many of the proteins encoded by these genes are stage-specific and expressed to perform certain functions at the particular lifecycle stage. Even the common 'house keeping' genes have varying levels of expression at various stages. All these factors demonstrate that gene expression is tightly controlled in the parasite (Bozdech et al., 2003, Le Roch et al., 2003).

1.4.4 The study of *P. falciparum* gene regulation

The study of gene regulation in *P. falciparum* has been limited in the past, but molecular techniques have recently been developed in *P. falciparum* (De Koning-Ward et al., 2000) and the release of the full *P. falciparum* genome sequence has

facilitated genome mining approaches to study gene regulation in the parasite (Gardner et al., 2002). The process of gene expression in *P. falciparum* occurs via the interaction of *cis*-regulatory sequences and *trans*-acting factors that drive gene transcription. Factors controlling gene regulation of the parasite lifecycle, as well as mechanisms to determine which copies of the multi-gene families are expressed during the different stages are slowly being elucidated (Deitsch, 2004).

Molecular techniques such as transfections, reporter gene analysis, deletion analysis of regions upstream of genes and electromobility shift assays (EMSA) have all been crucial in the discovery of DNA regulatory sequences and transcription factors. Bioinformatic genome mining has also revealed putative *cis*- and *trans*-acting factors involved in gene regulation that have been found in other organisms. As these elements are discovered, they need to be validated in *P. falciparum* using molecular techniques. Genome mining, however, is based on sequence homology and *P. falciparum* genes and proteins are quite divergent from higher eukaryotes, therefore it may not be as effective in the search for homologous proteins found in other eukaryotes (Deitsch, 2004). The use of proteomics (Florens et al., 2002), serial expression of gene analysis and microarrays (Bozdech et al., 2003, Llinas et al., 2006, Florent et al., 2004) has led to the study of the regulation of multiple gene families or groups of genes simultaneously.

1.5 Aspects of gene regulation in *P. falciparum*

The general transcription-associated proteins have been identified in *P. falciparum*, including RNA polymerase I, II and III, TFIIB, TFIIE and the TATA binding protein of TFIID (Callebaut et al., 2005, Coulson et al., 2004). Proteins involved in post-transcriptional gene regulation and chromatin remodelling have also been identified (Horrocks et al., 1998, Kumar et al., 2004). Therefore it is thought that a multi-level control of gene expression may take place in the parasite, using different control mechanisms.

1.5.1 Epigenetic gene regulation

The structure of the chromatin surrounding the genes of *P. falciparum* is likely to have an effect on transcriptional control of the genes. Protein complexes allow the DNA to dissociate from the histones in an ATP-dependent reaction. This renders the DNA promoter sites more accessible to transcription factors and therefore activates these promoters (Lodish et al., 2000). The genes encoding components of the enzyme complexes ISWI and SNF2L (involved in nucleosome remodelling), histone acetylase and deacetylase (involved in histone modification) have all been identified in the *P. falciparum* genome. It is thought that each lifecycle stage of the parasite may contain varying arrangements of chromatin organisation and that this may control which genes are expressed at the different developmental stages (Coulson et al., 2004).

The promoter structure of *P. falciparum* may also alter the form of the chromatin flanking the gene sequences. The homopolymeric dA:dT tracts found in some parasite promoters form a rigid structure in the DNA and may render the promoter site more accessible to transcription factors or increase its affinity for transcription factor binding. Homopolymeric tracts are usually unstable and lost during DNA replication, however there has been a strong selective advantage for these tracts in *P. falciparum* as they are found throughout the genome, generally flanking genes (Horrocks et al., 1998), implying that they are involved in gene regulation.

The *var* genes are a family of multi-copy genes that encode antigenic surface proteins (one *P. falciparum* parasite may contain between 50 to 150 *var* genes). Only one gene is expressed in a single parasite at a time, implying that the parasite is able to switch expression of the *var* genes, which helps to evade the host's immune response. *Var* genes are located at the subtelomeric regions of the chromosomes and epigenetic control of their expression has been demonstrated (Scherf et al., 1998). The silent information regulator 2 protein (*Pf*Sir2) binds to the subtelomeric region of the parasite's chromosome, altering chromatin structure through histone hypoaceteylation and the induction of a heterochromatin gradient in this region. This renders the *var* genes silent and the removal of *Pf*Sir2 results in their expression (Freitas-Junior et al., 2005).

1.5.2 Post-transcriptional control of gene regulation

Once RNA is produced it undergoes a number of processes before it is translated into protein. Post-transcriptional control mechanisms alter gene expression by manipulating various steps of the mRNA processing cycle and therefore alter the amount of protein produced (Lodish et al., 2000). Coulson et al. (2004) discovered through sequence alignments that the most abundant protein involved in post-transcriptional modification of pre-mRNA present in the *P. falciparum* genome is the CCCH-type zinc finger. The genome appears to contain twice the number of CCCH-type zinc fingers than higher eukaryotes. This motif functions at a post-transcriptional level by playing a role in mRNA stability, localisation and translation. Homologues of the CCR4-NOT coactivator were also discovered during this study. This complex has deadenylase activity, which affects the rate of mRNA translation and decay. This suggests that gene regulation in *P. falciparum* also relies on post-transcriptional controls through mRNA stability. Gunasekera et al. (2007) discovered three putative regulatory repeats upstream of gene clusters expressed in chloroquine-treated *P. falciparum* cultures. These motifs were able to drive reporter gene expression, however, gel mobility shift assays revealed that none of the motifs were bound by proteins at the DNA level but two were bound at the RNA level indicating post-transcriptional control of gene expression through mRNA interactions.

Post-transcriptional control of genes may be a more efficient way of controlling protein levels in *P. falciparum*. The parasite encounters rapid, but predictable changes in environment during its lifecycle. Therefore *P. falciparum* gene regulation can occur in an ordered, predictable manner using post-transcriptional control which allows a quick response to the new environment (Coulson et al., 2004). The gametocytes in the human host contain some mRNA which are not translated in the erythrocytic stage, but are rapidly activated by post-transcriptional mechanisms in the mosquito vector to produce proteins required for further development (Florens et al., 2002).

1.5.3 Transcriptional gene regulation

The gene structure in the parasite is similar to that of eukaryotes, however, some of the regulatory sequences found in the enhancer regions of the promoters share no

homology with those found in other eukaryotes. These promoters do not drive reporter gene expression in cultured mammalian COS-7 cells and conversely, viral promoters (such as the SV40) that are functional in eukaryotes do not drive reporter gene expression in *P. falciparum* (Horrocks et al., 1998). Therefore it is likely that the transcription factors that bind these enhancer sequences are specific to *Plasmodium* species and share no or very little homology with other eukaryotic transcription factors.

1.5.3.1 Regulatory sequences in *P. falciparum*

Promoter regions with homology to those found in eukaryotes have been identified in *P. falciparum*. TATA motifs have been demonstrated in the promoters of several *P. falciparum* genes and *P. falciparum* TBP has also been identified (Ruvalcaba-Salazar et al., 2005, McAndrew et al., 1993). Gene promoters containing the TATA box motif include proliferating cell nuclear antigen (*Pf*PCNA) (Horrocks and Kilbey, 1996), heat shock proteins (*hsp*) (Su and Wellems, 1994), calmodulin (Robson and Jennings, 1991) and two sexual stage specific genes: *pfs16* and *pfs25* (Dechering et al., 1999).

Other common regulatory sequences in eukaryotes are the CpG islands. These contain a 20-50 nucleotide stretch ~100bp from the transcriptional start site of genes and are recognized by the transacting protein, SP1, which drives transcription (Lodish et al., 2000). The *P. falciparum* homologue of this transcription factor has not been identified, however, CG-rich regions have been demonstrated in the parasite transcriptional start sites for the *hsp*86 (Wu et al., 1995) and *Pf*PCNA genes (Horrocks and Kilbey, 1996) and may have a role to play in transcriptional activation.

OCT1 has also been identified in promoters of *Plasmodium* species and the two regulatory regions, ATTCAAAT and ATGCATAT, are located 270bp and 230bp, respectively, upstream of the transcriptional start site of the *Pf*PCNA gene (Horrocks and Kilbey, 1996). Putative CAAT box motifs have been found in the promoters of *P. falciparum* sexual stage-specific *pfs25* and *pgf27* genes (Dechering et al., 1999, Olivieri et al., 2008). Their functional relevance, although well known as regulatory motifs to which transcription factors bind to in higher eukaryotes, still needs to be elucidated in *P. falciparum*.

A number of regulatory sequences that have no homology to higher eukaryotic promoters have also been identified in the *P. falciparum* genome. These were discovered through deletion analysis of upstream regions flanking the initiation start sites, EMSA studies and reporter gene expression. Horrocks and Lanzer (1999) were the first to discover a regulatory sequence essential for promoter activity for the glycophorin binding protein 130 (GBP130) gene. This was a 5bp sequence that was located in a region between 544-507bp from the transcriptional start site. It bound nuclear factors and was essential in *GBP130* expression.

Bioinformatics has also been utilised to discover *P. falciparum* specific *cis*-regulatory sequences that may be essential in gene expression. This approach identifies sequences that are over-represented and upstream of clustered or co-ordinately expressed genes. Militello et al. (2004) used this approach for heat shock protein (*hsp*) genes in *P. falciparum* and identified a G-box element which contained two copies of the sequence, (A/G)NGGGG(C/A), in palindromic form. These sequences were validated using transfection studies and reporter gene expression. The G-box was located 195bp from the transcriptional start site and other genes were also found to contain single G-boxes, suggesting that the palindromic form is not essential for gene expression. However, how this regulatory sequence controls transcription has still not been elucidated. A number of other regulatory elements have been demonstrated in a few *P. falciparum* promoters (Wickham et al., 2003, Osta et al., 2002, Lopez-Estrano et al., 2007, Tham et al., 2007). The most recent study was performed by Young et al. (2008) using an *in silico* method that takes into account the extreme AT-rich regions of the parasite genome. They discovered ~30 putative regulatory sequences upstream of a number of genes and gene families that function in various aspects of the *P. falciparum* lifecycle. These regions still need to be validated using molecular techniques. This study, however, emphasises the importance of bioinformatics in the discovery of factors pertaining to *P. falciparum* gene regulation.

1.5.3.2 Transcriptional control of *P. falciparum* gene expression

It is thought that *P. falciparum* has developed a unique set of transcription factors, which would not be identified in the genome through simple homologous sequence

alignments to other eukaryotic transcription factors. During sequence homology searches, Coulson et al. (2004) found that the *P. falciparum* genome encodes only one third of the proteins involved with transcriptional regulation when compared to other eukaryotic genomes. This could support the fact that unique transcription factors have evolved in the parasite. Another school of thought to explain the small number of identified transcription factors in *P. falciparum* is that transcriptional activation occurs using fewer *trans*-acting elements. A study on the identification of regulatory sequences by Van Noort and Huynen (2006) indicated that *P. falciparum* genes have four to five regulatory sequence sites per gene. They speculate that different combinations of transcription factors bind to these sequences and this combinatorial effect drives gene expression. They also speculate that transcription factors may contain overlapping binding sites for regulatory sequences. Therefore fewer *trans*-acting elements are needed as they can bind to promoters in different combinations to produce differential gene expression and this explains the relative lack of transcription factors in the *P. falciparum* genome.

Various studies using microarrays have demonstrated that gene expression appears well regulated during the various stages of the intraerythrocytic lifecycle, as only genes relating to particular functions for each stage are expressed when required (Le Roch et al., 2003, Bozdech et al., 2003, Llinas et al., 2006). This expression pattern also seems to be consistent in three different strains of *P. falciparum* (3D7, Dd2 and HB3) (Llinas et al., 2006). Unless the parasite differentiates into its sexual forms, it undergoes a repeating cycle of gene expression to complete the intraerythrocytic cycle. Therefore the parasite will most likely contain a rigid program of gene regulation and a single induction event may lead to gene expression in a cascade-like manner. This less sophisticated process of gene regulation would require fewer transcription factors to initiate RNA synthesis (Bozdech et al., 2003).

1.5.3.3 Specific transcription factors in *P. falciparum*

Specific transcription factors in *P. falciparum* are slowly being discovered using molecular and bioinformatic techniques, however, until these proteins are characterised they remain only putative transcription factors. Hybridisation studies performed by Kumar et al. (2004) revealed a number of putative transcription factors

in the nuclear extracts of *P. falciparum* parasite cultures (containing either asexual, sexual or stressed parasites) that bound to eukaryotic *cis*-regulatory elements. However, none of these were differentially expressed in the different parasite cultures. Eight of these putative transcription factors (c-Myb, CREB, EGR, MEF-1, NFkB, E2F1, Smad3/4 and HSE) were further analysed using EMSA. Only two, c-Myb and MEF-1, had high specific binding for their regulatory elements. It is thought that the others may have a broader range of DNA binding specificity. Gissot et al. (2004) have also discovered five putative *P. falciparum*-specific transcription factors that are developmentally regulated during the intraerythrocytic cycle. These are *Pfhmg2*, *pfphD2*, *pfmyb3* (PF10_0143), *pfphDB* (PFL1905w) and *pfkrox* (MAL13P1.76). Table 1.5.1 summarises information available on *P. falciparum* transcription factors from the literature as well as online databases, PlasmoDB (www.plasmodb.org) and DBD: Transcription factor prediction database (www.transcriptionfactor.org). Many of these proteins are only hypothetical, emphasising the unusual composition of the *P. falciparum* genome and the possibility of the parasite evolving its own unique set of transcription factors.

These studies on gene regulation in *P. falciparum* have shown that gene expression in the parasite lifecycle is driven by transcription initiation. These transcription factors, albeit specific or general, may in turn be controlled by other epigenetic or post-transcriptional factors regulation.

1.6 The *Myb*-gene family and proteins

The first *Myb* gene identified was the oncogene, *v-Myb*, from the avian myeloblastosis virus. This gene was formed through recombination when the retrovirus was inserted into chicken cellular *Myb* (*c-Myb*) DNA. This led to the discovery of the *Myb* genes in vertebrates and other eukaryotes. This gene family and the proteins they encode are evolutionarily conserved amongst eukaryotes and have been found in fungi, plants and animals. Myb proteins bind to DNA and play a role in transcription of genes involved in cell proliferation and differentiation. Three members of the vertebrate *Myb* gene exist: *A-myb*, *B-myb* and *c-myb* (Oh and Reddy, 1999, Lipsick, 1996). In vertebrates, the *c-Myb* gene and protein have been the most extensively studied of the *Myb* gene family.

Table 1.5.1: Putative transcription factors in *P. falciparum*.

This information was summarised from online resources (PlasmoDB:wwwplasmodb.org and DBD:www.transcriptionfactor.org) and the literature (Kumar et al., 2004, Gissot et al., 2004, Boschet et al., 2004, Meyersfeld, 2005).

Transcription factor	Possible protein function	Reference
CREB	cAMP response element	Kumar et al., 2004
E2F1	Cyclin E regulation	Kumar et al., 2004
EGR	Early grown response element	Kumar et al., 2004
HSE	Heat shock element	Kumar et al., 2004
Smad3/4	Involved in signaling	Kumar et al., 2004
MEF-1	Myogenic cell differentiation	Kumar et al., 2004
NFkB	Type 1 interferon system regulator	Kumar et al., 2004
PF13_0088 (*Pf*Myb1)	Cell cycle and differentiation	Boschet et al., 2004; PlasmoDB
PFL0465c/ MAL13P1.76	Zinc finger transcription factor (*pfkrox*)	Gissot et al., 2004; DBD
MAL13P1.302	Putative SUMO ligase	DBD
MAL8P1.131	Putative Gas41 homologue	PlasmoDB
PF08_0056	Putative zinc finger protein	DBD
PF10_0058	Putative DnaJ protein	DBD
PF10_0143 (*Pf*Myb3)	Putative transcriptional activator ADA2; Myb family proteins	Gissot et al., 2004; PlasmoDB
PF11_0192	Putative histone acetyltransferase	DBD
PF11_0293	Putative multiprotein bridging factor type 1	DBD
PF13_0043	Putative CCAAT-binding transcription factor	PlasmoDB
PFA0470c	Putative cold-shock protein	DBD
PFE0355c	Putative serine protease; subtilisin family	DBD
MAL13P1.213	Putative transcription activator	PlasmoDB
PF10_0293	Putative transcription factor	PlasmoDB
PF10_0327 (*Pf*Myb2)	Putative transcription factor	Meyersfeld, 2005; PlasmoDB
PF13_0054	Putative transcription factor	PlasmoDB
PFB0290c	Putative transcription factor	PlasmoDB
PFD0360w	Putative transcription factor	PlasmoDB
PFE1025c	Conserved hypothetical protein	DBD
PFE1515w	Conserved hypothetical protein	PlasmoDB
PFF0490w	Conserved hypothetical protein	PlasmoDB
PFI1425w	Conserved hypothetical protein	PlasmoDB
PF11_0049	Conserved hypothetical protein	DBD
PFA0210c	Conserved hypothetical protein	PlasmoDB
PFC1060c	Conserved hypothetical protein	DBD
PFL1905w (*pfphDB*)	Conserved hypothetical protein	Gissot et al., 2004
PF10_0091	Hypothetical protein	DBD
PF11_0469	Hypothetical protein	DBD
PF14_0058	Hypothetical protein	DBD
PF14_0170	Hypothetical protein	DBD
PF14_0608	Hypothetical protein	DBD



Figure 1.6.1 The general structure of Myb proteins.
Myb family proteins contain three amino terminal DNA-binding domains designated R1, R2 and R3 (red). A transactivation domain (yellow) is located downstream from the DNA-binding domains and the carboxyl terminus contains a negative repressor element (grey).

The general structure of the Myb proteins is illustrated in figure 1.6.1. The DNA-binding domain defines the Myb family proteins. These amino terminal domains consist of two to three tandem repeats designated R1, R2 and R3 respectively, however, some Myb proteins contain more repeats. Two of the three motifs, usually R2 and R3, form two helix-turn-helix (HTH) repeats which bind a specific DNA regulatory sequence (PyAAG/TG), while the other motif stabilises the DNA-protein complex (Howe et al., 1990, Oh and Reddy, 1999). Each repeat motif contains regularly spaced tryptophan residues every 18 or 19 amino acids, as well as one highly conserved cysteine residue. The conserved tryptophan residues are thought to form a hydrophobic scaffold to stabilise the HTH repeat (Zargarian et al., 1999). Myb proteins contain a transactivation domain further downstream from the DNA-binding domain, consisting of acidic amino acid residues, while the carboxyl terminus functions as a negative repressor as truncated forms of the protein cause transactivation of transcription. Phosphorylation of certain amino acid residues in the protein and the interaction of Myb proteins with other co-activators and transcription factors result in enhanced promoter activity. Other proteins may contain Myb-like DNA-binding domains, however, not all these proteins act in transcriptional regulation (Oh and Reddy, 1999). An example is the TTAGGG repeat binding factor 1, which contains a carboxyl Myb-like domain that facilitates DNA-binding at the telomeres (Konig et al., 1998).

1.7 A specific *P. falciparum* transcription factor, *Pf*Myb1

*Pf*Myb1 was the first specific transcription factor, not forming part of the general transcription machinery, described in *P. falciparum* (Boschet et al., 2004). It contains three imperfect Myb-like DNA-binding domains but unlike other eukaryotic Myb

proteins these motifs are located in the carboxyl terminus. It is able to bind Myb regulatory elements (MRE) such as the chicken *mim-1* gene promoter and two putative MREs present in the promoters of two *P. falciparum* genes: *pfmap1* and *pfcrk1*. This transcription factor binds DNA in a sequence specific manner and regulates genes involved in cell cycle and differentiation (Boschet et al., 2004, Gissot et al., 2005). Microarray studies by Gissot et al. (2005) showed that the highest expression levels of *Pf*Myb1 were detected in the trophozoite stage. Knockdown of the *Pf*Myb1 mRNA caused up-regulation of *PfPk5*, a gene involved in trophozoite to schizont transition, and therefore a reduction in growth of the parasite. Seven genes were down-regulated and these included *PCNA1* which plays a role in DNA repair and cell cycle control; phosphoglycerate kinase involved in the glucose pathway; and two histone genes: *H3* and *H2A*, that are controlled by a bi-directional promoter. Computational analysis revealed that only one of these genes did not contain a MRE in its promoter and chromatin immuno-precipitation assays followed by PCR amplification of gene promoters demonstrated that *Pf*Myb1 was able to bind the promoters of six of these genes (including the promoter that did not contain a MRE). Therefore *Pf*Myb1 directly regulated expression of these genes. Two genes (*Pf*Pk2 and *Pf*Tbp) did not have direct interaction at the promoter level to *Pf*Myb1 and therefore may have differential expression due to secondary effects as a result of the protein's knockdown. These studies on *Pf*Myb1 confirmed sequence homology data and indicated that it belongs to the well conserved family of Myb proteins found in other eukaryotes and plays a role in transcriptional regulation of genes (Boschet et al., 2004, Gissot et al., 2005).

1.8 A putative transcription factor, *Pf*Myb2

*Pf*Myb2 was identified as a putative transcription factor in the *Plasmodium* database (PlasmoDB). It contains two Myb-like DNA-binding domains and is therefore a homologue of *Pf*Myb1. Meyersfeld (2005) cloned the DNA-binding domains of this protein and analysed their function by EMSA studies. The domains were able to bind to the same MREs that *Pf*Myb1 bound to, therefore verifying it as a DNA-binding protein and a possible transcription factor.

The current study was initiated to further characterise the full length *Pf*Myb2 protein. This full length protein could be used further to prove its functionality as a *P. falciparum* transcription factor and may also reveal other roles the protein may play in the parasite's biology. Therefore, this would lead to a better understanding of *P. falciparum* gene regulation and possibly other areas of the parasite's biology.

1.9 Objectives of the study

The aim of this study was to:

Clone the full-length *Pf*Myb2 gene into expression vectors

Express and purify the recombinant *Pf*Myb2 protein

Perform bioinformatic analysis on the *Pf*Myb2 gene and protein.

2. MATERIALS AND METHODS

2.1 *P. falciparum* parasite culture

The *P. falciparum* 3D7 strain was cultured by adapting the culturing methods used in Ljungstrom et al. (2004). All culturing techniques were carried out under a sterile Bioflow laminar hood using aseptic techniques. The ASE Coolspin 34123-606 centrifuge was used for all centrifugations during parasite culture.

2.1.1 Preparation for parasite culture

2.1.1.1 Culture media

Incomplete medium was made up with 10.4g RPMI, 5.9g Hepes buffer, 4.0g glucose and 44mg hypoxanthine in one litre of autoclaved milliQ water. Gentamicin powder was added to a final concentration of 50ug/ml. The medium was stirred for one hour and then sterile filtered using an autoclaved filter unit and peristaltic pump. Incomplete medium was stored at 4°C. Complete medium consisting of 87.5ml incomplete medium, 10ml AB plasma and 2.5ml sterile filtered 5% sodium bicarbonate ($NaHCO_3$) was prepared, stored at 4°C and kept for no longer than 5 days.

2.1.1.2 Heat inactivation of plasma

AB plasma was obtained from the South African National Blood Transfusion Service and heat inactivated by placing the bag at 56°C in a preheated water bath for 2 hours. The plasma was then transferred, using aseptic techniques in a sterile hood, to 50ml Nunc tubes and centrifuged at 2500 revolutions per minute (rpm) for ten minutes at room temperature. The supernatant was transferred to 15ml Nunc tubes and stored at -70°C in 10ml aliquots.

2.1.1.3 Erythrocytes

Blood was collected from healthy volunteers in acid citrate/dextrose (ACD) tubes and centrifuged using the ASE Coolspin 34123-606 centrifuge at 2500 rpm for 10 minutes at 4°C to separate the blood into its three components –plasma, buffy coat and erythrocytes. The plasma and buffy coat were aspirated using a Pasteur pipette and peristaltic pump. The erythrocytes were washed in two volumes of

sterile filtered phosphate buffered saline (PBS, see below) and any remaining white blood cells were removed as before. The wash cycle was repeated three times and the erythrocytes were finally resuspended in one volume of incomplete medium and stored at 4°C for no longer than two weeks.

PBS pH 7.2

137mM NaCl

2.7mM KCl

10mM $Na_2HPO_4.2H_2O$

1.5mM KH_2PO_4

2.1.2 Preparation of culture from frozen stocks

Frozen stock cultures of *P. falciparum* 3D7 in 1.5ml cryo-tubes were thawed at 37°C in a preheated water bath. The thawed culture was then moved to a 15ml Nunc tube and 100ul of sterile filtered 12% NaCl for each millilitre of culture was added in a drop wise fashion while swirling gently and then left at room temperature for five minutes. Nine volumes of sterile filtered 1.6% NaCl were added, mixed in gently and then centrifuged at 1500rpm for five minutes at room temperature. The supernatant was aspirated and the pellet was resuspended in nine volumes of a sterile filtered 0.9% NaCl/0.2% glucose solution and centrifuged as before. The supernatant was aspirated and the pellet resuspended in 750ul freshly washed 1:1 erythrocyte suspension. This was transferred to a small Nunc tissue culture flask ($35cm^2$) and made up to 5ml with complete medium containing 20% AB plasma. The cultures were gassed for 20 seconds with a gas composition of 5% CO_2, 2% O_2 and 93% N_2 and the flask sealed tightly. The culture was incubated for 48 hours at 37°C.

2.1.3 Culture maintenance

2.1.3.1 Viewing parasite cultures

Parasite cultures were viewed daily by preparing a smear of the culture on a glass microscope slide which was stained using the Rapindiff set staining kit based on the Giemsa Stain. The slide was viewed using the Carl Zeiss Axiostar microscope

under 1000x magnification and oil immersion. The percentage of parasitaemia was calculated using the following formula:

$$\%\text{Parasitaemia} = \frac{\#\text{ Infected erythrocytes}}{\#\text{ Infected erythrocytes} + \#\text{ Uninfected erythrocytes}} \times 100$$

Infected erythrocytes and total erythrocytes were counted in 5 fields containing more than 100 erythrocytes per field. An average percentage was calculated using the percentages from five fields. Cultures were maintained at a 5% parasitaemia and if the parasitaemia increased above 10% the culture would be moved to a larger flask or divided into two flasks.

2.1.3.2 Changing medium of cultures

Culture medium was changed once a day by gently tilting the flask and aspirating the medium. Fresh complete medium, warmed to 37°C was then added to a final volume of 5ml, 15ml or 30ml for a small (35cm^2), medium (80cm^2) and large flask (175cm^2) respectively. Culture flasks were gassed for 20 seconds, sealed tightly and incubated at 37°C. As some of the erythrocytes are removed during the change of medium, washed erythrocytes were added when needed to maintain the culture at a 5% haematocrit.

2.1.3.3 Parasite culture synchronisation

Cultures were synchronised by sorbitol treatment (Lambros and Vanderberg, 1979), which lyses erythrocytes that are infected with parasites in all stages except for those in the early ring stages. Cultures in large flasks containing parasites in mainly the early ring stage of development and parasitaemia greater than 10% were transferred to a 50ml Nunc tube, centrifuged at 1500rpm for 5 minutes at room temperature. The supernatant was aspirated, ten volumes of 5% D-Sorbitol added to the erythrocyte pellet, incubated at room temperature for 30 minutes and centrifuged as before. The supernatant was aspirated, the pellet resuspended in complete medium and made up to 30ml. Washed erythrocytes were added to

maintain the 5% haematocrit. Cultures were gassed, sealed tightly and incubated at 37°C.

2.2 DNA extraction and analysis

2.2.1 DNA extraction protocol

DNA extraction techniques were adapted from Ljungstrom *et al*. (2004). This protocol uses proteinase-K and phenol/chloroform to purify the DNA from cellular proteins and ethanol to precipitate the DNA.

Parasite cultures with greater than 10% parasitaemia were transferred to a 30ml Beckman centrifuge tube, centrifuged in the Beckman J2-21 centrifuge using a JA-17 rotor at 1500rpm for 10 minutes at 4°C. The erythrocytes were resuspended in a total volume of 5ml PBS, lysed by adding 10ul of 5% saponin in PBS to each millilitre of suspension and left to stand for 5 minutes. The solution was centrifuged again at 3000rpm for 10 minutes at 4°C and the supernatant removed. The parasite pellet was washed with 10ml PBS and centrifuged as before and the wash cycle was repeated three times. After the third wash the pellet was resuspended in 300ul lysis buffer (see below) and 700ul water, incubated at 37°C for three hours while mixing by hand every 30 minutes. 400ul of a 1:1 ratio of phenol/chloroform was added. The contents were transferred to a 2ml Eppendorf tube, centrifuged using the desktop Eppendorf 5415R centrifuge at 13000rpm for 5 minutes at 4°C. The aqueous phase was collected and 100ul of Tris-EDTA (TE) buffer (see below) was added to the organic component to re-extract any nucleic acids that might still be present. The sample was centrifuged again and the aqueous phase collected and pooled with first extraction. An equal volume of chloroform was added and the sample was centrifuged as before. The aqueous phase was collected and divided into two 2ml Eppendorf tubes.

The DNA was precipitated by adding 1/10 volume of 3M sodium acetate pH 5.2 and 2.5 volumes 100% ice-cold ethanol to the aqueous phases. Precipitation was aided by placing the solution at –70°C for 30 minutes. The solution was

centrifuged at 13000rpm for 20 minutes at 4°C. The pellet was washed in 70% ethanol and centrifuged again. The supernatant was removed and the pellet left to air-dry overnight. The pellet was finally resuspended in 30ul nuclease-free water.

Lysis buffer pH 8.0

40mM Tris-HCl

80mM EDTA

2% SDS

0.1mg/ml Proteinase-K (add just before use)

TE buffer pH 8.0

10mM Tris-HCl

1mM EDTA

3M sodium acetate pH 5.2

24.61g sodium acetate in a final volume of 100ml milliQ water, pH with NaOH

2.2.2 Determining DNA concentration and purity

DNA concentration was determined spectrophotometrically using a Beckman DU-65 spectrophotometer. The absorbance of a 20x diluted sample of the DNA was read at a wavelength of 260nm. The following formula was used to calculate the concentration of the DNA sample:

$$\text{Concentration (ug/ul)} = \frac{\text{OD260 x extinction coefficient (E)* x dilution factor}}{1000}$$

*1 OD260 unit of double stranded DNA = 50ug/ml

Readings taken at a wavelength of 280nm indicated the purity of the DNA by calculating the OD 260/280 ratio. A ratio of 1.8 indicates a pure sample, less than

this indicates protein and/or phenol contamination and greater than 1.8 indicates RNA contamination.

The DNA was also visualised on a 1% agarose gel. 0.5g agarose was added to 50ml Tris-acetate-EDTA (TAE) buffer (see below) and dissolved by heating. 5ul of 10ug/ul ethidium bromide was added once the agarose had cooled slightly. The agarose was poured and left to set in a gel holder. The gel was transferred to a horizontal mini-gel chamber that had been filled with TAE buffer so that the gel was submerged. Between 1-5ul of DNA was mixed with 1ul Fermentas MassRuler loading dye solution and loaded into wells using a Gilson pipette. The Fermentas MassRuler™ High Range (or full range) DNA Ladder was used as a marker to indicate the DNA size according to the number of DNA base pairs. 10ul of 10ug/ul ethidium bromide was added to the buffer at the positive end of the gel chamber. The ethidium bromide intercalates in the double stranded DNA and fluoresces under UV light. Images of the gel were captured using a Syngene gel doc system.

The concentration of DNA was estimated using the Syngene Tools on the Syngene gel doc system. These tools compare the intensity of the bands between the sample of DNA and the known concentrations of the Fermentas MassRuler™ High Range DNA Ladder. The purity of the DNA was determined by the absence of rRNA contamination in the gel and the integrity of DNA was assessed by a single band present in the gel as opposed to a smear due to degraded DNA.

TAE Buffer pH 8.0
40mM Tris-Acetate
1mM EDTA

2.3 PCR amplification of *Pf*Myb2

2.3.1 PCR conditions

PCR primers were designed for cloning the 2700bp full length *Pf*Myb2 gene into three expression vectors- pET-41a, pGEX-4T-2 and pTriEx-3. The forward sequence contained a *Bam*H1 restriction site and the reverse sequence contained a *Xho*1 restriction site to facilitate directional cloning. The reverse primer was also designed to remove the stop codon of the gene to allow for a carboxyl his-tag to be added to the protein when the pET-41a and pTriEx-3 expression vectors were used. The pGEX-4T-2 expression vector attached an amino terminal glutathione-S-tranferase (GST) protein tag to the recombinant protein.

The primers were synthesized by Inqaba Biotech, made up to a 100uM stock in TE buffer, pH 8.0 (see section 2.2.1) and stored at –20°C. The sequences were as follows:

*Pf*Myb2wholeForward:

$^{5'}$GAATGGATCCATGAGGATTCAAATAAAAGGAGG$^{3'}$

*Pf*Myb2wholeReverse: $^{5'}$GGCGGCGGCCTCGAGTGGCACCTTTAACGT$^{3'}$

The restriction enzyme sites are underlined (*Bam*HI and *Xho*I for the forward and reverse primer respectively) and the bases flanking these sites were added to allow for the restriction enzyme to attach to the strand to allow for digestion.

PCR conditions were optimised using the Promega master mix system but inserts for cloning were generated using the Eppendorf TripleMaster PCR system, which uses a high fidelity Taq polymerase to minimize errors during amplification. All reactions were set up on ice. The final reaction volume for both master mixes was 50ul. The reactions contained the components seen in tables 2.3.1 and 2.3.2. The PCR conditions are listed in table 2.3.3.

Table 2.3.1: Components of the PCR reaction using the Promega Master Mix

Components	**Amount for 1 reaction**
DNA	50ng
2X Promega master mix	25ul
Forward Primer (10uM)	1ul
Reverse Primer (10uM)	1ul
Nuclease free water	Make up to final volume of 50ul
Total	50ul

Table 2.3.2: Components of the PCR reaction using the Eppendorf TripleMaster PCR system

Components	**Amount for 1 reaction**
DNA	50ng
10 X High fidelity buffer	5ul
dNTPs (2.5mM each)	4ul
Forward Primer (10uM)	1ul
Reverse Primer (10uM)	1ul
High Fidelity *Taq* polymerase	1ul
Nuclease free water	Make up to final volume of 50ul
Total	50ul

Table 2.3.3: Optimised PCR conditions for *Pf*My2 PCR

	Temperature	**Time**	
Initial Denaturation	94°C	2 min	
Primary Cycles			
Denaturation	94°C	1.5 min	X4
Annealing	56°C	1 min	
Extension	68°C	3 min	
Secondary Cycles			
Denaturation	94°C	1.5 min	X 36
Annealing	50°C	1 min	
Extension	68°C	3 min	

PCR products were viewed on a 1% agarose gel as indicated in section 2.2.2. The size of the amplified gene was estimated against the Fermentas MassRuler™ High Range DNA Ladder.

2.3.2 Phenol/chloroform purification of PCR product

Once PCR amplification of the *Pf*Myb2 gene was verified on an agarose gel the amplified gene was purified using phenol and chloroform. Four PCR reactions were pooled together to make up a volume of 200ul and to this 200ul 1:1 dilution of phenol/chloroform added. This was centrifuged on the desktop Eppendorf 5415R centrifuge at 13000rpm for 5 minutes at 4°C and the aqueous phase collected. 50ul TE buffer was added to the organic phase to extract any DNA that might still be present. This was centrifuged as before and the aqueous phase collected and pooled with the previously collected aqueous phase. To this an equal volume of chloroform was added followed by centrifugation as before and collection of the aqueous phase. The DNA was precipitated as in section 2.2.1 and the DNA pellet was resuspended in 10ul nuclease free water. The concentration of the DNA was determined using the Beckman DU-65 spectrophotometer or electrophoresis in a 1% agarose gel as mentioned in section 2.2.2.

2.4 Cloning *Pf*Myb2

2.4.1 Large-scale preparations of expression vectors

Alkaline lysis was used to isolate plasmid DNA. This method involves denaturing the bacterial chromosomal DNA and plasmid DNA. The plasmid DNA is circular and therefore the two stands remain connected to one another and can renature rapidly. The chromosomal DNA becomes single stranded and does not renature effectively. It is removed with the rest of the cell debris as single stranded DNA is insoluble in high salt concentrations.

Escherichi coli cells containing one of the Novagen expression vectors –pET-41a, pGEX-4T-2 and pTriEx-3 (appendix A-1, A-2 and A-5 for vector maps) were inoculated into 20ml Luria broth (LB) (see below) containing either 0.1mg/ml

kanamycin (for pET-41a) or 0.1mg/ml ampicillin (for pGEX-4T-2 and pTriEx-3) and left shaking at 37°C overnight. Cultures were centrifuged in the Beckman J2-21 centrifuge using a JA-17 rotor at 10000rpm for 5 minutes at room temperature and the supernatant removed. The bacterial cells were resuspended in 1ml ice-cold lysis buffer (see below), vortexed and left to stand for 5 minutes at room temperature. Two ml of NaOH and SDS solution and 1.5ml of potassium acetate solution (see below) were added to the lysed bacteria. The solution was left on ice for 5 minutes then centrifuged at 13000rpm for 10 minutes at 4°C and the supernatants collected in six 2ml Eppendorf tubes. Each tube was filled to capacity with 100% ice-cold ethanol and left to stand at room temperature for 5 minutes. These solutions were centrifuged in the Eppendorf desktop 5415R centrifuge at 13000rpm for 10 minutes at 4°C. The supernatants were removed and the plasmid DNA pellets left to air-dry for 30 minutes. Each pellet was then resuspended in 100ul TE buffer. One ul Fermentas Ribonuclease A (10mg/ml) was added to each tube and this was incubated in a water bath preheated to 37°C for 30 minutes. 300ul 5M potassium acetate and 500ul of a 1:1 dilution of phenol/chloroform were added to each tube and then centrifuged at 13000rpm for 5 minutes at 4°C. The aqueous phase was collected and the DNA was precipitated as in section 2.2.1. The DNA pellets were resuspended in 20ul nuclease free water.

The concentration of the DNA was determined using the Beckman DU-65 spectrophotometer and the DNA was visualised on a 1% agarose gel as described in section 2.2.2.

LB Medium

10g NaCl

10g Bacto™Tryptone

5g Yeast Extract

10ml 1M Tris-HCl pH 7.5

Made up in 1L milliQ water and autoclaved

Lysis Buffer

50mM Glucose

10mM EDTA

25mM Tris-HCl pH 8.0

NaOH and SDS solution

500ul 0.4M NaOH

100ul 10% SDS

400ul milliQ water

Made up just before use from stock solutions

Potassium acetate solution

2ml 5M CH_3COOH

1ml 5M CH_3COOK

Made up just before use from stock solutions

TE Buffer

10mM Tris-HCl pH 8

1mM EDTA

2.4.2 Enzyme digestion of *Pf*Myb2 PCR product and vectors

All expression vectors and the amplified full length *Pf*Myb2 gene were digested with *Bam*HI and *Xho*I to allow for site directed ligation of the gene into the vector. Two controls were set up with vector DNA, each one containing only one of the restriction enzymes to confirm that both enzymes were active. The reactions were set up using a final concentration of 2x Fermentas Tango™ buffer. According to the package insert of the enzymes, 2X Tango™ buffer is the optimal buffer for a double digest. Their activity in 2x Tango™ is 100% and 50-100% for *Bam*HI and *Xho*I respectively. Restriction enzyme digestion was verified by electrophoresis of the vector DNA digestions on a 1% agarose gel, as the digested, linearised vector DNA would separate at a different size to undigested circular

vector DNA. The components of the enzyme digestion can be seen in table 2.4.2.1.

Table 2.4.2.1: The components of the restriction enzyme digestion

Components	Expression vector			*Pf*Myb2 amplicon
	Experiment	***Bam*HI Control**	***Xho*I Control**	
DNA	20ul	5ul	5ul	10ul
Tango buffer™ (10X)	8ul	4ul	4ul	4ul
***Bam*HI (10U/ul)**	1ul	1ul	-	1ul
***Xho*I (10U/ul)**	1ul	-	1ul	1ul
Water	10ul	10ul	10ul	4ul
Total	40ul	20ul	20ul	20ul

After verification of digestion, the *Pf*Myb2 amplicon and vector (that had been cut by both restriction enzymes) were made up to 200ul with nuclease free water and the DNA was purified using the phenol/chloroform method as performed in section 2.3.2. The concentrations of the DNA samples were estimated using the Beckman DU-65 spectrophotometer or by electrophoresis in a 1% agarose gel (both methods described in section 2.2.2).

2.4.3 Ligation and subcloning of *Pf*Myb2 into pET-41a and pGEX-4T-2

Amplified *Pf*Myb2 was cloned into the expression vectors using the Roche Rapid Ligation Kit. The amount of vector to be used (in ng) was determined by the concentration of the purified vector following restriction enzyme digestion. The ratio of vector to insert size was calculated using the following equation:

$$\text{Ratio of vector to insert size } (m) = \frac{\text{Size of vector (Kb)}}{\text{Size of insert (Kb)}}$$

The Roche Rapid Ligation Kit suggests that a 1:3 vector:insert molar ratio should be used for ligation. To calculate how much insert is needed (in ng) for the liagtion the following equation was used:

$$\text{Amount of insert required for a 1:3 ratio (ng)} = \frac{\text{Amount of vector (ng)}}{m} \text{ X3}$$

No more than 200ng of total DNA was used per reaction. Ligation reactions were set up on ice and were then left at room temperature for 20 minutes. A control was set up with only vector and no insert to determine if re-ligation of vector occurred. The components of the ligation reaction are in table 2.4.3.1.

Table 2.4.3.1: Components of the ligation reaction

	Experiment	**Control**
Vector DNA	*x*ul	*x*ul
Insert DNA	*y*ul	-
5X DNA Dilution buffer	4ul	4ul
Water	6 - (*x* + *y*) ul	6-*x*ul
Sub-total	10ul	10ul
2X T4 DNA Ligation Buffer	10ul	10ul
T4 DNA ligase (5U/ul)	1ul	1ul
Total	21ul	21ul

2.4.4 Transformation of DH5α *E.coli* cells with pET-41a and pGEX-4T-2 constructs

Invitrogen DH5α *E. coli* bacterial cells were used for sub-cloning of the expression vector constructs. This strain contains no antibiotic resistance and acquires antibiotic resistance with the uptake of the expression vectors, which encode antibiotic resistance. Therefore transformed cells are selected under antibiotic stress.

Transformation reactions were performed according to the protocol provided with the competent DH5α *E. coli* cells. For the control and experiments half of the ligation reaction (~100ng) was transferred to 50ul competent DH5α *E. coli* cells and placed on ice for 30 minutes. To induce uptake of the DNA the cells were heat-shocked at 37°C for 20 seconds in a pre-heated water bath and placed back on ice for 2 minutes. 450ul LB was added to the bacterial cells and they were incubated on a shaking platform at 37°C for 1 hour therefore allowing the cells to

grow under no antibiotic stress. The culture was plated in 50ul, 100ul and 300ul aliquots onto agar plates (see below) containing either 0.1mg/ml kanamycin or ampicillin, depending on the expression vector that was used for cloning. Plates were incubated at 37°C overnight.

After transformation, if a similar number or more colonies grew on the control plate than on the experiment plate then it was assumed that the vector had re-ligated and that the majority of the colonies on the experiment plate would also only contain empty vectors. If there were fewer colonies on the control plate, the colonies on the experiment plate were more likely to contain vector and insert and were screened for the *Pf*Myb2 insert.

Colonies that were present on the experiment plates were picked off and resuspended in 10ul sterile water. 2ml of LB containing 0.1mg/ml of the appropriate antibiotics (section 2.4.1) in sterile Falcon® 14ml Polypropylene round-bottom tubes was inoculated with 5ul of the resuspended colony. This was incubated on a shaking platform at 37°C overnight. To verify the presence of the *Pf*Myb2 insert the remaining 5ul were analysed by PCR as described in section 2.3.1 with either the *Pf*Myb2-specific PCR primers or vector-specific primers.

Glycerol stocks were made of the overnight cultures by combining 500ul sterile 60% glycerol and 500ul of culture and storing at -70°C. Plasmid preparations were performed on the remaining 1.5ml overnight cultures using the Sigma GenElut™ plasmid miniprep kit. The concentration of plasmid DNA was determined using the Beckman DU-65 spectrophotometer or a 1% agarose gel as described in section 2.2.2. The plasmid DNA was used for a restriction enzyme digestion with *Bam*HI and *Xho*I as described in section 2.4.2 to confirm the presence of the *Pf*Myb2 insert.

Agar plates

200ml LB medium

3g Agar

0.1mg/ml specific antibiotics for each expression vector after autoclaving.

Pour into sterile Petri dishes (8cm diameter)

2.4.5 Sequencing of *Pf*Myb2 insert

Once it was verified that the pGEX-4T-2 and pET-41a vectors contained the *Pf*Myb2 insert, plasmid preparations of the clones were sent to Inqaba Biotech for automated sequencing. Initially vector primers were used to sequence the insert from both the 5' and 3' end.

Sequencing primers for pGEX-4T-2 bind to nucleotides 869-891 and 1041-1019 for the 5' and 3' ends respectively (appendix A-3). The inserted gene would be located between nucleotides 930-966 in the multiple cloning site (MCS). The pET-41a T7 Promoter Primer and T7 Terminator Primer bind to regions up stream and down stream of the MCS respectively. This would also cause amplification of the MCS or the inserted gene fragment if ligation was successful. Three internal forward primers were designed, that would bind to three regions along the gene insert, which was approximately 2700bp and too large to sequence in one reaction. The internal primers were synthesized by Inqaba Biotech and their sequences were as follows:

*Pf*Myb2WholeSeq1F: $^{5'}$ACTTGCAAATACGAAAGG$^{3'}$

*Pf*Myb2WholeSeq2F: $^{5'}$GATCTAAGTATACAACAAGCAG$^{3'}$

*Pf*Myb2WholeSeq3F: $^{5'}$CCAGAATGTGATACGGAC$^{3'}$

These three primers bind along the *Pf*Myb2 gene insert at nucleotides 465-482, 1150-1171 and 1753-1770 respectively (appendix A-3).

2.4.6 Ligation and subcloning of *Pf*Myb2 into pTriEx-3

The pGEX-4T-2-*Pf*Myb2 constructs that had been sequenced were digested as described in section 2.4.2 to cut out the insert. Electrophoresis on a 1% agarose gel of the digest sample separated the *Pf*Myb2 insert from the rest of the vector. The *Pf*Myb2 DNA bands were visualised on a UV trans-illuminator at a wavelength of 302nm and cut out of the gel using a sharp scalpel. The insert DNA was isolated from the gel using the Qiagen MiniElute Gel Extraction Kit. Ligation of the insert into the pTriEx-3 vector and transformation of DH5α *E. coli* cells were performed as described in sections 2.4.3 and 2.4.4.

2.4.7 Transformation of Rosetta 2(DE3) *E. coli* cells with expression vector constructs

Novagen Rosetta 2(DE3) *E. coli* were the choice of bacterial cells used to produce recombinant protein as they contain a chloramphenicol-resistant plasmid (pRARE) encoding rare tRNAs (AUA, AGG, AGA, CUA, CCC and GGA) that are not usually expressed in *E. coli* and help enhance the expression of eukaryotic malaria parasite proteins.

Rosetta 2(DE3) *E. coli* cells were transformed with the fully sequenced vector construct DNA obtained from the plasmid mini preparations. One hundred ng of plasmid DNA was added to 20ul competent Rosetta *E. coli* cells and left on ice for 30 minutes. The cells were then heat shocked at 37°C for 20 seconds and placed back on ice for 2 minutes. The cells were plated directly onto agar containing 0.1mg/ml of specific antibiotics (section 2.4.1) for the vector construct and 0.05mg/ml chloramphenicol to select for Rosetta 2(DE3) *E. coli* cells containing pRARE. The plates were incubated at 37°C, overnight. Transformed cells were analysed in the same manner as described for transformation of DH5α *E. coli* cells in section 2.4.4.

2.5 Expression of Recombinant *Pf*Myb2

2.5.1 Induction of protein expression

2.5.1.1 Induction with Overnight Express™ Instant TB Medium

10ul of the glycerol stocks of the transformed Rosetta 2(DE3) *E. coli* cells were used to inoculate 10ml LB containing 0.05mg/ml chloramphenicol and 0.1mg/ml of the appropriate antibiotics (section 2.4.1) for each expression vector. These cultures were grown overnight in autoclaved 50ml Erlenmeyer flasks on a shaking platform at 37°C. The absorbance of the cultures were read on the Beckman DU-65 spectrophotometer at a wavelength of 600nm and an absorbance of greater than 0.5 ensured that the cells had entered the log phase of growth. A control was also set up that contained the expression vector with no insert. The Novagen Overnight

Express™ Instant TB Medium (see below) was used as the medium of choice for autoinduction of cells to produce the recombinant protein. This system induces transcription of the cloned gene that is under control of a *lac* promoter. The medium components are metabolised and cause a build up of lactose, which in turn causes autoinduction of the *lac* promoter. 500ml of Overnight Express™ Instant TB Medium was inoculated with 10ml of the overnight log phase culture and this was divided into two autoclaved 1L Erlenmeyer flasks. Protein expression was induced at room temperature on a shaking platform for at least 18 hours or until the absorbance of each of the cultures was greater than one at a wavelength of 600nm. The control containing only the expression vector was set up in the same manner.

After induction the cultures were pooled and 1ml of the culture was removed and centrifuged in the Eppendorf desktop 5415R centrifuge at 13000rpm for 1 minute at room temperature. The bacterial pellet was resuspended in 200ul nuclease free water and this sample was used for total sample protein evaluation using SDS-PAGE. The rest of the culture was transferred to two 250ml Beckman centrifuge tubes and centrifuged in the Beckman J2-21 using a JA-14 rotor at 5000rpm for 10 minutes at room temperature. The supernatant was removed and the bacterial pellets were placed at -70°C for 20 minutes to aid bacterial cell lysis. Bacterial cells in each pellet were then lysed with 1ml Novagen BugBuster® containing 1ul Calbiochem Protease Inhibitor Cocktail Set III (details in table 2.5.1.1).

Table 2.5.1.1: Protease inhibitors contained in the Calbiochem Protease Inhibitor Cocktail Set III.

(www.merckbiosciences.co.uk/html/cbc/Protease_Inhibitor_Cocktail_Sets.htm)

Product	Concentration	Target protease
AEBSF, Hydrochloride	100mM	Serine proteases
Aprotinin, Bovine Lung, Lyophilized	80uM	Broad spectrum, serine proteases
Bestatin	5mM	Aminopeptidase B and leucine aminopeptidase
E-64, Protease inhibitor	1.5mM	Cysteine proteases
Leupeptin, Hemisulfate	2mM	Cysteine proteases and trypsin-like proteases
Pepstatin A	1mM	Aspartic proteases

The recommended amount of BugBuster® is 2.5ml per 50ml of culture but only 1ml per 500ml culture was used so that the induced protein was as concentrated as possible. Incomplete lysis of the bacterial cells, due to the smaller volume of BugBuster® used, could occur but to prevent this the suspensions were gently agitated on a shaking platform for 1 hour at room temperature before being centrifuged in the Eppendorf desktop 5415R centrifuge at 13000rpm for 20 minutes at 4°C. The soluble fraction of the samples (the supernatants) were collected and pooled and 150ul put aside for SDS-PAGE analysis. The rest of the soluble phase was stored at 4°C or -20°C for long term storage. As much of the insoluble fraction (pellet) as possible was resuspended in 300ul TE buffer and 150ul collected for SDS-PAGE analysis. The rest of the insoluble fraction was stored at –20°C. The same procedure was carried out for the control sample.

Overnight Express™ Instant TB Medium

Overnight Express™ Instant TB Medium EasyPak (60g)

10ml Glycerol

Make up to 1L

Autoclave and store at 4°C

2.5.1.2 Induction with IPTG

The induction of protein expression was also analysed using IPTG, which induces the utilisation of lactose and activation of the inserted gene under control of the *lac* promoter. The Rosetta *E. coli* stocks were grown overnight as described in section 2.5.1.1. 10ml of culture was used to inoculate 500ml TB medium containing various concentrations of IPTG ranging from 0.4-1mM (from a 100M IPTG stock made up in water and sterile filtered). Induction was performed at either 25°C or 37°C for various lengths of time ranging from 6 hours to 24 hours. Protein was extracted from these cultures as described in section 2.5.1.1.

2.5.2 SDS-PAGE analysis

Protein samples in 150ul aliquots from the total, soluble and insoluble fractions were mixed with 49ul of gel loading dye made up of 40ul 5X suspension solution

(see below), 4ul β-mercaptoethanol and 5ul sucrose and dye solution (see below), which resulted in a final concentration of 2.5% β-mercaptoethanol, 5% sucrose and 0.016% bromophenol blue. The samples were boiled for 3 minutes to denature the proteins. These samples were stored at 4°C for short-term storage or -20°C for long-term storage.

Protein samples were analysed on a 12% Laemmli gel (Laemmli, 1970) containing a 4% stacking gel. The gel was cast in a Hoefer Mighty Small™ SE 245 dual gel caster and the gel was made up from the components in table 2.5.2.1.

Table 2.5.2.1: Components of the Laemmli separating and stacking gels

	12% Separating gel	4% Stacking gel
30% Acrylamide	4ml	521ul
1% Bis-acrylamide	1ml	400ul
4X Lower Gel Buffer	2.5ml	-
4X Upper Gel Buffer	-	1ml
10%SDS	200ul	80ul
Water	2.195ml	1.946ml
10% Ammonium persulphate (made just before use)	100ul	40ul
TEMED	5ul	3ul
Final volume	10ml	4ml

The separating gel was poured between a 10cm X 8cm glass plate and notched aluminium plate that were separated with 1.5mm spacers and had been placed in the casting chamber. This was layered with isopropanol so that the gel surface was not exposed to oxygen, which ensured efficient cross-linking between the acrylamide and bis-acrylamide. The isopropanol was poured off after 30 minutes and the stacking gel was poured on top of the separating gel and a 12 well comb inserted at the top and left to set for 30 minutes. The gel was transferred to a vertical running chamber and the top and bottom chamber filled with running buffer (see below). Protein samples were loaded into the wells of the gel using a 50ul Hamilton syringe. The Fermentas PageRuler™ prestained protein ladder or an erythrocyte membrane sample (extracted by Kubendran Naidoo) were also loaded into wells and were used to determine the molecular weights of the

proteins. The running chamber was connected to a Hoefer MightySlim SX 250 power supply and the current set at 25mA. The gel was removed from the chamber after 1.5 hours, when the dye front would have reached the bottom of the gel, and placed in Coomassie Blue stain with gentle agitation on a shaker overnight. The gel was destained in a solution containing 10% acetic acid and 10% methanol.

5X suspension solution pH 8.0

50mM Tris-HCL

5mM EDTA

5%SDS

25% sucrose

Sucrose and dye solution

2.5% sucrose

0.5% Bromophenol blue

4X Lower gel buffer (LGB) pH 8.8

1.5M Tris-HCl

4X Upper gel buffer (UGB) pH 6.8

500mM Tris-HCl

Running buffer pH 8.5

25mM Tris

192mM Glycine

1% SDS

Coomassie Blue

1g Coomassie Blue R-250

500ml Isopropanol

200ml Acetic acid

Made up to 2L with milliQ water

2.5.3 Western blot analysis

Proteins were electrophoresed on a Laemmli gel as described in section 2.5.2. The gel was not placed in Coomassie Blue stain as the proteins in the gel were transferred to a nitrocellulose membrane. The gel and water-soaked Amersham HyBond™-C Extra nitrocellulose membrane were placed between two sheets of Biorad extra thick filter paper and these were placed between two sponges, all of which had been soaked in transblot buffer (see below). This was secured in a cassette and transferred to a Hoefer TE series Transphor electrophoresis unit filled with ice-cold transblot buffer so that the cassette was immersed in the buffer. In the chamber the gel faced the cathode and the nitrocellulose membrane faced the anode. The transfer chamber was connected to BioRad model 200/2.0 power supply and set to 35V (0.08Amps) in the 4°C cold room and left overnight. The negatively charged proteins moved towards the positive anode and bound to the positive nitrocellulose membrane.

The membrane was removed from the cassette and placed in TBS for 5 minutes. 1ml Ponceau S (see below) was added to the membrane to stain proteins and verify that they had been transferred from the gel to the membrane. The protein bands of the erythrocyte membrane were also marked out to indicate the molecular weight of the protein bands. Ponceau S stain was removed from the membrane as soon as the protein bands were visible by rinsing in milliQ water several times.

Membranes containing recombinant proteins linked to GST were blocked in 30ml 3% BSA in TBS (see below) for 1 hour and then washed three times in TBS for 5 minutes per wash. The membrane was incubated in 30ml 1% BSA in TBS containing a 1:120000 Amersham goat anti GST antibody linked to horseradish peroxidase (HRP). This was agitated gently for 1 hour at room temperature. The membrane was washed five times in TBS containing 0.05% Tween®-20 detergent for 5 minutes per wash and then in TBS only for 5 minutes.

Chemiluminescence detection was carried out in a dark room. The Pierce SuperSignal® West Pico Chemiluminescent Substrate Kit was used. The HRP attached to the antibodies, which were bound to the tag of the fusion *Pf*Myb2 protein, would cause fluorescence of the substrate. An X-ray film was exposed to the fluorescence revealing the presence of the fusion protein, *Pf*Myb2. One ml of a 1:1 dilution of SuperSignal® West Pico Enhancer Solution and SuperSignal® West Pico Peroxide Solution were added to the membrane and left for 5 minutes. The solution was poured off and the blot was exposed to AGFA CP-G Plus medical X-ray film in a cassette for 2-5 minutes. The X-ray film was developed in 1X Axim Medical X-ray developer until an image of the fluorescing protein bands was visualised. The film was fixed for 1 minute in 1X Perfix high speed X-ray fixer to stop further development of the signal.

Transblot buffer (pH 8.3)
25mM Tris
192mM Glycine
20% (v/v) Methanol

Tris Buffered Saline (TBS) (pH 7.5)
50mM Tris-HCl
154mM NaCl

Ponceau S
0.1g Ponceau S
700ul Glacial acetic acid
9.3ml milliQ water

2.6. Purification of recombinant *Pf*Myb2

2.6.1 Purification of inclusion bodies and refolding of insoluble protein

The bacterial inclusion body purification and protein refolding techniques were adapted from Vallejo and Rinas (2004). This involves purifying the inclusion

bodies that contain the insoluble recombinant protein, denaturing the protein using guanidine hydrochloride followed by refolding of the protein by reducing the concentration of the guanidine hydrochloride. The refolding buffer contains a redox couple (reduced and oxidised glutathione) to facilitate the formation of native disulphide bonds and L-arginine to prevent aggregation.

The insoluble pellets, containing bacterial inclusion bodies, obtained in section 2.5.1 were resuspended in 2.5ml Novagen BugBuster™. Merck L-Lysozyme was added to a final concentration of 1KU per millilitre and the solution left to stand for 5 minutes. Six volumes (15ml) of 1:10 dilution of BugBuster™ was added, the sample vortexed, then centrifuged in the Beckman J2-21 centrifuge using a JA-17 rotor at 5000rpm for 15 minutes at 4°C. The supernatant was removed and the pellet resuspended in three volumes (7.5ml) of 1:10 dilution of BugBuster™ and centrifuged as before. This step was repeated four times, but on the final step the sample was centrifuged at 16000rpm and the pellet resuspended in 1ml denaturing buffer (see below). The sample was left to denature on a shaking platform, overnight at 4°C. The sample was centrifuged at 17000rpm for 20 minutes at 4°C in the Beckman J2-21 centrifuge using a JA-17 rotor and the supernatant, containing the denatured protein, was collected.

The protein was refolded by diluting the guanidine hydrochloride with refolding buffer. 100ul of the denatured protein was used for protein refolding and the rest of the stock was stored at -20°C. The 100ul aliquot was diluted to bring the concentration of guanidine hydrochloride from 6M to 3M by adding 10ul refolding buffer (see below) to the sample every 5 minutes for 50 minutes, with a final volume of 200ul. The sample was divided into two aliquots of 100ul and placed into two Slide-A-Lyzer® mini dialysis units. The dialysis tubes were fitted into 2ml Eppendorf tubes containing 1.5ml refolding buffer, on ice and left for 5 minutes. This process was repeated with fresh wash buffer, seven times, to remove as much of the guanidine hydrochloride without precipitating the protein. The dialysed samples were pooled and centrifuged in the Eppendorf desktop

5415R centrifuge at 13000rpm for 10 minutes and the supernatant collected. This fraction was used for affinity purification of the recombinant *Pf*Myb2.

Denaturing buffer

6M guanidine hydrochloride

50mM Tris-HCl pH 8

10mM DTT

100mM NaCl

10mM EDTA

Refolding buffer

200mM Tris-HCl pH 8.0

10mM EDTA

1M L-arginine

0.1mM PMFS (400mM stock made up in DMSO, added gradually to the aqueous solution)

2mM reduced glutathione

0.2mM oxidised glutathione

2.6.2 Affinity purification of the GST-linked recombinant *Pf*Myb2

2.6.2.1 Affinity purification of *Pf*Myb2-GST from the soluble protein fraction

The soluble protein fractions (~2ml) (section 2.5.1) were added to 10ul Promega MagneGST™ Glutathione particles that had been washed five times, 1 minute per wash, in 200ul GST wash buffer (see below) and resuspended in 10ul GST wash buffer. The suspension was agitated overnight on a shaking platform at 4°C. The magnetic beads were separated from the rest of the solution using a particle separator and the unbound fraction was removed. The beads were washed five times with GST wash buffer. During the first 2 washes the wash buffer was incubated with the beads for 5 minutes and for the rest of the washes the beads were quickly rinsed in GST wash buffer and the buffer removed. The GST-bound recombinant *Pf*Myb2 was eluted with elution buffers containing 100mM glutathione or 200mM glutathione (see below) to deduce which concentration

resulted in the highest yield of protein. The sample was incubated with 60ul of either elution buffer for 1 hour while agitating on a shaking platform at 4°C, keeping the beads suspended for the full hour. This was repeated for a second elution but for only 15 minutes and the beads were finally stripped with 200ul GST-wash buffer containing 1% SDS. The eluted protein samples were analysed by SDS-PAGE and western blots (sections 2.5.2 and 2.5.3).

2.6.2.2 Affinity purification of refolded *Pf*Myb2

The refolded protein obtained in section 2.6.1 was affinity purified as in section 2.6.2.1. The final volume of refolded protein was added to 10ul glutathione particles and eluted as in section 2.6.2.1 but the second elution buffers contained 1% Tween®-20 detergent. The eluted protein samples were analysed as in sections 2.5.2, 2.5.3. The final yield of protein was estimated, using the Laemmli gels, by comparing the intensity of the refolded protein band to the known concentrations of the protein bands of the Fermentas protein ladder.

GST-wash buffer

4.2mM Na_2HPO_4

2mM K_2HPO_4

140mM NaCl

10mM KCl

Glutathione elution buffer

100mM Glutathione (pH with NaOH to between 7.0-8.0)

or

200mM Glutathione (pH with NaOH to between 7.0-8.0)

50mM Tris-HCl (pH 8.1)

Glutathione elution buffer (for the second elution of refolded proteins)

100mM Glutathione (pH with NaOH to between 7.0-8.0)

or

200mM Glutathione (pH with NaOH to between 7.0-8.0)

50mM Tris-HCl (pH 8.1)

1% Tween®-20 detergent

2.7 Bioinformatic analysis of *Pf*Myb2

The *Plasmodium* genome database (PlasmoDB version 5.4) was used to summarise predicted characteristics of *Pf*Myb2 such as gene and protein size, protein characteristics and expression profiles of the protein in the parasite. Hydropathy plots were calculated using the following website: http://www.vivo.colostate.edu/molkit/hydropathy/index.html, which used the Kyte-Doolittle scale to determine hydrophobic regions of the protein. Secondary structure predictions were performed on SCRATCH, using SS pro at the following website: www.igb.uci.edu/tools/scratch/. Basic Local Alignment Search Tool for proteins (BLASTP) searches were done to find homologous proteins to *Pf*Myb2 while the European Molecular Biology Open Software Suite (EMBOSS) tools were used to calculate identity and similarity percentages between these proteins and perform multiple alignments. These tools included the pairwise alignment algorithms and ClustalW2 available at the following respective websites: www.ebi.ac.uk/emboss/align/index.html and www.ebi.ac.uk/Tools/clustalw2/. The recombinant proteins molecular weight and extinction co-efficients were determined using Expert Protein Analysis System (ExPASy) ProtParam tool (http://au.expasy.org/tools/protparam.html).

3. RESULTS

3.1 DNA Extraction

DNA was extracted from *P. falciparum* cultures that contained parasites in mainly the late trophozoite stage with over 10% parasitaemia (see figure 3.1 for a mixed parasite culture blood smear). The total yield of DNA per 30ml parasite culture ranged from 2.7ug to 18.9ug and the concentrations of the DNA were 0.09ug/ul to 0.63ug/ul. The purity of the samples had OD260/280 ratios between 1.2 and 1.9, with the low ratio implying protein or phenol contamination. An example of a DNA extraction can be seen in figure 3.2. Samples with a high yield and a purity ratio close to 1.8 were used to amplify the *Pf*Myb2 gene during PCR.

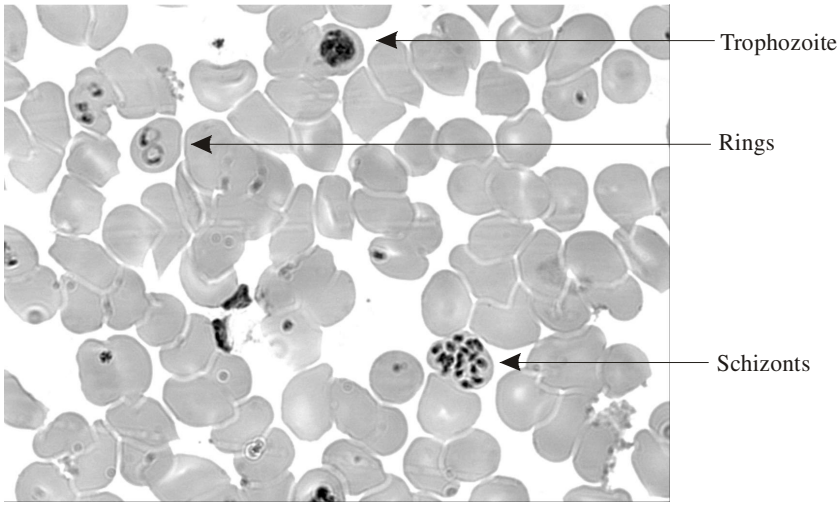


Figure 3.1: Mixed parasite culture
Blood smear, stained with Rapindiff staining kit, of a mixed parasite culture containing trophozoite, ring and schizont stages of the intraerythrocytic lifecycle. The image was captured under oil immersion at 1000x magnification.

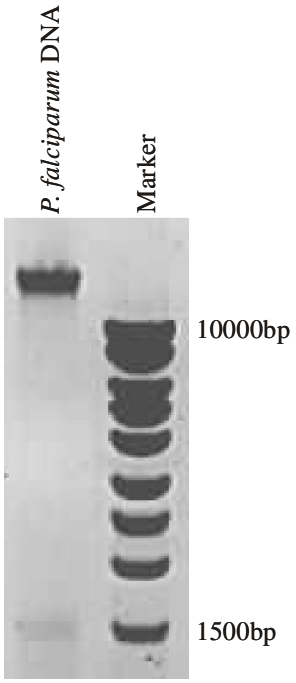


Figure 3.2: *P. falciparum* DNA extraction
A 1% agarose gel showing 1ul of a total of 30ul extracted *P. falciparum* DNA (first lane) next to 5ul Fermentas High Range MassRuler™. The chromosomal DNA has a high molecular weight and does not migrate as far as the marker in the gel.
The single band in the gel indicates that the DNA is not degraded.

3.2. PCR amplification of *Pf*Myb2

3.2.1 Optimisation of annealing temperatures for PCR segments

The PCR used to amplify the *Pf*Myb2 gene consisted of two segments, since the 5' restriction enzyme sites were not complementary to the *Pf*Myb2 gene but were incorporated into the *Pf*Myb2 PCR primers to allow directional cloning of the insert into the expression vectors. The first segment had four cycles and a low annealing temperature to allow the primers to pair up with complimentary bases on the *Pf*Myb2 gene. The second segment had 34 cycles and a higher annealing temperature since the first four cycles would have generated a sequence that would contain the restriction enzyme sites and therefore the primers were fully

complimentary and would have a higher melting temperature. The optimisation of the annealing temperatures was performed using the Promega master mix.

The two primers, *Pf*Myb2wholeForward and *Pf*Myb2wholeReverse, were designed using Gene Runner, version 3.05. These had calculated melting temperatures that were quite far apart at 65.9°C and 75.6°C respectively. This variation was due to the 5' flanking bases that were added to incorporate the restriction enzyme site as well as additional bases to allow the enzyme to bind to the fragment during digestion. *Pf*Myb2wholeReverse was unusual as it had only 15bp complimentary to the *Pf*Myb2 gene sequence. This design was decided upon as a longer primer was predicted (by Gene Runner) to form two possible hairpins which may have affected its efficacy during PCR. The additional 5' flanking region required more base pairs for the *Xho*I restriction enzyme site than for *Bam*HI site. Temperatures for both the primary and secondary annealing step were analysed from 50°C to 64°C. The best yield of pure *Pf*Myb2 amplicon was obtained when the primary and secondary annealing temperatures were 56°C and 50°C respectively, even though the secondary temperature was lower than the primary temperature (figure 3.2.1). These PCR conditions were decided on after setting up a gradient of temperatures (using the calculated melting temperatures as a guideline) on the Eppendorf gradient cycler machine that uses a range of primary and secondary annealing temperatures.

3.2.2 Optimisation of amount of DNA used for PCR

During optimisation of the annealing temperatures it was noted that the amount of DNA used in each PCR also had an influence on product yield and PCR success. Therefore while optimising the conditions for PCR, the reactions were set up with different amounts of DNA. 50ul Promega PCR reactions were set up with 50ng, 200ng and 500ng of DNA (figure 3.2.2). The reaction tube with 50ng DNA resulted in *Pf*Myb2 gene amplification while the other reactions only showed amplification of primer dimers. Therefore 50ng of DNA was used per reaction.

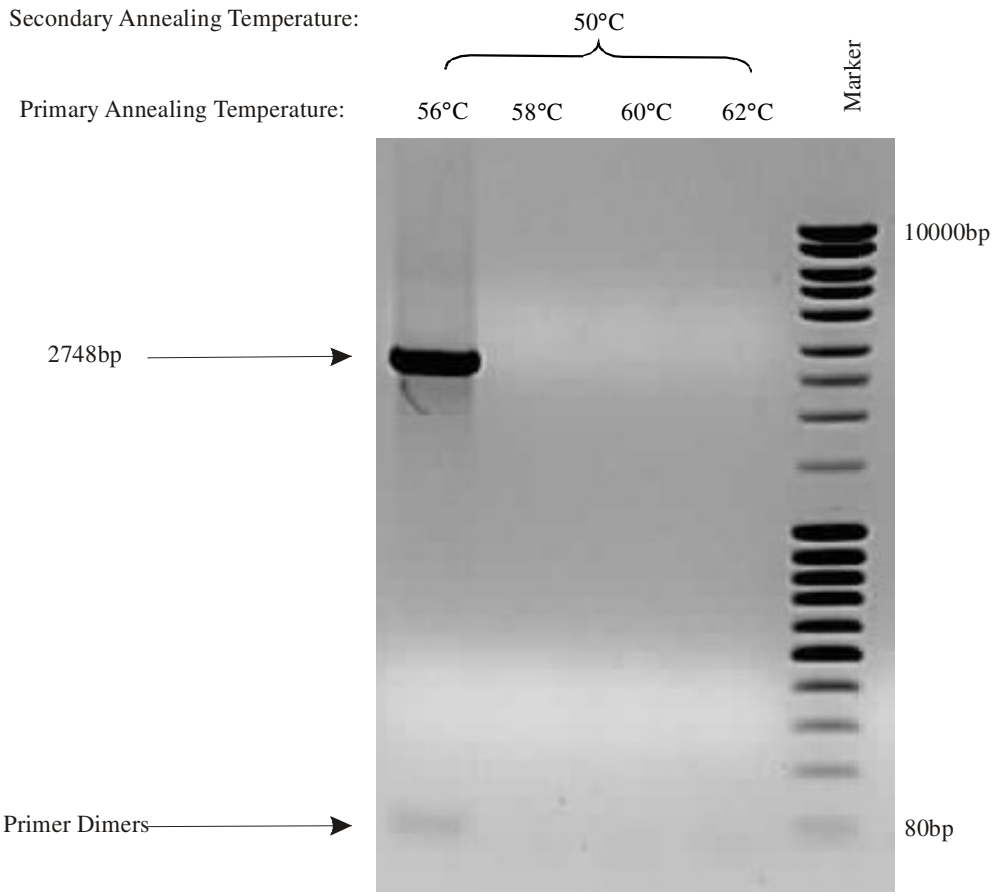


Figure 3.2.1: Optimisation of annealing temperatures for *Pf*Myb2 PCR
The 1% agarose gel shows an example of the PCR results to amplify the *Pf*Myb2 gene using different primary annealing temperatures. The fragment length of the *Pf*Myb2 gene is 2748bp.
Amplification of the *Pf*Myb2 gene was optimal when the primary and secondary annealing temperatures were 56°C and 50°C respectively, as seen in the first lane of the gel.
Amplification did not occur when higher primary annealing temperatures were used and this may be due to failed PCR as it is also noted that no primer dimers are seen in these lanes.
10ul of each reaction was loaded per well and 5ul of the Fermentas full range MassRuler™ was used as the marker.

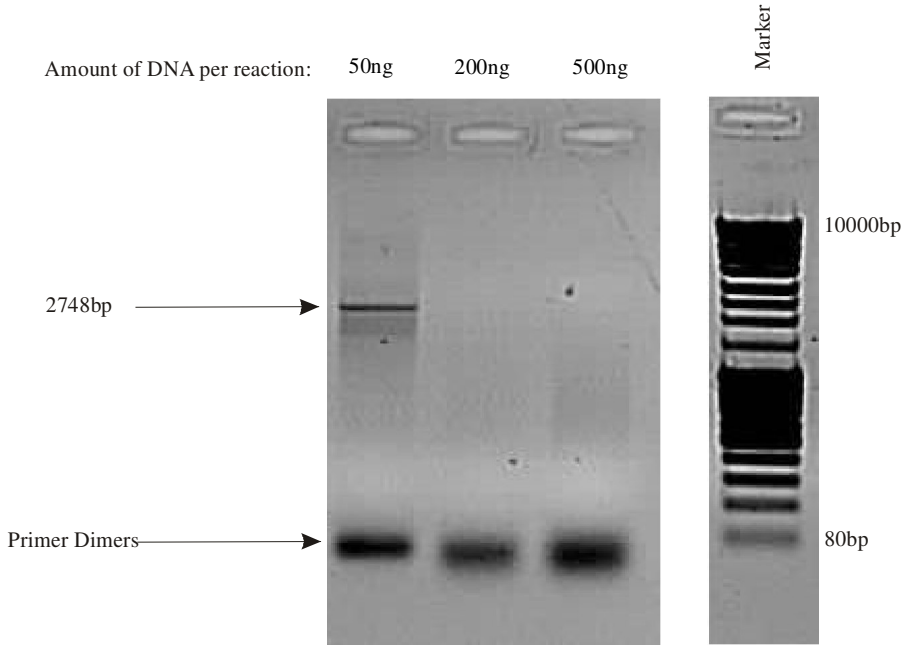


Figure 3.2.2: Optimisation of DNA concentration per PCR reaction
The 1% agarose gel shows that 50ng of *P. falciparum* DNA was optimal for PCR amplification of the *Pf*Myb2 gene as seen in the first lane. The other lanes only showed primer dimer amplification. Irrelevant lanes of the gel have been excluded.
10ul of each reaction was loaded per well and 5ul of the Fermentas full range MassRuler™ was used as the marker.

Once the PCR conditions had been optimised, the *Pf*Myb2 fragment was generated (using the high fidelity *Taq* polymerase system to minimise errors) so that it could be cloned into plasmid DNA and expressed in bacteria.

3.3 Cloning of *Pf*Myb2 into pET-41a

The pET-41a system adds a carboxyl terminus his tag to the inserted gene resulting in a his-tagged recombinant protein (appendix A-1 for vector map). Numerous attempts were made to clone *Pf*Myb2 into the pET-41a expression system without success. Restriction enzyme digest controls confirmed that both *Bam*HI and *Xho*I were active and therefore the vector and gene insert were cut as seen in figure 3.3.1. Although the double digest of the vector only shows one band at 5933bp, corresponding to linearised pET-41a, there must still have been a small amount of undigested/incompletely digested vector present that was not sufficient

to be visualised in an agarose gel as the vector appeared to be religating after transformation of *E. coli* cells. On one occasion three colonies grew on one of the experiment plates but these contained no gene insert as seen in figure 3.3.2. No further attempts were made to clone into pET-41a.

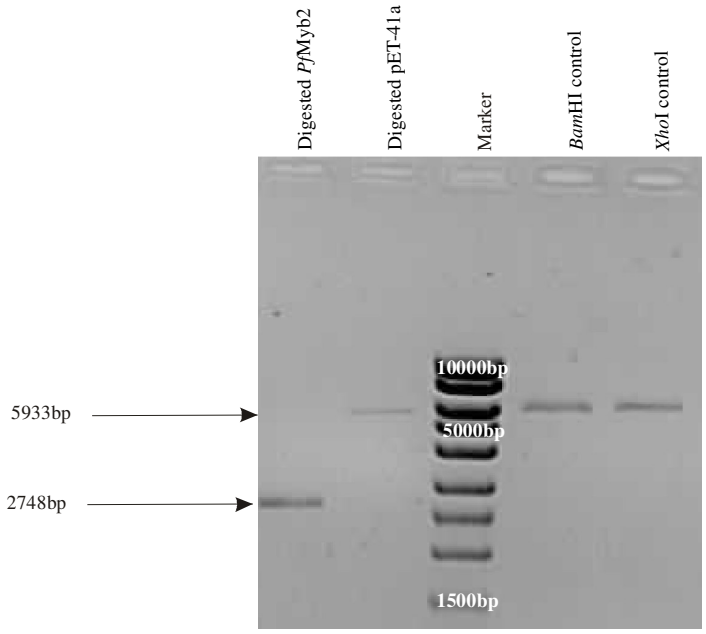


Figure 3.3.1: The pET-41a vector and *Pf*Myb2 gene insert digested with *Bam*HI and *Xho*I for cloning

The 1% agarose gel shows the *Pf*Myb2 gene insert and pET-41a vector in the first and second lanes respectively, both of which were digested with *Bam*HI and *Xho*I to facilitate directional cloning into pET-41a.

Digest controls, set up using only one restriction enzyme and vector DNA, can be seen in the last two lanes and indicate that the vector had been linearised to produce a single band of the expected size. Undigested vector is not shown.

1ul of the *Pf*Myb2 insert and pET-41a vector, both digested with restriction enzymes, were loaded and 5ul of each restriction enzyme control digest was loaded per well. 5ul of the Fermentas high range MassRuler™ was used as the marker.

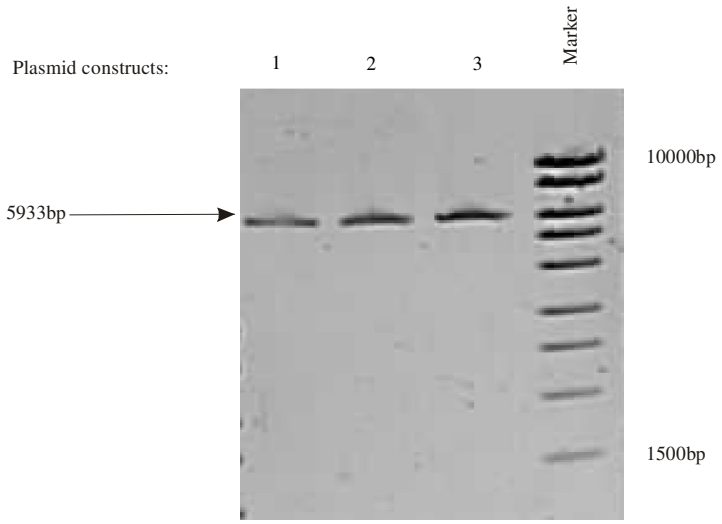


Figure 3.3.2: Digested pET-41a plasmid constructs
The 1% agarose gel shows the three pET-41a plasmid constructs digested with *Bam*HI and *Xho*I. None of the constructs contained the *Pf*Myb2 gene insert as no DNA was detected at 2748bp and the linearised construct migrated near to 6000bp, which is very close to 5933bp – the size of the pET-41a vector. Therefore the constructs represented re-ligated vector.
10ul of each digest was loaded per well and 5ul of the Fermentas high range MassRuler™ was used as the marker.

3.4. Cloning and expression of pGEX-4T-2-*Pf*Myb2 constructs

The pGEX-4T-2 expression system encodes a GST tag that is attached to the amino terminal of the fusion protein, *Pf*Myb2-GST (appendix A-2, A-3 and A-4 for vector map, gene sequence and fusion protein).

3.4.1 Cloning of *Pf*Myb2 into pGEX–4T-2

*Pf*Myb2 was successfully ligated into pGEX-4T-2 and DH5α *E. coli* bacterial cells were transformed with the plasmid constructs. The vector and insert, used in the ligation reaction, can be seen in figure 3.4.1.1. Experiment agar plates that had been spread with 300ul and 100ul of the transformation reaction contained 15 and 10 colonies respectively. The control agar plate that had been spread with 300ul had 5 colonies, indicating that some of the plasmid had religated. 11 colonies were picked off the experiment plates and analysed by an initial PCR screen, using vector primers. Four of these colonies were positive for the *Pf*Myb2 gene insert (figure 3.4.1.2) and these constructs were named pGEX2:2, pGEX2:3,

pGEX2:4 and pGEX2:6 after the order they had been taken from the plates. Restriction enzyme digestions on the plasmid preparations from the positive colonies verified the presence of the gene insert (figure 3.4.1.3)

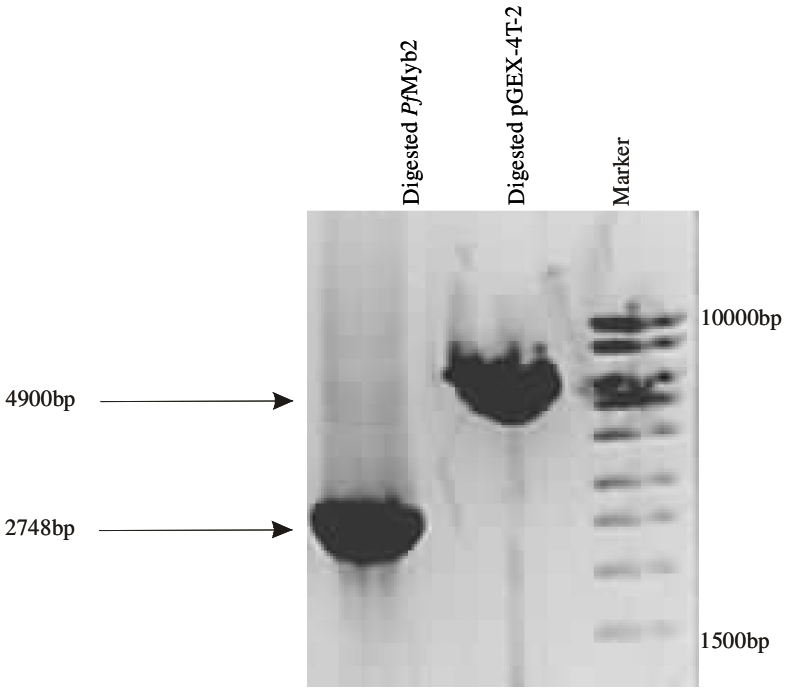


Figure 3.4.1.1: The pGEX-4T-2 vector and *Pf*Myb2 gene insert digested with *Bam*HI and *Xho*I for cloning
The 1% agarose gel shows the *Pf*Myb2 gene insert and pGEX-4T-2 vector digested with the restriction enzymes *Bam*HI and *Xho*I to facilitate directional cloning into the vector.
5ul of the *Pf*Myb2 insert and pET-41a vector, both digested with restriction enzymes, were loaded per well. 5ul of the Fermentas high range MassRuler™ was used as the marker.

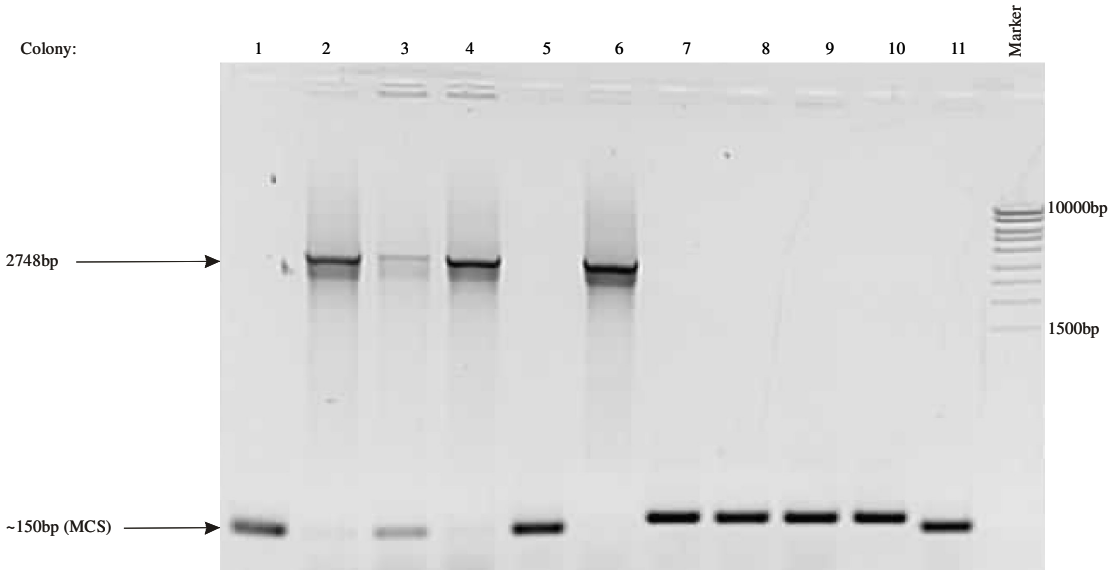


Figure 3.4.1.2: PCR amplification of pGEX-4T-2 transformed DH5α *E. coli* colonies
The 1% agarose gel shows the PCR amplification of the 11 colonies picked from the experiment plates after cloning into pGEX-4T-2.
The only colonies that were positive for the *Pf*Myb2 gene insert were 2,3,4 and 6. The smaller bands seen at the bottom of the gel are the amplified multiple cloning site (MCS) (~150bp) of pGEX-4T-2. The MCS was amplified as the primers used to screen the pGEX-4T-2-*Pf*Myb2 constructs were specific for the vector.
The third colony is positive for both the gene insert and a MCS and it was thought that two colonies may have been picked off the agar plate.
The size difference in the amplified MCS of some of the colonies cannot be explained.
10ul of each reaction was loaded per well and 5ul of the Fermentas high range MassRuler™ was used as the marker.

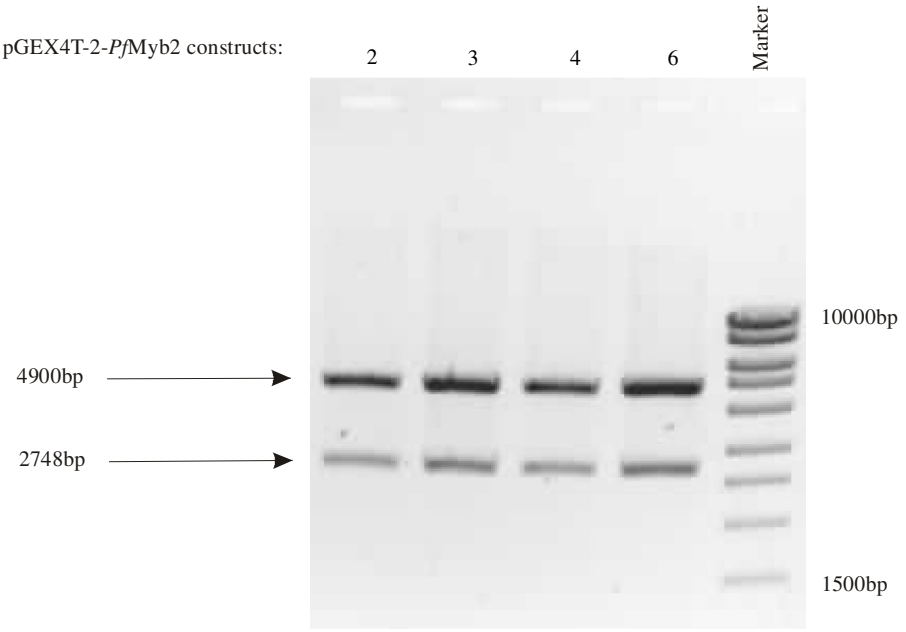


Figure 3.4.1.3: The pGEX-4T-2-*Pf*Myb2 constructs, digested with *Bam*HI and *Xho*I, from transformed DH5α *E. coli* cells
The 1% agarose gel shows the pGEX-4T-2-*Pf*Myb2 constructs, digested with *Bam*HI and *Xho*I, that were positive for the *Pf*Myb2 gene during the PCR screening.
The vector can be seen at 4900bp and the *Pf*Myb2 insert at 2748bp, which confirms the PCR screening results.
20ul of each digest was loaded per well and 5ul of the Fermentas high range MassRuler™ was used as the marker.

3.4.2 Induction of *Pf*Myb2-GST fusion protein in Rosetta *E. coli*

Rosetta *E. coli* cells were transformed successfully with each of these pGEX-4T-2-*Pf*Myb2 constructs. One colony from each plate containing either the pGEX2:2, pGEX2:3, pGEX2:4 or pGEX2:6 constructs was picked off and analysed for the gene insert by PCR, using the *Pf*Myb2 primers. The presence of the gene insert was verified with plasmid preparations and restriction enzyme digestions of these transformants (results not shown).

The pGEX2:2, pGEX2:3 and pGEX2:6 constructs were induced to produce recombinant *Pf*Myb2, fused to GST at the amino terminal end of the protein. The total and soluble protein fractions were analysed on Laemmli gels and Western blots. The expected molecular weight of the fusion protein was 134kDa (appendix

A-4) but it was not clear on the Laemmli gel whether the recombinant protein was being expressed, as there were *E. coli* proteins that also migrated between 170kDa and 130kDa. A Western blot detected the recombinant *Pf*Myb2-GST at 134kDa in the protein fractions from *E. coli* cells containing the pGEX2:2 and pGEX2:3 constructs, as seen in figure 3.4.2.1. The pGEX2:6 construct was positive for GST but at a lower molecular weight than the fusion protein. The band had a very similar molecular weight to the GST control. Therefore this construct presumably contained a mutation in the gene insert, resulting in a truncated protein or only the expression of the GST tag. The protein was soluble which further indicates that it is likely to be GST. No further studies were performed on pGEX2:6. Sequencing of pGEX2:4 was problematic and therefore this construct was not analysed for protein expression.

Initial sequencing of the insert was performed using vector specific primers, however, due to the large size of the gene insert internal forward primers were designed to sequence the full length of the *Pf*Myb2 gene insert (appendix A-3 and A-8). The DNA sequencing results for pGEX2:2 confirmed that the sequence of the amplified *Pf*Myb2 gene was correct. Since it had been determined that this construct was expressing *Pf*Myb2-GST in Rosetta *E. coli* cells, further experiments were performed on this construct only. The protein fractions of induced Rosetta *E. coli* cells containing pGEX2:2 were compared to induced Rosetta *E. coli* cells containing pGEX-4T-2 vector only. Induction of *Pf*Myb2-GST was tested with two induction media, LB containing IPTG and Overnight Express™ Instant TB Medium. Various conditions were tested including different concentrations of IPTG, induction temperatures and the length of time of induction. These variables were altered individually and together but resulted in very little protein expression each time (results not shown). The best induction results were achieved using Overnight Express™ Instant TB Medium, however most of the recombinant protein was in the insoluble fraction (figure 3.4.2.2). With each induction a faint protein band was visible at 134kDa in the total and insoluble fractions but was not as intense in the soluble fraction. This protein band was not seen in the soluble fraction of induced Rosetta *E. coli* containing pGEX-

4T-2 only. However, other *E. coli* proteins could migrate at a similar molecular weight to the fusion protein so Western blots (figure 3.4.2.3) were performed to confirm the presence of *Pf*Myb2-GST. The total and insoluble fractions were positive for GST at 134kDa for all inductions, while the soluble fraction was not always positive reflecting a variable yield of fusion protein. Even when a smaller volume of BugBuster® was used to lyse the cells and concentrate any soluble fusion protein, the recombinant protein could still not be detected consistently in the soluble fraction.

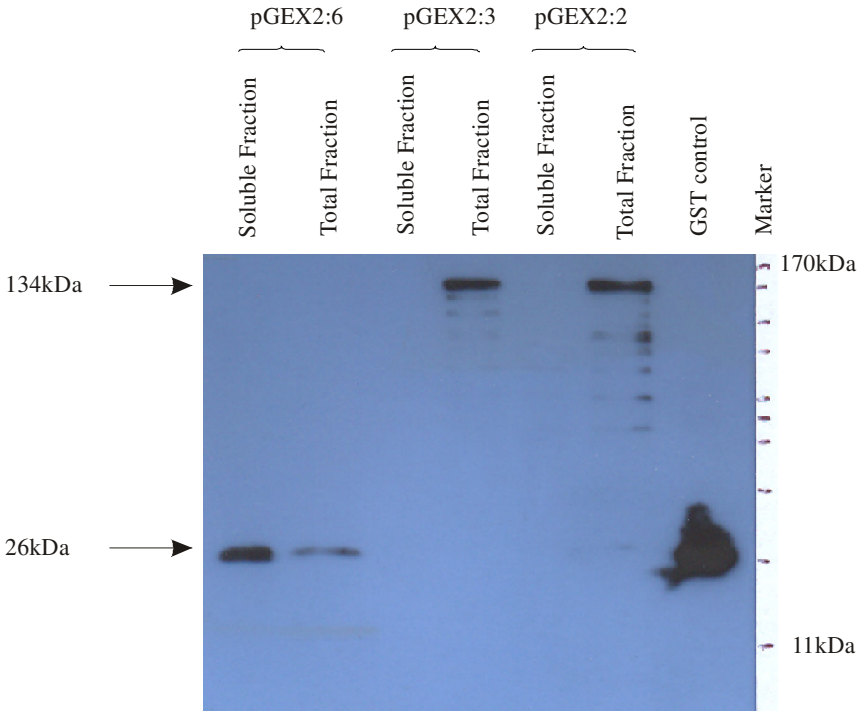


Figure 3.4.2.1: Western blot of *Pf*Myb2-GST protein expressed from three of the pGEX-4T-2-*Pf*Myb2 constructs
A Western blot to detect *Pf*Myb2-GST at 134kDa using anti GST-HRP conjugate. The fusion protein was detected in the total fractions of *E. coli* containing constructs pGEX2:2 and pGEX2:3 but not in the soluble fractions.
Faint signals below the fusion protein were also detected in the total fractions. This may be due to truncated protein being expressed, as the GST is attached to the amino terminal end of the protein. It may also be non specific binding of the anti GST-HRP conjugate to *E. coli* proteins.
E. coli containing the pGEX2:6 construct expressed a protein tagged with GST at a much lower molecular weight than 134kDa. This protein migrated at the same level as the GST control and therefore the insert presumably contained a mutation resulting in expression of the GST tag only.
Protein bands of the Fermentas prestained protein ladder have been marked off and placed next to the blot, as the GST-HRP conjugate did not detect the protein markers.

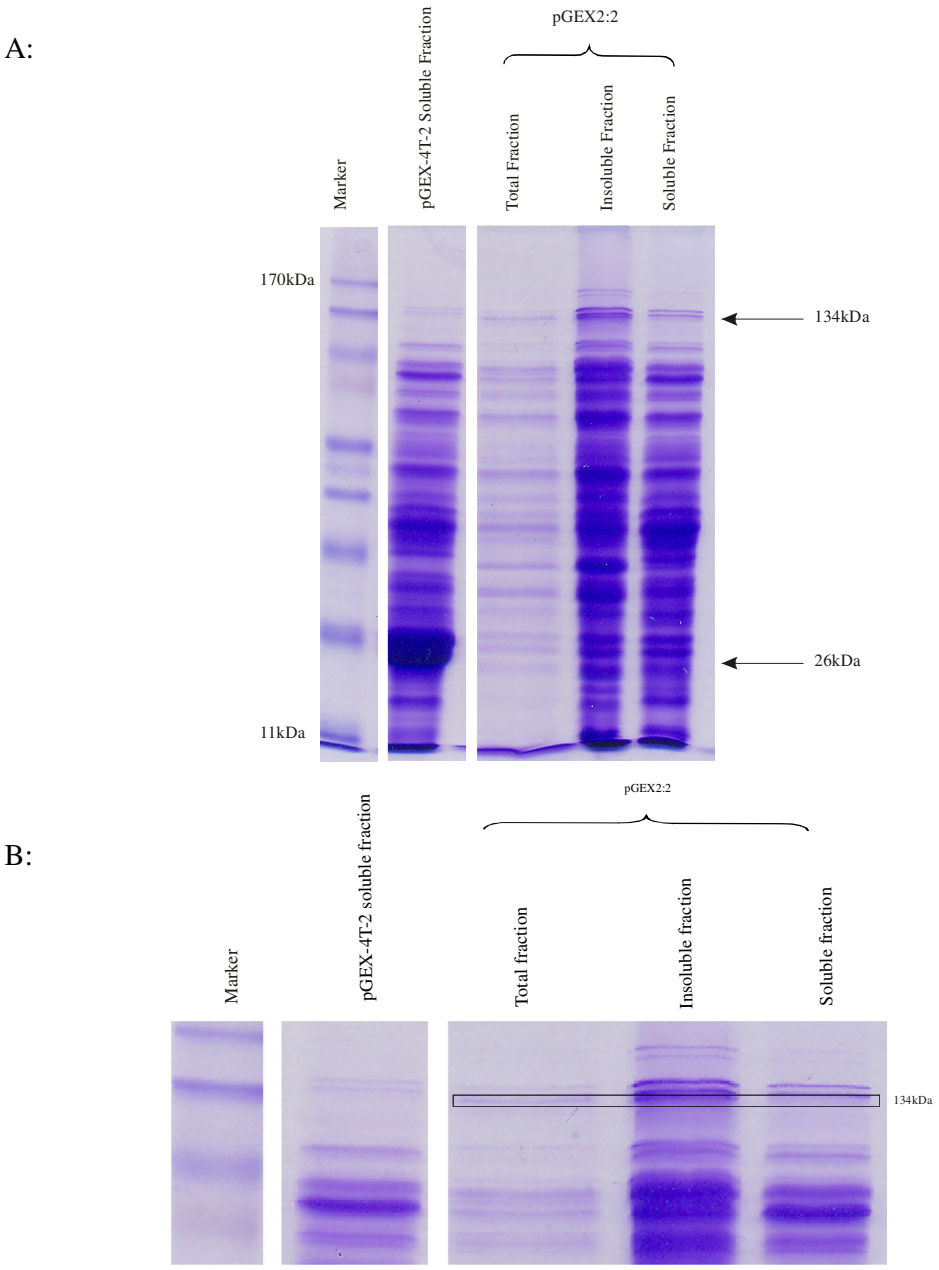


Figure 3.4.2.2: Laemmli gel of induced Rosetta *E. coli* containing either pGEX2:2 or pGEX-4T-2

A: The Laemmli gel shows the pGEX-4T-2 soluble fraction (10ul/2ml) and total (5ul/1ml), insoluble (10ul/300ul) and soluble (15ul/2ml) fractions of induced Rosetta *E coli* containing pGEX2:2.

B: Enlargement of the region where *Pf*Myb2-GST should occur (highlighted with a black box). There is much less protein loaded in the total pGEX2:2 fraction than the other fractions. Even with such a small amount of protein in the total fraction there is a band at 134kDa (presumably *Pf*Myb2-GST) that is also present in the insoluble fraction but absent from the soluble fraction of pGEX2:2. This band is also absent in the soluble pGEX-4T-2 fraction.

Irrelevant lanes of the gel have been excluded.

5ul of the Fermentas prestained protein ladder was used as the marker.

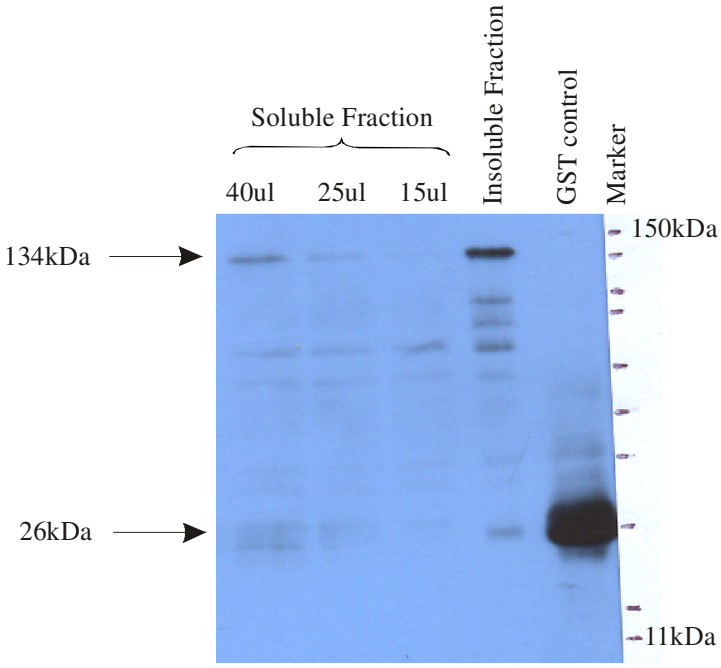


Figure 3.4.2.3: Western blot of soluble and insoluble fractions of Rosetta *E. coli* expressing the pGEX2:2 construct
A Western blot to detect *Pf*Myb2-GST using anti GST-HRP conjugate. 10ul/300ul of the insoluble fraction was loaded while 15ul/2ml, 25ul/2ml and 40ul/2ml of the soluble fraction were loaded.
There are large amounts of *Pf*Myb2-GST in the insoluble fraction compared to the amount of soluble *Pf*Myb2-GST.
The GST control also serves as an *E. coli* control (20ul/2ml soluble fraction loaded). The band above GST between molecular weights of ~30kDa - 50kDa may indicate non-specific binding of the anti-GST antibody to *E. coli* proteins as these bands are present in all fractions.
The higher molecular weight bands (> ~50kDa) are presumably truncated forms of *Pf*Myb2-GST as they are not present in the *E. coli* control fraction.
Protein bands of the Fermentas prestained protein ladder have been marked off and placed next to the blot, as the GST-HRP conjugate did not detect the protein markers.

Affinity purification of the soluble fraction using MagneGST™ Glutathione particles showed that no protein was eluted with elution buffer containing either 100mM or 200mM glutathione. The absence of protein was due to a lack of starting material and not due to ineffective elution as no protein was detected when the glutathione particles were stripped (results not shown).

As very little soluble *Pf*Myb2-GST was expressed, the insoluble fusion protein from the bacterial inclusion bodies was extracted and denatured in guanidine hydrochloride. After denaturation, the guanidine hydrochloride concentration was decreased through dialysis and the protein refolded. The refolded fusion protein

was purified using 5ul of the MagneGST™ Glutathione particles and eluted in 60ul elution buffer with 50mM glutathione.

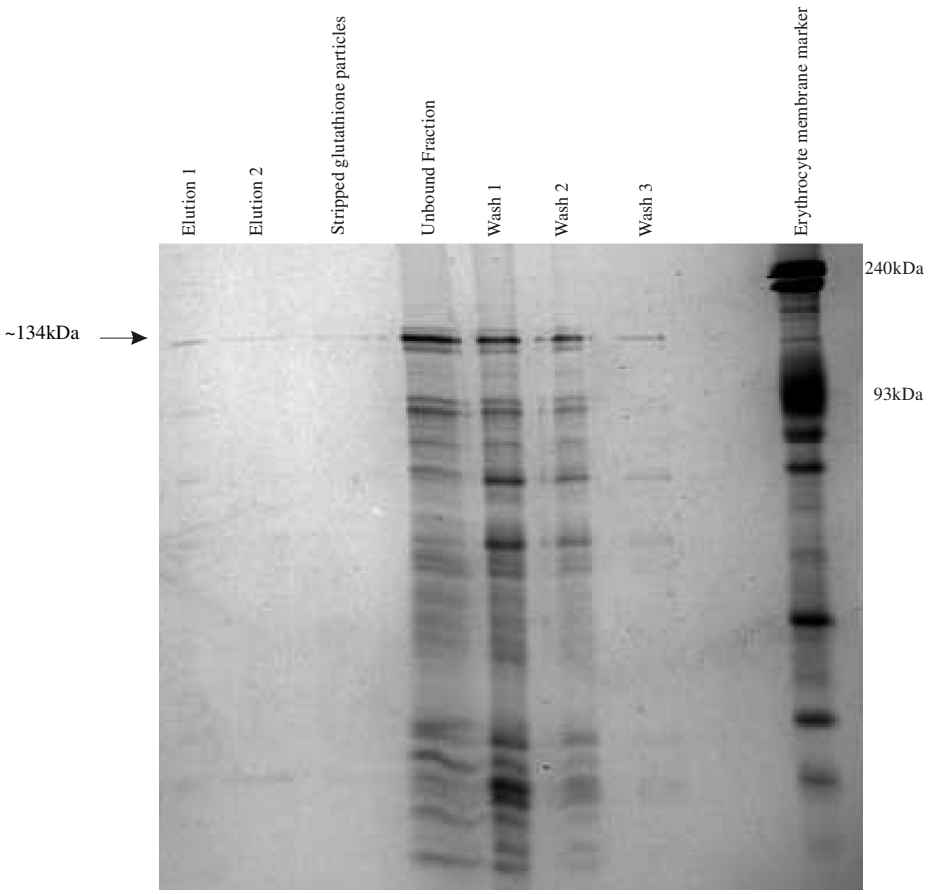


Figure 3.4.2.4: Laemmli gel of refolded *Pf*Myb2-GST
The Laemmli gel shows the results of refolding the 134kDa *Pf*Myb2-GST fusion protein and elution with 50mM glutathione.
Protein bands at 134kDa can be seen in all the protein fractions, however the bands are very faint in the eluted samples.
30ul/60ul of the elution and 30ul/200ul stripped particle fraction were loaded per well and 20ul/200ul of the unbound and wash fractions were loaded per well. 5ul of erythrocyte membrane proteins was loaded as a molecular weight marker.

When the samples were analysed by SDS-PAGE very faint protein bands were seen in the eluted fractions at ~134kDa (figure 3.4.2.4). Protein bands were also seen at this molecular weight in the unbound and stripped particle fractions. The western blots on these samples, however, were only positive for GST at ~134kDa in the stripped particle fraction (figure 3.4.2.5). This implied that the elution of the

protein was not effective and that it remained bound to the glutathione particles. Therefore varying elution conditions were tested, including the use of a non-ionic detergent.

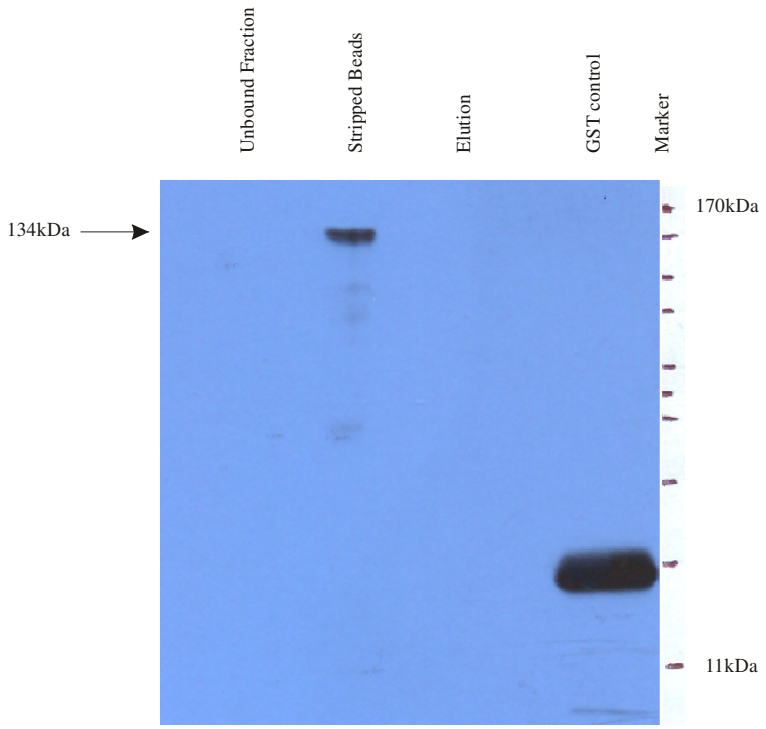


Figure 3.4.2.5: Western blot of refolded *Pf*Myb2-GST
A Western blot using anti-GST HRP conjugate shows that *Pf*Myb2-GST was not eluted with 50mM glutathione but was still bound to the MagneGST™ Glutathione particles as the only detection of GST was at 134kDa in the stripped particle fraction.
20ul/200ul of the unbound fraction and 30ul/60ul of the elution and 30ul/200ul stripped particle fraction were loaded per well. Protein bands of the Fermentas prestained protein ladder have been marked off and placed next to the blot, as the GST-HRP conjugate did not detect the markers.

The glutathione concentration was increased and the protein was eluted in buffer containing 100mM or 200mM glutathione. A second elution was performed on the same sample using the same concentrations of glutathione and 1% Tween®-20 detergent. The eluted samples were analysed on a Laemmli gel (figure 3.4.2.6). Protein bands were present at ~134kDa when the elution buffer contained a final concentration of 100mM and 200mM glutathione, but not in the secondary elutions. Elution was inefficient since protein bands at ~134kDa were still present when the beads had been stripped with SDS. These results were confirmed with a Western blot, figure 3.4.2.7, with the most intense signal detected when the

elution buffer had a final concentration of 200mM glutathione. The final yield of protein was estimated from the Laemmli gel by comparing the *Pf*Myb2-GST protein band with the Fermantas prestained protein ladder that had a known concentration. There appeared to be ~1-2ug of *Pf*Myb2-GST on the gel and therefore ~2-4ug per elution of refolded protein. This would result in a yield of ~20-40ug *Pf*Myb2-GST per 500ml *E. coli* culture.

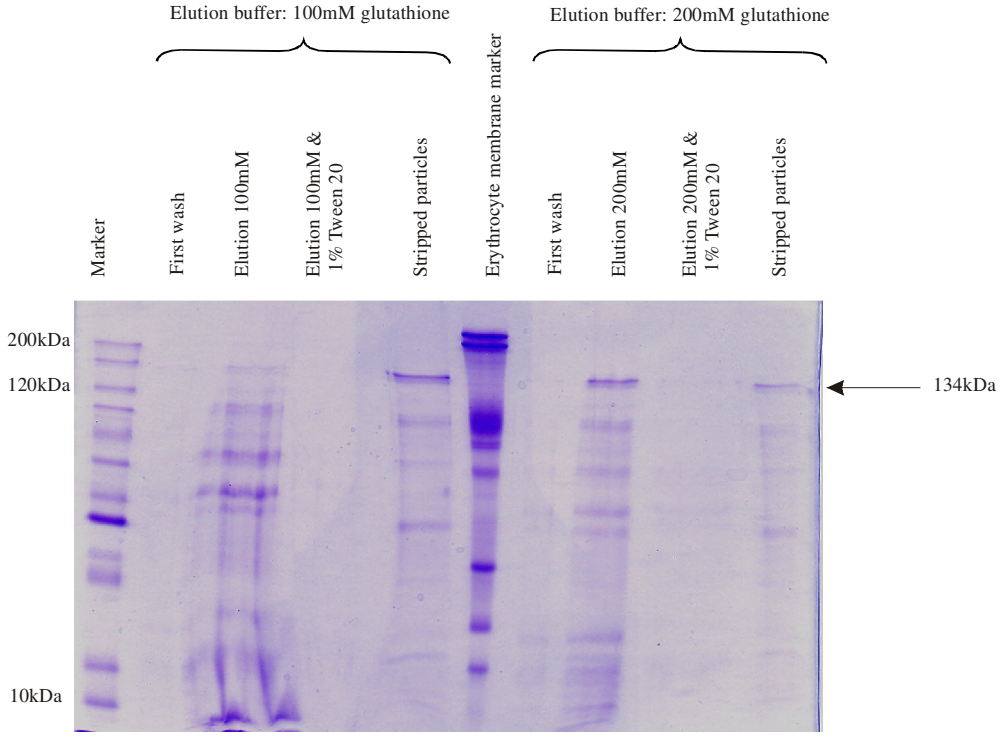


Figure 3.4.2.6: Laemmli gel of refolded *Pf*Myb2-GST using varying elution conditions
The Laemmli gel shows the eluted, refolded *Pf*Myb2-GST using different elution conditions in two experiments. Protein bands can be seen when the elution buffer contained 100mM and 200mM glutathione. No bands were detected when the elution buffers contained 1% Tween®-20. 30ul/60ul of the elution sample was loaded per lane. Protein bands can also be seen in the stripped particle fractions (40ul/200ul per lane).
There is a band at the correct size (134kDa) of *Pf*Myb2-GST and the lower molecular weight bands may be truncated forms of the fusion protein or *E.coli* proteins. 5ul of the Fermentas unstained protein ladder and 5ul of an erythrocyte membrane were loaded as molecular weight markers.

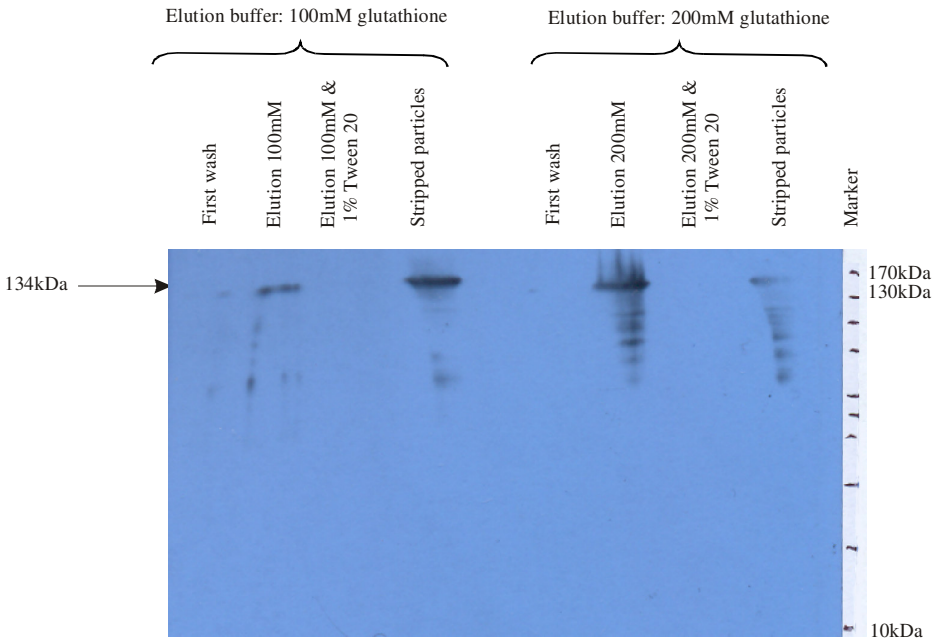


Figure 3.4.2.7: Western blot of refolded *Pf*Myb2-GST using varying elution conditions
The Western blot using anti-GST HRP conjugate shows the eluted, refolded *Pf*Myb2-GST using different elution conditions.
A signal at 134kDa was detected when the elution buffer contained 100mM and 200mM glutathione and when the particles were stripped. No signal was detected when Tween-20 was added to the elution buffer. GST was also detected in various bands that are smaller than the *Pf*Myb2-GST fusion protein, which may be due to truncated proteins or degradation.
20ul/200ul of the wash fractions, 40ul/200ul of the stripped particle fractions, and 30ul/60ul of the eluted sample were loaded. Protein bands of the Fermentas prestained protein ladder have been marked off and placed next to the blot, as the GST-HRP conjugate did not detect the markers.

3.5. Cloning of *Pf*Myb2 into pTriEx-3

The yield of soluble *Pf*Myb2-GST from the pGEX-*Pf*Myb2 constructs was very low and the purification using glutathione-linked magnetic beads was inefficient and yielded minute amounts of protein, which prompted the use of the pTriEx-3 expression vector. This vector encodes a his tag that is attached to the carboxyl end of the fusion protein (appendix A-5 and A-6) and purification of *Pf*Myb2-his, using nickel coated beads may have been more successful. This vector can also be used to express recombinant proteins in mammalian and insect cell lines, as well as in *E. coli*.

3.5.1 Cloning into pTriEx-3

The pGEX-4T-2 and pTriEx-3 vectors both have *Bam*HI and *Xho*I restriction enzyme sites in the same orientation. Therefore the fully sequenced *Pf*Myb2 gene insert was successfully cut out from pGEX2:2 and ligated into the pTriEx-3 vector which was used to transform DH5α *E. coli* cells. Religation of the vector did occur as the control plate contained 33 colonies and the experiment plate contained 45 colonies. 10 colonies from the experiment plates were chosen and screened for the *Pf*Myb2 gene using PCR and *Pf*Myb2 primers. PCR screening revealed that only two out of the 10 colonies were positive for the gene insert, as seen in figure 3.5.1.1. These two pTriEx-3-*Pf*Myb2 constructs were named pTriEx-3:5 and pTriEx-3:8 after the order they had been picked from the agar plate. Plasmid preparations and restriction enzyme digestion with *Bam*HI and *Xho*I verified that these constructs contained the gene insert (figure 3.5.1.2). The plasmid preparations were used to transform Rosetta *E. coli* bacterial cells and the presence of the gene insert was verified as described above.

3.5.2 Induction of pTriEx-*Pf*Myb2 constructs

Prior to inducing *Pf*Myb2-his protein the vector sequence was scrutinised more closely, which revealed that the *Bam*HI site was unfortunately not in frame with the vector initiation codon. This would create a premature stop codon at the first amino acid of *Pf*Myb2 (appendix A-7) and no recombinant protein would be expressed which was confirmed by an induction experiment (results not shown).

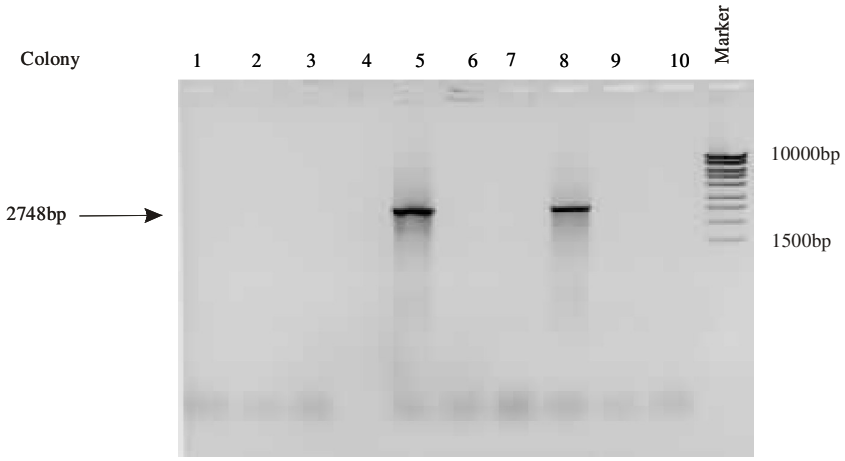


Figure 3.5.1.1: PCR amplification of pTriEx-3 transformed DH5α *E. coli* colonies
The 1% agarose gel shows PCR amplification of *Pf*Myb2 insert from 10 colonies using *Pf*Myb2 specific primers.
Only colonies 5 and 8 were positive for the 2748bp *Pf*Myb2 amplicon.
10ul of each reaction tube was loaded per well and 5ul of the Fermentas high range MassRuler™ was used as the marker.

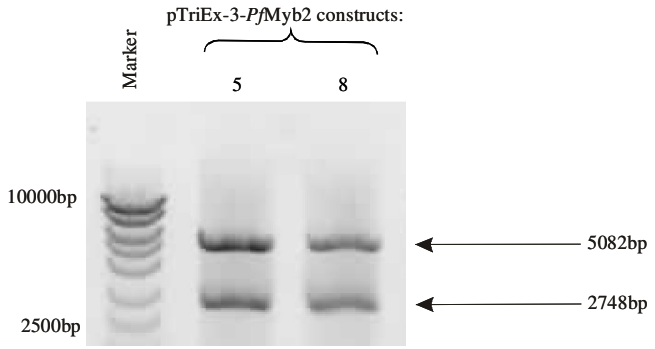


Figure 3.5.1.2: Restriction enzyme digest of pTriEx-3-*Pf*Myb2 constructs
The 1% agarose gel shows the pTriEx-3-*Pf*Myb2 constructs, digested with *Bam*HI and *Xho*I.
The vector can be seen at 5082bp and the *Pf*Myb2 insert at 2748bp for both constructs.
20ul of each digest was loaded per well and 5ul of the Fermentas high range MassRuler™ used as the marker.

3.6 Bioinformatic analysis of *Pf*Myb2

3.6.1 General features of *Pf*Myb2

The gene (Pf10_0327) is located on chromosome 10 and has only one exon of 2748bp, which encodes 915 amino acids (see figure 3.6.1). *Pf*Myb2 contains two Myb-like DNA-binding domains, which fall between residues 7-53 (R1) and 59-103 (R2). These domains have homology to the R1 and R2 domains of *Pf*Myb1. Seven low complexity regions are found throughout the protein but only one (KTEWNKEEEEKLLHLAKL) falls in part of the R2 domain, as it is located between residues 59-76. This low complexity region does not appear to be specific to *P. falciparum* as it is also found in Myb2 proteins of other *Plasmodium* species and in human cell division cycle 5 (CDC5), a homologue of *Pf*Myb2. Hydropathy plots revealed that the protein is mainly hydrophilic with no transmembrane domain and secondary structure predictions showed mainly helices.

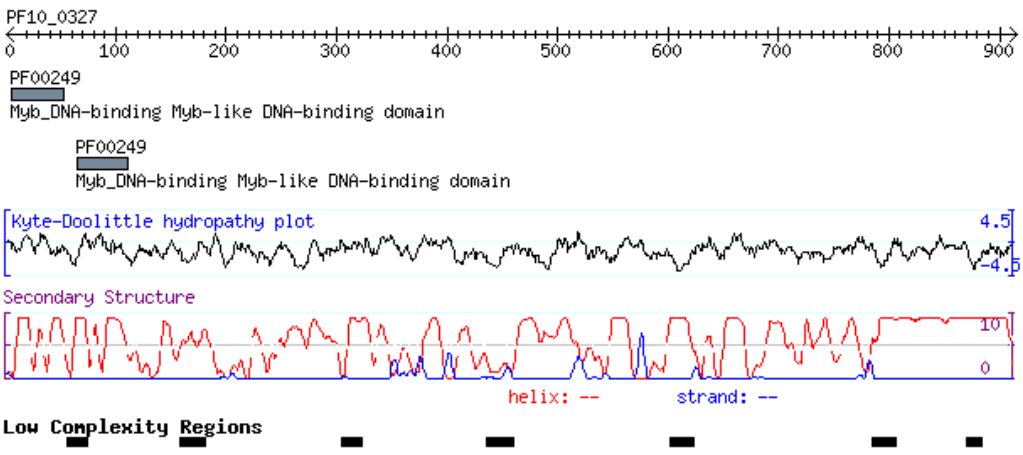
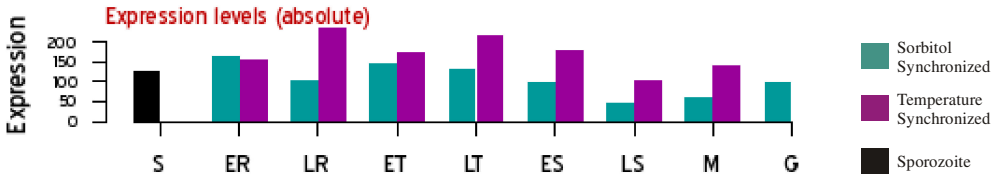


Figure 3.6.1: Bioinformatic predictions of *Pf*Myb2
Annotation of *Pf*Myb2 taken from PlasmoDB, version 5.4.
The protein contains two amino-terminal, Myb-like DNA-binding domains. A Kyte-Doolittle hydropathy plot revealed that the protein is mainly hydrophilic as most regions of the protein in the plot fall below 0. Secondary structure predictions show mainly helices. One low complexity region out of seven falls within the second Myb-like DNA binding domain.

Expression studies by Le Roch *et al.* (2003) and by Bozdech *et al* (2003) have been summarised on PlasmoDB. These studies revealed that *Pf*Myb2 mRNA is expressed throughout the intraerythracytic lifecycle (figure 3.6.2). Le Roch *et al* (2003) showed that temperature-synchronised and sorbitol-synchronised cultures have the highest mRNA expression at late ring and early ring stages respectively and the lowest expression level occurred in the late schizont stage. Work done by Florens *et al* (2004) which analysed proteomic data in *P. falciparum* using mass spectrometry revealed that peptide fragments were detected in trophozoites and gametocytes.



(ER = Early Rings | LR = Late Rings | ET = Early Trophs | LT = Late Trophs | G = Gametocytes ES = Early Schizonts | LS = Late Schizonts | M = Merozoites | S = Sporozoites)

Figure 3.6.2: Expression levels of *Pf*Myb2 mRNA in the intraerythrocytic life stages of *P. falciparum*
The graph, taken from PlasmoDB version 4.5, shows *Pf*Myb2 mRNA expression levels in the parasite at all stages of the intraerythrocytic lifecycle (Le Roch *et al*., 2003).

3.6.2 *Plasmodium* homologues of *Pf*Myb2

BLASTP searches of *Pf*Myb2 showed that the protein had homology to Myb2 proteins found in other *Plasmodium* species. The alignments between *Pf*Myb2 and Myb2 found in *P. yoelii*, *P. berghei* and *P. vivax* were performed using ClustalW2 and can be seen in figure 3.6.3. Overall these three homologues had a 59.4%, 58.6% and 57% identity to *Pf*Myb2 respectively. The amino terminal ends of these four proteins are highly conserved and within the first 300 residues there is ~80% similarity. This area also contains the two Myb-like DNA-binding domains, which are highly conserved throughout these homologues (table 3.6) and contain the highly conserved tryptophan and cysteine residues.



Figure 3.6.3: Alignment of *Pf*Myb2, *P. berghei* Myb2 and *P. vivax* Myb2
The alignment of *Pf*Myb2 (P.f) and Myb2 proteins found in *P. yoelii* (P.y), *P. berghei* (P.b) and *P. vivax* (P.v) was performed using ClustalW2. The R1 and R2 Myb-like DNA-binding domains of all four proteins have been highlighted in yellow and grey respectively. Within these domains the conserved tryptophan (W) and cysteine (C) residues have been highlighted in red and green respectively. These regions are conserved between all four Myb2 proteins.
Identical (*), conserved (:), semi-conserved regions (.) and gaps to fit the alignment (-) are indicated.

Table 3.6: Comparison of *Pf*Myb2 whole protein and DNA-binding domains with other homologues proteins.

The table shows the percentage identity and similarity between the whole *Pf*Myb2 protein and its two DNA-binding domains (R1 and R2) with four other *Plasmodium* homologues (*Py*Myb2, *Pb*Myb2, *Pv*Myb2 and *Pf*Myb1) and a human protein (CDC5). Identity is calculated using exact matching base pairs while similarity takes into account both exact and conserved base pairs for the two proteins.

	Whole Protein		R1		R2	
	Identity	**Similarity**	**Identity**	**Similarity**	**Identity**	**Similarity**
***Py*Myb2**	59.4%	73.1%	100%	100%	97.8%	100%
***Pb*Myb2**	58.6%	72.2%	100%	100%	95.6%	100%
***Pv*Myb2**	57%	72%	100%	100%	93.3%	100%
***Pf*Myb1**	9.4%	17.1%	22.4%	44.9%	31.2%	50%
CDC5	31.7%	49.1%	75%	82.7%	76.5%	84.3%

3.6.3 Comparison between *Pf*Myb2 and *Pf*Myb1

Alignments between *Pf*Myb1 and *Pf*Myb2 were performed using ClustalW2. The *Pf*Myb1 protein is a *Plasmodium*-specific transcription factor that contains three carboxyl terminus Myb-like DNA-binding domains (R1, R2 and R3), which were the most conserved regions between the proteins (figure 3.6.4), however, there was still a large degree of variability. The *Pf*Myb2 R1 and R2 domains align with the *Pf*Myb1 R1 and R2 domains respectively (see table 3.6). Overall the two proteins had an identity and similarity of 9.4% and 17.1% respectively. *Pf*Myb2 is a large protein and is more than double the size of *Pf*Myb1, which may account for the low identity between the proteins.

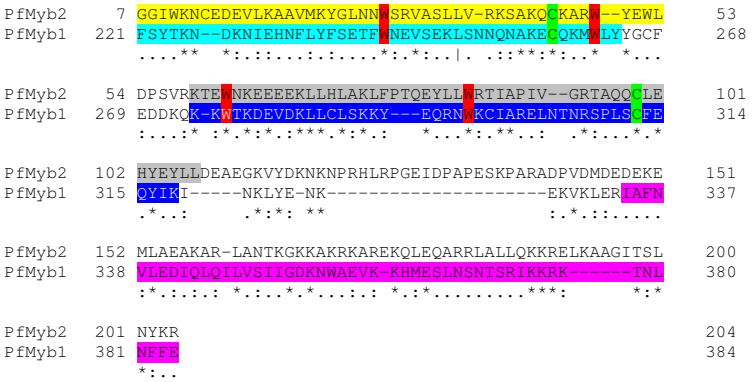


Figure 3.6.4: The conserved DNA-binding domains in *Pf*Myb2 and *Pf*Myb1
The alignment between the DNA-binding domains of *Pf*Myb2 and *Pf*Myb1 was performed using ClustalW2. The *Pf*Myb2 R1 and R2 domains are highlighted in yellow and grey respectively and align with *Pf*Myb1 R1 (turquoise) and R2 (blue) domains. R3, the third DNA binding domain of *Pf*Myb1 is highlighted in purple, but is not present in *Pf*Myb2. The conserved tryptophan (W) and cysteine (C) residues, that fall within the DNA-binding domains, have been highlighted in red and green respectively.
Identical (*), conserved (:), semi-conserved regions (.) and gaps to fit the alignment (-) are indicated.

3.6.4 Homology between *Pf*Myb2 and human CDC5

BLASTP searches revealed that *Pf*Myb2 had a 38% and 33% identity to the DNA-binding domains of c-Myb and B-Myb respectively, but it had an unusually high homology to the human CDC5 protein, which is the closest human homologue. Alignments between these two proteins using ClustalW2 revealed an identity and similarity of 31.7% and 49.1% respectively over the whole molecule (table 3.6). Like *Pf*Myb2, CDC5 only contains two amino-terminal DNA-binding domains, R1 and R2, (figure 3.6.5) and there is an incredibly high similarity and identity (table 3.6) between these domains between the two proteins, considering the evolutionary distance between humans and the parasite. This contrasts with the low homology to *Pf*Myb1 and other Myb proteins (table 3.6).

BLASTP analysis of *Pf*Myb2 and CDC5 revealed an additional conserved domain called REB1. This domain is found in Myb super-family proteins, which play a role in transcription as well as mRNA splicing. It was located at the amino terminal end of the proteins and spanned up to 289 residues in *Pf*Myb2 and up to

257 residues in CDC5 (figure 3.6.5). The alignment also revealed that *Pf*Myb2 contains *Plasmodium* specific inserts, indicated by large gaps in the CDC5 sequence (figure 3.6.5).

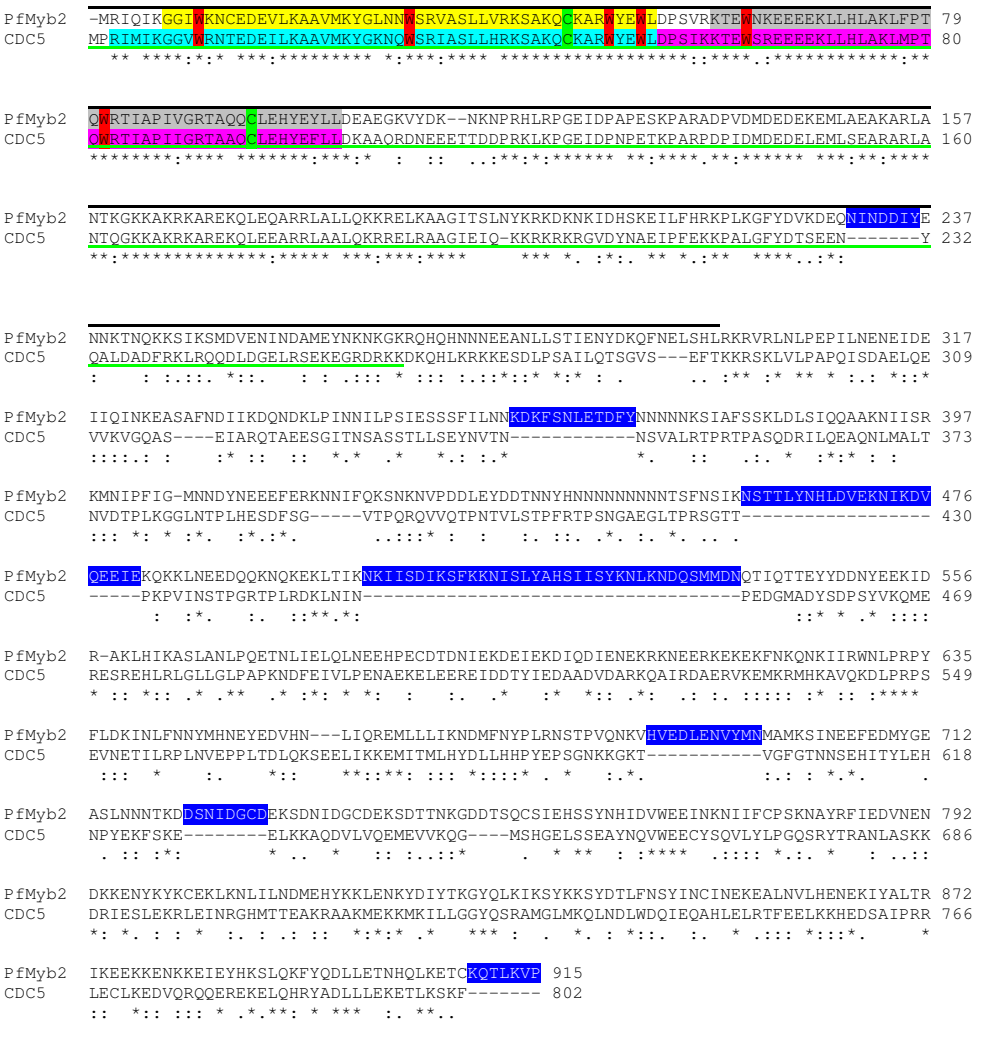


Figure 3.6.5: Alignment of *Pf*Myb2 and human CDC5 protein
The alignment between *Pf*Myb2 and CDC5 was performed using ClustalW2. The R1 and R2 DNA-binding domains of *Pf*Myb2 have been highlighted in yellow and grey respectively. The R1 and R2 DNA-binding domains of CDC5 have been highlighted turquoise and pink respectively. The conserved tryptophan (W) and cysteine (C) residues, which fall within the DNA-binding domains, have been highlighted in red and green respectively. The REB1 domain has been highlighted by a black over-line and green underline for *Pf*Myb2 and CDC5 respectively. *Plasmodium* specific regions are highlighted in blue.
Identical (*), conserved (:), semi-conserved regions (.) and gaps to fit the alignment (-) indicated.

4. DISCUSSION

4.1 Recombinant *Pf*Myb2-GST protein expression in *E. coli*

The pGEX-4T-2 vector was induced in *E. coli* to express *Pf*Myb2, a putative *P. falciparum* transcription factor, with an amino terminal GST tag. Auto induction of *Pf*Myb2-GST using Overnight Express™ Instant TB Medium resulted in mostly insoluble protein production, which confirms studies by various other groups showing that expression of recombinant *P. falciparum* proteins in *E. coli* is notoriously difficult.

4.1.1 Expression of *P. falciparum* proteins in *E. coli*

Mehlin et al. (2006) analysed the expression of 1000 *P. falciparum* genes in *E. coli* of which only 33.7% were able to express protein, and 6.3% of the proteins were soluble. Another study by Vedadi et al. (2007) also demonstrated a low success rate of producing soluble *P. falciparum* proteins in *E. coli*. Insolubility of *P. falciparum* proteins in a prokaryotic host is due to a number of factors pertaining to the physical characteristics of the protein and its gene. A high (>6) isoelectric point (pI) resulted in insoluble expression of proteins while a lower pI (<6) yielded less protein but higher solubility (Mehlin et al., 2006). The molecular weight of the recombinant proteins also played a role in solubility and a high molecular weight (>60kDa) resulted in low expression and low solubility of proteins (Mehlin et al., 2006, Vedadi et al., 2007). Tag technology, for example the addition of a GST tag, can improve solubility (http://www.emdbiosciences .com/html/NVG/solubility.html), but this does not seem to be the case for *P. falciparum* proteins (Mehlin et al., 2006). The current study also demonstrated that *Pf*Myb2-GST was insoluble despite the presence of a GST tag. The *Pf*Myb2-GST fusion protein has a pI >6 and a molecular weight of 134kDa and therefore its insolubility corroborates the Mehlin et al (2006) study.

The lack of homology of the target proteins to *E. coli* proteins impacted negatively on protein expression (Mehlin et al., 2006). Myb proteins have been identified in slime moulds and fungi (Lipsick, 1996), however, these proteins do not seem to be present in bacteria. During the BLASTP searches on *Pf*Myb2, no

homology to any possible endogenous *E. coli* Myb was found and this could have contributed to the insolubility of the recombinant protein.

4.1.2 Strategies to improve expression of soluble *Pf*Myb2-GST

The high AT content of the parasite genome results in the prevalence of certain codons that are not utilised frequently in *E. coli* and large stretches of adenosines and thymidines may also cause problems in expression of *P. falciparum* genes in bacteria. Two codons, ATA and AGA expressing tRNAs for isoleucine and arginine respectively, are common in *P. falciparum* and are particularly rare in *E. coli* (Mehlin et al., 2006). To correct this potential problem, Rosetta *E. coli* cells containing the pRARE plasmid were used in this study. This plasmid expresses tRNAs for arginine, isoleucine, glycine, leucine and proline and therefore facilitates the production of *P. falciparum* proteins.

The pGEX-4T-2 vector contains a *lac* promoter, which in the presence of lactose induces expression of the inserted gene fragment. Induction of *Pf*Myb2-GST was initially tested using IPTG. This molecule activates beta-galactosidase, which binds the lac repressor and induces the use of lactose resulting in expression of genes under control of the *lac* promoter. A number of varying induction conditions were tested including IPTG concentration, temperature and the length of time of induction. Altering these variables independently and together did not improve the expression or solubility of *Pf*Myb2-GST. Therefore Overnight Express™ Instant TB Medium was utilised. This medium does not require the addition of IPTG and allows the bacterial cells to grow to a high density initially, followed by autoinduction of protein expression due to a build up of lactose. It resulted in improved recombinant protein expression although ~90% of the protein was insoluble. The amount of soluble protein was too small to purify using affinity chromatography and it was therefore decided to isolate the insoluble protein from the bacterial inclusion bodies.

4.1.3 Purification of *Pf*Myb2-GST from inclusion bodies

When the protein induction rate is very high and large quantities of recombinant protein is produced, the protein is not folded correctly and is packaged into

inclusion bodies within the bacterial cell (Vallejo and Rinas, 2004). Although the recombinant protein does not have the correct conformation it can be isolated from the inclusion bodies, solubilised with a strong denaturant (such as guanidine hydrochloride) and then refolded into its presumed native shape by slowly decreasing the concentration of the denaturant. Inclusion bodies in the cell are a different size to the rest of the cellular components and can easily be isolated based on their density and size. They contain very few cellular components other than the insoluble recombinant protein, which results in a high yield of pure protein. The inclusion bodies also protect the protein from degradation and cellular proteases (Singh and Panda, 2005).

The refolding method and affinity purification of the refolded *Pf*Myb2-GST had to be optimised. Although some of the *Pf*Myb2-GST did not bind to the glutathione particles, that which did remained tightly bound even after elution with a buffer containing 50mM glutathione. This indicated that the dissociation of bound *Pf*Myb2-GST was not efficient and the amount of glutathione in the buffer was increased to a final concentration of 200mM, four times that of the recommended concentration.

Minor protein bands at lower molecular weights than that expected for *Pf*Myb2-GST were detected in Western blots. These were presumably truncated forms of *Pf*Myb2-GST, as the GST tag is attached to the amino terminal end of the protein and would therefore also be attached to any truncated recombinant protein. The *Pf*Myb2-GST construct is very large (3.5kb) and this would increase the likelihood of premature termination of translation. The majority of the recombinant protein purified from the inclusion bodies represented the desired full length *Pf*Myb2-GST. The yield was however, still relatively low (~20-40ug per 500ml culture of *E. coli*). Larger amounts of recombinant protein could be obtained if the volume of the culture is increased, even though the expression is poor. It was not known if the protein had refolded into its native state but the fact that it did not precipitate during refolding implied that it was in its correct conformation. The verification of correct conformation of the refolded protein

could be done using circular dichroism (CD). This measures the difference in a molecule's absorption of left- and right-handed polarised light. Secondary structures of proteins including helices, coils and β-sheets all exhibit different CD spectra (Johnson, 1990). Therefore, the secondary structure of a protein can be determined by comparing its CD spectrum to that of helices, coils and β-sheets. In this case the reference curve obtained for the soluble recombinant protein can be compared to that obtained for the refolded protein, however, CD requires relatively large amounts of soluble protein to obtain the reference curve and there may not be enough soluble *Pf*Myb2 available. Another method to determine correct refolding of the insoluble recombinant protein would be to test its ability to perform its function of DNA binding, using DNA-binding studies.

4.1.4 Additional strategies to enhance soluble expression of *Pf*Myb2

There are a number of other options that may be tested in the future to try and enhance the soluble expression of *Pf*Myb2.

Instead of producing the *Pf*Myb2 gene through PCR amplification of genomic *P. falciparum* DNA, a gene consisting of codons optimised for *E. coli* can be synthesized (Burgess-Brown et al., 2008). This will eliminate the codon bias present in *P. falciparum* genes due to their high AT content. However, Mehlin et al. (2006) did not find any differences in soluble expression using this technique although other studies have shown its efficacy (Zhou et al., 2004, Yadava and C.F., 2003). *Pf*Myb2 is a large gene (2.7kb) and codon optimisation may not be cost effective especially since there is no guarantee of success.

Molecular chaperones facilitate the correct folding of proteins in the cell (Cole, 1996) and this can be used to increase soluble protein expression by transforming *E. coli* cells that contain the pGEX-*Pf*Myb2 constructs with a plasmid that expresses chaperones. Another modification that may result in more soluble protein is the use of a different soluble tag such as N utilization substance A (NusA), maltose-binding protein (MBP) (Nallamsetty and Waugh, 2006), thioredoxin (TRX), Z-domain from protein A (ZZ) and Gb1-domain from protein

G (Gb1) (Hammarstrom et al., 2002). These amino-linked fusion tags, have different effects on recombinant protein expression and solubility in *E. coli* (Hammarstrom et al., 2002, Braun et al., 2002, Chelur, 2008) and thus the most suitable tag for a protein has to be assessed on an individual basis. Chelur (2008) also report that fusion tags may hinder protein expression and solubility as well as alter its conformation which can cause a loss of the proteins function. In the current study, fusion of an amino-terminal GST to *Pf*Myb2 produced very little soluble protein.

Finally, the baculovirus expression system could be utilised for the expression of *Pf*Myb2. In this system, vector DNA containing the gene of interest is inserted into baculoviral DNA through recombination and transfected into insect cells where the gene is induced to express recombinant protein (Ausubel et al., 1995). The advantage of this system is that insect cells are eukaryotic and this may facilitate correct folding of *Pf*Myb2. Mehlin et al. (2006) had good success with this system. Out of 17 proteins that were insoluble in *E. coli*, they were able to produce seven soluble proteins using the baculovirus expression system. The pTriEx-3 vector was available in Prof Coetzer's laboratory and can be used to express protein in bacteria, mammalian, as well as insect cells utilising the baculovirus system. The *Pf*Myb2 insert was cut out of the pGEX-*Pf*Myb2 vector constructs and inserted into pTriEx-3. However, the insert is not in frame and new primers will have to be designed to clone *Pf*Myb2 into this and other vectors. Other pTriEx vectors that are available contain a signal for export, which secretes the recombinant protein from the insect cell into the culture supernatant allowing for easy retrieval of the protein (Novagen insect cell expression catalogue).

4.2 Possible cellular functions of *Pf*Myb2

4.2.1 Role of *Pf*Myb2 as a transcription factor

Expression studies (Le Roch et al., 2003, Bozdech et al., 2003) and mass spectrometry analysis (Florens et al., 2004, Florens et al., 2002) indicated that *Pf*Myb2 is expressed at low levels throughout the intraerythrocytic lifecycle of the

parasite, while peptide fragments have been detected in gametocytes and trophozoites. Sequence alignments of the Myb-like DNA binding regions between *Pf*Myb2, Myb2 in other *Plasmodium* species, as well as with *Pf*Myb1 showed that these domains contain the characteristic Myb HTH motif with regularly spaced tryptophan residues and a highly conserved cysteine residue. These conserved residues play a role in stabilising the DNA-binding interaction. EMSA studies performed on the DNA-binding domains of *Pf*Myb2 showed that they bound to a consensus MRE, a MRE from the *mim1* gene in chicken and two putative MREs in the promoters of two *P. falciparum* genes: *pfmap1* and *pfcrk1* (Meyersfeld, 2005). The *pfcrk1* gene is expressed mainly in gametocytes and encodes a kinase that is important in sexual differentiation of the parasite (Doerig et al., 1995), while the *pfmap1* gene encodes a kinase that plays a role in cell cycle control (Doerig et al., 1996). This supports the fact that *Pf*Myb2 may activate genes involved in cell cycle control and that its function overlaps with *Pf*Myb1, a *P. falciparum*-specific transcription factor that regulates genes involved in cell cycle and differentiation (Boschet et al., 2004). Immunofluorescence localisation assays with an antibody specific for *Pf*Myb2 showed that the protein is present in the parasite nucleus (Meyersfeld et al., 2008), which further corroborates its role as a transcription factor.

4.2.1.1 Possible interactive role of *Pf*Myb2 and *Pf*Myb1

A combinatorial effect of transcription factor binding has been suggested by Van Noort and Huynen (2006). They speculate that transcription factors can bind in different combinations to different regulatory *cis*-sequences, which may account for the apparent scarcity of transcription factors in the *P. falciparum* genome. The combinatorial effect may be a mechanism employed by *Pf*Myb2 as it binds to the same MREs as *Pf*Myb1. *Pf*Myb2 contains only two Myb-like DNA-binding domains (R1 and R2), while in most cases the Myb family proteins contain three of these domains (R1, R2 and R3, see figure 4.2). Studies in c-Myb have shown that R2 and R3 interact directly with the DNA while the R1 motif is not essential for DNA-binding (Howe et al., 1990). Therefore the loss of R3 in *Pf*Myb2 suggests that DNA binding could occur by a different mechanism to that of

*Pf*Myb1, which has all three binding motifs. Due to the difference in their DNA-binding domains, these proteins may also display variable affinity for the MREs, further implementing the combinatorial effect.

*Pf*Myb2 mRNA is expressed throughout the intraerythrocytic lifecycle and *Pf*Myb1 mRNA is also present in all stages except late schizonts. Both have differential levels of mRNA expression (measured in Affymetrix MOID expression values) with *Pf*Myb2 peaking in the early ring stage at 200 and *Pf*Myb1 in the early trophzoite stage at 70 (Le Roch et al., 2003). This may suggest that *Pf*Myb2 transcriptionally activates more genes than *Pf*Myb1 or that it has other roles in the parasite (discussed in sections 4.2.2 and 4.2.3).

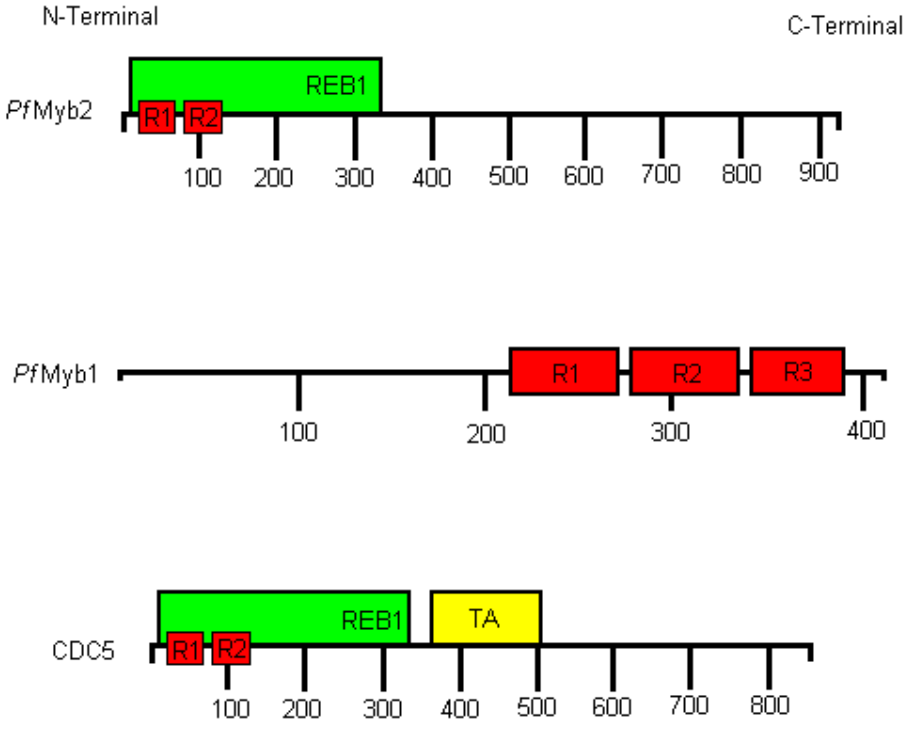


Figure 4.2 The general structure of *Pf*Myb2, *Pf*Myb1 and human CDC5 proteins
The diagram depicts the general structure of *Pf*Myb2 and two homologues proteins, *Pf*Myb1 and human CDC5. The DNA-binding domains (R1-R3) are indicated in red. *Pf*Myb2 and CDC5 only have the R1 and R2 motifs at the amino terminal end, while *Pf*Myb1 contains all three motifs at the carboxyl end of the protein. *Pf*Myb2 and CDC5 also have an amino terminal REB1 domain, indicated in green, that is not present in *Pf*Myb1. A transactivation domain, indicated in yellow, is present in CDC5 only. The number of amino acids in each protein is indicated in each diagram.

4.2.1.2 Role of *Pf*Myb2 based on its homology to human CDC5

The human CDC5 protein and *Pf*Myb2 contain a ~50% similarity to one another and have a marked conservation in their amino terminal ends. The overall homology between these two proteins is much greater than that of *Pf*Myb2 and *Pf*Myb1, where conserved residues are only apparent in the DNA-binding regions. The amino terminal ends of these two proteins contain the R1 and R2 DNA-binding domains, which have up to ~85% similarity, as well as the REB1 domain, which incorporates R1 and R2 (figure 4.2). Based on this homology *Pf*Myb2 may have functions more similar to CDC5 rather than to *Pf*Myb1.

The CDC5 protein is highly conserved both structurally and functionally amongst eukaryotes and appears to be a multifunctional protein (Lei et al., 2000). Many studies have investigated yeast and human CDC5 to ascertain their functions in the cell. CDC5 positively regulates the cell cycle G_2/M progression, which involves the mitotic division of the cell after DNA replication has taken place (Bernstein and Coughlin, 1998). In addition to the two amino terminal Myb-like tandem repeat motifs, CDC5 also contains a central transcriptional activation domain but this is not present in *Pf*Myb2. CDC5 is able to bind a double stranded DNA motif (GATTTANCATAA, where N can be any base), which has a core HTH binding motif (ANCA) (Bernstein and Coughlin, 1998, Bernstein and Coughlin, 1997, Lei et al., 2000). These factors and the similarity of CDC5 to Myb family proteins, which function as transcriptional regulators, imply that it regulates the cell cycle by transcriptional activation. The high sequence similarity between *Pf*Myb2 and CDC5, as well as the fact that *Pf*Myb2 bound to MREs, supports the notion that *Pf*Myb2 is a transcription factor, which may control expression of genes in the cell cycle as does CDC5 and *Pf*Myb1.

4.2.2 Possible role of *Pf*Myb2 in RNA processing

CDC5 has additional functions in the cell that the Myb family proteins are not involved in. This may explain the differences noted between CDC5 and c-Myb mRNA expression levels. In humans, CDC5 has a greater level of expression compared to c-Myb (Bernstein and Coughlin, 1997) and this pattern is similar in

P. falciparum, where more *Pf*Myb2 mRNA is present when compared to *Pf*Myb1 (Le Roch et al., 2003). One other function of CDC5 is its involvement in mRNA processing. It forms part of the spliceosome (Ajuh et al., 2001), a protein complex that processes pre-mRNA into mature mRNA by splicing out introns and joining exons together. CDC5 is required for the first biochemical step in processing pre-mRNA and if its levels are disrupted it results in an accumulation of pre-mRNA and defective mRNA splicing (Ajuh et al., 2001, Burns et al., 1999). Therefore the involvement of CDC5 in the cell appears to be multifunctional and it plays a role in regulating the cell cycle by controlling gene expression through transcriptional as well as post-transcriptional mechanisms. It would be interesting to ascertain if *Pf*Myb2, like CDC5, plays a role in pre-mRNA splicing by forming part of the spliceosome in the malaria parasite.

4.2.3 Possible role of *Pf*Myb2 in chromatin remodelling

The RNA polymerase I enhancer binding (REB1) domain may also provide clues to additional functions of *Pf*Myb2. This domain is common to both *Pf*Myb2 and CDC5 and was identified as a conserved domain during BLASTP searches of these two proteins. REB1 is a multifunctional protein associated with Myb family proteins and it has been studied mainly in yeasts. It is one of the general regulatory proteins that function in transcription (Ju et al., 1990) by binding to DNA consensus motifs (CCGGGTA and CGGGTRRPuPu) and enhancing the binding of other specific transcription factors. REB1 appears to play a role in transcriptional termination by causing pol I (responsible for the expression of ribosomal RNA genes (Fox et al., 1993)) or pol II, to pause along the DNA strand (Lang and Reeder, 1993, Lang et al., 1994, Paule and White, 2000). In addition to these functions REB1 is also involved in chromatin remodelling (Angermayr and Bandlow, 1997). REB1 binds to DNA and partitions it into active euchromatin by altering the shape of the chromosome or inhibiting nucleosome binding to flanking regions at the REB1 binding site, producing a nucleosome free region. This may stop the spread of silent chromatin encroaching on active chromatin (Fourel et al., 2002, Raisner et al., 2005, Yu et al., 2003). In yeast the REB1 protein is one of the factors that controls the expression of a cyclin needed for

G_2/M transition (Van Slyke and Grayhack, 2003), which is the same point in the cell cycle that CDC5 regulates. It would be interesting to evaluate the *P. falciparum* genome for REB1 DNA-binding motifs and perform DNA-binding studies with recombinant *Pf*Myb2, if any such sequences are identified. If *Pf*Myb2 were to bind to these regions, additional studies to assess whether it controls gene expression in a transcriptional or an epigenetic manner would provide valuable insight into the role of this protein in the parasite.

4.3 Future studies

Bioinformatic analysis of *Pf*Myb2 has revealed conserved motifs, which provide clues to the possible functions of this protein. However, until these are validated using molecular techniques they still remain as speculations. This study has produced a vector construct containing the full length *Pf*Myb2 gene that can be used to express recombinant *Pf*Myb2 protein to validate these putative functions. Since the soluble expression of this protein is very low, use of the baculovirus system in insect cells may enhance the production of soluble recombinant protein to be utilised in these studies.

The DNA binding regions of recombinant *Pf*Myb2 have been used in our laboratory for EMSA studies which showed that these domains bound MREs from two *P. falciparum* genes (Meyersfeld, 2005). The protein has also been localised to the nucleus (Meyersfeld et al., 2008) and all this evidence is strong support that *Pf*Myb2 is a transcription factor. The full length recombinant *Pf*Myb2 protein can be used in DNA-binding studies to MREs to further validate these data. EMSAs could also be utilised to determine if *Pf*Myb2 is able to bind to REB1 consensus repeats, indicating the protein's ability to regulate gene expression through possible epigenetic means. Reporter gene assays would be required to verify that *Pf*Myb2 is able to trans-activate and drive expression of a downstream reporter gene, such as luciferase. Studies using gene knockdown and knockout technologies would also provide insight into the role of *Pf*Myb2 in the parasite.

Based on the high homology to CDC5 it would be interesting to see if *Pf*Myb2 also forms part of the parasite's spliceosome. *Pf*Myb2 may play a role in pre-mRNA splicing by forming part of this complex or by directly interacting with pre-mRNA molecules. Direct protein-RNA interactions in *P. falciparum* have been demonstrated by Gunasekera et al. (2007). Similar experiments to demonstrate the interaction of *Pf*Myb2 with RNA will highlight its possible role in post-transcriptional control of gene expression.

4.4 Conclusion

This study has resulted in the production of the full length, putative transcription factor, *Pf*Myb2 as a recombinant GST fusion protein. Future studies to elucidate the function of this protein in *P. falciparum* will provide valuable insight into mechanisms of gene regulation, which are currently not well understood. By studying the parasite's biology we will slowly gain understanding of the complex lifecycle of this organism and find novel means of intervention to help control the parasite and the burden of malaria.

5. LIST OF SUPPLIERS

Acrylamide	Promega, USA
AB plasma	South African National Blood Transfusion Service
ACD tubes	Beckton Dickinson, UK
Agar	Merck, Germany
Agarose	Hispanagar, Spain
Ammonium chloride	Sigma, USA
Ammonium persulphate	Promega, USA
Ampicillin	Roche, Germany
β-mercaptoethanol	Merck, USA
Bactotryptone	Beckton Dickinson, UK
*Bam*H1 restriction enzyme	Fermentas, RSA
Beckman J2-21 centrifuge and JA-17 rotor	Beckman, USA
Bis-acrylamide	Promega, USA
Blue loading dye	Fermentas, RSA
Bovine serum albumin	Roche, Germany
Bromophenol blue	BDH, UK
BRL model S2 electrophoresis apparatus	Amersham, UK
BugBuster®	Novgen, USA
Centrifuge tubes	Nunc, Denmark
Chloramphenicol	Sigma, USA
Chloroform	Associated Chemical Enterprises
Coomassie brilliant blue R-250	Merck, Germany
Cryotubes	Nunc, Denmark
Culture flasks	Nunc, Denmark
DC power supply, PS 500X	Hoefer, USA
DH5α competent cells	Invitrogen, USA
Dialysis tubing	Sigma, USA
DNA MassRuler™	Fermentas, RSA
DTT	Behringer Mannheim GmbH
EDTA	BDH, RSA
Eppendorf 5415R centrifuge	Merk, USA
Eppendorf TripleMaster PCR system	Merk, USA
Ethanol	Saarchem, RSA
Ethidium bromide	Sigma, USA
Filter paper (extra thick)	Biorad, USA

Gentamicin	Roche, Germany
GenElut™ plasmid miniprep kit	Sigma, USA
Glucose	Saarchem, RSA
Glutathione	Sigma, USA
Glycerol	Saarchem, RSA
Glycine	BDH, RSA
Gyanidine-hydrochloride	Calbiochem, Germany
Hepes buffer	Invitrogen, USA
HyBond™-C Extra nitrocellulose membrane	Amersham, UK
Hypoxanthine solution	Sigma, USA
IPTG	Invitrogen, USA
Isopropanol	Saarchem, RSA
Kanamycin	Roche, Germany
Laminar flow hood	Labotec, RSA
L-Lysozyme	Mecrk, USA
Methanol	Saarchem, RSA
Mighty Small duel casting tray	Hoefer, USA
MiniElute Gel Extraction Kit	Qiagen
Mouse anti-his antibody	Amersham, UK
Oligonucleotide primers	Inquaba biotech, RSA
Overnight Express™ Instant TB Medium EasyPak	Novagen, USA
PageRuler™ prestained protein ladder	Fermentas, RSA
Para film	Whatman, UK
PCR Triple master mix kit	Roche, Germany
PCR Master mix	Eppendorf, Germany
PCR tubes	QSP, USA
Peristaltic pump SJ-1211	SIS, USA
pET-41a	Novagen, USA
PGEX-4T-2 expression vector	Novagen, USA
pTriEx-3	Novagen, USA
Phenol	Sigma, USA
Ponceau S	Sigma, USA
Potassium acetate	Saarchem, RSA
Potassium chloride	Saarchem, RSA
Potassium di-hydrogen orthophosphate	Saarchem, RSA
Proteinase-K	Roche, Germany

Protease inhibitor cocktail set III	Calibiochem, USA
Rapid DNA ligation kit	Roche, Germany
Rapid Haematomolgy staining kit	SAIMR, RSA
Ribonuclease A	Fermentas, RSA
PMFS	Roche, Germany
RPMI culture medium	Invitrogen, USA
Rosetta 2(DE3) competent cells	Novagen, USA
Saponin	USB, UK
SDS	BDH, UK
Sorbitol	Sigma, USA
Sodium acetate	Saarchem, RSA
Sodium bicarbonate	Saarchem, RSA
Sodium chloride	Saarchem, RSA
Sodium citrate	Saarchem, RSA
Sodium di-hydrogen orthophosphate	Saarchem, RSA
Sodium hydroxide	Saarchem, RSA
Spectrophotometer (SU65)	Beckman, USA
Syngene gel doc system	Synoptics, UK
Tango™ buffer	Fermentas, RSA
TEMED	Promega, USA
Tris	BDH, UK
Triton X-100	Sigma, USA
Tween® 20 detergent	Calibiochem, USA
Yeast extract	Oxoid, UK
*Xho*1 restriction enzyme	Fermentas, RSA
Zeiss Axiostar microscope	Zeiss, Germany

6. APPENDIX

A-1: pET-41a

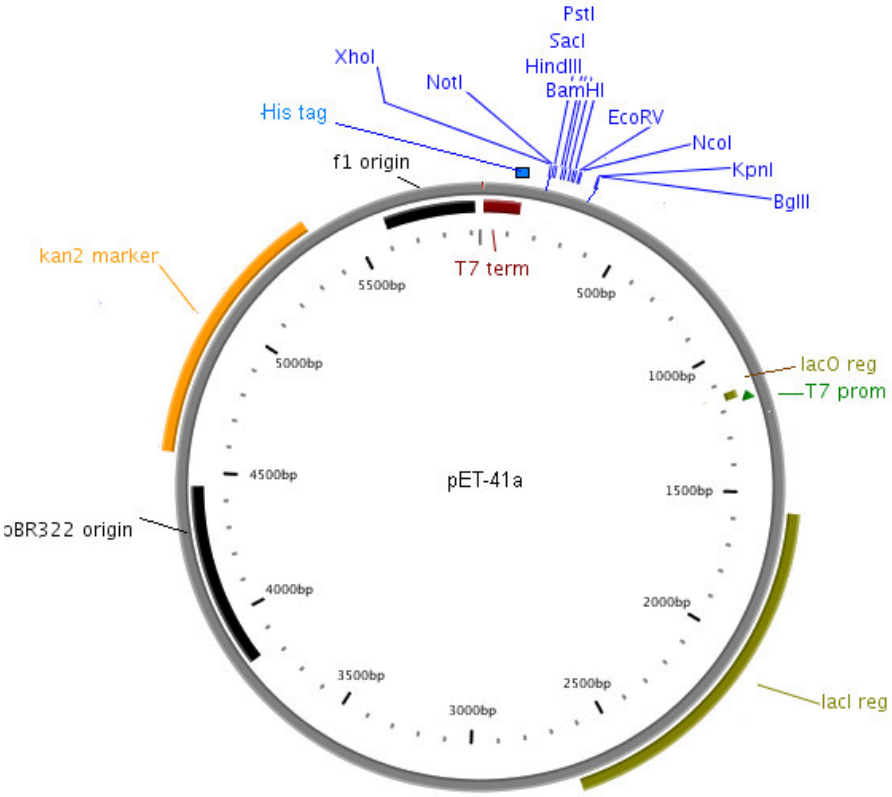
The expression vector adds eight codons for histidine to the 3' end of the inserted gene which results in a carboxyl terminal his tag in the recombinant protein.

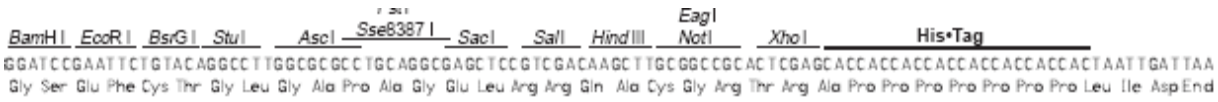
Vector map drawn with PlasMapper 2.0

(http://wishart.biology.ualberta.ca/PlasMapper/)

The cloning/expression region was taken from the pET-41a vector map on the Novagen website

(http://www.merckbiosciences.com/g.asp?f=NVG/pETtable.html#pet41a).





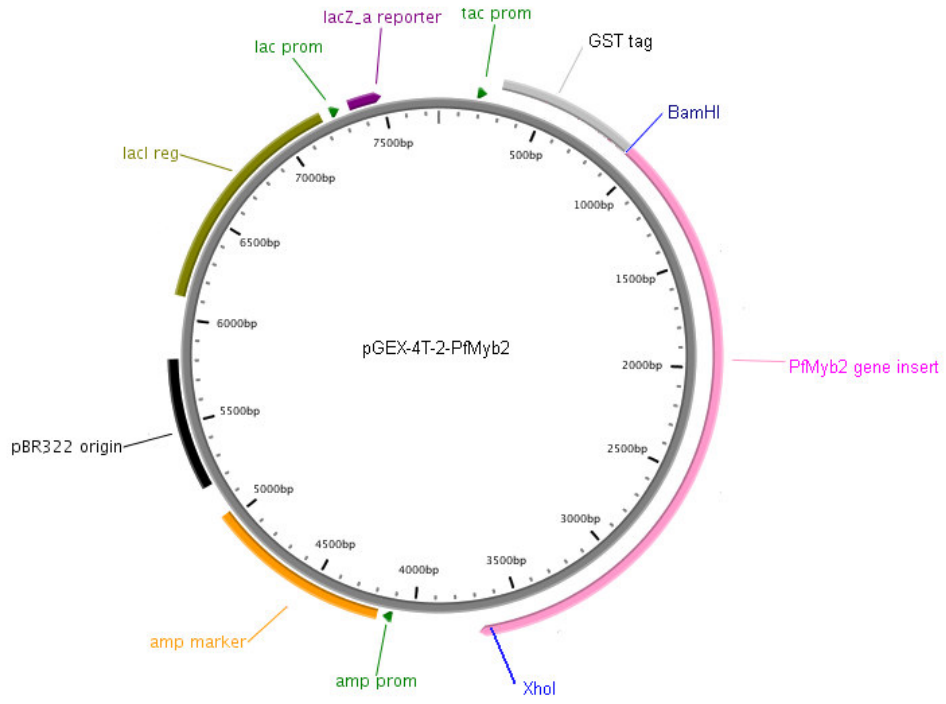
A-2: pGEX-4T-2

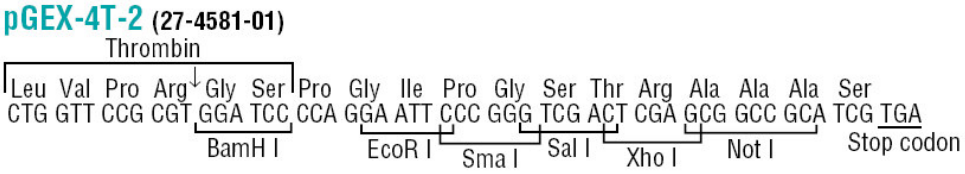
The pGEX-4T-2 vector with *Pf*Myb2 gene inserted into the *Bam*HI and *Xho*I sites of the MCS. The expression vector adds a 5' GST sequence to the inserted gene which results in an amino terminus GST tag in the recombinant protein.

Vector map drawn with PlasMapper 2.0

(http://wishart.biology.ualberta.ca/PlasMapper/)

The cloning/expressing region was taken from the vector map available on the GE-healthcare life sciences website (www1.gelifesciences.com).





A-3: *Pf*Myb2-GST gene sequence in pGEX-4T-2

GST start Codon: ATG; Stop codon: TGA

GST gene: Black; *Pf*Myb2 gene insert: Green; vector sequence: Purple

Restriction enzyme sites: *Bam*HI: GGATCC

*Xho*I: CTCGAG

pGEX Sequencing primers:

5' primer, nucleotides 869-891: 5'GGGCTGGCAAGCCACGTTTGGTG3'

3' primer, nucleotides 1042-1020: 5'CCGGGAGCTGCATGTGTCAGAGG 3'

*Pf*Myb2 sequencing forward primers:

*Pf*Myb2WholeSeq1F: 5'ACTTGCAAATACGAAAGG3'

*Pf*Myb2WholeSeq2F: 5'GATCTAAGTATACAACAAGCAG3'

*Pf*Myb2WholeSeq3F: 5'CCAGAATGTGATACGGAC3'

ATGTCCCCTATACTAGGTTATTGGAAAATTAAGGGCCTTGTGCAACCCACTCGACTTCTTTTGGAATATCTTGAAGAAAAATATGA
AGAGCATTTGTATGAGCGCGATGAAGGTGATAAATGGCGAAACAAAAAGTTTGAATTGGGTTTGGAGTTTCCCAATCTTCCTTATT
ATATTGATGGTGATGTTAAATTAACACAGTCTATGGCCATCATACGTTATATAGCTGACAAGCACAACATGTTGGGTGGTTGTCCA
AAAGAGCGTGCAGAGATTTCAATGCTTGAAGGAGCGGTTTTGGATATTAGATACGGTGTTTCGAGAATTGCATATAGTAAAGACTT
TGAAACTCTCAAAGTTGATTTTCTTAGCAAGCTACCTGAAATGCTGAAAATGTTCGAAGATCGTTTATGTCATAAAACATATTTAA
ATGGTGATCATGTAACCCATCCTGACTTCATGTTGTATGACGCTCTTGATGTTGTTTTATACATGGACCCAATGTGCCTGGATGCG
TTCCCAAAATTAGTTTGTTTTAAAAAACGTATTGAAGCTATCCCACAAATTGATAAGTACTTGAAATCCAGCAAGTATATAGCATG
GCCTTTGCAGGGCTGGCAAGCCACGTTTGGTGGTGGCGACCATCCTCCAAAATCGGATCTGGTTCCGCGTGGATCCATGAGGATTC
AAATAAAAGGAGGTATATGGAAAAATTGTGAGGATGAAGTTCTTAAAGCAGCTGTTATGAAATATGGTTTAAATAATTGGTCAAGG
GTCGCATCTTTATTAGTTCGTAAATCAGCTAAACAGTGTAAGGCGAGGTGGTATGAATGGCTTGATCCATCAGTTAGAAAAACCGA
ATGGAATAAAGAAGAAGAAGAAAAACTTTTACATCTAGCCAAATTGTTTCCAACACAATGGAGAACCATAGCACCTATAGTAGGAA
GAACAGCACAACAATGTTTAGAACATTATGAATATCTACTTGATGAAGCGGAAGGAAAGGTTTATGATAAAAATAAGAACCCACGA
CATCTAAGGCCAGGTGAAATAGACCCTGCTCCTGAATCCAAACCAGCACGTGCAGATCCTGTAGATATGGATGAAGATGAAAAAGA
AATGCTTGCAGAAGCTAAGGCAAGACTTGCAAATACGAAAGGTAAAAAAGCAAAAAGAAAAGCTAGAGAAAAACAACTGGAACAAG
CTAGGAGATTAGCTTTATTACAAAAAAAGAGAGAATTAAAAGCAGCAGGGATAACATCATTAAATTATAAAAGAAAAGATAAAAAT
AAAATTGATCATTCGAAAGAAATATTATTTCATAGGAAACCATTAAAAGGTTTTTATGATGTTAAAGATGAACAAAATATTAATGA
TGATATATATGAAAATAATAAGACAAATCAGAAGAAAAGTATAAAATCAATGGATGTTGAAAATATTAATGATGCTATGGAATATA
ATAAAAATAAAGGTAAACGACAACATCAACATAATAATAATGAAGAAGCAAATTTGTTATCTACCATAGAAAATTATGATAAACAA
TTTAACGAATTAAGTCATTTAAGAAAAAGGGTTCGATTGAATTTACCAGAACCTATATTAAATGAAAATGAAATAGATGAAATAAT
ACAAATAAATAAAGAAGCATCAGCATTTAACGATATTATAAAAGATCAAAACGATAAATTACCAATAAATAATATTTTACCAAGTA
TTGAAAGCTCCTCATTTATATTAAACAATAAAGATAAATTTTCAAACTTGGAAACAGATTTTTATAATAATAATAATAATAAATCA
ATTGCTTTTTCTAGTAAACTAGATCTAAGTATACAACAAGCAGCAAAAAATATCATTTCTCGAAAAATGAATATTCCATTTATAGG
AATGAACAATGATTATAATGAAGAAGAATTCGAAAGAAAAAATAATATATTTCAAAAGAGTAATAAGAATGTTCCTGATGATTTAG
AATATGATGACACGAACAATTATCATAATAATAATAATAATAATAATAATACTTCTTTTAATTCGATCAAAAATTCCACAACC
CTTTATAATCATTTGGATGTAGAAAAAAATATAAAAGACGTACAAGAAGAAATTGAAAAACAAAAGAAATTAAATGAAGAAGATCA
ACAAAAGAATCAAAAAGAAAAGTTAACTATTAAAAATAAAATTATTAGTGATATTAAAAGTTTTAAAAAAAATATTAGTTTATATG
CTCATTCTATTATTTCTTATAAAAATTTAAAAAATGATCAATCTATGATGGATAATCAAACTATACAAACAACTGAATATTATGAT
GATAATTATGAAGAGAAAATCGATCGAGCAAAATTACATATTAAAGCATCGTTAGCCAATTTACCCCAAGAAACGAATCTTATAGA
ACTTCAATTAAATGAAGAACATCCAGAATGTGATACGGACAATATAGAAAAAGATGAAATAGAAAAAGATATACAAGATATTGAAA
ATGAAAAAAGAAAAAACGAAGAAAGAAAAGAAAAAGAAAAATTTAATAAGCAAAATAAAATTATTAGATGGAATTTACCTAGACCA
TATTTCTTAGATAAAATTAATCTCTTTAATAATTATATGCACAATGAATATGAAGATGTCCATAATTTAATACAAAGAGAGATGCT
CCTCTTAATAAAAAATGATATGTTTAATTATCCCCTCAGAAATTCAACACCAGTACAAAATAAAGTTCATGTGGAAGATCTCGAAA
ACGTATACATGAATATGGCTATGAAGAGTATAAATGAGGAGTTCGAAGATATGTATGGCGAAGCATCATTAAATAACAACACAAAG
GATGATTCTAATATTGATGGGTGTGATGAAAAGAGTGATAATATTGATGGTTGTGATGAAAAAAGTGATACAACTAATAAAGGTGA
TGATACGTCCCAATGTAGTATCGAGCATAGTTCATATAATCATATCGATGTGTGGGAAGAAATTAACAAAAATATAATTTTTTGCC
CATCAAAAAATGCATATCGCTTTATTGAAGATGTTAATGAAAACGATAAAAAAGAAAATTACAAATACAAATGTGAAAAATTAAAA
AATTTAATTTTAAATGATATGGAACATTATAAAAAATTAGAAAATAAATATGATATTTATACAAAAGGATATCAACTAAAAATTAA
AAGTTATAAAAAATCATATGATACTCTATTTAATTCATATATTAATTGTATTAACGAAAAAGAAGCTTTAAATGTTCTACATGAAA
ATGAAAAAATATATGCATTAACAAGAATTAAAGAAGAAAAAAAGGAAAATAAAAAAGAAATCGAATATCATAAATCATTACAAAAG
TTTTATCAAGACCTTTTAGAAACAAATCATCAATTAAAAGAAACATGTAAACAAACGTTAAAGGTGCCACTCGAGCGGCCGCATCG
TGACTGACTGACGATCTGCCTCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGG

A-4: *Pf*Myb2-GST fusion protein sequence

Molecular weight: 134 kDa (*Pf*Myb2: 108 kDa; GST: 26kDa)

Number of residues: 1148 residues

pI: 6.59

Extinction coefficient: 146460M^{-1} cm^{-1}

GST Start Codon: M; Stop codon: -

GST sequence: Black

*Pf*Myb2 sequence: Green

Vector sequence: Purple

Restriction enzyme sites: *Bam*HI: GS

*Xho*I: LE

MSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFELGLEFPNLPYYIDGDVKLTQSM
AIIRYIADKHNMLGGCPKERAEISMLEGAVLDIRYGVSRIAYSKDFETLKVDFLSKLPEMLKMFEDRLC
HKTYLNGDHVTHPDFMLYDALDVVLYMDPMCLDAFPKLVCFKKRIEAIPQIDKYLKSSKYIAWPLQGWQ
ATFGGGDHPPKSDLVPRGSMRIQIKGGIWKNCEDEVLKAAVMKYGLNNWSRVASLLVRKSAKQCKARWY
EWLDPSVRKTEWNKEEEEKLLHLAKLFPTQWRTIAPIVGRTAQQCLEHYEYLLDEAEGKVYDKNKNPRH
LRPGEIDPAPESKPARADPVDMDEDEKEMLAEAKARLANTKGKKAKRKAREKQLEQARRLALLQKKREL
KAAGITSLNYKRKDKNKIDHSKEILFHRKPLKGFYDVKDEQNINDDIYENNKTNQKKSIKSMDVENIND
AMEYNKNKGKRQHQHNNNEEANLLSTIENYDKQFNELSHLRKRVRLNLPEPILNENEIDEIIQINKEAS
AFNDIIKDQNDKLPINNILPSIESSSFILNNKDKFSNLETDFYNNNNNKSIAFSSKLDLSIQQAAKNII
SRKMNIPFIGMNNDYNEEEFERKNNIFQKSNKNVPDDLEYDDTNNYHNNNNNNNNNTSFNSIKNSTTLY
NHLDVEKNIKDVQEEIEKQKKLNEEDQQKNQKEKLTIKNKIISDIKSFKKNISLYAHSIISYKNLKNDQ
SMMDNQTIQTTEYYDDNYEEKIDRAKLHIKASLANLPQETNLIELQLNEEHPECDTDNIEKDEIEKDIQ
DIENEKRKNEERKEKEKFNKQNKIIRWNLPRPYFLDKINLFNNYMHNEYEDVHNLIQREMLLLIKNDMF
NYPLRNSTPVQNKVHVEDLENVYMNMAMKSINEEFEDMYGEASLNNNTKDDSNIDGCDEKSDNIDGCDE
KSDTTNKGDDTSQCSIEHSSYNHIDVWEEINKNIIFCPSKNAYRFIEDVNENDKKENYKYKCEKLKNLI
LNDMEHYKKLENKYDIYTKGYQLKIKSYKKSYDTLFNSYINCINEKEALNVLHENEKIYALTRIKEEKK
ENKKEIEYHKSLQKFYQDLLETNHQLKETCKQTLKVPLERPHRD-

A-5: pTriEx-3

The pTriEx-3 vector with *Pf*Myb2 gene inserted into the *Bam*HI and *Xho*I sites of the MCS.

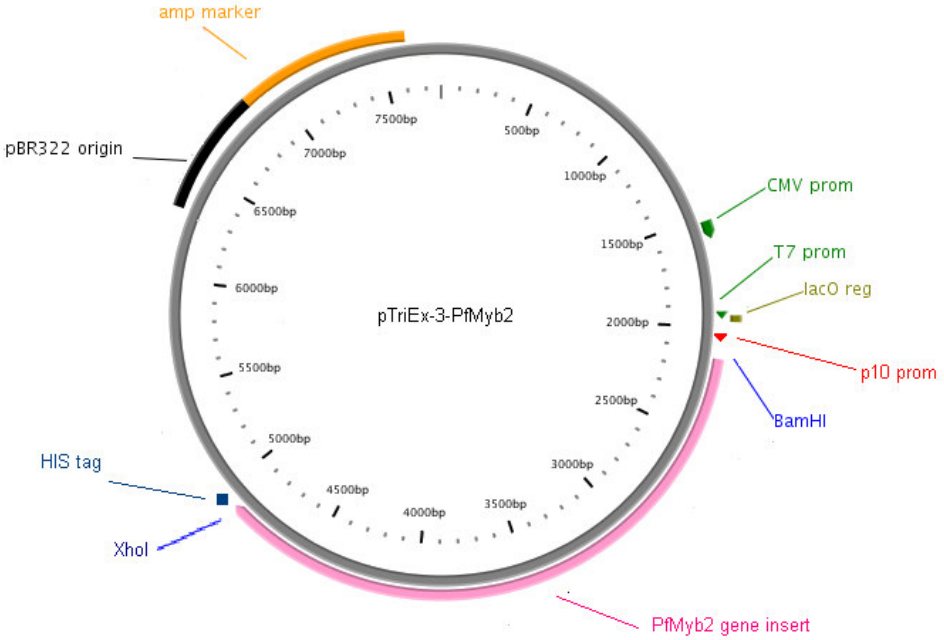
The expression vector adds eight codons for histidine to the 3' end of the inserted gene which results in a carboxyl terminal his tag in the recombinant protein.

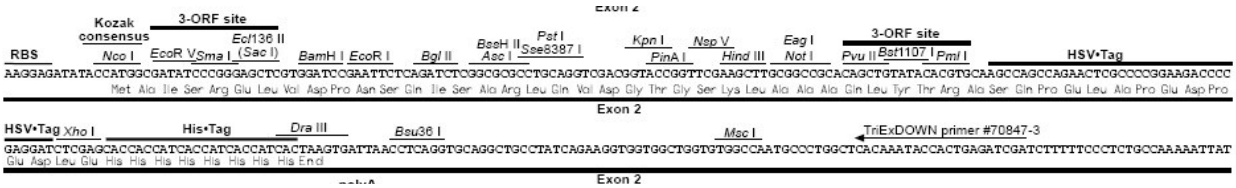
Vector map drawn with PlasMapper 2.0

(http://wishart.biology.ualberta.ca/PlasMapper/)

The cloning/expression region was taken from form the pTriEx3 vector map on the Novagen website

(http://www.merckbiosciences.com/html/NVG/pTriExTable.html).





A-6: *Pf*Myb2-his gene sequence in pTriEx-3

Vector start codon: ATG; *Pf*Myb2 start codon: ATG;

Stop codon: TAA; Premature stop codon: TGA

*Pf*Myb2 gene insert: Green; Vector sequence: Black; his tag: H

Restriction enzyme sites: *Bam*HI: GGATCC

*Xho*I: CTCGAG

```
ATGGCGATATCCCGGGAGCTCGTGGATCCATGAGGATTCAAATAAAAGGAGGTATATGGAAAAATTGTGAGGATGAAGTTCTTAAA
GCAGCTGTTATGAAATATGGTTTAAATAATTGGTCAAGGGTCGCATCTTTATTAGTTCGTAAATCAGCTAAACAGTGTAAGGCGAG
GTGGTATGAATGGCTTGATCCATCAGTTAGAAAAACCGAATGGAATAAAGAAGAAGAAGAAAAACTTTTACATCTAGCCAAATTGT
TTCCAACACAATGGAGAACCATAGCACCTATAGTAGGAAGAACAGCACAACAATGTTTAGAACATTATGAATATCTACTTGATGAA
GCGGAAGGAAAGGTTTATGATAAAAATAAGAACCCACGACATCTAAGGCCAGGTGAAATAGACCCTGCTCCTGAATCCAAACCAGC
ACGTGCAGATCCTGTAGATATGGATGAAGATGAAAAAGAAATGCTTGCAGAAGCTAAGGCAAGACTTGCAAATACGAAAGGTAAAA
AAGCAAAAAGAAAAGCTAGAGAAAAACAACTGGAACAAGCTAGGAGATTAGCTTTATTACAAAAAAAGAGAGAATTAAAAGCAGCA
GGGATAACATCATTAAATTATAAAAGAAAAGATAAAAATAAAATTGATCATTCGAAAGAAATATTATTTCATAGGAAACCATTAAA
AGGTTTTTATGATGTTAAAGATGAACAAAATATTAATGATGATATATATGAAAATAATAAGACAAATCAGAAGAAAAGTATAAAAT
CAATGGATGTTGAAAATATTAATGATGCTATGGAATATAATAAAAATAAAGGTAAACGACAACATCAACATAATAATAATGAAGAA
GCAAATTTGTTATCTACCATAGAAAATTATGATAAACAATTTAACGAATTAAGTCATTTAAGAAAAAGGGTTCGATTGAATTTACC
AGAACCTATATTAAATGAAAATGAAATAGATGAAATAATACAAATAAATAAAGAAGCATCAGCATTTAACGATATTATAAAAGATC
AAAACGATAAATTACCAATAAATAATATTTTACCAAGTATTGAAAGCTCCTCATTTATATTAAACAATAAAGATAAATTTTCAAAC
TTGGAAACAGATTTTTATAATAATAATAATAATAAATCAATTGCTTTTTCTAGTAAACTAGATCTAAGTATACAACAAGCAGCAAA
AAATATCATTTCTCGAAAAATGAATATTCCATTTATAGGAATGAACAATGATTATAATGAAGAAGAATTCGAAAGAAAAAATAATA
TATTTCAAAAGAGTAATAAGAATGTTCCTGATGATTTAGAATATGATGACACGAACAATTATCATAATAATAATAATAATAATAAT
AATAATACTTCTTTTAATTCGATCAAAAATTCCACAACCCTTTATAATCATTTGGATGTAGAAAAAAATATAAAAGACGTACAAGA
AGAAATTGAAAAACAAAAGAAATTAAATGAAGAAGATCAACAAAAGAATCAAAAAGAAAAGTTAACTATTAAAAATAAAATTATTA
GTGATATTAAAAGTTTTAAAAAAAATATTAGTTTATATGCTCATTCTATTATTTCTTATAAAAATTTAAAAAATGATCAATCTATG
ATGGATAATCAAACTATACAAACAACTGAATATTATGATGATAATTATGAAGAGAAAATCGATCGAGCAAAATTACATATTAAAGC
ATCGTTAGCCAATTTACCCCAAGAAACGAATCTTATAGAACTTCAATTAAATGAAGAACATCCAGAATGTGATACGGACAATATAG
AAAAAGATGAAATAGAAAAAGATATACAAGATATTGAAAATGAAAAAAGAAAAAACGAAGAAAGAAAAGAAAAAGAAAAATTTAAT
AAGCAAAATAAAATTATTAGATGGAATTTACCTAGACCATATTTCTTAGATAAAATTAATCTCTTTAATAATTATATGCACAATGA
ATATGAAGATGTCCATAATTTAATACAAAGAGAGATGCTCCTCTTAATAAAAAATGATATGTTTAATTATCCCCTCAGAAATTCAA
CACCAGTACAAAATAAAGTTCATGTGGAAGATCTCGAAAACGTATACATGAATATGGCTATGAAGAGTATAAATGAGGAGTTCGAA
GATATGTATGGCGAAGCATCATTAAATAACAACACAAAGGATGATTCTAATATTGATGGGTGTGATGAAAAGAGTGATAATATTGA
TGGTTGTGATGAAAAAAGTGATACAACTAATAAAGGTGATGATACGTCCCAATGTAGTATCGAGCATAGTTCATATAATCATATCG
ATGTGTGGGAAGAAATTAACAAAAATATAATTTTTTGCCCATCAAAAAATGCATATCGCTTTATTGAAGATGTTAATGAAAACGAT
AAAAAAGAAAATTACAAATACAAATGTGAAAAATTAAAAAATTTAATTTTAAATGATATGGAACATTATAAAAAATTAGAAAATAA
ATATGATATTTATACAAAAGGATATCAACTAAAAATTAAAAGTTATAAAAAATCATATGATACTCTATTTAATTCATATATTAATT
GTATTAACGAAAAAGAAGCTTTAAATGTTCTACATGAAAATGAAAAAATATATGCATTAACAAGAATTAAAGAAGAAAAAAAGGAA
AATAAAAAAGAAATCGAATATCATAAATCATTACAAAAGTTTTATCAAGACCTTTTAGAAACAAATCATCAATTAAAAGAAACATG
TAAACAAACGTTAAAGGTGCCACTCGAGCACCACCATCACCATCACCATCACTAA
```

A-7: *Pf*Myb2-his fusion protein sequence

Start codon encoded by the vector: M; Premature top codon: -

Vector Sequence: Black

Due to the *Bam*HI site occurring out of frame a premature stop codon (TGA in A-6) is created and no *Pf*Myb2 protein is produced.

```
MAISRELVDP-
```

A-8: Sequence analysis of the *Pf*Myb2 gene inserted into pGEX-4T-2

Black sequence is from PlasmoDB while the coloured sequences represent the gene insert.

The pGEX-4T-2 primer indicated as: 5'

The three designed, internal primers indicated as: INT1, INT2 and INT3

The pGEX-4T-2 primers indicated as: 3'pGEX and 3'pGEX2

```
Myb          1 ATGAGGATTCAAATAAAAGGAGGTATATGGAAAAATTGTGAGGATGAAGT     50
               ||||||||||||||||||||||||||||||||||||||||||||||||||
5'           1 ATGAGGATTCAAATAAAAGGAGGTATATGGAAAAATTGTGAGGATGAAGT     50

Myb         51 TCTTAAAGCAGCTGTTATGAAATATGGTTTAAATAATTGGTCAAGGGTCG    100
               ||||||||||||||||||||||||||||||||||||||||||||||||||
5'          51 TCTTAAAGCAGCTGTTATGAAATATGGTTTAAATAATTGGTCAAGGGTCG    100

Myb        101 CATCTTTATTAGTTCGTAAATCAGCTAAACAGTGTAAGGCGAGGTGGTAT    150
               ||||||||||||||||||||||||||||||||||||||||||||||||||
5'         101 CATCTTTATTAGTTCGTAAATCAGCTAAACAGTGTAAGGCGAGGTGGTAT    150

Myb        151 GAATGGCTTGATCCATCAGTTAGAAAAACCGAATGGAATAAAGAAGAAGA    200
               ||||||||||||||||||||||||||||||||||||||||||||||||||
5'         151 GAATGGCTTGATCCATCAGTTAGAAAAACCGAATGGAATAAAGAAGAAGA    200

Myb        201 AGAAAAACTTTTACATCTAGCCAAATTGTTTCCAACACAATGGAGAACCA    250
               ||||||||||||||||||||||||||||||||||||||||||||||||||
5'         201 AGAAAAACTTTTACATCTAGCCAAATTGTTTCCAACACAATGGAGAACCA    250

Myb        251 TAGCACCTATAGTAGGAAGAACAGCACAACAATGTTTAGAACATTATGAA    300
               ||||||||||||||||||||||||||||||||||||||||||||||||||
5'         251 TAGCACCTATAGTAGGAAGAACAGCACAACAATGTTTAGAACATTATGAA    300

Myb        301 TATCTACTTGATGAAGCGGAAGGAAAGGTTTATGATAAAAATAAGAACCC    350
               ||||||||||||||||||||||||||||||||||||||||||||||||||
5'         301 TATCTACTTGATGAAGCGGAAGGAAAGGTTTATGATAAAAATAAGAACCC    350

Myb        351 ACGACATCTAAGGCCAGGTGAAATAGACCCTGCTCCTGAATCCAAACCAG    400
               ||||||||||||||||||||||||||||||||||||||||||||||||||
5'         351 ACGACATCTAAGGCCAGGTGAAATAGACCCTGCTCCTGAATCCAAACCAG    400

Myb        401 CACGTGCAGATCCTGTAGATATGGATGAAGATGAAAAAGAAATGCTTGCA    450
               ||||||||||||||||||||||||||||||||||||||||||||||||||
5'         401 CACGTGCAGATCCTGTAGATATGGATGAAGATGAAAAAGAAATGCTTGCA    450

Myb        451 GAAGCTAAGGCAAGACTTGCAAATACGAAAGGTAAAAAAGCAAAAAGAAA    500
               ||||||||||||||||||||||||||||||||||||||||||||||||||
5'         451 GAAGCTAAGGCAAGACTTGCAAATACGAAAGGTAAAAAAGCAAAAAGAAA    500

Myb        501 AGCTAGAGAAAAACAACTGGAACAAGCTAGGAGATTAGCTTTATTACAAA    550
               ||||||||||||||||||||||||||||||||||||||||||||||||||
5'         501 AGCTAGAGAAAAACAACTGGAACAAGCTAGGAGATTAGCTTTATTACAAA    550

Myb        551 AAAAGAGAGAATTAAAAGCAGCAGGGATAACATCATTAAATTATAAAAGA    600
               ||||||||||||||||||||||||||||||||||||||||||||||||||
INT1         1 AAAAGAGAGAATTAAAAGCAGCAGGGATAACATCATTAAATTATAAAAGA     50

Myb        601 AAAGATAAAAATAAAATTGATCATTCGAAAGAAATATTATTTCATAGGAA    650
               ||||||||||||||||||||||||||||||||||||||||||||||||||
INT1        51 AAAGATAAAAATAAAATTGATCATTCGAAAGAAATATTATTTCATAGGAA    100

Myb        651 ACCATTAAAAGGTTTTTATGATGTTAAAGATGAACAAAATATTAATGATG    700
               ||||||||||||||||||||||||||||||||||||||||||||||||||
INT1       101 ACCATTAAAAGGTTTTTATGATGTTAAAGATGAACAAAATATTAATGATG    150
```

```
Myb              701 ATATATATGAAAATAATAAGACAAATCAGAAGAAAAGTATAAAATCAATG  750
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
INT1             151 ATATATATGAAAATAATAAGACAAATCAGAAGAAAAGTATAAAATCAATG  200

Myb              751 GATGTTGAAAATATTAATGATGCTATGGAATATAATAAAAATAAAGGTAA  800
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
INT1             201 GATGTTGAAAATATTAATGATGCTATGGAATATAATAAAAATAAAGGTAA  250

Myb              801 ACGACAACATCAACATAATAATAATGAAGAAGCAAATTTGTTATCTACCA  850
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
INT1             251 ACGACAACATCAACATAATAATAATGAAGAAGCAAATTTGTTATCTACCA  300

Myb              851 TAGAAAATTATGATAAACAATTTAACGAATTAAGTCATTTAAGAAAAAGG  900
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
INT1             301 TAGAAAATTATGATAAACAATTTAACGAATTAAGTCATTTAAGAAAAAGG  350

Myb              901 GTTCGATTGAATTTACCAGAACCTATATTAAATGAAAATGAAATAGATGA  950
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
INT1             351 GTTCGATTGAATTTACCAGAACCTATATTAAATGAAAATGAAATAGATGA  400

Myb              951 AATAATACAAATAAATAAAGAAGCATCAGCATTTAACGATATTATAAAAG  1000
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
INT1             401 AATAATACAAATAAATAAAGAAGCATCAGCATTTAACGATATTATAAAAG  450

Myb             1001 ATCAAAACGATAAATTACCAATAAATAATATTTTACCAAGTATTGAAAGC  1050
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
INT1             451 ATCAAAACGATAAATTACCAATAAATAATATTTTACCAAGTATTGAAAGC  500

Myb             1051 TCCTCATTTATATTAAACAATAAAGATAAATTTTCAAACTTGGAAACAGA  1100
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
INT1             501 TCCTCATTTATATTAAACAATAAAGATAAATTTTCAAACTTGGAAACAGA  550

Myb             1101 TTTTTATAATAATAATAATAATAAATCAATTGCTTTTTCTAGTAAACTAG  1150
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
INT1             551 TTTTTATAATAATAATAATAATAAATCAATTGCTTTTTCTAGTAAACTAG  600

Myb             1151 ATCTAAGTATACAACAAGCAGCAAAAAATATCATTTCTCGAAAAATGAAT  1200
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
INT1             601 ATCTAAGTATACAACAAGCAGCAAAAAATATCATTTCTCGAAAAATGAAT  650

Myb             1201 ATTCCATTTATAGGAATGAACAATGATTATAATGAAGAAGAATTCGAAAG  1250
                     |||||||||||||||||||||||||||||||||||||||
INT1             651 ATTCCATTTATAGGAATGAACAATGATTATAATGAAGAA               689
                                         ||||||||||||||||||||||||||||||
INT2               1                     ATGATTATAATGAAGAAGAATTCGAAAG    28

Myb             1251 AAAAAATAATATATTTCAAAAGAGTAATAAGAATGTTCCTGATGATTTAG  1300
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
INT2              29 AAAAAATAATATATTTCAAAAGAGTAATAAGAATGTTCCTGATGATTTAG   78

Myb             1301 AATATGATGACACGAACAATTATCATAATAATAATAATAATAATAATAAT  1350
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
INT2              79 AATATGATGACACGAACAATTATCATAATAATAATAATAATAATAATAAT   128

Myb             1351 AATACTTCTTTTAATTCGATCAAAAATTCCACAACCCTTTATAATCATTT  1400
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
INT2             129 AATACTTCTTTTAATTCGATCAAAAATTCCACAACCCTTTATAATCATTT   178

Myb             1401 GGATGTAGAAAAAAATATAAAAGACGTACAAGAAGAAATTGAAAAACAAA  1450
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
INT2             179 GGATGTAGAAAAAAATATAAAAGACGTACAAGAAGAAATTGAAAAACAAA   228

Myb             1451 AGAAATTAAATGAAGAAGATCAACAAAAGAATCAAAAAGAAAAGTTAACT  1500
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
INT2             229 AGAAATTAAATGAAGAAGATCAACAAAAGAATCAAAAAGAAAAGTTAACT   278

Myb             1501 ATTAAAAATAAAATTATTAGTGATATTAAAAGTTTTAAAAAAAATATTAG  1550
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
INT2             279 ATTAAAAATAAAATTATTAGTGATATTAAAAGTTTTAAAAAAAATATTAG   328

Myb             1551 TTTATATGCTCATTCTATTATTTCTTATAAAAATTTAAAAAATGATCAAT  1600
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
INT2             329 TTTATATGCTCATTCTATTATTTCTTATAAAAATTTAAAAAATGATCAAT   378
```

```
Myb          1601 CTATGATGGATAATCAAACTATACAAACAACTGAATATTATGATGATAAT 1650
                  ||||||||||||||||||||||||||||||||||||||||||||||||||
INT2          379 CTATGATGGATAATCAAACTATACAAACAACTGAATATTATGATGATAAT  428

Myb          1651 TATGAAGAGAAAATCGATCGAGCAAAATTACATATTAAAGCATCGTTAGC 1700
                  ||||||||||||||||||||||||||||||||||||||||||||||||||
INT2          429 TATGAAGAGAAAATCGATCGAGCAAAATTACATATTAAAGCATCGTTAGC  478

Myb          1701 CAATTTACCCCAAGAAACGAATCTTATAGAACTTCAATTAAATGAAGAAC 1750
                  ||||||||||||||||||||||||||||||||||||||||||||||||||
INT2          479 CAATTTACCCCAAGAAACGAATCTTATAGAACTTCAATTAAATGAAGAAC  528

Myb          1751 ATCCAGAATGTGATACGGACAATATAGAAAAAGATGAAATAGAAAAAGAT 1800
                  ||||||||||||||||||||||||||||||||||||||||||||||||||
INT2          529 ATCCAGAATGTGATACGGACAATATAGAAAAAGATGAAATAGAAAAAGAT  578

Myb          1801 ATACAAGATATTGAAAATGAAAAAAGAAAAAACGAAGAAAGAAAAGAAAA 1850
                  ||||||||||||||||||||||||||||||||||||||||||||||||||
INT2          579 ATACAAGATATTGAAAATGAAAAAAGAAAAAACGAAGAAAGAAAAGAAAA  628
                                      ||||||||||||||||||||||||||||
INT3            1                     AAGAAAAAACGAAGAAAGAAAAGAAAA   27

Myb          1851 AGAAAAATTTAATAAGCAAAATAAAATTATTAGATGGAATTTACCTAGAC 1900
                  ||||||||||||||||||||||||||||||||||||||||||||||||||
INT3           28 AGAAAAATTTAATAAGCAAAATAAAATTATTAGATGGAATTTACCTAGAC   77

Myb          1901 CATATTTCTTAGATAAAATTAATCTCTTTAATAATTATATGCACAATGAA 1950
                  ||||||||||||||||||||||||||||||||||||||||||||||||||
INT3           78 CATATTTCTTAGATAAAATTAATCTCTTTAATAATTATATGCACAATGAA  127

Myb          1951 TATGAAGATGTCCATAATTTAATACAAAGAGAGATGCTCCTCTTAATAAA 2000
                  ||||||||||||||||||||||||||||||||||||||||||||||||||
INT3          128 TATGAAGATGTCCATAATTTAATACAAAGAGAGATGCTCCTCTTAATAAA  177

Myb          2001 AAATGATATGTTTAATTATCCCCTCAGAAATTCAACACCAGTACAAAATA 2050
                  ||||||||||||||||||||||||||||||||||||||||||||||||||
INT3          178 AAATGATATGTTTAATTATCCCCTCAGAAATTCAACACCAGTACAAAATA  227

Myb          2051 AAGTTCATGTGGAAGATCTCGAAAACGTATACATGAATATGGCTATGAAG 2100
                  ||||||||||||||||||||||||||||||||||||||||||||||||||
Int3          228 AAGTTCATGTGGAAGATCTCGAAAACGTATACATGAATATGGCTATGAAG  277

Myb          2101 AGTATAAATGAGGAGTTCGAAGATATGTATGGCGAAGCATCATTAAATAA 2150
                  ||||||||||||||||||||||||||||||||||||||||||||||||||
INT3          278 AGTATAAATGAGGAGTTCGAAGATATGTATGGCGAAGCATCATTAAATAA  327

Myb          2151 CAACACAAAGGATGATTCTAATATTGATGGGTGTGATGAAAAGAGTGATA 2200
                  ||||||||||||||||||||||||||||||||||||||||||||||||||
INT3          328 CAACACAAAGGATGATTCTAATATTGATGGGTGTGATGAAAAGAGTGATA  377

Myb          2201 ATATTGATGGTTGTGATGAAAAAAGTGATACAACTAATAAAGGTGATGAT 2250
                  ||||||||||||||||||||||||||||||||||||||||||||||||||
Int3          378 ATATTGATGGTTGTGATGAAAAAAGTGATACAACTAATAAAGGTGATGAT  427

Myb          2251 ACGTCCCAATGTAGTATCGAGCATAGTTCATATAATCATATCGATGTGTG 2300
                  ||||||||||||||||||||||||||||||||||||||||||||||||||
INT3          428 ACGTCCCAATGTAGTATCGAGCATAGTTCATATAATCATATCGATGTGTG  477

Myb          2301 GGAAGAAATTAACAAAAATATAATTTTTTGCCCATCAAAAAATGCATATC 2350
                  ||||||||||||||||||||||||||||||||||||||||||||||||||
INT3          478 GGAAGAAATTAACAAAAATATAATTTTTTGCCCATCAAAAAATGCATATC  527

Myb          2351 GCTTTATTGAAGATGTTAATGAAAACGATAAAAAAGAAAATTACAAATAC 2400
                  ||||||||||||||||||||||||||||||||||||||||||||||||||
INT3          528 GCTTTATTGAAGATGTTAATGAAAACGATAAAAAAGAAAATTACAAATAC  577

Myb          2401 AAATGTGAAAAATTAAAAAATTTAATTTTAAATGATATGGAACATTATAA 2450
                  ||||||||||||||||||||||||||||||||||||||||||||||||||
INT3          578 AAATGTGAAAAATTAAAAAATTTAATTTTAAATGATATGGAACATTATAA  627

Myb          2451 AAAATTAGAAAATAAATATGATATTTATACAAAAGGATATCAACTAAAAA 2500
                  ||||||||||||||||||||||||||||||||||||||||||||||||||
INT3          628 AAAATTAGAAAATAAATATGATATTTATACAAAAGGATATCAACTAAAAA  677
```

```
Myb          2501 TTAAAAGTTATAAAAAATCATATGATACTCTATTTAATTCATATATTAAT   2550
                  ||||||||||||||||||||||||||||||||||||||||||||||||||
INT3          678 TTAAAAGTTATAAAAAATCATATGATACTCTATTTAATTCATATATTAAT    727
                                                 ||||||||||||||||||||
3'Pgex          1                                ATTTAATTCATATATTAAT     19

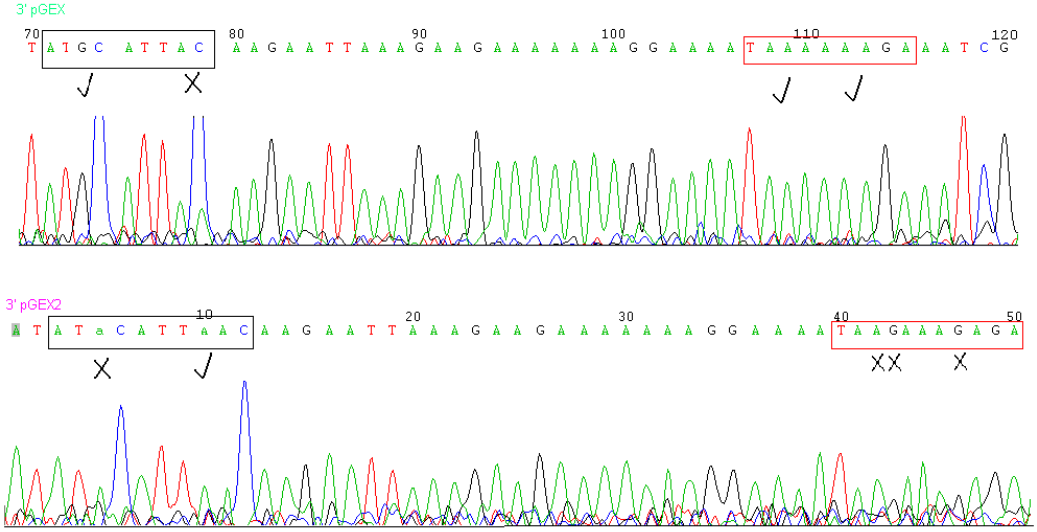
Myb          2551 TGTATTAACGAAAAAGAAGCTTTAAATGTTCTACATGAAAATGAAAAAAT   2600
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||
3'Pgex         20 TGTATTAACGAAAAAGAAGCTTTAAATGTTCTACATGAAAATGAAAAAAT     69

Myb          2601 ATATGCATTAACAAGAATTAAAGAAGAAAAAAAGGAAAATA--AAAAAGA   2648
                  | |||||||  ||||||||||||||||||||||||||||||  |||||||
3'Pgex         70 ATATGCATT-ACAAGAATTAAAGAAGAAAAAAAGGAAAATA--AAAAAGA    116
                  | ||.||||||||||||||||||||||||||||||||||  |||.|||
3'Pgex2         1 ATATaCATTAACAAGAATTAAAGAAGAAAAAAAGGAAAATAAGAAAGAGA     50

Myb          2649 AATCGAATATCATAAATCATTACAAAAGTTTTATCAAGACCTTTTAGAAA   2698
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||
3'Pgex2        51 AATCGAATATCATAAATCATTACAAAAGTTTTATCAAGACCTTTTAGAaA    100

Myb          2699 CAAATCATCAATTAAAAGAAACATGTAAACAAACGTTAAAGGTGCCATAA   2748
                  ||||||||||||||||||||||||||||||||||||||||||||||||...
3'Pgex2       101 CAAATCATCAATTAAAAGAAACATGTAAACAAACGTTAAAGGTGCCACTC    150
```

Two sequencing reactions were carried out with the 3' pGEX-4T-2 primer as there appeared to be a base change and insertion at positions 74 and 79 of the pGEX sequence respectively (in black box). When compared to the other sequencing reaction (pGEX2) using the same primer this substitution was not noted but there were other base insertions and changes downstream of this second sequence (in red box). In the figure below, the ticks and crosses represent what is correct and incorrect respectively. If one of either of the sequencing reactions had the correct base the sequence was considered correct.



7. REFERENCES

Ajuh, P., Sleeman, J., Chusainow, J. and Lamond, A.I. (2001) *A Direct Interaction between the carboxyl-terminal region of CDC5L and the WD40 domain of PLRG1 is essential for Pre-mRNA Splicing*. Journal of Biological Chemistry, **276,** 42370–42381.

Angermayr, M. and Bandlow, W. (1997) *The general regulatory factor Reb1p controls basal, but not Gal4p-mediated, transcription of the GCY1 gene in yeast*. Molecular and General Genetics, **256,** 682-689.

Aravind, L., Iyer, L.M., Wellems, T.E. and Miller, L.H. (2003) *Plasmodium biology: Genomic gleanings*. Cell, **115,** 771-785.

Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J. G., Smith, J.A. and Struhl, K. (1995) *Overview of the baculovirus expression system*. Current Protocols in Molecular Biology, **Unit 5.4,** 5.4.1-5.4.2.

Bernstein, H.S. and Coughlin, S.R. (1997) *Pombe Cdc5-related Protein: A putative human transcription factor implicated in mitogen-activated signalling*. Journal of Biological Chemistry, **272,** 5833–5837.

Bernstein, H.S. and Coughlin, S.R. (1998) *A mammalian homolog of fission yeast cdc5 regulates G2 progression and mitotic entry*. Journal of Biological Chemistry, **273,** 4666–4671.

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,** 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,** E5.

Braun, P., Hu, Y., Shen, B., Halleck, A., Koundinya, M., Harlow, E. and and LaBaer, J. (2002) *Proteome-scale purification of human proteins from bacteria*. Proceedings of the National Academy of Sciences USA. **99,** 2654-2659.

Brocchieri, L. (2001) *Low-complexity regions in Plasmodium proteins: In search of a function*. Genome Research, **11,** 195-197.

Burgess-Brown, N.A., Sharma, S., Sobott, F., Loenarz, C., Oppermann, U., and Gileadi, O. (2008) *Codon optimization can improve expression of human genes in Escherichia coli: A multi-gene study*. Protein Expression and Purification, **59,** 94–102.

Burns, C.G., Ohi, R., Krainer, A.R. and Gould, K.L. (1999) *Evidence that Myb-related CDC5 proteins are required for pre-mRNA splicing*. Proceedings of the National Academy of Sciences USA, **96,** 13789-13794.

Callebaut, I., Prat, K., Meurice, E., Mornon, J.P. and Tomavo, S. (2005) *Prediction of the general transcription factors associated with RNA polymerase II in Plasmodium falciparum: Conserved features and differences relative to other eukaryotes*. BMC Genomics, **6,** 100.

Chelur, D., Unal, O., Scholtyssek, M. and Strickler, J. (2008) *Fusion tags can improve the yield and solubility of many recombinant proteins*. BioPharm International Supplements.

Cole, P.A. (1996) *Chaperone-assisted protein expression*. Structure, **4,** 239-242.

Coulson, R.M., Hall, N. and Ouzounis, C.A. (2004) *Comparative genomics of transcriptional control in the human malaria parasite Plasmodium falciparum*. Genome Research, **14,** 1548-1554.

De Koning-Ward, T.F., Janse, C.J. and Waters, A.P. (2000) *The development of genetic tools for dissecting the biology of malaria parasites*. Annual Review of Microbiology, **54,** 157-185.

Dechering, K.J., Kaan, A.M., Mbacham, W., Wirth, D.F., Eling, W., Konings, R.N. and Stunnenberg, H.G. (1999) *Isolation and functional characterization of two distinct sexual-stage-specific promoters of the human malaria parasite Plasmodium falciparum*. Molecular Cell Biology, **19,** 967-978.

Deitsch, K.W. (2004) *Gene Expression in* Malaria Parasites: Genomes and molecular biology, pg 205-227. Ed: Waters, A.P. and Janse, C.J. Caister Academic Press, UK.

Doerig, C., Horrocks, P., Coyle, J., Carlton, J., Sultan, A., Arnot, D. and Carter, R. (1995) *Pfcrk-1, a developmentally regulated cdc2-related protein kinase of Plasmodium falciparum*. Molecular Biochemical Parasitology, **70,** 167-174.

Doerig, C.M., Parzy, D., Langsley, G., Horrocks, P., Carter, R. and Doerig, C.D. (1996) *A MAP kinase homologue from the human malaria parasite, Plasmodium falciparum*. Gene, **177,** 1-6.

Florens, L., Liu, X., Wang, Y., Yang, S., Schwartz, O., Peglar, M., Carucci, D. J., Yates, J. R., 3rd and Wub, Y. (2004) *Proteomics approach reveals novel proteins on the surface of malaria-infected erythrocytes*. Molecular Biochemical Parasitology, **135,** 1-11.

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,** 520-526.

Florent, I., Charneau, S. and Grellier, P. (2004) *Plasmodium falciparum genes differentially expressed during merozoite morphogenesis*. Molecular Biochemical Parasitology, **135,** 143-148.

Fourel, G., Miyake, T., Defossez, P.A., Li, R. and Gilson, E. (2002) *General regulatory factors (GRFs) as genome partitioners*. Journal of Biological Chemistry, **277,** 41736–41743.

Fox, B.A., Li, W.B., Tanaka, M., Inselburg, J. and Bzik, D.J. (1993) *Molecular characterization of the largest subunit of Plasmodium falciparum RNA polymerase I*. Molecular Biochemical Parasitology, **61,** 37-48.

Freitas-Junior, L.H., Hernandez-Rivas, R., Ralph, S.A., Montiel-Condado, D., Ruvalcaba-Salazar, O.K., Rojas-Meza, A.P., Mancio-Silva, L., Leal-Silvestre, R.J., Gontijo, A.M., Shorte, S. and Scherf, A. (2005) *Telomeric heterochromatin propagation and histone acetylation control mutually exclusive expression of antigenic variation genes in malaria parasites*. Cell, **121,** 25-36.

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.

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,** 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.

Greenwood, B.M., Bojang, K., Whitty, C.J. and Targett, G.A. (2005) *Malaria*. Lancet, **365,** 1487-1498.

Gunasekera, A.M., Myrick, A., Militello, K.T., Sims, J.S., Dong, C.K., Gierahn, T., Le Roch, K., Winzeler, E. and Wirth, D.F. (2007) *Regulatory motifs uncovered among gene expression clusters in Plasmodium falciparum*. Molecular Biochemical Parasitology, **153,** 19-30.

Hammarstrom, M., Hellgren, N., van Den Berg, S., Berglund, H. and Hard, T. (2002) *Rapid screening for improved solubility of small human proteins produced as fusion proteins in Escherichia coli*. Protein Science, **11,** 313-321.

Horrocks, P., Dechering, K. and Lanzer, M. (1998) *Control of gene expression in Plasmodium falciparum*. Molecular Biochemical Parasitology, **95,** 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 Biochemical Parasitology, **82,** 207-215.

Horrocks, P. and Lanzer, M. (1999) *Mutational analysis identifies a five base pair cis-acting sequence essential for GBP130 promoter activity in Plasmodium falciparum*. Molecular Biochemical Parasitology, **99,** 77-87.

Howe, K.M., Reakes, C.F. and Watson, R.J. (1990) *Characterization of the sequence-specific interaction of mouse c-myb protein with DNA*. European Molecular Biology Organization Journal, **9,** 161-169.

Johnson, W.C. (1990) *Protein secondary structure and circular dichroism: A practical guide*. Proteins, **7**, 205-214.

Ju, Q.D., Morrow, B.E. and Warner, J.R. (1990) *REB1, a yeast DNA-binding protein with many targets, is essential for growth and bears some resemblance to the oncogene myb*. Molecular Cell Biology, **10,** 5226-5234.

Konig, P., Fairall, L. and Rhodes, D. (1998) *Sequence-specific DNA recognition by the myb-like domain of the human telomere binding protein TRF1: A model for the protein-DNA complex*. Nucleic Acids Research, **26,** 1731-1740.

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 of Parasitology, **34,** 1451-1458.

Laemmli, U.K. (1970) *Cleavage of structural proteins during 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,** 418-420.

Lang, W.H., Morrow, B.E., Ju, Q., Warner, J.R. and Reeder, R.H. (1994) *A model for transcription termination by RNA polymerase I*. Cell, **79,** 527-534.

Lang, W.H. and Reeder, R.H. (1993) *The REB1 site is an essential component of a terminator for RNA polymerase I in Saccharomyces cerevisiae*. Molecular Cell Biology, **13,** 649-658.

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.

Lei, X.H., Shen, X., Xu, X.Q. and Bernstein, H.S. (2000) *Human Cdc5, a regulator of mitotic entry, can act as a site-specific DNA binding protein*. Journal of Cell Science, **113 Pt 24,** 4523-4531.

Lipsick, J.S. (1996) *One billion years of Myb*. Oncogene, **13,** 223-235.

Ljungstrom, I., Perlmann, H., Schlichtherle, M., Scherf, A. and Wahlgren, M. (2004) *Methods in malaria research; 4th Edition*. MR4/ATCC Virginia.

Llinas, M., Bozdech, Z., Wong, E.D., Adai, A.T. and DeRisi, J.L. (2006) *Comparative whole genome transcriptome analysis of three Plasmodium falciparum strains*. Nucleic Acids Research, **34,** 1166-1173.

Lobo, C.A. and Kumar, N. (1998) *Sexual differentiation and development in the malaria parasite*. Parasitology Today, **14,** 146-150.

Lodish, H., Berk, A., Zipursky, L.S., Matsudaira, P., Baltimore, D. and Darnell, J. (2000) *Molecular Cell Biology; 4th edition.* W. H. Freeman.

Lopez-Estrano, C., Gopalakrishnan, A.M., Semblat, J.P., Fergus, M.R., Mazier, D. and Haldar, K. (2007) *An enhancer-like region regulates hrp3 promoter stage-specific gene expression in the human malaria parasite Plasmodium falciparum*. Biochimica et Biophysica Acta, **1769,** 506-513.

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.

Mehlin, C., Boni, E., Buckner, F.S., Engel, L., Feist, T., Gelb, M.H., Haji, L., Kim, D., Liu, C., Mueller, N., Myler, P.J., Reddy, J.T., Sampson, J.N., Subramanian, E., Van Voorhis, W.C., Worthey, E., Zucker, F. and Hol, W.G. (2006) *Heterologous expression of proteins from Plasmodium falciparum: Results from 1000 genes*. Molecular Biochemical Parasitology, **148,** 144-160.

Meyersfeld, D. (2005) *Aspects of gene expression and regulation in Plasmodium falciparum gametocytogenesis*. PhD Thesis**,** University of the Witwatersrand.

Meyersfeld, D., Baker, G.L., Vaquero, C. and Coetzer, T.L. (2008) *Manuscript in preperation*. Open Parasitology**,** Still to be published.

Militello, K.T., Dodge, M., Bethke, L. and Wirth, D.F. (2004) *Identification of regulatory elements in the Plasmodium falciparum genome*. Molecualar Biochemical Parasitology, **134,** 75-88.

Miller, L.H., Baruch, D.I., Marsh, K. and Doumbo, O.K. (2002) *The pathogenic basis of malaria*. Nature, **415,** 673-679.

Nallamsetty, S. and Waugh, D.S. (2006) *Solubility-enhancing proteins MBP and NusA play a passive role in the folding of their fusion partners*. Protein Expression and Purification, **45,** 175-182.

Oh, I.H. and Reddy, E.P. (1999) *The myb gene family in cell growth, differentiation and apoptosis*. Oncogene, **18,** 3017-3033.

Olivieri, A., Silvestrini, F., Sanchez, M. and Alano, P. (2008) *A 140-bp AT-rich sequence mediates positive and negative transcriptional control of a Plasmodium falciparum developmentally regulated promoter*. International Journal Parasitology, **38,** 299-312.

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,** 87-98.

Paule, M.R. and White, R.J. (2000) *Survey and summary: transcription by RNA polymerases I and III*. Nucleic Acids Research, **28,** 1283-1298.

Raisner, R.M., Hartley, P.D., Meneghini, M.D., Bao, M.Z., Liu, C.L., Schreiber, S.L., Rando, O.J. and Madhani, H.D. (2005) *Histone variant H2A.Z marks the 5' ends of both active and inactive genes in euchromatin*. Cell, **123,** 233-248.

Robson, K.J. 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., del Carmen Ramirez-Estudillo, M., 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,** 183-196.

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*. European Molecular Biology Organization Journal, **17,** 5418-5426.

Singh, S.M. and Panda, A.K. (2005) *Solubilization and refolding of bacterial inclusion body proteins*. Journal of Bioscience and Bioengineering, **99,** 303-310.

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.

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,** 225-230.

Tham, W.H., Payne, P.D., Brown, G.V. and Rogerson, S.J. (2007) *Identification of basic transcriptional elements required for rif gene transcription*. International Journal of Parasitology, **37,** 605-615.

Vallejo, L.F. and Rinas, U. (2004) *Strategies for the recovery of active proteins through refolding of bacterial inclusion body proteins*. Microbial Cell Factories, **3**, 11.

Van Noort, V. and Huynen, M.A. (2006) *Combinatorial gene regulation in Plasmodium falciparum*. Trends in Genetics, **22,** 73-78.

Van Slyke, C. and Grayhack, E.J. (2003) *The essential transcription factor Reb1p interacts with the CLB2 UAS outside of the G2/M control region*. Nucleic Acids Research, **31,** 4597-4607.

Vedadi, M., Lew, J., Artz, J., Amani, M., Zhao, Y., Dong, A., Wasney, G.A., Gao, M., Hills, T., Brokx, S., Qiu, W., Sharma, S., Diassiti, A., Alam, Z., Melone, M., Mulichak, A., Wernimont, A., Bray, J., Loppnau, P., Plotnikova, O., Newberry, K., Sundararajan, E., Houston, S., Walker, J., Tempel, W., Bochkarev, A., Kozieradzki, I., Edwards, A., Arrowsmith, C., Roos, D., Kain, K. and Hui, R. (2007) *Genome-scale protein expression and structural biology of Plasmodium falciparum and related Apicomplexan organisms*. Molecular and Biochemical Parasitology, **151,** 100-110.

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,** 147-156.

Wu, Y., Sifri C.D., Lei, H.H., Su, X.Z. and Wellems, T.E. (1995) *Transfection of Plasmodium falciparum within human red blood cells*. Proceedings of the National Academy of Sciences USA, **92,** 973-977.

Yadava, A. and Ockenhouse C.F. (2003) *Effect of codon optimization on expression levels of a functionally folded malaria vaccine candidate in prokaryotic and eukaryotic expression systems*. Infection and Immunity, **71,** 4961-4969.

Young, J.A., Johnson, J.R., Benner, C., Yan, S.F., Chen, K., Le Roch, K.G., Zhou, Y. and Winzeler, E.A. (2008) *In silico discovery of transcription regulatory elements in Plasmodium falciparum*. BMC Genomics, **9,** 70.

Yu, Q., Qiu, R., Foland, T.B., Griesen, D., Galloway, C.S., Chiu, Y.H., Sandmeier, J., Broach, J.R. and Bi, X. (2003) *Rap1p and other transcriptional regulators can function in defining distinct domains of gene expression*. Nucleic Acids Research, **31,** 1224-1233.

Zargarian, L., Le Tilly, V., Jamin, N., Chaffotte, A., Gabrielsen, O. S., Toma, F. and Alpert, B. (1999) *Myb-DNA recognition: Role of tryptophan residues and structural changes of the minimal DNA binding domain of c-Myb*. Biochemistry, **38,** 1921-1929.

Zhou, Z., Schnake, P., Xiao, L. and and Lal, A.A. (2004) *Enhanced expression of a recombinant malaria candidate vaccine in Escherichia coli by codon optimization.* Protein Expression and Purification, **34,** 87-94.