An analysis of non-coding RNAs in *Plasmodium falciparum* and their potential role in antigenic variation.

Zóe Christodoulou

Thesis submitted for the degree of Doctor of Philosophy.

Michaelmas Term 2012

St. Catherine's College,

University of Oxford.

TABLE OF CONTENTS

Declaration

The majority of work in this thesis is my own. However, for the long term switching experiment in Chapter 6, maintaining 12 *P. falciparum* cultures, daily for 180 days was a communal effort between Bob Pinches, Sue Kyes and me. Sue Kyes also prepared some of the RNA from these cultures. The cDNA, qRT-PCR and data analysis was my own work. Bob Pinches supplied parasites for me to make the protein extracts from in chapter 8. And from time to time, when the number of plasmid preps waiting to be done was overwhelming, Bob assisted me.

Apart from the instances stated above, I declare all other work in this thesis to be my own.

Acknowledgements

My sincerest thanks go to Bob Pinches and Sue Kyes without whom I would never have got this far. To Bob, for teaching me parasite culturing after I had managed to avoid it for 13 years; for supplying parasites, thus enabling me to get on with other areas; and for being a willing pair of hands when confronted with numerous wizard preps. To Sue, for always being there for advice on all things molecular biological; for proofreading this thesis and giving comments; and for never (well rarely) complaining that she had to do her own experiments as her RA was busy. I would like to take this opportunity to thank both of them for their support, encouragement and friendship over the past years.

I have other people to thank; my supervisor Chris Newbold for allowing me to indulge myself in studying for a D. phil, whilst continuing to keep me gainfully employed; my college advisor Kate Carr for her support, her refusal to let me give up, and her endless supply of Kleenex; Rob Noble for helpful discussions and advice on qRT-PCR data analysis; Benedikt Kessler and Joanna M^{c}Gouran for their advice on proteomics experiments. I would also like to thank all the students and colleagues over the years who have told me that I could do a D. phil, maybe you were right. Finally I would like to thank Mark, Ross and Alex for their love and support, and never complaining that I wasn’t around.

Abbreviations

A	Amplification value
AGO	Argonaute
ARMS	Amplification Refractory Mutation System
asRNA	Antisense RNA
B+W	Binding and washing buffer
BLAST	Basic local alignment search tool
bp	base pairs
CAT	Chloramphenicol acetyltransferase
CCLR	Cell culture lysis reagent
cDNA	complementary DNA
CDS	Coding sequence
CF	Correction factor
CIDR	Cysteine-rich interdomain region
CSA	Chondroitin sulphate A
DBL	Duffy binding like domain
DC	Domain cassette
dhfr-ts	dihydrofolate reductase-thymidylate synthase
dNTP	Deoxyribonucleotide triphosphate
dsDNA	Double stranded DNA
dsRBD	double-stranded RNA binding domain
EF1α	Elongation factor 1-α
EMSA	Electrophoretic mobility shift assay
ETC	Extra TFIIIC sites
gDNA	genomic DNA
GRE	GC rich element
H3k4me2	Di methylation at lysine 4 of histone 3
H3k4me3	Trimethylation at lysine 4 of histone 3
H3k9	Lysine 9 of histone 3
H3k9ac	Acetylation at lysine 9 of histone 3
H3k9me3	Trimethylation at lysine 9 of histone 3
HRM	High resolution melt
ICAM-1	Intercellular adhesion molecule-1
IE	Infected erythrocytes
IRES	Internal ribosome entry site
KAHRP	Knob assciated histidine rich protein
KH	K homology domain
lncRNA	Long non-coding RNA
MESA	mature-parasite-infected erythrocyte surface antigen
MIQE	Minimum information for publication of quantitative real-time PCR experiments
miRNA	micro RNA

NAI	Naturally acquired imunity
ncRNA	non-coding RNA
nt	nucleotide
PAGE	Poly acrylamide gel electrophoresis
PAM	Pregnancy associated malaria
PAP	Poly(A) polymerase
PASA	PCR Amplification of Specific Alleles
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PfEMP1	*P. falciparum* erythrocyte membrane protein-1
PFGE	Pulsed-field gel electrophoresis
PfHP1	*P. falciparum* heterochromatin protein 1
PNK	Poly nucleotide kinase
POL lll	RNA polymerase lll
qRT-PCR	Quantitative reverse-transcription polymerase chain reaction
RACE	Rapid amplification of cDNA ends
RBC	Red blood cells
RBD	RNA binding domain
RBP	RNA binding protein
RISC	RNA-induced silencing complex
RNA-FISH	RNA-fluorescent in situ hybridization
RNAi	RNA interference
RNP	Ribonucleoprotein
RRM	RNA recognition motif
RT-PCR	Reverse-transcription polymerase chain reaction
SAP	Shrimp alkaline phosphatase
SDS	Sodium dodecyl sulphate
siRNA	small interfering RNA
snoRNA	small nucleolar RNA
SNP	Single nucleotide polymorphism
snRNA	small nuclear RNA
SRP	Signal recognition particle
SSC	Salt and sodium citrate buffer
ssDNA	Single stranded DNA
T.O.	Take off
TAPS	Transcription Associated Proteins
TARES	Telomere associated repeat elements
TBE	Tris, borate EDTA buffer
TE	Tris, EDTA
TSP	Thrombospondin
Ups	5' upstream sequence
X-gal	5-bromo-4-chloro-indolyl-β-D-galactopyranoside

Abstract

A major virulence factor of the human malaria parasite *Plasmodium falciparum* is *Plasmodium falciparum* erythrocyte membrane protein 1(PfEMP-1). This protein is inserted into the erythrocyte membrane, giving cytoadherence properties. A family of genes called *var*, located sub-telomerically and in chromosome central clusters encode this protein. *Var* genes are expressed in a mutually exclusive manner, how this is controlled is unclear. A non-coding RNA (ncRNA) termed the GC-rich element (GRE) had been identified that is only located at the central clusters and is transcribed throughout the parasite lifecycle. A screen of the *P. falciparum* genome for novel ncRNAs identified ncRNAs from known classes. Novel transcripts were identified, but none in the proximity of *var* genes.

We have investigated the role of the GRE in *var* gene regulation. A set of qRT-PCR primers have been designed and tested to follow *var* gene expression in the HB3 isolate, these are not cross-reactive with a published set for the 3D7 isolate. Alterations were made to the 3D7 set to remove cross-reactivity with HB3. *Var* gene expression was studied in 31 HB3 clones and progeny of the 3D7xHB3 genetic cross. Following *var* switching over five months in eleven HB3 clones showed that all of the clones ended up expressing *var* genes from the same central cluster on chromosome 4. GRE Transcription in these clones is linked to a specific class of *var* gene. Transcription from a single GRE locus occurs only when

a *var* gene of the central UpsC class is expressed from the same cluster. Expression of other classes of *var* gene gives multiple transcripts from different GRE loci.

Investigations into the *in vitro* binding properties of the GRE revealed an RNA:protein complex that can be resolved by electrophoresis. Proteomic analysis of the complex revealed predominantly ribosome proteins and translation factors.

Chapter 1

Introduction

Disease burden

The apicomplexan parasite *Plasmodium falciparum* is the causative agent of the most severe, and often fatal, form of malaria. Globally, in 2010, 655,000 deaths were due to the disease (1). Although there has been a 25% reduction in deaths from the disease over the past decade, it remains a major killer (1). This reduction, led by the efforts of the World Health Organization (WHO) is due to several factors including increases in use of bednets, access to accurate diagnosis and financial resources, estimated at US$ 2 billion in 2010 (1). In 2010, 99 countries reported malaria, but its effects are felt most severely in Africa, where 91% of worldwide deaths due to the disease occur (1). Malaria accounts for 22% of childhood deaths in Africa (2). Children under the age of 5 years are most at risk as immunity to infection is acquired with age and increased exposure to infection. Pregnant women and people infected with HIV also have increased risk of contracting the disease. Although preventable by medication, tourists returning from malarious areas are also at risk. This is due to several factors, but primarily, as tourists are immunologically naïve to the parasite, any infectious bite is likely to cause serious disease.

Secondly, the prepatent stage, during which patients remain asymptomatic, is 9 days so they can return home feeling fine. At this point patients assume they have escaped uninfected and stop taking their medication, only for malaria to occur sometime later. This is a major problem, as annually approximately 1,500 travellers return to the UK with malaria and in 2008 there were 6 deaths (3).

Plasmodium biology

Plasmodium species belong to the phylum Apicomplexa. Apicomplexa refers to the apical complex, an organelle present in all members of this entirely parasitic phylum. Apicomplexa are the causative agents of many human and animal diseases including toxoplasmosis, cryptosporidiosis and babesiosis. Members of the genus *Plasmodium* infect birds, reptiles, primates and other mammals. There are four species of malaria that infect humans: *P. malariae*, *P. ovale*, *P. vivax* and *P. falciparum*. Increasingly, infection rates of a fifth species, *P. knowlesi*, are rising (4). The rise is likely due to mis-diagnosis previously, as microscopically it resembles *P. malariae*, but is now distinguishable by PCR methods (4,5). As a parasite of long tailed macaques, *P. knowlesi* is a zoonosis, but there are calls for it to be classed as the fifth species that infect humans (6,7). Murine malarias, especially *P. berghei*, *P. yoellii* and *P. chabaudi*, have been utilised by the scientific community to study the biology of the parasite.

Life cycle

Plasmodium species have a two-host life cycle (figure 1.1(8)). The diploid sexual stages occur in the *Anopheles* mosquito. Human infection follows a bite from an infected mosquito (1), when the haploid sporozoites are

released into the blood from the mosquito salivary gland. Sporozoites invade hepatocytes (2) where they undergo primary schizogeny (3). From this approximately 30,000 merozoites are released into the bloodstream (4), where they invade erythrocytes (5) and enter an asexual cycle of development, through ring, trophozoite and schizont forms. Schizonts rupture after 48 hours releasing an average of 24 merozoites to continue the cycle (6). This erythrocytic cycle continues until triggers, thought to be stress related, cause development of gametocytes (7). These are the haploid sexual forms, and following ingestion by another mosquito (8), will go on to complete the sexual stage of the lifecycle.

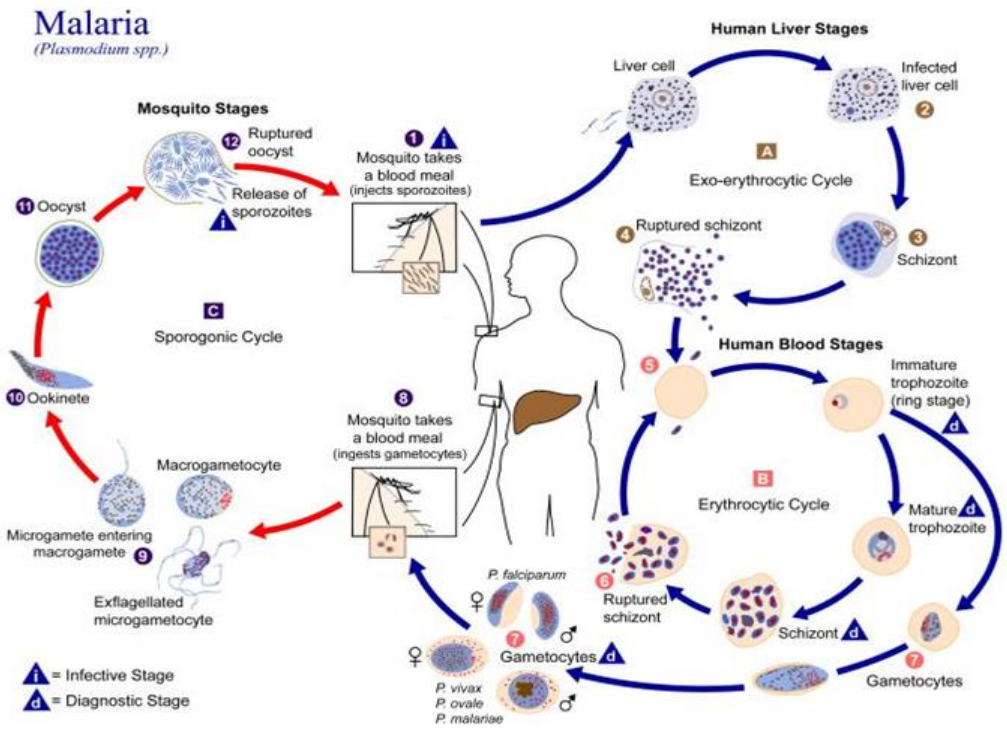


Figure 1.1 Malaria Lifecycle. A description of the stages indicated numerically is given in the text. Reproduced from (8).

The sexual stage begins in the mosquito gut. Here, female gametocytes mature to macrogametes, and males to microgametes (9). Fertilization of a macrogamete by a microgamete leads to a diploid zygote,

which develops into an ookinete (10). The ookinete is motile and passes from the mosquito stomach into the midgut. Once in the midgut, the ookinete forms an oocyst (11), in which the nuclei divide, forming motile haploid sporozoites. Eventually the increasing size of the ookinete causes it to rupture, releasing the sporozoites (12). The sporozoites move to the salivary glands, where they are ready to infect a new host during a blood meal.

This thesis focuses on the asexual, disease causing, stages of *P. falciparum,* images of which are in figure 1.2. It concentrates on a family of genes coding for known virulence factors, and the resulting proteins.

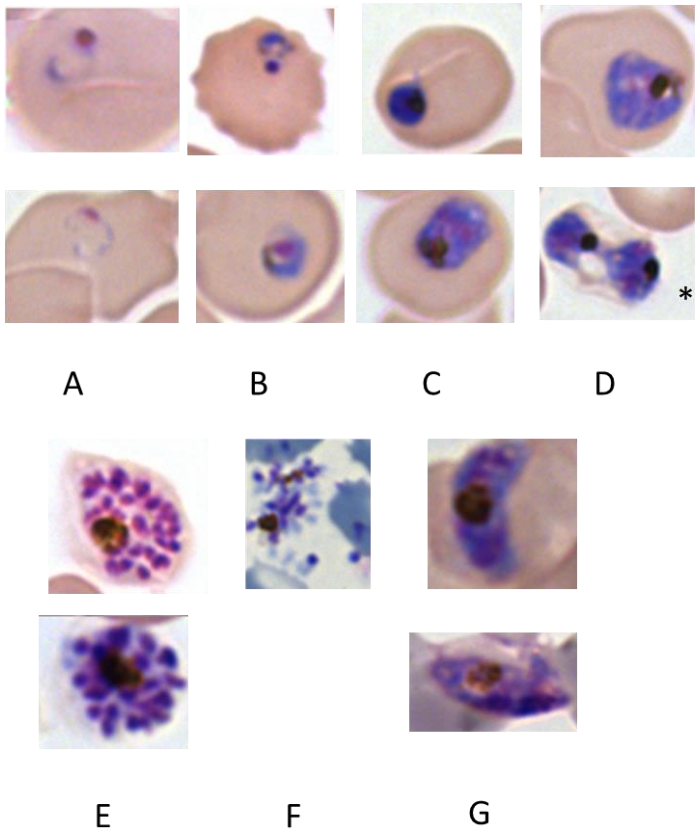


Figure 1.2 Intraerythrocytic stages of *P. falciparum*. Giemsa stained thin blood smears, showing the following stages; A, ring forms; B-D, trophozoites of increasing maturity. (* multiply infected); E, schizont; F, rupturing schizont releasing merozoites; G, gametocytes.

Malaria pathology

Traditionally, malaria is recognised by recurrent episodes of fever with asymptomatic intervals. This is not the case however in early stages of *P. falciparum* malaria. Early symptoms are non-specific and include dizziness, nausea, headache and fever. If untreated, the disease progresses rapidly in non-immune patients and death has been reported as early as 24 hours from first symptoms (9).

There are several conditions attributed to *P*. *falciparum* infection, although cerebral malaria is the most severe. Symptoms of malaria infection vary widely, and differ between adults and children, and depending on the level of naturally acquired immunity. In endemic areas, infants under 6 months rarely develop malaria. A recent study suggests this is due to a combination of foetal haemoglobin impairing binding of parasitized erythrocytes to host cells, and maternal antibodies (10). As this passive immunity is lost, they are most at risk from severe malaria infection, until they have developed their own naturally acquired immunity (NAI). In non-immune children infection from *P. falciparum* may rapidly progress to severe disease, although they may have low levels of parasitaemia. Children initially present with a high fever, often history of a cough, and coma, following convulsions (11). Metabolic acidosis presenting as respiratory distress has also been shown to be a likely predictor of death (12) and hypoglycaemia is common. As NAI develops as a result of successive infections, the risk of severe disease reduces. At this stage they may have high parasite levels, and anaemia is the biggest disease risk (13). Non-immune adults, like non-immune children, if left

untreated may rapidly progress to severe disease. Many symptoms are shared between adults and children, but renal failure and jaundice are common in adults, not children (11). Respiratory distress and hypoglycaemia are seen less often in adults than children (11).

Sequestration

One contributing factor to the pathology of severe malaria is the sequestration of infected erythrocytes (IE) in the microvasculature. Post-mortem samples show capillaries congested with IE in many tissue types (figure 1.3 (14)). The effect on multiple organs is due to the blockage of blood vessels in specific tissues, leading to anaerobic metabolism (15), competition of the parasites for substrates and production of toxic products by the parasites (16). Sequestration is of benefit to the parasite, as whilst the IE are residing in organs, and not circulating, they avoid clearance by the spleen.

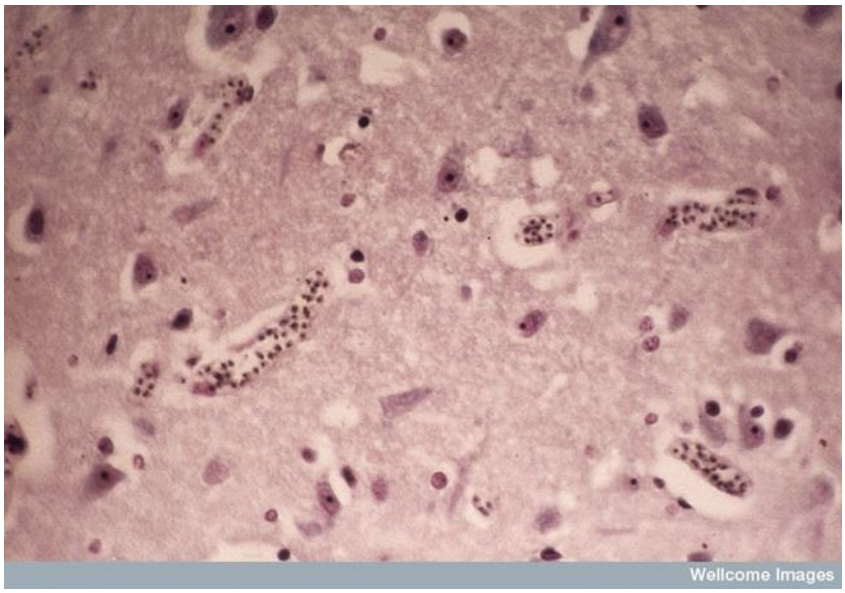


Figure 1.3 Sequestration of *P. falciparum* in blood vessels. Photomicrograph of cerebral malaria, showing parasites sequestered in blood vessels. From Wellcome Images (14).

Cytoadherence

Sequestration occurs as a result of interactions between endothelial surface molecules and parasite derived proteins inserted into the IE membrane. Electron microscopy reveals the connections to be at knob like protrusions on the IE surface (17). The knobs contain the parasite protein *Plasmodium falciparum* erythrocyte membrane protein 1 (PfEMP-1) (18), the major variant surface antigen of *P. falciparum*. PfEMP-1 is encoded by the multicopy *var* gene family (19-21). Whilst PfEMP-1 is expressed on the surface of the knobs, underneath the major component is knob associated histidine rich protein (KAHRP) (22). Other components are include mature-parasite-infected erythrocyte surface antigen (MESA) (23), spectrin and actin (24).

Host cell ligands for endothelial cell cytoadherence

The first host cell ligand to be identified was thrombospondin (TSP). TSP is a glycoprotein that was shown to bind erythrocytes infected with trophozoite or schizont stage parasites (25). Several other ligands have been identified since TSP, (reviewed in (26)). Another glycoprotein, CD36 is found on many cell types, including platelets. Ockenhouse and co-workers demonstrated that like TSP, IE bound to CD36 coated onto plastic (27). Using COS cells expressing either CD36 or the glycoprotein Intercellular adhesion molecule 1 (ICAM-1), Berendt and co-workers demonstrated that IT04 parasites could bind to either ligand whereas FCR3 parasites were only able to bind CD36 (28). It has been demonstrated that human brain endothelial cells bind significantly more IE than amelanotic melanoma or human umbilical vein endothelial cells (29),

and that ICAM-1 expression on endothelial cells is stimulated when incubated with IE (30). Immunohistochemistry on brain sections of cerebral malaria fatalities showed enhanced levels of ICAM-1 (31). Although a recent study did not find increased binding of IE from cerebral malaria patients to ICAM-1in a static assay, there was a significant increase in binding under flow conditions (32). This is more akin to the conditions in the microvasculature. Like ICAM-1, Platelet-endothelial cell adhesion molecule-1(33), vascular cell adhesion molecule-1 and E-selectin (34) are also upregulated during infections. In the placenta, the main receptor on syncytiotrophoblasts is chondroitin sulphate A (CSA) (35), but IgG may also have a role (36).

Rosetting

A specific form of cytoadherence is rosetting. Rosettes are formed from IE adhering to uninfected erythrocytes (37,38). The erythrocyte receptors involved are complement receptor 1, A and B blood group antigens and heparin sulphate like molecules (39-41). In the parasite PfEMP-1 is the main protein involved in rosetting (40), although recent microarray experiments identified 377 upregulated genes in rosetting parasites (42). The rosetting phenotype has been implicated in many severe malaria infections (43-45).

Platelet-mediated clumping

A novel cytoadherent phenotype was described by Roberts et al (46) when agglutinates of IE were formed during *in vitro* culture. These differ from the rosetting phenotype because no uninfected erythrocytes are involved. Further investigation of these clumps found them to be comprised of IE

bound to platelets (47). The platelet ligands involved are CD36 and gC1qR/HABP1/p32 (47,48) however the parasite ligand has yet to be identified. Several studies have linked platelet-mediated clumping to severe disease (47,49,50). One such recent study however (51), when analysed further showed that the association was between clumping and parasitaemia. The authors have since reported that differences in experimental conditions, such as haematocrit and parasitaemia can affect the results of clumping assays (52). They have suggested a standardised method for use in subsequent studies.

Antigenic variation

The *P. falciparum* genome has 50-60 different *var* genes depending on the isolate. The *var* genes are mutually exclusively expressed, so that only one variant is expressed as protein on the surface of one IE. This clonal expression of PfEMP-1, coupled with the ability of the parasite to switch expression to different *var* genes, allows the parasite to evade the host's immune response. Marsh and Howard demonstrated that antigens were expressed on the IE surface, and that children develop antibodies to these during the course of an infection (53). The predominant variant surface antigen has since been shown to be PfEMP-1 (19).The large repertoire of *var* genes can go some way to explaining the age related increase in immunity seen in endemic regions. Using a longitudinal study, and agglutination assays, Bull and colleagues showed that a child's sera taken at the start of the study was less likely to recognise the surface of infected red cells from a new infection acquired during the study, than community controls (54). The same study also that demonstrated that as the age of

the group of children increased, sera from Kenyan children recognised more isolates. As children are exposed to new infections and as infections switch expression, they slowly gain immunity to the disease.

Pregnancy associated malaria

NAI is retained during continual exposure to parasites in endemic areas. It changes however during pregnancy. Pregnancy associated malaria (PAM) occurs in first and second pregnancies, but reduces during successive ones. PAM is a clear example of tissue specific sequestration, when IE sequester in the placenta, binding to CSA on the syncytiotrophoblast (35). Primagravid women lack antibodies to the CSA binding parasites, as they have had limited exposure to these. However, sera from multigravid women inhibit binding to CSA as their antibodies to it increase with successive exposure (55). Salanti and colleagues showed that a highly conserved and structurally distinct *var* gene, *var2csa* was upregulated in CSA-binding parasites (56).

Var genes

Genomic location

Many of the members of the *var* gene family have a sub-telomeric location next to the Telomere Associated Repeat Elements (TAREs), figure 1.4. Other multicopy gene families such as *rif* and *stevor* are also found in close proximity to the *var* genes in the sub-telomeric regions. Of the 14 chromosomes in the parasite clone 3D7, sequenced by the Malaria Genome Sequencing Project, only chromosome 14 has no *var* genes (57,58). Four chromosomes, 4, 7, 8 and 12 also have clusters of *var*

genes in central locations. These genes are arranged in head to tail orientation, with only one known exception on chromosome 12 in the *P. falciparum* HB3 clone. Although this was originally assumed to be misassembled, mapping data from our laboratory confirm this unique arrangement. In the sub-telomeric locations, *var* genes are predominantly in either head to tail, or tail to tail arrangements.

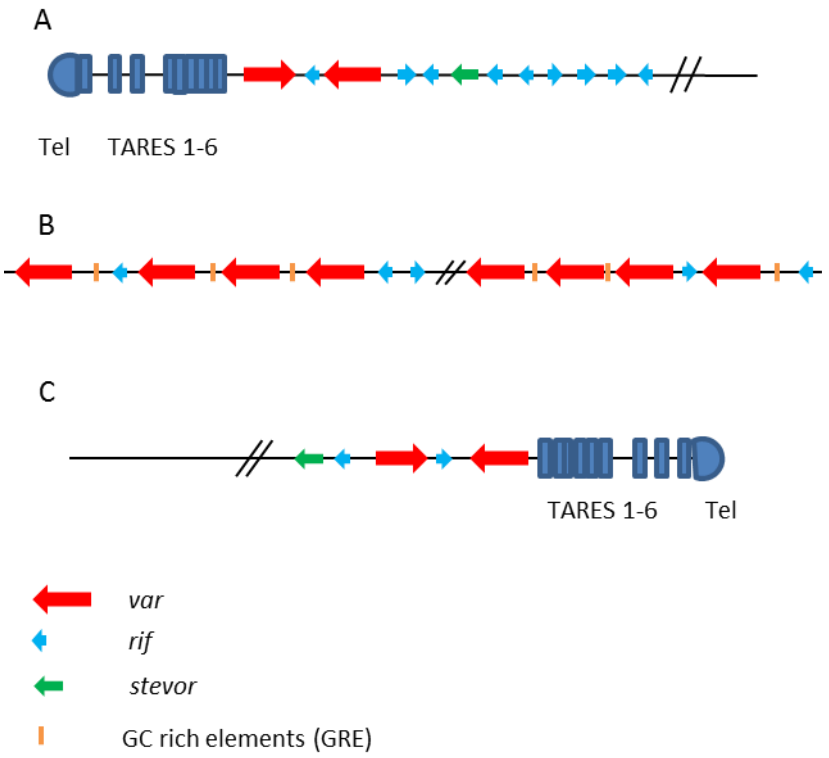


Figure 1.4 Arrangement of *var* gene clusters on chromosome 4. Arrangement of surface variant antigen genes on *P. falciparum* 3D7 chromosome 4. The position of telomeres (Tel) and telomere associated repeat elements (TAREs) at the chromosome ends are shown in A and C. *var*, *rif* and *stevor* genes, and GC rich elements are shown in A, B and C. A and C, either end of chromosome 4. B The two central *var* gene clusters.

Structure

Structurally, *var* genes share common features, figure 1.5. They have two exons; exon 1 codes for the transmembrane and surface exposed portions of the protein, exon 2 for the cytoplasmic tail. Exon 1 is highly diverse, with variable numbers of cytoadherent domains, the Duffy binding-like domain (DBL) and cysteine-rich interdomain region (CIDR). The term DBL derives from their homology to cysteine-rich domains of Duffy antigen binding proteins, the receptors for *P. knowlsei* and *P. vivax* invasion of RBC (20,59,60). Although variable, the DBLs were originally grouped into five main types, and the CIDR into three types based on their sequence similarity (61). A recent study analysing seven *P. falciparum* genomes re-defined the domains in all the elements of var gene structure (62). They described six major and 5 minor DBL classes as well as five major CIDR classes. Further alignment revealed “domain cassettes” (DC) where the order of multiple consecutive domains are conserved between multiple genomes, figure 1.6. The exon 2 sequence is semi conserved. The conserved intron has bi-directional promoter activity (63), and long non-coding RNAs have recently been described that originate from this intron (64).

Classification

The first comparison of the 5’ upstream (Ups) sequences of *var* genes by Voss and colleagues reported that they sorted two different groups (65). Gardner and colleagues then classified *var* genes into three groups based on alignments of 1.5 Kb of their Ups sequences (57). Further analysis of genome sequences of many isolates by Lavstsen (66) classified the *var*

genes into five groups based on 2 Kb of Ups sequences and chromosomal location. The three main groups are: UpsA, UpsB and UpsC, details of

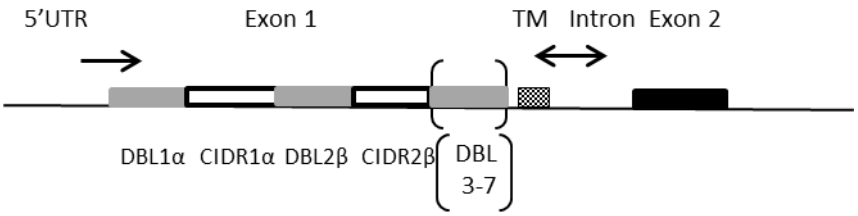


Figure 1.5 *var* gene structure.
Generic structure of *var* genes in *P. falciparum*. Arrows indicate the 2 promoters in the 5'UTR and the bi-directional promoter in the intron. Exon 1 contains DBL and CIDR domains. The majority have 4 alternating DBL and CIDR domains as shown, some also have extra DBL domains. Three 3D7 var genes have only 2 DBLs and no CIDRs. TM is the short transmembrane sequence. Following the intron is a fairly conserved exon 2.

their classification are shown in Table 1. The DC described by Rask and colleagues are each predominantly associated with a specific Ups class. This association is clearest for UpsA where seven DC are specific to UpsA *var* genes (62).

Studies looking at *var* genes in field isolate parasites have demonstrated a link between disease severity and Ups class of the expressed gene. Jensen and colleagues selected 3D7 parasites on plasma pools from semi-immune African children (67). They examined *var* gene expression by qRT-PCR from both unselected and selected lines. Comparing expression between the two groups, they found that 5/7 upregulated genes were UpsA, one was UpsB/A, and the other UpsE. Further evidence supporting UpsA genes as the predominant *var* genes in

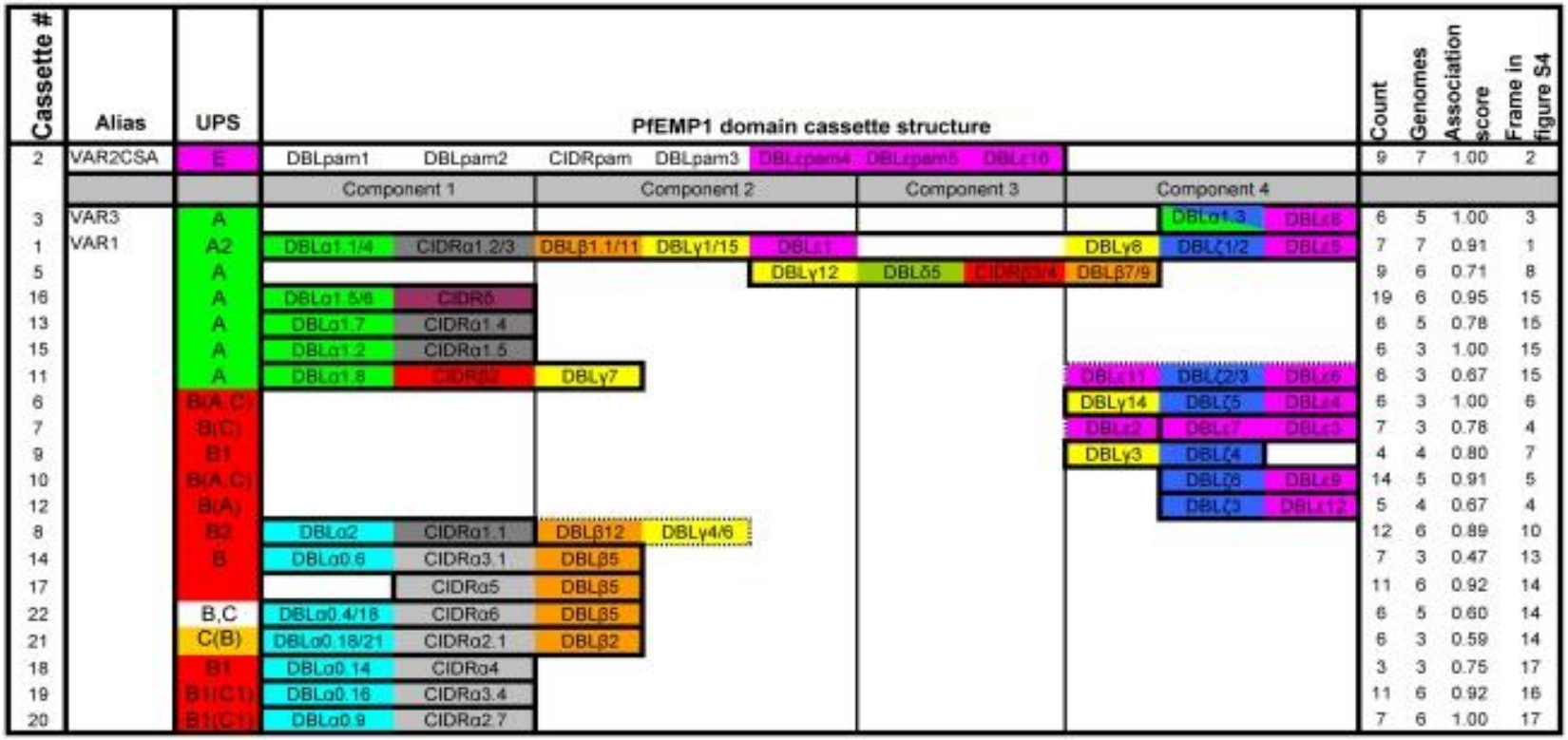


Figure 1.6 Domain Cassettes of PfEMP-1. Domain cassettes as determined from alignment of PfEMP-1 sequences from 7 *Plasmodium* genomes by Rask et al (62). Domain cassettes were called when the consecutive alignment of two or more conserved domains were found in at least three of the seven genomes. Figure taken from (62) Rask TS, et al. (2010) *Plasmodium falciparum* Erythrocyte Membrane Protein 1 Diversity in Seven Genomes – Divide and Conquer. PLoS Comput Biol 6(9): e1000933. doi:10.1371/journal.pcbi.1000933

cerebral malaria was described by Kyriacou and colleagues (68). Using generic primers to DBL1α domains, they found that those sequences transcribed from cerebral malaria patients were characteristic of UpsA or B/A genes. Rosetting parasites, shown to be common in severe malaria infections (69), were found to have significantly more UpsA transcripts than non-rosetting infections.

Studies have recently been published analysing transcriptome and qRT-PCR data from parasites selected for binding to human brain microvascular endothelial cell lines (70,71). These have identified a subset of UpsA *var* genes that are highly up-regulated and contain either DC8 or DC13. Similar results were also obtained from a different study, analysing *var* transcript levels by qRT-PCR in patients with differing severe malarial syndromes (72). UpsA *var* genes were associated with severe malaria, with the DC8 containing genes the most prevalent. From the body of evidence now reported, it is clear that UpsA *var* genes are associated with severe and cerebral malaria.

Var genes in laboratory isolates

Since their first description in 1995 (19-21), the repertoire of *var* genes has been studied in laboratory isolates. This was further aided by the publication of the *P. falciparum* 3D7 genome sequence in 2002 (58). The Broad Institute has since sequenced the HB3 and Dd2 clones (73). In 2007, Kraemer and colleagues used the Broad data and their own sequence data from the IT04 clone to publish schematics of the *var* repertoire of the three genomes (74). The three parasite lines are thought to originate from different continents; 3D7 from Africa (75,76), HB3 from

Central America (77) and IT04 which genotyping studies have grouped with Asian isolates (46,78). This spatial difference, along with differences in their cytoadherence properties made them ideal to compare *var* gene sequence.

Class	5' promoter	location on chromosome	Orientation of transcription	Number in 3D7 genome	Number in HB3 genome *	Notes
A	UpsA	Telomeric and sub-telomeric	Towards telomere	10	7	HB3 has only known case of an UpsA in a central cluster
B	UpsB	Telomeric	Towards centromere	22	18	
C	UpsC	Central	Towards telomere	13	11	
D	UpsD	Telomeric	Towards telomere	1	1	*Var1csa*
E	UpsE	Sub-telomeric	Towards telomere	1	2	*Var2csa*
B/C	UpsB like	Central	Towards telomere	9	9	
B/A	UpsB like	Telomeric and sub-telomeric	Towards centromere	4		

*A further 6 HB3 *var* genes have not been classified by Ups type

Table 1. Classification of *var* types.
Classification of major *var* gene groups in 3D7 and HB3 genomes. Compiled from data by Lavstsen, and Kraemer (66,74)

Kraemer et al found many similarities between the isolates. The telomeric location of genes, with central genes only occurring on chromosomes 4, 6, 7, 8, and 12 was the same in 3D7 and HB3. Central

var genes were described on chromosomes 4, 5-8, 9 and 13 in IT04. However the recently updated IT sequence on Plasmodb has central *var* genes on identical chromosomes to HB3 and 3D7 (79). The Ups classification system is valid for HB3 and IT04, with one notable exception of an UpsA gene in a central cluster in HB3. This is the same gene on chromosome 12 that we had confirmed by mapping.The tandem domain pattern was consistent, as were associations e.g DBLα adjacent to CIDR1. All three strains have copies of *var1csa* and *var2csa*, with HB3 having two copies of *var2csa*. These genes are semi-conserved, having >75% sequence similarity. As both 3D7 and HB3 are used in this work, their *var* repertoires are shown in figures 1.7 and 1.8.

Gene expression in *Plasmodium*

The regulation of gene expression in *P. falciparum* is highly organized. This was described by Bozdech et al, studying the transcriptome by micro-array analysis (80). Their data showed that genes are switched on and off in a highly organized cascade pattern, at different stages of the intra-erythrocytic cycle. Early comparison of the *P. falciparum* genome sequence with other organisms revealed a lack of transcription factors (57,81). This was surprising, given the tightly regulated expression. In 2005, Balaji and colleagues reported a family of apicomplexan specific transcription factors, ApiAP2 (82). Domain searches revealed 43 copies in the *P. falciparum* genome. More recently, Bischoff *et al* have done a comprehensive search of the *Plasmodium* genome with newly reported Transcription Associated Proteins (TAPS) from other organisms (83). They reported 202 TAPs, including identification of 34 new TAPs.

3D7 gene name	Ups	Extracellular Domain Structure (Predicted)									ATS	Chr Loc
PFA0015c	A1	DBL1α1	DBL2ε								exon2A1	1L-STt
PFF0020c	A1	DBL1α1	DBL2ε								exon2A1	6L-SSTt
PFI1820w	A1	DBL1α1	DBL2ε								exon2A1	9R-STt
PF11_0008	A1	DBL1α1	CIDR1γ	DBL2γ	DBL3δ	CIDR2β	DBL4β	C2			exon2A1	11L-STt
PF08_0141	A1	DBL1α1	CIDR1γ	DBL2β	C2	DBL3γ	DBL4β	DBL5ε			exon2A1	8L-STt
PF13_0003	A1	DBL1α1	CIDR1γ	DBL2β	C2	DBL3γ	DBL4δ	CIDR2β	DBL5β	C2	exon2A1	13L-STt
PFD0020c	A1	DBL1α1	CIDR1α1	DBL2β	C2	DBL3γ	DBL4γ	DBL5δ	CIDR2β		exon2A1	4L-STt
PF11_0521	A1	DBL1α1	CIDR1α1	DBL2β	C2	DBL3β	C2	DBL4δ	CIDR2γ		exon2A1	11R-Tt
MAL8P1.207	A1	DBL1α1	CIDR1α1	DBL2β	C2	DBL3β	C2	DBL4γ	DBL5δ	CIDR2β	exon2A1	8R-STt
PFD1235w	A1	DBL1α1	CIDR1α1	DBL2β	C2	DBL3β	C2	DBL4γ	DBL5δ	CIDR2β	exon2A1	4R-STt
PFE1640w-ps	A2	DBL1α1	CIDR1α1	DBL2β	C2	DBL3γ	DBL4ε	DBL5γ	DBL6β	DBL7ε		5R-Tt
PF08_0140	B1	DBL1α	CIDR1α1	DBL2β	C2	DBL3γ	DBL4δ	CIDR2β			exon2B	8L-SSTc
PFF1580c	B1	DBL1α	CIDR1α	DBL2β	C2	DBL3γ	DBL4δ	CIDR2β	DBL5ε	DBL6ε	DBL7ε	exon2B	6R-STc
PFL0020w	B1	DBL1α	CIDR1α	DBL2β	C2	DBL3γ	DBL4β	DBL5ε			exon2B	12L-STc
PFF0010w	B2	DBL1α	CIDR1γ	DBL2β	C2	DBL3γ	DBL4β	DBL5ε			exon2B	6L-STc
PFB0010w	B1	DBL1α	CIDR1α	DBL2γ							exon2B	2L-Tc
PFA0005w	B1	DBL1α	CIDR1α	DBL2δ	CIDR2β						exon2B	1L-Tc
PFA0765c	B1	DBL1α	CIDR1α	DBL2δ	CIDR2β						exon2B	1R-Tc
PFB1055c	B1	DBL1α	CIDR1α	DBL2δ	CIDR2β						exon2B	2R-Tc
PFC0005w	B1	DBL1α	CIDR1α	DBL2δ	CIDR2β						exon2B	3L-Tc
PFC1120c	B1	DBL1α	CIDR1α	DBL2δ	CIDR2β						exon2B	3R-Tc
PFD1245c	B1	DBL1α	CIDR1α	DBL2δ	CIDR2β						exon2B	4R-Tc
PFE0005w	B1	DBL1α	CIDR1α	DBL2δ	CIDR2β						exon2B	5L-Tc
PFF1595c	B1	DBL1α	CIDR1α	DBL2δ	CIDR2β						exon2B	6R-Tc
MAL7P1.212	B1	DBL1α	CIDR1α	DBL2δ	CIDR2β						exon2B	7L-Tc
PF08_0142	B1	DBL1α	CIDR1α	DBL2δ	CIDR2β						exon2B	8L-Tc
MAL8P1.220	B1	DBL1α	CIDR1α	DBL2δ	CIDR2β						exon2B	8R-Tc
PF10_0406	B1	DBL1α	CIDR1α	DBL2δ	CIDR2β						exon2B	10R-Tc
PF11_0007	B1	DBL1α	CIDR1α	DBL2δ	CIDR2β						exon2B	11L-Tc
PFL0005w	B1	DBL1α	CIDR1α	DBL2δ	CIDR2β						exon2B	12L-Tc
PFL2665c	B1	DBL1α	CIDR1α	DBL2δ	CIDR2β						exon2B	12R-Tc
MAL13P1.1	B1	DBL1α	CIDR1α	DBL2δ	CIDR2β						exon2B	13L-Tc
MAL13P1.356	B1	DBL1α	CIDR1α	DBL2δ	CIDR2β						exon2B	13R-Tc
PFI0005w	B1	DBL1α	CIDR1α	DBL2δ	CIDR2γ						exon2B	9L-Tc
PFI1830c	B1	DBL1α	CIDR1α	DBL2δ	CIDR2γ						exon2B	9R-Tc
PF10_0001	B1	DBL1α	CIDR1α	DBL2δ	CIDR2γ						exon2B	10L-Tc
MAL7P1.187	B1	DBL1α	CIDR1α	DBL2δ	CIDR2β	DBL3ε					exon2B	7R-Tc
PFD0005w	B1	DBL1α	CIDR1α	DBL2γ	DBL3δ	CIDR2β					exon2B	4L-Tc
PF08_0106	B1	DBL1α	CIDR1α	DBL2δ	CIDR2β						exon2B	8-C
PFL1955w	B1	DBL1α	CIDR1α	DBL2δ	CIDR2β						exon2B	12-C
PFD1005c	B1	DBL1α	CIDR1α	DBL2δ	CIDR2γ						exon2B	4-C
MAL7P1.50	B1	DBL1α	CIDR1α	DBL2δ	CIDR2γ						exon2B	7-C
MAL7P1.55	B1	DBL1α	CIDR1α	DBL2δ	CIDR2γ						exon2B	7-C
PFL0935c	B1	DBL1α	CIDR1α	DBL2δ	CIDR2γ						exon2B	12-C
PF08_0103	B3	DBL1α	CIDR1α	DBL2δ	CIDR2β						exon2B	8-C
PFD0635c	B3	DBL1α	CIDR1α	DBL2δ	CIDR2γ						exon2B	4-C
PF07_0050	B3	DBL1α	CIDR1α	DBL2β	C2	DBL3γ					exon2B	7-C
PFL1950w	B4	DBL1α	CIDR1α	DBL2β	C2	DBL3δ	CIDR2β				exon2B	12-C
PFD0615c	C1	DBL1α	CIDR1α	DBL2δ	CIDR2β						exon2B	4-C
PFD0625c	C1	DBL1α	CIDR1α	DBL2δ	CIDR2β						exon2B	4-C
PFD1015c	C1	DBL1α	CIDR1α	DBL2δ	CIDR2β						exon2B	4-C
PF07_0048	C1	DBL1α	CIDR1α	DBL2δ	CIDR2β						exon2B	7-C
PFL1960w	C1	DBL1α	CIDR1α	DBL2δ	CIDR2β						exon2B	12-C
PF07_0051	C1	DBL1α	CIDR1α	DBL2δ	CIDR2β						exon2A2	7-C
PFD0630c	C1	DBL1α	CIDR1α	DBL2δ	CIDR2γ						exon2B	4-C
PFD1000c	C1	DBL1α	CIDR1α	DBL2δ	CIDR2γ						exon2B	4-C
PF07_0049	C1	DBL1α	CIDR1α	DBL2δ	CIDR2γ						exon2B	7-C
PF08_0107	C1	DBL1α	CIDR1α	DBL2δ	CIDR2γ						exon2B	8-C
PFD0995c	C1	DBL1α	CIDR1α	DBL2δ	CIDR2γ						exon2A2	4-C
PFF0845c	C1	DBL1α	CIDR1α	DBL2β	C2	DBL3δ	CIDR2β				exon2A2	6-C
MAL7P1.56	C2	DBL1α	CIDR1α	DBL2δ	CIDR2β						exon2B	7-C
PFL0030c	E	DBL1	DBL2	DBL3	DBL4ε	DBL5ε	DBL6ε				exon2E	12L-SSTt

Figure 1.7 3D7 *var* genes.
Nomenclature, domain structure and chromosomal location of *var* genes in 3D7. Adapted from Kraemer et al (74).

HB3 gene Name	Broad locus name	HB3 Contig Name	Super-contig	Ups	Extracellular Domain Structure (Predicted)									ATS	Chr Loc
HB3 var6	PFHG_02274.1	699-3	1.47	A1	DBL1α1	CIDR1γ	DBL2γ	DBL3δ	CIDR2γ	DBL4ε	DBL5β	C2	DBL6ε	exon2A1	12-C2
HB3 var3	PFHG_05052.1	1737	1.265	A1	DBL1α1	CIDR1α1	DBL2β	C2	DBL3γ	DBL4δ	CIDR2β	DBL5β	C2		11L-STt
HB3 var1	PFHG_03234.1	1000-2	1.88	A1	DBL1α1	CIDR1γ1	DBL2β	C2	DBL3γ	DBL4γ	DBL5δ	CIDR2β		exon2A1	9R-STt
HB3 var46Ψ	PFHG_03997.1	1278Ψ	1.138	A1	DBL1α1	CIDR1γ	DBL2γ	DBL3δ	CIDR2β	DBL4β	C2				7L-STt
HB3 var2	PFHG_03840.1	1210	1.125	A1	DBL1α1	CIDR1α1	DBL2γ	DBL3δ	CIDR2β	DBL4β	C2			exon2A1	7L-STt
HB3 var4	PFHG_04861.1	1703	1.253	A1	DBL1α1	CIDR1γ	DBL2γ	DBL3δ	CIDR2β	DBL4γ	DBL5β	C2	DBL6ε		STt
HB3 var5	PFHG_03671.1	1235	1.130	A2	DBL1α1	CIDR1γ	DBL2γ	DBL3δ	CIDR2β	DBL4β	C2				5L-STt
HB3 var1csa	PFHG_03521.1	1108	1.105	A2	DBL1α1	CIDR1α1	DBL2β	C2	DBL3γ	DBL4ε	DBL5γ	DBL6β	DBL7ε		STt
HB3 var7	PFHG_04769.1	1604-1	1.226	B1	DBL1α	CIDR1α	DBL2δ	CIDR2γ	DBL3γ	DBL4β	C2	DBL5ε			SSTc
HB3 var14	PFHG_04035.1	1308	1.144	B1	DBL1α	CIDR1α	DBL2δ	CIDR2β						exon2B	10R-Tc
HB3 var16	PFHG_03232.1	1000-1	1.188	B1	DBL1α	CIDR1α	DBL2δ	CIDR2β						exon2B	9R-Tc
HB3 var18	PFHG_04593.1	1523	1.203	B1	DBL1α	CIDR1α	DBL2δ	CIDR2β						exon2B	11R-Tc
HB3 var12	PFHG_03416.1	1040	1.196	B1	DBL1α	CIDR1α	DBL2δ	CIDR2β						exon2B	9L-Tc
HB3 var47Ψ	PFHG_04057.1	1161Ψ	1.114	B1	DBL1α	CIDR1α	DBL2δ	CIDR2βp							4R-Tc
HB3 var13	PFHG_03516.1	1107	1.105	B1	DBL1α	CIDR1α	DBL2δ	CIDR2β							Tc
HB3 var15	PFHG_04035.1	1426	1.171	B1	DBL1α	CIDR1α	DBL2δ	CIDR2β							Tc
HB3 var20	PFHG_04770.1	1604-2	1.226	B1	DBL1α	CIDR1α	DBL2δ	CIDR2β						exon2B	Tc
HB3 var11	PFHG_04491.1	1499	1.193	B1	DBL1α	CIDR1α	DBL2δ	CIDR2β	DBL3ε					exon2B	Tc
HB3 var10	PFHG_04749.1	1587	1.220	B1	DBL1α	CIDR1α	DBL2β	C2	DBL3δ	CIDR2β					13R-Tc
HB3 var48Ψ	PFHG_04277.1	1479Ψ	1.189	B1	DBL1α	CIDR1α	DBL2β	C2	DBL3γ	DBL4β	C2				Tc
HB3 var8	PFHG_04081.1	1334	1.153	B1	DBL1α	CIDR1α	DBL2δ	CIDR2γ	DBL3γ	DBL4β	C2	DBL5ε		exon2B	Tc
HB3 var9	PFHG_04057.1	1408	1.168	B1	DBL1α	CIDR1α	DBL2δ	CIDR2γ	DBL3β	C2	DBL4ε			exon2B	Tc
HB3 var49Ψ	PFHG_03718.1	1181Ψ	1.118	B1	DBL1α	CIDR1α	DBL2	CIDR2γ	DBL3β	C2	DBL4ε			exon2B	12R-Tc
HB3 var38p	PFHG_05280.1	1966p	1.350	B1	DBL1α	CIDR1α									Tc?
HB3 var39p	efbe	2007p	1.368	B1	DBL1α	CIDR1α									10L-Tc
HB3 var37p	PFHG_04277.1	2064p	1.395	B1	DBL1α	CIDR1α	DBL2β	C2	DBL3γ						
HB3 var19	PFHG_04620.1	1514	1.200	B1	DBL1α	CIDR1α	DBL2δ	CIDR2γ						exon2B	12-C1
HB3 var17	PFHG_04014.1	1296-2	1.142	B1	DBL1α	CIDR1α	DBL2β	C2	DBL3γ	DBL4δ	CIDR2β			exon2B	7-C
HB3 var22	PFHG_02425.1	752-4	1.54	B1	DBL1α	CIDR1α	DBL2δ	CIDR2γ	DBL3ε	DBL4ε	DBL5ε			exon2B	8-C
HB3 var21	PFHG_04910.1	1671_1	1.243	B1	DBL1α	CIDR1α	DBL2β	C2	DBL3γ	DBL4δ	CIDR2γ	DBL5ε		exon2B	7-C
HB3 var26	PFHG_02276.1	699-4	1.47	B3	DBL1α	CIDR1α	DBL2δ	CIDR2β						exon2B	12-C2
HB3 var27	PFHG_03476.1	1074-1	1.100	B3	DBL1α	CIDR1α	DBL2δ	CIDR2β						exon2B	4-C1
HB3 var24	PFHG_02421.1	752-2	1.54	B3	DBL1α	CIDR1α	DBL2δ	CIDR2γ	DBL3β	C2	DBL4ε			exon2B	8-C
HB3 var25	PFHG_02423.1	752-3	1.54	B3	DBL1α	CIDR1α	DBL2δ	CIDR2β	DBL3β	C2	DBL4ε			exon2B	8-C
HB3 var23	PFHG_02272.1	699-1	1.47	B4	DBL1α	CIDR1α	DBL2δ	CIDR2γ	DBL3β	C2	DBL4ε			exon2B	12-C2
HB3 var28	PFHG_03478.1	1074-2	1.100	C1	DBL1α	CIDR1α	DBL2δ	CIDR2β						exon2B	4-C1
HB3 var29	PFHG_03480.1	1074-3	1.100	C1	DBL1α	CIDR1α	DBL2δ	CIDR2β						exon2B	4-C1
HB3 var30	PFHG_02273.1	699-2	1.47	C1	DBL1α	CIDR1α	DBL2δ	CIDR2β						exon2B	12-C2
HB3 var31	PFHG_02277.1	699-5	1.47	C1	DBL1α	CIDR1α	DBL2δ	CIDR2β						exon2B	12-C2
HB3 var35	PFHG_04015.1	1296-3	1.142	C1	DBL1α	CIDR1α	DBL2δ	CIDR2β						exon2B	7-C
HB3 var32	fbadf	1459	1.183	C1	DBL1α	CIDR1α	DBL2δ	CIDR2γ						exon2A2	4-C2
HB3 var33	PFHG_02429.1	752-5	1.54	C1	DBL1α	CIDR1α	DBL2δ	CIDR2γ						exon2A2	8-C
HB3 var51Ψ	PFHG_04911.1	1671_2Ψ	1.243		DBL1α	CIDR1α	DBL2δ	CIDR2γ							7-C
HB3 var50Ψ	PFHG_02419.1	752_1Ψ	1.54	C1	DBL1α	CIDR1α	DBL2δ	CIDR2γ	DBL3β	C2	DBL4ε				8-C
HB3 var34	PFHG_00592.1	209	1.8	C1	DBL1α	CIDR1α	DBL2β	C2	DBL3γ	DBL4δ	CIDR2β			exon2A2	6-C
HB3 var36	PFHG_04012.1	1296-1	1.142	C2	DBL1α	CIDR1α	DBL2δ	CIDR2γ							7-C
HB3 var2csaA	PFHG_05046.1	1727	1.262	E	DBL1	DBL2	DBL3	DBL4ε	DBL5ε	DBL6ε				exon2E	12L-STt
HB3 var2csaB	PFHG_05155.1	1817	1.295	E	DBL1	DBL2	DBL3	DBL4ε	DBL5ε	DBL6ε				exon2E	STt
HB3 var40p	PFHG_04928.1	1678p	1.245		DBL1α	CIDR1α	DBL2δ	CIDR2γ	DBL3β	C2	DBL4ε			exon2B	12R-STt?
HB3 var41p	PFHG_05421.1	2124p	1.426		DBL1α	CIDR1α	DBL2δ								
HB3 var42p	PFHG_05521.1	2463p	1.684		DBL1α	CIDR1α									
HB3 var43p	PFHG_04990.1	1941p	1.337		DBL1α										
HB3 var44p	PFHG_05458.1	2342p	1.566		DBL1α										
HB3 var45p	PFHG_05575.1	2622p	1.841		DBL1α										

Figure 1.8 HB3 *var* genes.
Nomenclature, domain structure and chromosomal location of *var* genes, where known, in HB3. Adapted from Kraemer et al (74)

Chromatin and modifications of histones are known to have a crucial role in gene regulation (reviewed in (84)). One study of histones revealed 44 different modifications (85). These include acetylation or methylation of the histone tail, resulting in binding of different factors that hence regulate transcription. The involvement of known histone modifications in *var* gene regulation is discussed below.

Regulation of *var* gene expression

Like other *falciparum* genes, *var* genes are expressed during a specific time window, 3-18 hours post-invasion. It is the mechanisms that regulate the mutually exclusive expression that are the subject of much investigation. Although regularly described as poorly understood, our understanding is increasing rapidly.

Elements involved in the regulation of *var* gene expression can be loosely assigned to three groups: 1.Epigenetic marks and heterochromatin modifications, 2. Sub-nuclear position, 3. DNA and RNA elements. These can be divided on paper, for the purposes of discussion. However, it will become clear that the division within the cell is not so, as *var* gene regulation involves co-operation and interaction between all of these elements.

Epigenetic marks and heterochromatin modification

The epigenetic marks that differentiate the transcriptional state of an individual *var* gene within the gene family were first described in 2007 (86). The 5'UTR and exon 1 of silent *var* genes are enriched in H3K9me3. In the active gene, the H3K9me3 is removed and H3K9ac and H3K4me3 are enriched in the 5'UTR. One member of the family is poised during late

stages to become the active gene in the following cycle; this gene is marked by enrichment of H3K4me2 in the 5'UTR (86).

Since 2007, there has been an ever increasing identification of factors that have a role in these changes. Some are involved with all classes of *var*, but several are associated with a specific class of *var* genes. Those currently thought to be class independent include *P. falciparum* heterochromatin protein 1(PfHP1), PfMYST and PfSET10. PfHP1 binds to H3K9me3, the silencing mark, in both sub-telomeric and central *var* genes. It is removed from the upstream region before activation of a specific gene (87,88). A histone acetyltransferase, PfMYST is enriched at the active *var* gene promoter during ring stage (89). It preferentially acetylates H4, which had previously been shown to correlate with *var* gene activity (89,90). Associated with the poised state of a *var* gene, PfSET10 is a histone lysine 4 methyltransferase with activity on H3. It is localised in the perinuclear expression site, and pull down experiments identified actin as a protein interacting with PfSET10 (91).

Some factors have been shown to interact with a specific class of *var* genes. A major group of class-specific factors was uncovered following the identification of homologues of the yeast Sir2 protein, which is involved in chromatin modification and gene repression. PfSIR2 was shown to bind to the promoter of the UpsE *var* gene, *var2csa*, only when the gene was silent (90). ChIP studies determined that PfSIR2 binds to the promoter regions of both UpsE and UpsB genes, when an UpsC gene is active. This PfSIR2 silencing appears to be class specific, owing to the location of those classes in a sub-telomeric location. Further work by Tonkin and

colleagues, identified two distinct Sir2 homologues (92). Using 3D7 parasites containing either a disrupted PfSIR2A or PfSIR2B gene, they performed global transcription analysis. Looking specifically at *var* genes, they found a correlation between the disrupted PfSIR2 gene and the direction of transcription and location of the activated gene. UpsA, UpsE and UpsC genes were activated with disrupted PfSIR2A and UpsB genes by disruption of PfSIR2B. They concluded that PfSIR2B has a role in silencing UpsB *var* genes, and PfSIR2A in silencing UpsA, UpsC and UpsE. Another protein, PfORC1 has been shown to co-localise with PfSIR2 in the telomeric clusters (93).

Sub-nuclear position

FISH studies have shown that the chromosome ends, and hence sub-telomeric *var* genes, are arranged in 4-7 distinct clusters at the nuclear periphery (94). Central *var* genes have also been shown to localise to the nuclear periphery, but not to co-localize with telomeric genes (95). Evidence for a *var* transcription site has come from several studies. Ralph et al showed that activation of a *var* gene resulted in its movement from a cluster, into a region of uncondensed euchromatin within the nuclear periphery. (95). A separate study showed that silent and active *var* loci occupy different perinuclear locations, and confirmed that activation of a *var* gene requires movement to a different site (96). Whilst investigating mutually exclusive expression, Dzikowski and colleagues, reported dual *var* expression from both an episomal plasmid and an endogenous gene (97). FISH analysis showed co-localization of both active *var* promoters, again supporting a *var* transcription site.

Further evidence for *var* gene transcription being at the nuclear periphery comes from FISH studies by Lopez-Rubio et al (98). Antibodies to PfKMT, the *falciparum* homologue of the histone lysine 9 methyltransferase KMT1, localised to the nuclear periphery. KMT1 is responsible for trimethylation of histone H3 lysine 9 (H3K9) to H3K9me3, an epigenetic mark of silent *var* genes. Central *vars* were randomly distributed within the periphery in relation to telomeric clusters. RNA FISH to the active *var2csa* gene confirmed that it was separate from the clusters. The authors postulate the existence of “perinuclear repressive centers”, containing the telomeric end clusters and silencing factors.

Recently, a 25 base pair *var* intron motif (iNPE) was shown to have nuclear protein binding properties with a role in perinuclear tethering (99). Proteomic studies identified actin as one of the components of the intron:protein complex. That actin has a role in spatial organisation was confirmed using Jasplakinolide (JAS), a compound which stabilizes actin filaments. Treatment with JAS led to a reduction in co-localisation of *var* foci and an episome containing iNPE.

DNA and RNA elements

From a genome wide perspective, the most obvious DNA elements involved in *var* gene regulation are the Ups regions. We have seen that PfSIR2A has a role in silencing UpsA, UpsC and UpsE genes (92). As these classes are both central and sub-telomeric, the specificity cannot just be due to their chromosomal location. The intron promoter also has an important role in the counting mechanism for *var* gene silencing. Strict pairing of both the upstream and intron promoters is required for silencing

(100). On a more localized level, the SPE2 element, identified by Voss et al is a sequence motif consisting of repeats of (T/G)GTGC(A/G) separated by four nucleotides (101). SPE2 are located in the promoter region of UpsB var genes and form a complex with proteins when incubated with nuclear extracts. An ApiAP2 protein, PfSIP2, was identified by Fluek et al, using concatemerised SPE2 elements to isolate SPE2 binding proteins (102). Genome wide ChIP analysis revealed that PfSIP2 binding only occurs at TARE 2-3 and a region with a cluster of SPE2 elements upstream of UpsB *var* genes. No binding was found at chromosome-internal sites (102).

Recently, a family of 22 long ncRNAs (lncRNA) have been implicated in the regulation of UpsB *var* genes due to the presence of SPE2 elements within them (103). Called lncRNA-TARE owing to their location in the Telomere Associated Repeat Region, they are expressed at 40 hours post invasion from TARE 2-3. Intriguingly, each lncRNA-TARE is located adjacent to an UpsB var. The authors propose 4 models for the action of lncRNA-TARE, including one that involves lncRNA-TARE recruiting chromatin remodelling factors to provide silencing memory marks at UpsB var genes. Other non-coding RNAs (ncRNA) implicated in *var* gene expression have been described from the bi-directional intron promoter (64).

Although *var* gene regulation is not fully understood, some facets are now established. The evidence for localisation of *var* genes at the nuclear periphery (94,95), and a var transcription site is compelling (95,96), as is the requirement for both promoters in silencing (100). The

involvement of histone modifications and proteins involved is becoming clearer on a family wide scale. H3K9me3 modification is a silencing mark, to which PfHIP1 binds, and is removed upon activation (86-88). H3K4me2 at the 5’ UTR marks a gene poised for transcription, and this is associated with PfSET10 in the expression site (86,91). Finally, the 5’UTRs of active genes are enriched with H3K9ac and H3K4me3 and H4 (86,90). What still needs to be established are factors that are involved in regulation of specific Ups classes. For example, the PfSIP2 transcription factor binds to UpsB regions, but only those at sub-telomeric locations, not central ones (102). It is reasonable to assume therefore that there are elements, be they protein, ncRNA or both, yet to be discovered that are specific for other Ups types.

Other IE variant surface antigens

Although the major variant surface antigen is PfEMP-1, other multigene families have been identified that may have a role in antigenic variation. As shown in figure 1.4, these genes are primarily located in the TAREs, with a few *rif* genes also in the central *var* clusters. Witmer and colleagues have recently suggested that none of these families are expressed in a mutually exclusive manner (104). Using similar constructs, each replacing a housekeeping promoter with a promoter from either UpsA, UpsB or UpsC var or other multicopy family, they measured expression levels of the endogenous genes. With the Ups constructs, no *var* expression was detected, whereas it was detected with other constructs. In contrast, there was no change in endogenous transcription levels with the other promoters (104). These families will be discussed briefly below.

Pir family

Comparative genome analysis has revealed the existence of a multigene family in both primate and rodent malarias, termed the *Plasmodium* interspersed repeat family (*pir*) (105). It was known before this, that the *vir* genes from *P. vivax* had homologues in the rodent malarias, but this was the first publication to link these to *P. falciparum rif* and *stevor* families, described below. As a multigene family,the characteristics of the *pir* genes are diverse, (reviewed in (106)). There are between 36 and 838 copies of each family, with sizes ranging from 93-3399 nucleotides, and 1-7 exons (106). A smaller family of proteins, PfMC-2TM, have structural similarities to other membrane proteins, and are located in the Maurer's clefts (107). These have been described as surface proteins (108), but immunoelectron microscopy experiments have shown them to be located under the surface of knobs in trophozoites (109).

Rifins

The repetitive interspersed family or *rif* genes were first described in 1998, and the expression of RIFIN proteins confirmed in 1999 (110-112). There are 149 full length *rif* copies in the 3D7 genome (57), each having a 2 exon structure; a short exon 1 and an exon 2 of approximately 1kb (112). During the asexual cycle most *rif* genes are transcribed at 18-23 hours post invasion, although two have been reported to be expressed during late trophozoite and early ring stages (111,113). RIFIN proteins are 35-45 kDa, and were predicted to have an N-terminal signal sequence, two transmembrane domains flanking an external polymorphic region, and a conserved C-terminal cytoplasmic tail (111,112). More recently, sequence analysis has divided the *rif* gene family into two groups, A and B types

(114). The B types are predicted to have the original structure, described above, whereas most A types have no signal peptide and only one transmembrane domain (115) FISH and immunoblotting have shown the A-type RIFINS localised in Maurer's clefts, the vesicles involved in protein trafficking to the IE membrane. B-type RIFINS were retained within the parasite. Both RIFINS were found to be expressed in merozoites, A-types at the apical region and B-types in the cytosol (116). RIFINS are also expressed in gametocytes, with A-types expressed at the surface of the IE, and B-types again retained within the parasite (117).

Stevor

The *stevor* gene family has 28 copies, all sub-telomerically located in 3D7 (57). Similar in structure to *rif*, they have a 50-75 bp exon 1, and an exon 2 of approximately 1 kb. The proteins are smaller at 30-40 kDa, but have the same predicted domain regions (112). They are transcribed at 22-32 hours post invasion in early trophozoites and are also found in the Maurer's clefts (118). The protein is expressed in both schizonts and merozoites in IE (119-121). The protein is also expressed throughout the gametocyte stages (122). The function of STEVOR proteins is unknown, but recent evidence suggests that they contribute to the rigidity of the IE membrane (123).

Surfins

The final IE surface antigen gene family are the *surfins*. There are 10 subtelomeric members, encoding proteins of 285 kDa (124). More akin to PfEMP-1 in structure, the proteins have a variable extracellular region, a transmembrane region and a semi-conserved intracellular domain. The

protein is located on the IE membrane and at the apex of merozoites (124).

GC-rich elements in the Plasmodium genome

The publication of the *P.falciparum* 3D7 genome sequence in 2002 identified 15 GC rich sequence elements (GRE) (58). These same sequences were described as RNAs of unknown function-6, (RUF-6) by Chakrabarti and colleagues in 2007 (125).BLAST searches of NCBI and Rfam databases confirmed that they have no homologues in other genera, and have only been found in *P. falciparum* and *P. reichenowi*. The sequences are highly conserved between GRE, and are 85% G-C, as opposed to the *P. falciparum* genome which is 20% G-C. They are all located in close proximity to central *var* genes on chromosomes 4, 7, 8 and 12, (figure 1.8 (58)). The location of the GREs in the central *var* clusters was of interest to a laboratory whose primary research interest is *var* genes.

Experiments in our laboratory have shown that the GRE are transcribed as 135 nucleotide ncRNA throughout the life cycle, figure 1.9 (our unpublished data). Sequence analysis of the GRE identified a conserved RNA polymerase lll promoter, figure 1.10. This was substantiated by run on experiments with the RNA polymerase ll inhibitor, alpha –amanitin, which had no effect on GRE transcription, figure 1.11 (our unpublished data). The 5’ end of the transcript was determined by 5’ RACE, figure 1.10. The 3’ end was determined by GeneRacer cloning, as 3’RACE was unsuccessful owing to the lack of a poly(A) tail, figure 1.10. Northern blots of several laboratory isolates have confirmed GRE

expression in all those tested with the exception of TM180, figure 1.11 (our unpublished data). RNA folding programmes such as Mfold (126) predict that the GRE RNAs have a hairpin structure, figure 1.12.

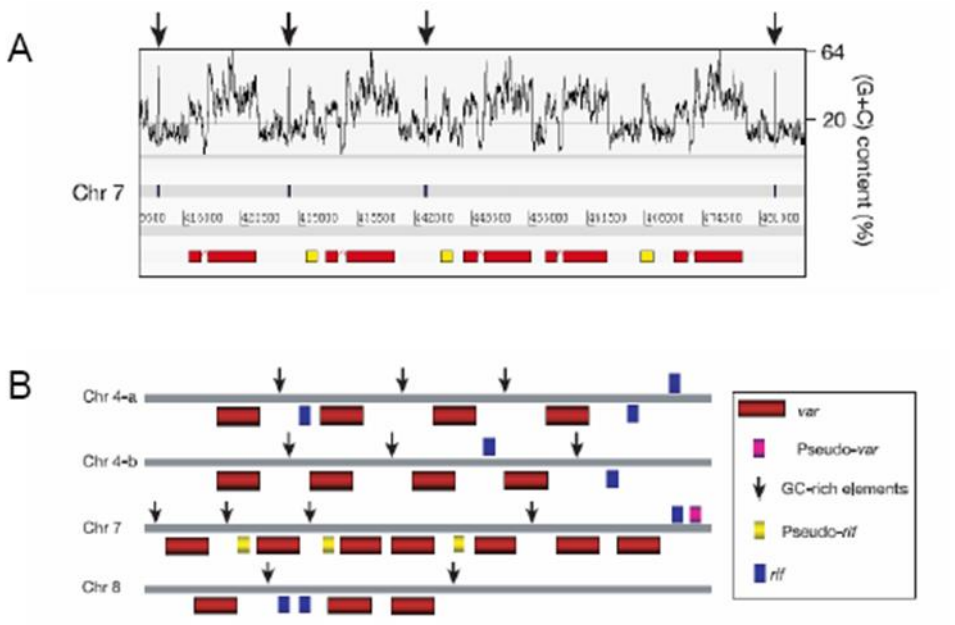


Figure 1.8. Identification of GREs in the *P. falciparum* genome
A. Arrows indicating GC peaks in intergenic regions of internal *var* cluster on chromosome 7. B. Location of GRE in internal *var* clusters on chromosomes 4, 7 & 8. (GRE also identified on chromosome 12) (58).

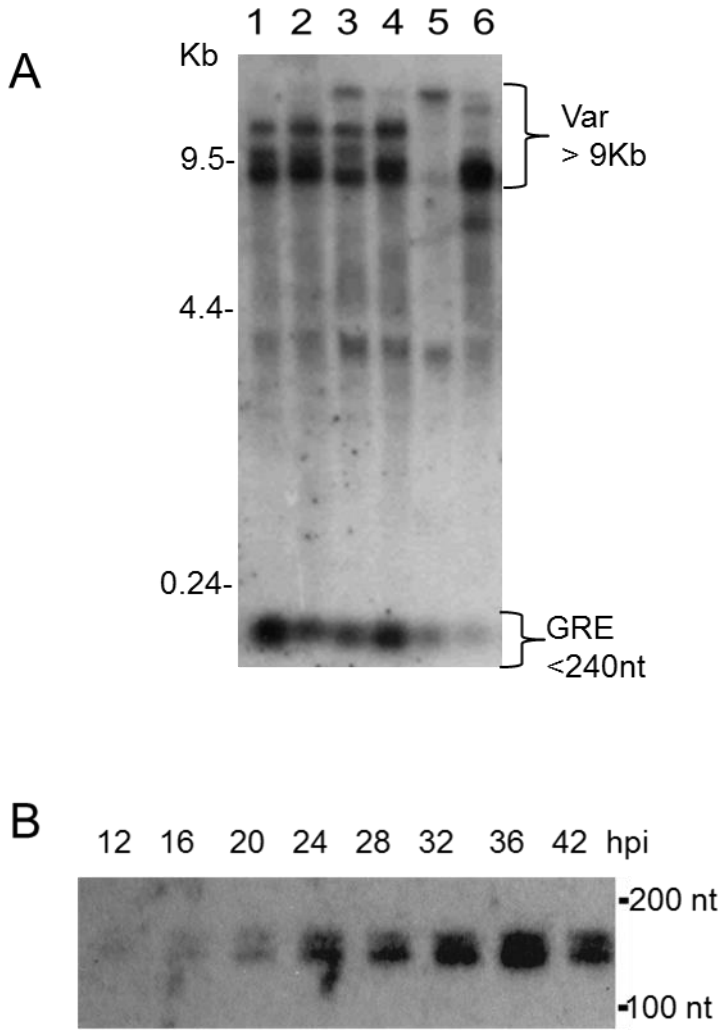


Figure 1.9 GRE is transcribed as a 135 nucleotide transcript during the asexual stages of *P. falciparum*.
A. Northern blot of total RNA from ring stages of six 3D7 clones. A small transcript is detected in all clones. 1% agarose gel, hybridized with the generic exon 2 probe, varC. This was then hybridized with a GRE probe.
B. Northern blot of total RNA extracted at 4 hour intervals from a *P. falciparum* IT culture. A transcript of approximately 140 nucleotides is detected at every time point in the erythrocytic cycle. 15% Acrylamide / 7M urea gel and GRE probe.

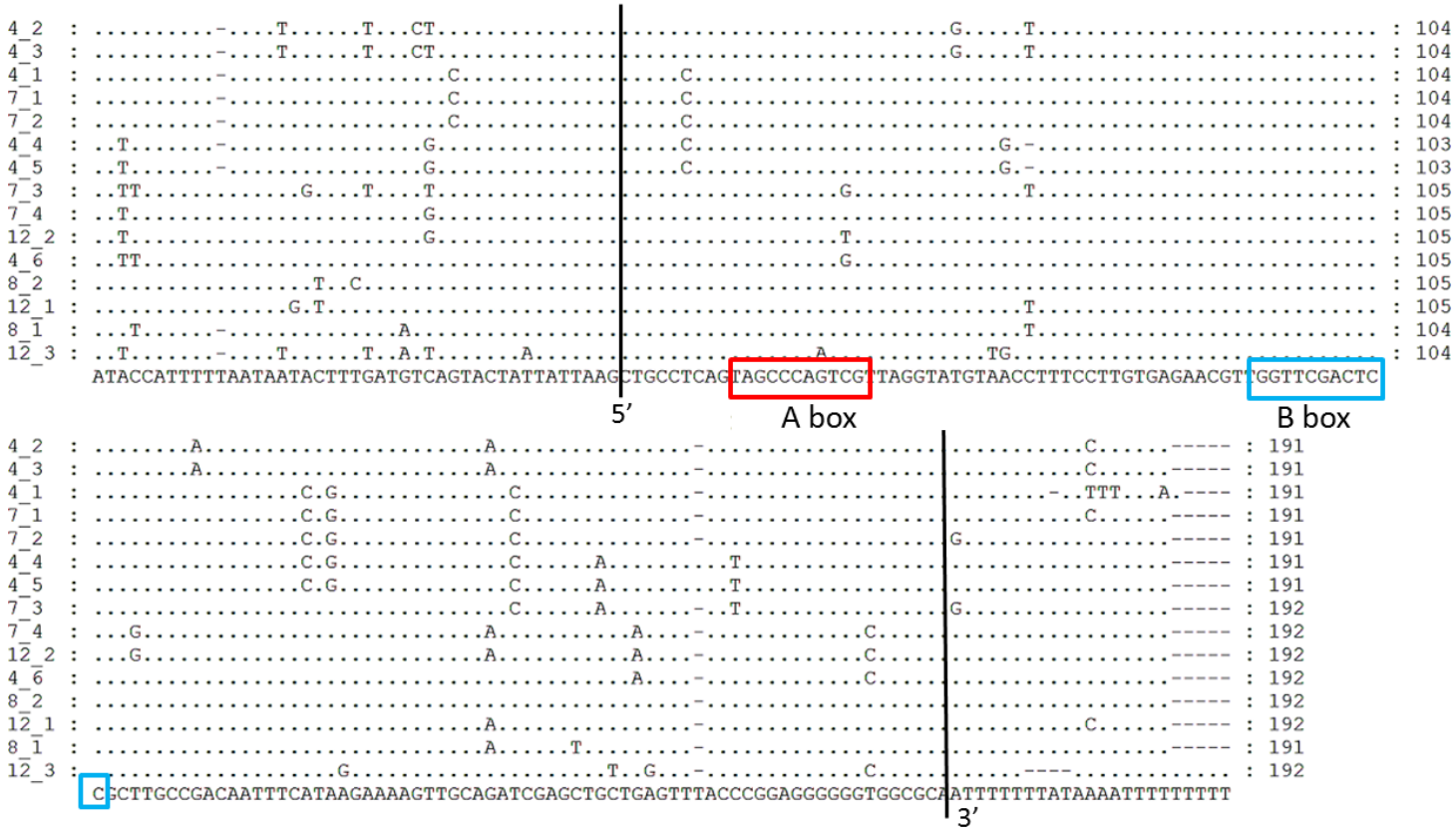


Figure 1.10 GRE sequence features

Alignment of 3D7 GRE sequences showing consensus sequence underneath, and deviations from the consensus sequence. The 5’ and 3’ ends of the transcript are shown as determined by RACE and GeneRacer. The RNA polymerase III promoter sequences A and B box are shown in red (A) and blue (B) boxes.

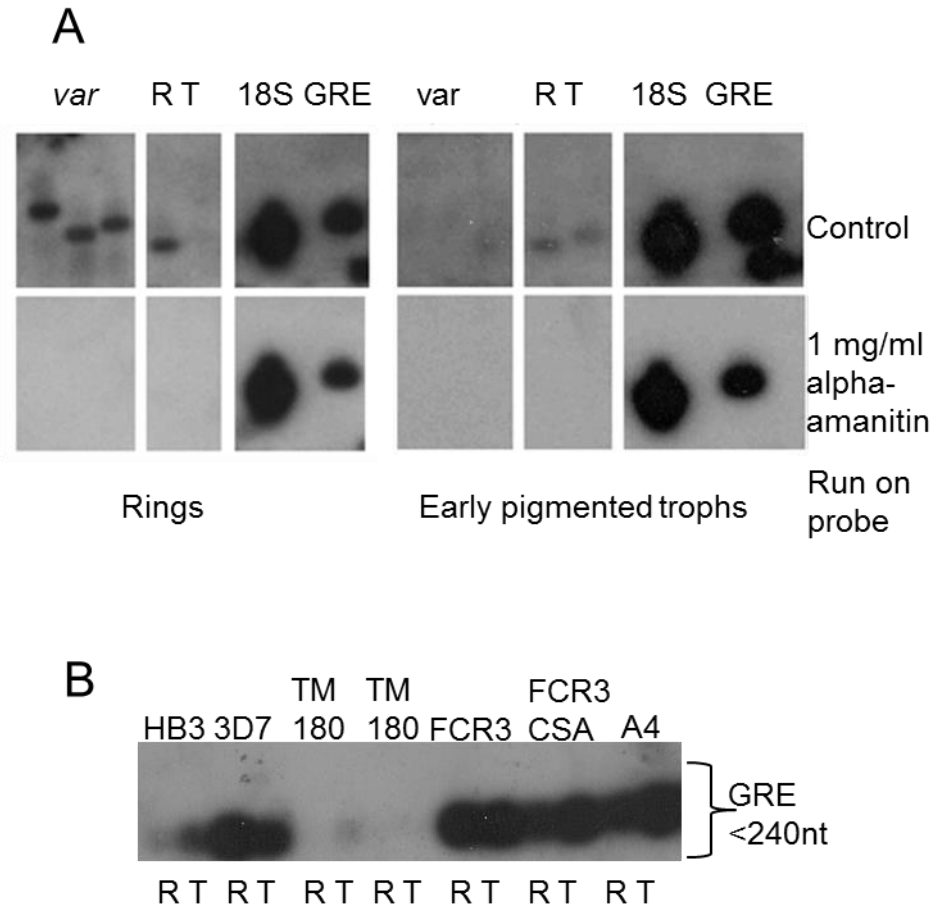


Figure 1.11. GRE is not transcribed by RNA polymerase ll, and is transcribed in many *P. falciparum* isolates.
A. The RNA polymerase ll inhibitor alpha-amanitin has no effect on GRE transcription. Nuclear run on probes made at rings and early troph stages hybridized to Southern blots containing the following PCR fragments; *var* = A4 var ICAM exon1(RNA pol ll transcript); R = KAHRP, ring-stage control; T = MSP-1, troph-stage control; 18S = 18S rRNA (RNA pol l transcript); GRE = GC rich element. B. GRE is transcribed from several *P.falciparum* isolates, but not TM180. Northern blot of total RNA from several isolates, on 1% gel, hybridized with a GRE probe.

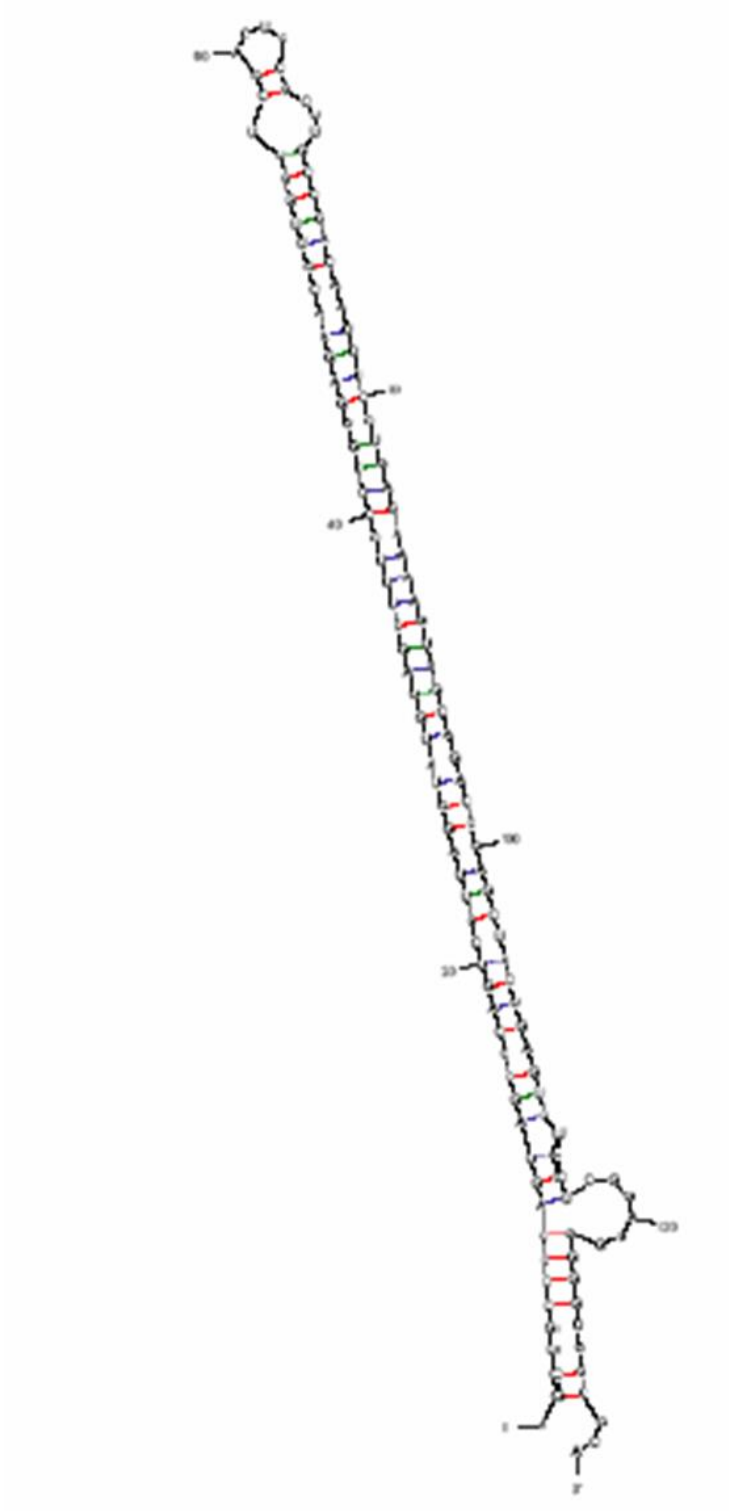


Figure 1.12 Mfold prediction of GRE structure

Although only 3D7 4_5 is shown, all 3D7 GRE predict similar structures.

Initial experiments comparing *var* gene and GRE expression found one case where a GRE and an UpsC *var* are transcribed from the same cluster on chromosome 12 (our unpublished data). These preliminary data suggest that the GRE may have a role in *var* gene regulation. Increasingly, ncRNAs are reported to have roles in a range of processes including RNA processing, protein trafficking and regulation of gene expression (reviewed in (127,128)). As they are a central theme of this thesis, the ncRNA field will be discussed below.

Non-coding RNA

One of the central dogmas in molecular biology has been that DNA is transcribed to make RNA, which in turn is translated to make protein. It has long been known that there are exceptions to this rule, in the form of tRNAs and rRNAs whose biological role do not require them to be translated into protein. Originally thought of as transcriptional noise, since the late 1990s, there has been a growing number of RNA transcripts discovered that are not transcribed into protein, but have a biological function as RNA. These are the non-coding RNAs (ncRNA). The discovery of these transcripts which number in an order of magnitude greater than that of predicted genes, for example in the mouse genome, revealed a possible role for the DNA previously assumed to be untranscribed (129). Many functions have now been assigned to ncRNAs and figure 1.13 gives an overview of their roles in gene regulation in eukaryotic cells.

Although ncRNAs are a structurally diverse group, they do collectively share some broad similarities in their activity. As biological molecules, their specificity is due to standard base pairing. They function,

not as an effector directly, but as a bridge, between a target and a protein or protein complex. Collectively these proteins are referred to as RNA binding proteins (RBP). When associated with ncRNA molecules, the complex is called a ribonucleoprotein (RNP).

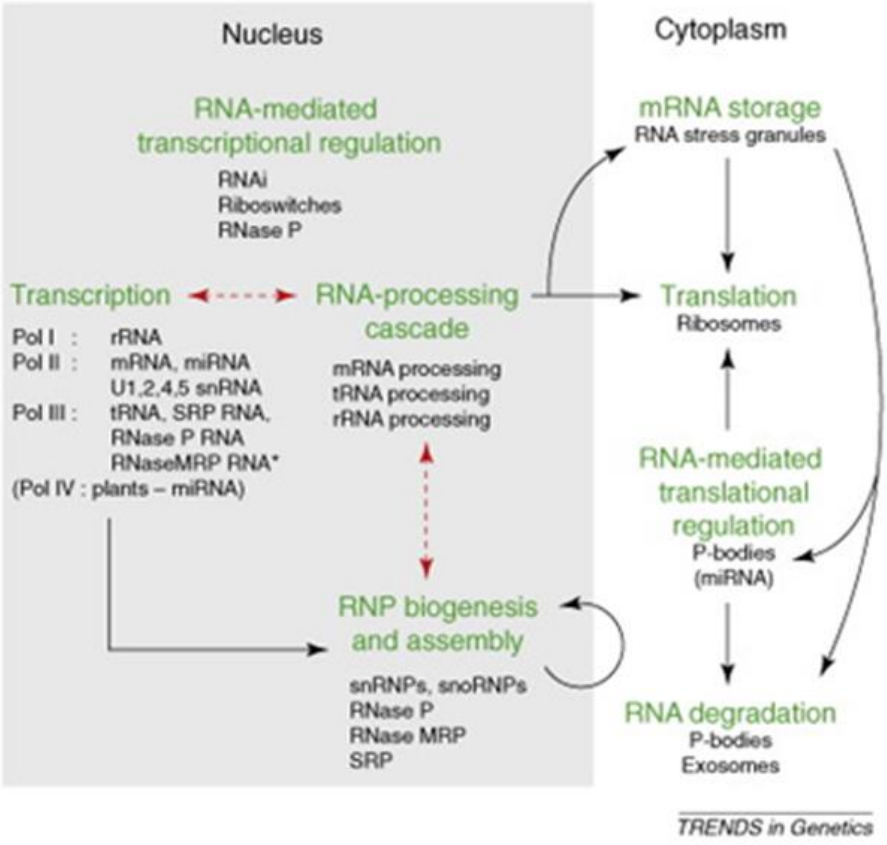


Figure 1.13 RNA infrastructure in Eukaryotes. Roles of ncRNA transcripts in gene regulation in the nucleus and cytoplasm of eukaryotic cells. *MRP RNA is transcribed by Pol III in humans but by Pol II in *S. cerevisiae*. Reprinted from Trends in Genetics, **25,** Collins**,** L. J. and Penny, D.,The RNA infrastructure: dark matter of the eukaryotic cell?, 120-128, 2009, with permission from Elsevier (130).
http://dx.doi.org/10.1016/j.tig.2008.12.003

As more transcripts are described, and more roles are attributed to ncRNAs, classification systems expand and change. At the time of writing, Rfam, an ncRNA database, divides ncRNA into families based on a

common ancestor. Currently it has 1973 families (131). A second database, NONCODE 3.0, classifies its 411,500 entries from 134 traditional classes into 26 process function families (132). A description of the total ncRNA field would be beyond the scope of this thesis, but there are several classes of well described ncRNA which will be discussed below.

RNAi

Collectively, the regulation of gene expression by small ncRNAs is known as RNA interference (RNAi). The RNA components of RNAi are either endogenous microRNA (miRNA), or exogenous small interfering RNA (siRNA), both of which were originally described in *Caenorhabditis elegans* (133,134). RNAi by micro RNA (miRNA) is a major source of regulation in the human genome, as computational estimates predict miRNAs regulating up to a third of all genes (135). RNAi as an experimental regulator of gene expression has been a growing field, as endogenous siRNAs can be introduced to knock out expression of a specific gene. This has moved on as siRNAs are now in clinical trials to treat a variety of conditions including; age- related macular degeneration, viral infections and tumours (136).

There are similarities between RNAi involving siRNA and miRNA. Both are processed in the cytoplasm, by Dicer from longer precursors, and are then assembled with Ago and other proteins into the RNA-induced silencing complex (RISC) (for review (137)). The main components of RISC are the RNA and an argonaute (AGO) protein, but other proteins are also involved. AGOs are a family of proteins found in most eukaryotes

and some prokaryotes. Many proteins involved in the RNAi pathway such as Dicer and AGO are highly conserved within eukaryotes, however these components are not detected in the *P. falciparum* genome (81). That RNAi was not a method of gene regulation in *P. falciparum* was confirmed in 2006 (138). Rathjen and colleagues were unable to detect any *Plasmodium* specific sequences when cloning miRNA from IRBC. The lack of endogenous components of RNAi renders this gene specific knock-down tool ineffective in *Plasmodium* research.

MicroRNA (miRNA)

miRNAs are endogenous RNAs, transcribed as 60-100 nucleotide pre-miRNAs in the nucleus where they are cleaved to 70 nt pre-mRNAs. They are then transported to the cytoplasm where they are processed by Dicer into 22 nt miRNAs (139). These miRNAs assemble into RISC, leading to degredation of the mRNA, or translational repression. For example, *Drosophila melanogaster* has two AGO proteins, Ago-1 and -2. Ago1 acts by two independent processes, both of which involve the P-body GW182 protein. GW182 recruits either the mRNA poly(A) deadenylation complex or mRNA decapping complexes leading to mRNA degradation (140,141). However, when Ago2/RISC is complexed with mRNA it inhibits the interaction of initiation factor eIF4E with eIF4G. This prevents translational initiation (141).

Small interfering RNA (siRNA)

siRNAs are 21-23 nucleotide RNAs processed from double stranded exogenous RNAs (for review (139)). These exogenous RNAs can be introduced experimentally, or arise from transposons or viruses. Within

RISC, the assembly of siRNA onto AGO acts as a guide for AGO which then assembles a target mRNA into the complex and cleaves it. The guide/target assembles via a 2-8 nucleotide sequence in the siRNA.

Small ncRNA

snRNA

Small nuclear RNAs are the RNA component of small nuclear ribonucleoproteins (snRNP). The snRNP form the core of the spliceosome, catalysing the removal of introns from mRNA. Each snRNP consisits of a uridine rich RNA, a core of 7 “sm” proteins and other proteins, each specific for each snRNA (reviewed in (142)). Eighty-nine proteins were identified in a study of the yeast spliceosome (143). Eighteen putative protein components of the *P. falciparum* spliceosome have been annotated in Genedb (79). There are two classes of snRNA, the sm class of U1, U2, U4 and U5 are RNA polymerase ll transcripts; the lsm class, containing U6, are RNA polymerase lll transcripts. The *Plasmodium* snRNAs were first identified by Upadhayay, and were confirmed as 105-201 nucleotide transcripts by Chakrabarti and colleagues (125,144).

snoRNA

Small nucleolar RNAs are located in the nucleolus where their function is to process and modify other ncRNA, most notably rRNAs (reviewed in (142)). There are two classes of snoRNA, C/D and H/ACA RNAs. C/D RNAs direct 2’-O-ribose methylation. C (UGAUGA) and D (CUGA) are conserved sequence elements that interact to restructure the RNA, creating a new binding site which is recognised by specific proteins. Thirty-

three C/D snoRNAs have been identified in *P. falciparum* using bioinformatics screens (125,145,146). The H/ACA snoRNAs are more difficult to determine using bioinformatics as the H box element is not as conserved (ANANNA). Eleven have been detected so far in *P. falciparum* (125,146) These RNA direct pseudouridylation of rRNAs. In common with all ncRNA, both classes of snoRNA can only function in partnership with RNP. Homologues of the majority of the main protein components of eukaryotic snoRNP complexes have been identified in *P. falciparum* (145).

tRNA

Transfer RNAs are a well-established group of ncRNA that are aminoacylated with a specific amino acid corresponding to the anticodon site of the tRNA. Binding of the tRNA to mRNA by sequence specificity catalyses the addition of the accompanying amino acid from the tRNA onto the growing peptide. Forty-four tRNAs have been identified in *P. falciparum (57).* tRNAs are transcribed by RNA polymerase lll.

rRNA

Ribosomal RNAs are a second well characterised group of ncRNA that form the RNA component of the ribosome. *P. falciparum* has two types of rRNA genes, S and A expressed in the sexual and asexual stages respectively. They are transcribed by RNA polymerase l into 18s, 5.8s and 28s rRNA (57).

Antisense transcripts

Although not classified into other ncRNA groups by size or structural elements, antisense transcripts (asRNA) are an important part of the transcriptome. In other eukaryotes the pre-miRNAs are a large part of the

asRNA. Antisense RNA have been reported in *P. falciparum* from several loci including *msp-2* and *var* (64,147) and the first report of widespread asRNA in *P. falciparum* was in 2004 (148). As molecular biological techniques have advanced, so has the analysis of both the transcriptome and asRNA transcriptome. Raabe and co-workers reported 328 asRNA from genes involved in metabolic pathways, transcriptional and translational regulators and both *var* and *rif* genes (146). This contrasts with a more recent report which detected many asRNAs, the majority of which were from proteins of unknown function (149). The discrepancy could be due to the lifecycle stages sampled. The first report looked solely at erythrocyte stages, whereas the second also looked at gametocyte stages. They found that the GV stage had the largest number of antisense transcripts, but only looked at the function of 312 genes with high levels of antisense transcripts, many of which would be from gametocytes. Antisense RNA have been detected from both introns and exons, but the majority of antisense transcripts reportedly come from UTR on the opposite strand. It is possible however, that some of these are overlapping UTRs. An alternative study using RNA-seq reported the higher AT content of UTRs leads to them being poorly mapped (150). In humans, 87% of asRNA are from overlapping UTRs, mostly from overlapping 3'UTRs (151,152).

Long ncRNA

Telomerase RNA

Telomerase RNA is the RNA component of RNP complexes that protect the end of telomeres. The *P. falciparum* telomerase RNA was identified in Chakrbarti's screen of ncRNA (125). The RNA is 2.2 Kb long with an

undefined structure, but containing 2 regions that could potentially form a pseudoknot, a conserved feature from other species. The *P. falciparum* telomerase reverse transcriptase (PfTERT), the enzymatic component of the complex, is a 280 kDa protein (153).

Other long ncRNA

In *P. falciparum*, the lncRNA-TARES, associated with UpsB var genes, described previously, are predicted by RACE to produce two transcripts 1.5 Kb and 3.1 Kb in length (103). Other novel long ncRNA transcripts up to 8.9 Kb in length were detected by RNA-seq sequencing of the *P. falciparum* transcriptome, but have yet to be characterised (150). Long ncRNAs are found in many species and many are well characterised. They act by recruiting chromatin activating or repressor complexes to target *loci*. An example of this are the dosage compensation *roX1* and *roX2* ncRNAs in *Drosophila*, these RNAs form a complex with 5 proteins which bind to multiple sites on the male X chromosome. This binding results in increased transcription from X-linked genes (154).

Potential roles of GRE

Clearly the roles of ncRNA are many, and wide-ranging. As the GRE is transcribed as a ncRNA we would like to identify its function within the parasite. Comparisons of *Plasmodium* genomes that have been sequenced to date reveal that the GRE are only present in the *P. falciparum* and *P. reichenowi* genomes. Similarly, *var* genes are only present in these two species (155,156) Evolutionary analysis based on rRNA sequences hypothesise that the two species diverged 5-9 million years ago, around the same time as the divergence of humans and

chimpanzees (157). As both of these features are specific to the two genomes we can speculate that they came into the genome together. This must have occurred after the split of *P. falciparum* and *P. reichenowi* from the other species, but before their divergence from each other. Because of its location solely in central *var* clusters, and its transcription from the same cluster as a dominant *var* gene, we speculate that it may have a role in *var* gene regulation. However, *var* gene regulation is not the only potential function for GRE. Another theory is a possible role as a recombination hotspot. In yeast, there are three classes of meiotic recombination hotspots. The γ class is associated with peaks of high G+C content (158). Additionally, whilst looking for repetitive sequences with the potential to be involved in recombination events at syntenic breakpoints between *P. falciparum* and rodent malaria species, Kooij and colleagues reported a new family of genes associated with synteny breakpoints (159). These *var* internal cluster associated repeat *(vicar)* genes are the GRE. Incidentally, although they identified putative signal peptides and transmembrane domains to create open reading frames, our data support only a ncRNA product.

Tools for experimental analysis

Widely available tools

Several experimental tools make it possible to investigate these theories. Work by Trager and Jensen (160) established *in vitro* culture methods for the blood stages of *P. falciparum*, allowing long term maintenance of these parasites. Reported in 1987 (161), the progeny of a genetic cross between the 3D7 and HB3 strains of *P. falciparum* are a useful tool to study malaria

genetics. Analysis of the progeny makes it possible to look at meiotic recombination. The availability of sequence data for the 3D7 and HB3 clones allows for comparison between them. More importantly, it enables testing of hypotheses in more than one laboratory clone (57,73). A set of qRT-PCR primers exists for 3D7, allowing *var* analysis in this isolate (56).

In house tools

Within our laboratory we have a set of HB3 clones derived by limiting dilution. As clonal populations, they should initially be expressing a single *var* gene so are ideal for studying *var* and GRE expression. To investigate mitotic recombination we have two clones of NF54. One was selected for rosetting at cycle 0, the other, was maintained in culture for 100 generations.

Experimental Plans

As a general approach to evaluating the importance of ncRNA in malaria, and more specifically in *var* gene regulation, genome wide screens for novel RNAs will be undertaken.

To follow *var* expression in both the 3D7xHB3 progeny and the HB3 clones, it will be necessary to establish a set of qRT-PCR primers for HB3 *var* genes. Using the HB3 clones which should be expressing a single *var* gene will enable me to accurately correlate GRE expression. I will be able to follow any *var*/GRE relationship over time, to see if GRE expression alters as a result of *var* gene switching, or vice versa. If such a relationship exists, I will investigate it further by generating transfection constructs containing GRE and *var* promoter sequences.

By using clones that have undergone meiosis alongside clones that have been maintained for ~100 cycles *in vitro* we will be able to assess GRE as both meiotic and mitotic recombination hotspots.

To complement investigations into the role of GRE, attempts will be made to identify its biological function. This will be done using electrophoretic mobility shift assays (EMSAs). If binding activity is found, proteomic techniques will be used to identify the components.

Aims

The main aim of this work is to investigate the possible role of the GRE in *var* gene regulation or in recombination. To do this will involve the following elements.

- Conduct a genome wide screen of ncRNAs.
- Establish a set of qRT-PCR primers for HB3 *var* genes.
- Investigate GRE analysis techniques.
- Investigate the role of GRE in *var* gene expression and switching.
- Investigate the role of GRE in both meiotic and mitotic recombination.
- Identification of GRE RNA binding proteins.

Chapter 2

Materials and Methods

Some techniques used in this thesis are relevant for specific chapters only, and are described therein. Many methods however are used across some or all of the chapters. These common techniques are described below. Unless otherwise stated, chemicals were sourced from Sigma and plastic ware from BD Falcon.

Culture techniques for *P. falciparum*

Parasites were cultured *in vitro* using modifications of the techniques described in Trager and Jensen (160), Haynes (162), and Lambros and Vandenberg (163).

Media and solutions

Unless otherwise stated, all solutions were pre-warmed to 37°C before use.

Serum free RPMI (RPMI-SF): RPMI-1640 Medium supplemented with 37.5 mM HEPES; 20 mM glucose; 2 mM glutamine; 25 µg/ml gentamicin sulphate and 100 µM hypoxanthine.

Complete RPMI (RPMI-C) was RPMI-SF supplemented with 10% pooled human serum.

Washed human erythrocytes (RBC): 15 mls of white cell depleted type O blood (National Blood Service) was pelleted for 15 minutes at 3000 rpm. The pellet was washed with 10 mls RPMI-SF, to remove any remaining

buffy coat, and centrifuged again at 3000 rpm for 10 minutes. The pellet was then resuspended in an equal volume of RPMI-SF to give a 50% suspension of RBC.

Glycerolyte solution: 6.2M glycerol; 0.14M sodium lactate; 4 mM potassium chloride; 25 mM sodium phosphate. pH to pH 6.8 with sodium hydroxide. Make volume up to 250 mls with sterile water. Filter sterilise and store at 4°C.

Gassing of cultures

A mixture of 1% oxygen, 3% carbon dioxide, 96% nitrogen was passed over the surface of the culture for 1 minute, whilst gently agitating the flask.

Pelleting *P. falciparum* cells

Unless otherwise stated, cells were pelleted by centrifugation (1800 rpm, 5 minutes, room temperature) and the supernatant aspirated.

Thawing *P. falciparum* stabilates from liquid nitrogen storage

To culture a *P. falciparum* isolate**,** a cryovial was removed from liquid nitrogen and thawed at 37°C. The contents were transferred to a 50 ml tube and 1/5th volume of 12% NaCl was added slowly over 5 minutes, then allowed to stand for 5 minutes. Five volumes of 1.8% NaCl were added slowly and again allowed to stand for 5 minutes. Five volumes of 0.9% NaCl / 0.2% glucose were then added, and the cells pelleted. The cells were then washed with 30 mls RPMI-SF and centrifuged again. The pellet was resuspended in a small volume of RPMI-C, and transferred to a tissue culture flask, containing 50 mls RPMI-C and 500 µl RBC. Finally the culture was gassed, and then incubated at 37°C.

Culture conditions

To maintain *P. falciparum* in continuous culture, every 24 or 48 hours a 200 µl sample was removed and examined microscopically by thin smear. Thin smears were air dried, fixed with methanol, and then stained for 10 minutes with Giemsa's staining solution. Two parameters were assessed; growth by percentage parasitaemia and stage in the erythrocytic cycle. Cultures were then diluted with fresh RPMI-C and washed RBC to 1-2% haematocrit and 1-2% parasitaemia, then gassed and incubated at 37°C.

Synchronization of cultures

During continuous culture, *P. falciparum* becomes asynchronous. This results in a culture containing a mixture of erythrocytic stages. Two methods are available to enrich for specific developmental stages. Plasmagel flotation is used to enrich for trophozoites but was not used in this study. To synchronise a culture leaving predominantly ring stage parasites, sorbitol synchronization was routinely used (163). A culture was pelleted in a 50 ml centrifuge tube and the supernatant aspirated. The cell pellet was resuspended in 10 volumes of 5% sorbitol solution and incubated at 37°C for 20 minutes. The cells were pelleted and washed in RPMI-SF. Following a final centrifugation, the cells were resuspended in RPMI-C and RBC in culture as above.

Preparing culture stabilates

To prepare a stabilate of *P. falciparum* for long term storage in liquid nitrogen, it is necessary to use ring stage parasites. Cultures were examined by Giemsa stain to confirm the stage of growth. Ring stage cultures were transferred to a 50 ml centrifuge tube and pelleted by

centrifugation. When aspirating the supernatant, 500 µl of supernatant was left and used to resuspend the cell pellet. Glycerolyte solution was added in a ratio of 5 volumes of glycerolyte to 3 volumes of cell pellet. The first volume was added over 5 minutes then left to stand for 5 minutes. The remaining 4 volumes of glycerolyte were added slowly. The mixture was then transferred to cryovials in 800 µl aliquots and stored at -80°C before moving to liquid nitrogen.

Parasite isolates

The *P. falciparum* isolates and clones used in this study are detailed in table 2.

Molecular biology techniques

Many of the following molecular biology techniques are standard methods that are found in “Molecular Cloning – a laboratory manual” by Sambrook *et al* (164).

Common solutions

20x SSC	3 M NaCl, 0.3 M Na Citrate, pH 7
1x TBE	0.089 M Tris, 0.089 M Boric acid, 2 mM EDTA
TE	10 mM Tris pH8, 1 mM EDTA
LB Media	10 g Tryptone, 10 g NaCl, 5 g Yeast extract per litre
LB Agar	LB plus 20 g Agar per litre
Ampicillin	Stock is 50 mg/ml in 50/50 water/ethanol. Use at 50 µg/ml

isolate	origin	experiment	reference
3D7	Cloned from NF54,an isolate from Schipol airport, (groups with African isolates) (74)	ncRNA cloning GRE*/*var* expression	(76)
HB3	Honduras I/CDC (Central America) (74)	GRE/*var* expression	(77)
HB3 clones	Cloned by limiting dilution from HB3	GRE/*var* expression	
X-series	Progeny of cross between 3D7 & HB3	GRE/*var* expression	(161)
NR13-0	Cloned by limiting dilution from rosetting selected NF54. Cycle 0	Pfu testing	(165)
NR13-F	As NR13-0, but grown on for 100 cycles.	Pfu testing	(165)

*GRE G-C rich element

Table 2.1 *P. falciparum* clones and isolate used in this study

Genomic DNA (gDNA)

Liquid genomic DNA was prepared from trophozoite stage parasites. Parasite cultures were pelleted in 15 ml tubes, and the resulting pellet frozen. Haemoglobin was removed, after thawing, by washing the pellet with PBS and pelleting the cells at 3,000 rpm for 5 minutes. The washing was repeated until the supernatant was no longer red. After the final supernatant removal, the pellet was resuspended in PBS to 600 µl final volume. The following was added to the pellet, and mixed gently: 90 µl 10X TN9 (500 mM Tris pH 9; 2 M NaCl); 180 µl 0.5 M EDTA and 18 µl 50 mg/ml proteinase K. Finally, 30 µl of 20% SDS was added and mixed gently for 30 seconds. This was incubated at 50°C overnight.

On day 2, the solution was extracted with 1 volume phenol (equilibrated with TE pH 8) for 1 hour on a rotator (Stuart Scientific). The phases were separated by centrifugation in a microfuge at 9000 rpm for 5 minutes. The aqueous phase was re-extracted with 1 volume phenol as above. It was then extracted with 1 volume of 24:1 Chloroform: Isoamyl alcohol and spun as above. The aqueous phase was removed to a new tube and precipitated with 0.1 volume 3M Sodium acetate and 0.6 volumes isopropanol. The DNA was spooled out from the liquid with a plastic bacteriological loop and allowed to drain. The DNA was then resuspended in 0.5 ml TE overnight at 4°C and the concentration measured in a Genequant spectrophotometer (Pharmacia).

Block DNA

To preserve intact chromosomes for Pulse field gel electrophoresis (P.F.G.E.) parasites must be immobilised in agarose blocks prior to lysis. The following protocol produces two concentrations, high for restriction digestion and low for chromosome separation. Parasites at a packed cell volume of 500 µl and 10-15% parasitaemia were pelleted by centrifugation. The pellet was resuspended in 500 µl of 1.5% Low melting point agarose (LMP) (GIBCO) in PBS and 900 µl was put into 100 µl block formers (BioRad). A further 500 µl of the LMP solution was added to the remainder and 6 more blocks were poured. The blocks were set at 4°C for 1 hour then incubated twice overnight in 1% sarkosyl, 0.5 M EDTA, 0.125 g protease K at 50°C. Blocks were washed 4 times in 10 mM Tris, 50 mM EDTA (TE50) at 50°C and stored in TE50 at 4°C.

Saponin lysis for human erythrocyte microRNA removal

Preliminary experiments for small RNA cloning showed that human microRNA was the main species identified. To remove this prior to cloning, saponin was used to lyse erythrocyte cell membranes. This releases the parasites and the human miRNA is washed away.

Parasite culture was pelleted in 15 ml tubes and washed with phosphate buffered saline (PBS). Following a second spin the cell pellets were resuspended in 5 volumes of 0.01% saponin in PBS and mixed by inversion. The cell solution was then transferred into 1.5 ml tubes and spun in an Eppendorf microfuge at 13.2Krpm (30 seconds for trophozoites or 1 minute for ring stage). The supernatant was aspirated and pellets resuspended in 0.01% saponin in PBS. This was repeated, pooling the pellets together, until most of the haemoglobin had been washed away. Trophozoites were pooled as 1 ml starting cell pellet volume per final tube, and 1 ml TriZol reagent (Invitrogen) added per tube. Rings were pooled as 3 ml starting pellet volume per final tube and 1 ml TriZol. RNA preparation continues after the TriZol stage in the method below.

RNA preparation

RNA and northern blots were prepared by the method of Kyes et al (166). Parasites were harvested at 5-12% parasitaemia by centrifugation at 1800rpm for 4 minutes and the supernatant aspirated. For ring stages, 10 pellet volumes of prewarmed TriZol reagent were added and for trophozoites. 20 pellet volumes were used. The tubes were mixed then incubated at 37°C for 5 minutes. 0.2 volumes of bromochloropropane were added per volume of TriZol and the tube shaken for 15 seconds. Samples

were centrifuged either in a Sorvall centrifuge, at 3000 rpm, 4°C for 30 minutes, or in a microfuge at 13.2 Krpm at 4°C for 25 minutes. The aqueous supernatant was removed to a new tube and precipitated with 0.5 volumes isopropanol per volume of TriZol. Following an overnight precipitation at 4°C, the tubes were centrifuged in a microfuge at 13.2 Krpm, 4°C for 30 minutes. The pellets were resuspended in 100 µl formamide per ml of starting pellet volume, then heated at 60°C for 10 minutes. The concentration of RNA was measured in a Genequant spectrophotometer. Samples were stored at -80°C.

Electrophoresis of RNA and northern Blotting

RNA species were resolved by electrophoresis and blotted onto nylon membrane to enable specific sequences to be identified by hybridization with radiolabelled probes, Electrophoresis tanks (Anachem model H-3) were soaked in 3% hydrogen peroxide solution for 10 minutes then rinsed with distilled water prior to use. Agarose gels (Invitrogen), 0.8%, were prepared in 1xTBE with a final concentration of 5 mM guanidine thiocyanate added before casting.

RNA samples were diluted to 2-10 µg in formamide and denatured at 65°C before loading. Electrophoresis conditions were 110 V for 10 minutes followed by 80 V for 3 hours. Gels were stained with ethidium bromide and visualised on a GelDoc system (BioRad). Gels were soaked twice in 7.5 mM sodium hydroxide for 15 minutes. RNA was blotted onto HybondN+ (GE Healthcare) by capillary transfer in 7.5 mM sodium hydroxide overnight. Following transfer, filters were neutralised in 2xSSC buffer for 15 minutes then air dried.

Hybridization with radiolabelled probes

Two methods were used to generate radioactively labelled probes for use in hybridization experiments.

Random Prime Labelling

PCR products larger than 350 base pairs were labelled using the Megaprime DNA Labelling System (GE Healthcare). With this method, complimentary DNA is synthesised by Klenow DNA polymerase from nonamer primers hybridized to a denatured template. By the addition of a labelled dNTP with three unlabelled dNTPs, the resulting DNA fragments are radioactive (167,168) .Reactions were as follows. 25-40 ng DNA in 23 µl water was denatured at 95°C for 5 minutes with 5 µl randon nonamer primer solution. Ten times buffer (5 µl), 4 µl each of dCTP, dGTP and dTTP and 2 µl Klenow were added, along with 3 µl α^{32}P dATP (specific activity 370Mbq/ml) (GE Healthcare or Perkin Elmer). This was incubated at 37°C for 30 minutes before a final denaturing step and addition to the hybridization solution.

End Labelling of Oligonucleotides

The end labelling reactions were as follows: 0.5 µl of each oligonucleotide (10 µM stocks), 4 µl 5X T4 Polynucleotide kinase (PNK) buffer, 1 µl T4 PNK (New England Biolabs), 1 µl γ^{32}P ATP (specific activity 370Mbq/ml) (GE Healthcare or Perkin Elmer) and water to 20 µl then incubated at 37°C for 1 hour. The probe was then added to the hybridization solution.

Hybridization

Specific conditions for each probe will be given in detail in the following chapters, but they follow the same outline. Filters were prehybrized for 2

hours at the same temperature as hybridization. Oligonucleotide probes were hybridized at 37^{o}C and generic probes at 50^{o}C.The hybridization temperatures were established empirically for single copy probes.The hybridization buffer contained: 0.25 M NaH_2PO_4, 0.25 M NaCl, 1 mM EDTA, 7% SDS and 5% dextran sulphate. Either 5 ml or 12 ml of buffer was used depending on the size of the hybridization bottles. Hybridizations were overnight in glass bottles in a rotating oven (Hybaid). Filters were then washed in SSC solution, wrapped in Saran wrap (Appleton Woods) and exposed to film overnight, or as described in results.

The exon-2 probe is a generic probe used to detect all the *var* genes in a blot and is used routinely in this work. Originally described by Rubio et al as Var-C, it was generated by PCR primers to the exon 2 sequence of 3D7-*var*1. The Var-C sequence is the most highly conserved region in exon 2, consequently when gDNA is used as the template the majority of *var* genes are amplified (169).

PCR

Routine PCR

$MgCl_2$ concentration, annealing and extension temperature and extension time vary between primer pairs and will be specified when required. Unless specified other PCR conditions were as follows:

Each 50 µl reaction contains: 1 µM of each primer; 1x PCR buffer ll (Applied biosystems); 1.5-3.5 mM $MgCl_2$; 200 µM each of dATP, dCTP, dGTP and dTTP; 1.25 units Amplitaq (Applied biosystems); 100ng gDNA and water to 50 µl.PCR cycling conditions were 94°C for 4 minutes, followed by 35 cycles of 94°C for 30 seconds, annealing temperature for

30 seconds and 65°C for 30 seconds. Extension temperature is reduced to 65°C for A+T rich DNA (170). For PCR reactions involving GRE this is increased to 72°C as it has a higher G+C content.

High-fidelity PCR

Where sequence fidelity was required, particularly for expression analysis of GRE and generating components for transfection constructs, high-fidelity DNA polymerases were used. KAPA HiFi (Labtech) was used for GRE studies and transfection constructs. Phusion (NEB) was also used for transfection constructs. In both cases specific details of the kit components are withheld by the manufacturer.

KAPA HiFi

A 50 µl reaction contained the following: 10 µl 5x KAPA HiFi fidelity buffer, 1.5 µl dNTP mix (10 mM each dNTP), 1.5 µl each 10uM primer, 1 µl KAPA HiFi DNA polymerase (1U/ µl), template 2 µl cDNA or 100 ng gDNA and water to 50 µl. Cycling conditions for GRE PCR were 95°C for 2 minutes, followed by 25 cycles of 98°C for 20 seconds, 45°C for 15 seconds and 68°C for 20 seconds. Changes to these conditions will be described in Chapter 5.

Phusion

A 50 µl reaction contained the following: 10 µl 5x Phusion HF buffer, 1 µl dNTP mix (10 mM each dNTP), 1 µl each 10uM primer, 0.5 µl Phusion DNA polymerase (2U/µl), template 125 ng gDNA or 0.2 µl PCR product and water to 50 µl. Cycling conditions were 98°C for 1 minute followed by 30 cycles of 98°C for 20 seconds, annealing temperature for 20 seconds

and 68°C for the required extension time. Annealing temperature and extension time will be described for each primer pair.

All PCR products were analysed on agarose gels.

Electrophoresis of DNA

To separate DNA fragments or to visualise PCR products, DNA was subjected to electrophoresis on agarose gels. The concentration of agarose used depends on the size of the fragments/products. These are detailed in table 2.2. Agarose was melted in 1xTBE and gels were run in 1xTBE. Small gels (Anachem H1-set) were 50ml and run at 70 V. Larger gels (Anachem H3-set or BioRad wide mini-sub cell GT) were 100 ml and run at 130V. Gels were stained either with ethidium bromide, or WebGreen (Web scientific), then visualised on a GelDoc system.

Southern blotting

DNA was denatured by soaking gels twice in 0.4 M sodium hydroxide for 15 minutes. DNA was blotted onto HybondN+ (GE Healthcare) by capillary transfer in 0.4 M sodium hydroxide overnight. Following transfer, filters were neutralised in 2xSSC buffer for 15 minutes then air dried.

Fragment or product size	Agarose concentration
0-200 base pairs (bp)	2% MetaPhor agarose (Lonza) 1% Agarose (Invitrogen)
200 – 500 bp	2% Agarose
500 bp+	1% Agarose

Table 2.2 Agarose concentrations required to separate ranges of DNA sizes.

Cloning and Sequencing

Ligation

PCR products were cloned either into pCR2.1 or pCR4-TOPO TA cloning vectors (both Invitrogen). These vectors are supplied linearized with single 3' thymidine overhangs at the cloning site. The cloning reaction utilises the single deoxyadenosine added to the 3' end of a PCR product by *Taq* polymerase. With pCR2.1 T4 Ligase is used to ligate the PCR product to the vector but with pCR4-TOPO topoisomerase l is used to perform the cloning. Structurally both vectors are similar, having ampicillin and kanamycin resistance genes and T7 promoters. In general pCR4-TOPO was used to clone libraries of small PCR products, such as GRE clones. However if this was unsuccessful it was repeated with pCR2.1 due to the option to quantify the insert and optimize the reaction. All solutions, excluding PCR products, were supplied by the manufacturer. To clone into pCR2.1, the concentration of the PCR products was estimated by comparison to known standards on agarose gels. The amount of PCR product required to ligate into 25 ng of vector with a 1:1 ratio was determined by:

$$X \text{ ng PCR product} = \frac{Y \text{bp product} \times 25}{3900}$$

The ligation reactions contained: X ng PCR product, 1 µl 10x buffer, 1 µl pCR2.1, 1 µl T4 DNA Ligase and water to 10 µl. The reaction was incubated at 14°C overnight.

To clone into pCR4-TOPO, 0.5 – 4 µl PCR product, 1 µl vector, 1 µl salt solution and water to 6 µl were incubated at room temperature for 30 minutes.

Transformation

Ligated DNA was transformed into One Shot TOP10 competent cells (Invitrogen). DNA, 2 µl, was mixed with a 50 µl aliquot of cells and incubated on ice for 30 minutes. The cells were heat shocked at 42°C for 30 seconds, then returned to ice for 2 minutes. Next, 250 µl S.O.C. medium (supplied) was added and the cells shaken at 37°C for 1 hour. Cells containing pCR4-TOPO were spread onto LB (Agar)+Ampicillin plates. Cells containing pCR2.1 were spread onto LB (Agar)+Ampicillin plates that had been previously spread with 40 µl of 40 mg/ml 5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside (X-gal). Plates were incubated at 37°C overnight.

Screening of Transformants and DNA mini preps

Transformed colonies were screened for inserts before sequencing. Both vectors contain M13 sequence on either side of the cloning site. M13 primers were used to screen by PCR (Appendix i). 10 µl reactions containing 10 ng of each primer, 3 mM $MgCl_2$, 1x buffer, 200 µM each of dATP, dCTP, dGTP and dTTP and 0.25 units Amplitaq. Colonies were picked with a single pipette tip, first into the PCR reaction, then into 100 µl LB+Amp. Cycling conditions were 94°C for 4 minutes followed by 35 cycles of 94°C for 20 seconds, 45°C for 20 seconds and 65°C for 20 seconds. PCR products were run on 2% agarose gels.

The 100 µl LB+Amp inoculated with colonies were incubated at 37°C. PCR positive cultures were used to seed 3ml LB+Amp cultures then shaken at 37°C overnight.

DNA was extracted from the 3 ml cultures using Wizard *Plus* SV Minipreps (Promega) following the manufacturer's instructions. DNA was eluted in 75 µl water and quantified by spectrophotometry or gel analysis.

Sequencing

DNA was sequenced using an in-house service. The service uses BigDye terminator v.1 cycle sequencing chemistry and analyses reactions on an ABI-3730 DNA analyser (Both Applied biosystems). Results were visualised using Chromas (171).

Ethanol precipitation of RNA and DNA

DNA

To precipitate DNA, 1/10 volume of 3M sodium acetate and 2 volumes ethanol were added to DNA. These were mixed and placed at -20°C overnight.

RNA

Precipitation of RNA depends on the size of the RNA. For RNA>500 nucleotides, 1/10 volume of 3M sodium acetate and 2.5 volumes ethanol were added to the RNA. For smaller RNA species, the ethanol is increased to 3 volumes. In both cases, 20 µg glycogen (Roche) was added as a carrier. Precipitation was at -80°C overnight.

To pellet the nucleic acid, the samples were microfuged at 13.2 Krpm, 4°C for 30 minutes. The supernatant was removed and the pellet

air-dried. RNA was resuspended in water and heated to 65°C to ensure complete resuspension. DNA was resuspended in TE at room temperature.

DNase treatment

Before use in RT PCR, RNA was recovered from formamide by ethanol precipitation. Again, component details of the DNase kit were withheld by the manufacturer. For GRE expression 1 µg RNA was resuspended in 8 µl water, 1 µl 10x TURBO DNase buffer and 1 µl TurboFree DNase (Ambion) added, then incubated for 30 minutes at 16^{o}C. For qPCR 4.4 or 5 µg RNA was resupended in 17 µl water. 2 µl buffer and 1 µl DNase was added and the reaction incubated for 20 minutes at 16^{o}C, then a further 1 µl enzyme added for a further 20 minutes. For both treatments 2 µl DNase inactivation reagent was added then microfuged at 10.4 Krpm for 2 minutes. The RNA was then transferred to a fresh tube. Following DNase treatment all RNA samples were tested by PCR to ensure complete removal of gDNA.

cDNA synthesis

For GRE expression, 9 µl RNA was reversed transcribed in a 20 µl reaction. HB3 RNA samples for qPCR were divided into 3 and each 6 µl aliquot used in a 20 µl reaction. 13.75 µl RNA + water were denatured at 70°C for 5 minutes with 1 µl random hexamers (100ng/ µl) (Invitrogen). After 1 minute on ice, 4 µl 5x buffer, 0.25 µl BioScript (200u/ µl) (both Bioline), and 1 µl 10mM dNTPs were added. This was incubated at 40°C for 40 minutes. The enzyme was then heat denatured at 70°C for 10 minutes. cDNA was stored at -20°C in single-use aliquots.

The above method for cDNA synthesis is applicable in several of the following chapters; any alterations to this will be described therein.

Real Time PCR (qPCR)

Generation of a primer set for the HB3 *var* gene repertoire is a significant section of this thesis, and will be described in chapter 4. The conditions for qPCR in 3D7 had however been determined first by Salanti and colleagues (56) and then amended by Andrew Serazin, a D. Phil student in our laboratory (165). Because initial work was to study progeny of a cross between 3D7 and HB3, it was necessary to ensure that the HB3 primer set worked under the same conditions as those for 3D7. As these conditions are identical for all qPCR experiments, they will be described here.

qPCR reactions were assembled using a Corbett CAS1200 liquid handling system, and run on a Corbett Rotor-Gene 6000 (172). Each 10 µl reaction contained 1 µl template, 5 µl SybrPlus (Quantace), 3 µl water, and 1 µl primer mix. The primer mix contained 1.6 µM of each primer, giving a final concentration in the reaction of 0.16 µM of each. The cycling conditions were as follows: Hold, 95°C/10 minutes; Cycling, (95°C/25 seconds, 58°C/25 seconds, 68°C/30 seconds) X 45; Melt, 55°C-99°C. For primer testing, the template was gDNA used at a concentration of 20 ng/µl. For *var* gene expression experiments, an aliquot of cDNA was diluted before use, as will be described in the relevant chapters.

Primer Design

Many primers used in this thesis have been designed “by eye” from genomic sequence. This includes those for High resolution melt PCR,

those to make PCR products to use as probes, and those for oligonucleotide hybridization. The primers for qPCR were designed using Primer3plus (173). The parameters were as follows; product size range 100-150, primer length 20, Tm 60°C, GC clamp 1, salt correction formula and thermodynamic parameters SantaLucia 1998, all other parameters were default settings.

Bioinformatics

Several bioinformatic tools were used routinely in this work. Genomic sequences for *P. falciparum* 3D7 were obtained from PlasmoDB or GeneDB (79,156). Sequences for HB3 were from the Broad Institute (73). Sequences were viewed in the genome browser Artemis (174). Artemis allows features such as gene annotation, CDS, ncRNA to be viewed alongside both nucleotide and amino acid sequences. GRE sequences can be easily identified using the %GC graph function in Artemis.

Searching for sequence similarity was done using the Basic Local Alignment Search Tool (BLAST). The BLAST feature was used in the PlasmoDB website, or the NCBI website (175) to check for similarity with non-*Plasmodium* species. Direct comparison of multiple sequences used ClustalW2 sequence aligner (176). The GeneDoc program was used to edit multiple aligned sequences from ClustalW2 (177).

Chapter 3

Identification of novel non-coding RNA in *P. falciparum*

Introduction

It has long been known that some RNA species, e.g. tRNA and rRNA have specific functions that do not require them to be translated into proteins. During the 1990s a challenge to the DNA→RNA→Protein dogma grew, with the identification of ever increasing numbers of RNAs that are not processed into proteins. These transcripts are collectively known as non-coding RNAs with microRNAs being a class with the striking function of gene specific regulation of expression. Unlike many other eukaryotes, it was known that *P. falciparum* does not have the necessary machinery to utilise microRNA. This suggested the possibility of novel species or genus specific classes of ncRNA.

Following the identification of the GRE in *P. falciparum* (57) and our work showing it to be a ncRNA it was anticipated that many more ncRNAs might be discovered. Here I describe work towards identifying novel *P. falciparum* specific ncRNAs from a genome wide screen.

During the course of this study, seven groups have published results of screens of ncRNAs in *P. falciparum* (125,138,144-146,150,178). The first bioinformatic search by Upadhyay (144) was based on the

principle of a screen of ncRNA in hyperthermophiles (179). These organisms have an A+T rich genome and use increased hydrogen bonding, via higher G+C content, to stabilize rRNA and tRNA secondary structure. The authors successfully demonstrated that other ncRNAs in hyperthermophiles may also have this feature. In *P. falciparum,* the G+C content of tRNAs (56%) and rRNAs (31%) is higher than the overall average of 20% G+C content in the genome. Upadhyay and colleagues identified 18 novel ncRNAs in *P. falciparum* using this same principle. Chakrabarti also looked for regions with high G+C content in their screen (125).

Other bioinformatic approaches involve searches with known sequence motifs from other organisms or using algorithms designed for detecting ncRNA. The rationale followed by Mourier *et al* (178) included only intergenic and intronic sequences, whereas Chakrabarti *et al* (125) analysed the whole genome. Reliance on common sequence-specific motifs might not be expected to identify *P. falciparum*-specific sequences. These bioinformatic methods require experimental confirmation of transcription. Practical approaches start with the entire transcriptome, but size select either the RNA or cDNA before cloning. Highly expressed RNAs such as rRNA are easily identified. Raabe and colleagues sequenced approximately 25,000 clones whilst searching for novel transcripts, of which nearly half were rRNA (146). Transcripts with low expression are less likely to be identified experimentally. A major advantage of a computational screen is that it is not limited by expression levels. If a transcript is expressed at 1 copy per cell it would still be detected if its sequence falls within the given parameters. Stage specificity

is another factor that must be considered. When the sequence of the genome is screened, all stages of the life cycle are automatically included. With a practical approach, only transcripts expressed at the stage of RNA collection are considered. This was addressed by Otto (150) and Raabe (146), but was limited to analysis of asexual stages.

Comparisons of the different methodologies used for ncRNA identification are shown for bioinformatic approaches (table 3.1) and the practical screens (table 3.2). No two methods have detected the same ncRNA repertoire in *P.falciparum,* but there is significant overlap. It is clear that each method has merit, but none is able to identify all of the potential ncRNA species.

Here I describe a practical approach to cloning ncRNAs in asexual stages of *P. falciparum*. This is based on a screen of ncRNA in the social amoeba *Dictyostelium discoideum* using a modification of the GeneRacer (Invitrogen) technique (180). *D. discoideum* has an A+T rich genome of 77.57% (181), making it comparable to *P. falciparum.* Technical problems with cloning A+T rich inserts into *E. coli* appear to have been overcome as this technique was successful. For this reason this method was chosen to clone ncRNAs from *P. falciparum* that would not be identified using bioinformatic methods

Author	Upadhyay (144)	Chakrabarti (125)	Mourier (178)	Mishra (145)
Year	2005	2007	2008	2009
Isolate	3D7	3D7	3D7	3D7
Data source	Plasmodb	Plasmodb	Plasmodb	Plasmodb
Regions screened	Intergenic	Genome	Intergenic and intronic >100bp	Genome
Other exclusions	tRNA, rRNA		Known ncRNA	rRNA
Screening method	GC content of 70bp windows >35%	GC content and RNA folding. Sequence search for conserved classes e.g spliceosomal RNA	RNAz (182)	SNOSCAN (183) Searching for snoRNA only.
Other genome comparison	*P .yoelii, P. knowlesi, P. berghei*	*P .yoelii, P. knowlesi, P. berghei, P. chabaudi, P. vivax, P. reichenowi, P. gallinaceum*	*P .yoelii, P. knowlesi, P. berghei, P. chabaudi, P. vivax, P. reichenowi, P. gallinaceum*	After identification in *P. falciparum* BLAST used to find variants in other *Plasmodium* species.
Final selection	Candiates selected if homologous to sequence in *P.yoelii.*	Candidates selected if conserved in several genomes.	Candidates selected if >70% identity to other genomes and ≥ 50bp.	SNOSCAN score >20(parameter considered to be a positive candidiate)
Verification	Northern analysis	Northern analysis	Northern analysis	Northern analysis and RT PCR.
New sequences identified	18	40 (+ validated expression of 7 identified by other groups)	33	18 boxC/D snoRNAs

Table 3.1 Comparison of bioinformatic screens of *P.falciparum* RNA

Author	Raabe (146)	Otto (150)
Year	2010	2010
Isolate	3D7	3D7
Growth stages	Total erythrocytic stages	7 time points during erythrocytic stage
Pre cloning treatments	Poly acrylamide gel electrophoresis.	rRNA depletion
Size selection	10-60 nt and 60-500 nt	cDNAs between 200-250 bp sequenced.
Cloning method	C-tail, RNA oligo ligation (Sal I site), cDNA synthesis from C-tail (primer with Not I site). PCR, then double digestion and ligation into pSPORT1 vector	cDNA synthesis with oligo(dT) and random primers.
Sequencing method	Not disclosed	Illumina GA ll
	25,000 clones sequenced and assembled into 3588 contigs	
Screening	rRNA, tRNA, snRNA and mitochondrial RNA sequences removed.	rRNA depleted previously
mRNA	mRNA excluded by C-tail strategy	Screen of total transcriptome, includes mRNA
Comments	Strategy to detect ncRNA	Small ncRNAs not expected due to size selection of cDNA
Novel	630 contigs	107 transcripts

Table 3.2 Comparison of experimental approaches to identifying ncRNA

Material and methods

Size fractionation of RNA

Three methods were tested to size fractionate RNA from 3D7 trophozoite parasites.

Microcon columns

Microcon YM-100 columns (Millipore) are designed for removal of salts and other contaminants from samples and to concentrate samples. YM-100 columns have a published cut-off of 300 nucleotides for single stranded and 125 nucleotides for double-stranded nucleic acids. The filtrate from the column should only contain RNA shorter than 300 nucleotides.

Thirty µg RNA in 200 µl water was heated to 65°C and applied to a Microcon YM-100 column. The column was centrifuged in a microfuge at 10,400 rpm for 12 minutes at 17°C. A further 200 µl water was applied and the centrifugation repeated. The flow-through from the column containing the small RNA was precipitated for small RNA as previously described (Chapter 2).

Extraction from acrylamide gels

Small species of nucleic acid diffuse readily from gel matrices when not subjected to electrophoresis. Using this property, RNA can be extracted from gel slices by allowing the diffusion to take place in an appropriate buffer.

Total RNA was loaded onto denaturing mini acrylamide gels. The top half of the gel was 4.5% acrylamide, and the bottom 15% acrylamide (for details see figure 3.1). RNA (25 µg) in 45 µl formamide and 50 µl gel

loading buffer ll (Ambion) was heated to 65°C for 5 minutes prior to loading. For each extraction two gels (50 µg total) were run in parallel at 200 V in 1XTBE buffer. Gels were run until the bromophenol blue had run approximately 2 mM into the 15% acrylamide fraction.

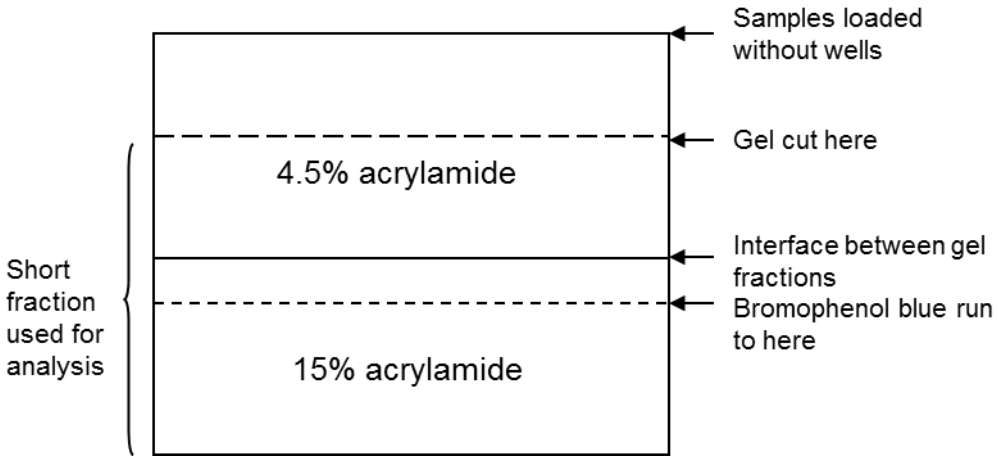


Figure 3.1 Gel layout for small RNA extraction
Total RNA was run on denaturing acrylamide gels. Gel percentages are shown. The sample was run at 200V until the bromophenol blue reached the line as shown. The gel was cut as indicated and the lower fragment extracted. 15% gel mix contains 50 ml ProtoFLOWgel (Flowgen), 10 ml 10XTBE, 42 g Urea and water to 100 ml. 4.5% gel mix contains 15.15 ml ProtoFLOWgel, 10 ml 10XTBE, 42 g Urea and water to 100 ml. Before pouring 1ul TEMED and 15 ul 10% Ammonium persulphate were added per ml of gel mix.

In trial gels to assess the run conditions, the gels were stained with ethidium bromide and visualised with UV light. For preparative gels a small gel slice was stained. The gel slices used for elution were not subjected to UV as this nicks the RNA. The gels were cut as shown in figure 3.1. Each lower gel slice containing the short RNA was cut into smaller pieces and placed in 7 ml 0.3 M sodium acetate (Ambion) at 4°C overnight. The solution was replaced with a further 1.5 ml 0.3 M sodium acetate and placed on a spiramix (Denley) at room temperature for 1 hour. The 7 ml and 1.5 ml were pooled and precipitated with 3 volumes of ethanol and 60

µg glycogen at -20°C overnight. Following centrifugation at 13,200 rpm, 4°C for 30 minutes, pellets were resuspended in 500 µl of water and heated at 65°C for 10 minutes to fully resuspend the RNA. The tubes were then placed on ice for 30 minutes to solidify any acrylamide that had been carried through, and the centrifugation repeated for 20 minutes. The supernatant was removed to a fresh microfuge tube and the pellet re-extracted with a further 150 µl water. The resulting supernatants were pooled and quantified by spectrophotometry. Finally the sample was precipitated for small RNA.

FlashPAGE fractionator

The flashPAGE fractionator (Ambion) (184) is a small electrophoresis unit that runs preformed acrylamide columns, primarily for isolating microRNA. The dye in the loading buffer co-migrates with 40 nucleotide RNA and this was used to calibrate the column. Fifty µg RNA in 12 µl water was mixed with 12 µl loading buffer and denatured at 95°C for 5 minutes. The sample was loaded onto the column and run at 80 V. When the dye ran off the column a sample of dye was reapplied to the top and the electrophoresis recommenced. After the third dye volume had been eluted, the sample in the bottom reservoir was collected. Fresh top and bottom buffers were applied and after the next dye elution the second sample was collected. Three more samples were collected after a 75 minute elution; a 30 minute elution and a final 30 minute elution. All fractions were precipitated for small RNA, then estimated for RNA concentration by spectrophotometry and for size range by PAGE.

Modified GeneRacer™ cloning

Following the method of Aspegren and colleagues (180) size fractionated RNA was treated using a modification of the GeneRacer protocol (Invitrogen) (185) (Figure 3.2). DNA was removed from the sample by treating with Turbo DNA Free DNase in a 50 µl reaction, and then ethanol precipitated at -80°C. C-tails were added to the RNA by poly(A) polymerase (PAP)(Ambion) in an 80 µl reaction containing 16 µl 5X PAP buffer, 8 µl 25 mM $MnCl_2$, 1.6 U PAP, 8 µl 10 mM CTP, incubated at 37°C for 1 hour. The sample was then ethanol precipitated. The method then follows the GeneRacer protocol with a few technical modifications. Following tobacco acid pyrophosphatase (TAP) treatment, the enzyme was inactivated with a phenol/chloroform extraction. 1 µl DMSO was added to the ligation reaction of the RNA oligo to the RNA.

For the cDNA sysnthesis, 1 µl Generacer dC oligo mix replaced the oligos supplied with the kit. The Generacer dC oligo mix consists of 1:1:1 mix of 50 µM stocks of the three Generacer dC3' oligos (Appendix i). The reaction was incubated as follows; 42°C for 30 minutes, 50°C for 30 minutes; 70°C for 15 minutes. The 25 µl RT PCR reaction contained 1.5 µl 5' GeneRacer primer, 1.5 µl 3'GeneRacer primer, 2.5 µl each of 10X PCR buffer and 25 mM $MgCl_2$, 0.5 µl 10 mM dNTPs, 0.25 µl Biotaq (Bioline) and 1 µl cDNA. The cycling conditions were 94°C for 4 minutes, followed by 35 cycles of 94°C for 30 seconds, 60°C for 30 seconds and 65°C for 1 minute). A 50 µl nested PCR reaction contained 1 µl of each GeneRacer nested oligo, 5 µl each of buffer and $MgCl_2$, 1 µl dNTPs and 0.5 µl Biotaq (all as above). 1 µl PCR product was added and the cycling conditions

were 94°C for 4 minutes, followed by 30 cycles of 94°C for 30 seconds, 65°C for 30 seconds and 68°C for 30 seconds.

The resulting PCR products were cloned and sequenced following the methods in Chapter 2.

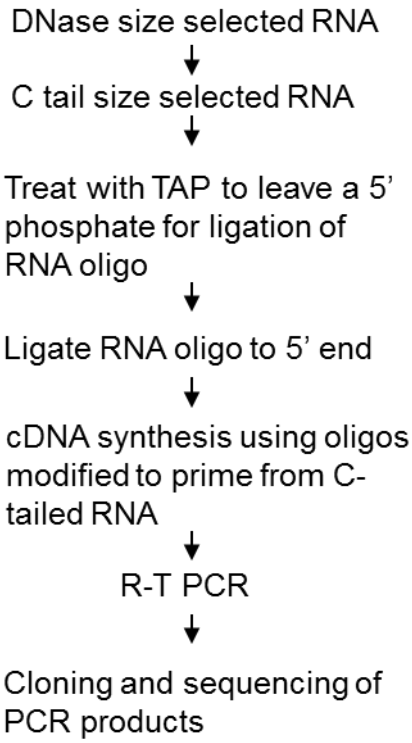


Figure 3.2 Flow chart of modified GeneRacer™ technique.The important modification of this technique from the standard GeneRacer method is the C-tailing of the size selected RNA. C-tailing allows priming cDNA synthesis with oligonucleotides to the polyC. This means that mRNA species which have a polyA tail will not be included in the priming reaction.

Reduction of rRNA species

Removal by restriction digestion prior to cloning

To reduce the number of clones that contained rRNA sequences, rRNA sequences were analysed for a restriction enzyme site to use to digest the PCR products, thereby removing them from the cloning reaction. RsaI was chosen as its restriction site is present in both 5.8SrRNA and 5SrRNA.

The PCR products were precipitated using SureClean (Bioline) and resuspended in water. They were digested prior to cloning with 1/10 volume of 10x buffer and 1/50 volume *RsaI* (both New England Biolabs) at 37°C for 1 hour. The reaction was precipitated again with SureClean and resuspended in 10 µl water. The resulting products were cloned as in chapter 2.

The Microcon products and the 5.8S and 200+ fractions from the FlashPAGE purification were treated with *Rsa*I. In all these cases, the PCR products were cloned with and without digestion. This was to avoid missing any non rRNA products that contained *Rsa*I sites.

Screening of clones by hybridization

To reduce the number of clones to be sequenced, the colonies from the cloning reactions were subjected to PCR (see Chapter 2). The PCR products were run on 2% agarose gels in 1XTBE. After staining with ethidium bromide and UV visualisation, the DNA was capillary transferred onto Hybond N+ membrane (GE Healthcare) (described in Chapter 2). Hybridization was performed using γ^{32}P ATP end labelled oligonucleotides, overnight at 37°C in C+G hybridization buffer (Chapter 2). Washing was initially at 37°C in 0.5XSSC/0.1% SDS buffer, if necessary the washing temperature was increased to 40°C.
Initially, 5.8SF and mir451 oligonucleotides (Appendix i) were used. The mir451 oligonucleotide was included because an initial trial of the method used RNA from parasites that had not been treated with saponin. The clones obtained from this were either rRNA or human microRNA 451 (data not shown). Once it was established that saponin lysis reproducibly removed the human microRNA component, the mir451 probe was no

longer included. A mixture of 5.8SF/5.8SMid/5.8SEnd was devised in order to cover more regions of the prevalent 5.8S sequence, as many initial clones appeared to be 5.8S fragments. The flashPAGE samples were screened with the 5.8S mixture, as well as SSFU mix and INT mix (Appendix i), designed to detect RNAs that were regularly found, including several mitochondrial RNAs.

rRNA depletion with modified oligonucleotides

To remove rRNA transcripts prior to GeneRacer cloning, a magnetic bead depletion method was developed. Briefly, biotinylated complementary rRNA oligonucleotides were bound to streptavidin beads. The biotinylated oligonucleotides were designed following alignment of the 5SrRNA and 5.8SrRNA, with the most commonly sequenced rRNAs from GeneRacer clones. For 18SrRNA and 28SrRNA, the oligonucleotides were designed to the most commonly obtained rRNA fragments from GeneRacer. The oligonucleotide coated beads were then mixed with size selected RNA and allowed to anneal. Ribosomal RNA annealed to the oligonucleotides was then removed by magnetic separation of the beads from the sample, leaving an rRNA depleted fraction.

For the rRNA oligonucleotide mix, two oligonucleotides complementary to 5.8S rRNA and one each to 5S, 28S and 18S rRNA were designed (Appendix i). These were synthesised with biotin at the 5' end (Operon). A mix of 10% each of 100pmols/µl 5s_BIO, 18s_BIO and 28s_BIO, plus 40% 5.8sa_BIO and 30% 5.8sb_BIO was used. The 5.8S rRNA sequences were included at higher proportions in the mixture because they were the most prevalent in initial experiments. 5 nmol of oligonucleotide mix was made to a final volume of 250 µl in water.

To prepare the oligonucleotide-bound beads, 5 mg M-270 streptavidin Dynabeads (Invitrogen) were pre-treated for RNA work following the manufacturer's instructions. The beads were resuspended in 250 µl B+W buffer (10 mM Tris pH7.5, 1 mM EDTA, 2 M NaCl). The oligonucleotide mix was added to the Dynabeads and allowed to bind via the biotin/streptavidin interaction for 15 minutes on a spiramix. The beads were washed three times with B+W buffer (by placing on a magnet) then resuspended in 250 µl binding buffer (20 mM Tris pH7.5, 2 mM EDTA, 1 M NaCl).

Before rRNA depletion, total RNA (50 µg) was size selected by the acrylamide gel method previously described. Of the gel-purified RNA, 1.5 µg was kept for controls and gels. The remaining 4.3 µg was denatured at 65°C, then incubated with the beads for 10 minutes at room temperature on the spiramix. After magnetic separation, the supernatant, containing rRNA-depleted RNA, was precipitated as small RNA then processed using the modified GeneRacer technique described above.

Results

Following our confirmation of the expression of the GRE as a ncRNA in *P. falciparum,* a cloning and sequencing strategy was employed to identify additional novel ncRNAs. A modified GeneRacer technique was used for cloning small RNAs (180). To enrich for small transcripts, size selected RNA was prepared using different methods. The first attempt, using Microcon spin columns, yielded many coding sequence (CDS) fragments (Table 3.3 and Appendix ii). This suggested either an interesting new species of CDS derived small RNA, or an artefact. To investigate this,

these results were compared to those from two alternative acrylamide-based methods. For the acrylamide gel extraction, RNA was eluted from gel slices as indicated in figure 3.1. For the FlashPAGE method, RNA was size fractionated and analysed on acrylamide gels (Figure 3.3). Eluted fractions were labelled as 0-150, 200+ and 5.8s. The 5.8s fraction containing material between 150-200 nucleotides should include the 5.8s rRNA of 158 nucleotides. The data combined from the alternative methods are shown in Table 3.3 and Appendix ii. A comparison of the classes of RNA found by the different purification methods is shown in table 3.4. CDS fragments were significantly reduced with the acrylamide gel based techniques. None of the CDS found with the Microcon method matched known ncRNA species.

For all three purification methods, many known ncRNAs were represented, such as snoRNAs, tRNAs, rRNAs and mitochondrial RNAs. The 5.8S rRNA sequence MAL5_5.8S and MAL7_5.8S in the rRNA clusters on chromosomes 5 and 7 were found most frequently. Of the novel intergenic RNAs, INTER-3 on chromosome 14 and the 2 copies of INTER-8 on chromosome 11 are in G+C peaks. INTER-6 on chromosome 9 has an A+T content of 91.5%. CDS-8 has 5 copies in the genome; 8a is in the CDS of MAL8P1.31, 8d and 8e are rRNA sequences. Although not rRNA, 8b and 8c are located next to rRNA genes. One novel intronic RNA was found in the first intron of PFI1450c, a conserved protein of unknown function. All other intronic sequences had been previously identified as snoRNAs.

Table 3.3 Details of sequences obtained from cloning of small RNA from *P. falciparum* (Sequences available in Appendix ii). rRNA dep = rRNA depletion. Matches:- Mi = Mishra (145); C = Chakrabarti (125) C* = Chakrabarti matches with website but not with paper (did originally but probably due to new assembly); M = Mourier (178); O = Otto (150); R = Raabe (146);U= Upadhayay (144).

CDS-8a*-8e* are the same sequence found in different regions.

Sequence number	length	Chromosome	start	end	RNA extraction method	Matches	notes
CDS							
CDS-1	121	2	127931	128051	rRNA depletion		spans 5' end PFB0120w early transcribed membrane protein
CDS-2	53	3	78819	78870	microcon		PFC0065c putative alpha/beta hydrolase exon2
CDS-3	45	4	843932	843976	microcon		PFD0900w unknown function
CDS-4	77	4	1100624	1100697	microcon		PFD1155w (exon 3) erythrocyte binding antigen-165
CDS-5	59	5	935827	935871	microcon		3' end PFE1120w
CDS-6	46	6	445101	445146	microcon		3' end CDS histone h3 PFF0510w
CDS-7	70	8	1238863	1238930	gel		PF08_0008 GPI-anchored micronemal antigen
CDS-8a*	26	8	1289448	1289473	gel		MAL8P1.310 (unknown function)

Sequence number	length	Chromosome	start	end	RNA extraction method	Matches	notes
CDS-9	45	9	1203017	1203061	microcon		5' end PFI1475w merozoite surface protein 1
CDS-10	112	9	1177837	1177948	microcon		10 PFI1445w (5th exon) high molecular weight rhoptry protein 2
CDS-11	35	10	1413422	1413456	microcon		PF10_0348 duffy binding-like merozoite surface protein
CDS-12	49	11	1119678	1119719	microcon		PF11_0300
CDS-13	94	11	1005446	1005539	microcon		PF11_0267 (exon 1) putative kelch protein
CDS-14	199	11	1461297	1461494	flashPAGE		PF11_0381a SnRNA encoding
CDS-15	45	12	1921198	1921235	microcon		PFL2215w actin1
CDS-16	18	13	2165679	2165696	microcon		MAL13P1.271 (5th exon) putative v-type ATPase
CDS-17	37	13	2057744	2057780	microcon		MAL13P1.260 (exon 3) alveolin putative
CDS-18	43	13	2030186	2030228	microcon		MAL13P1.256 (exon 1) phosphatydlinositol transfer protein putative
CDS-19a	65	13	1445835	1445899	microcon		PF13_0198 reticulocyte binding protein 2 homologue a exon2

Sequence number	length	Chromosome	start	end	RNA extraction method	Matches	notes
CDS-19b	65	13	1434128	1434192	microcon		MAL13P1.176 reticulocyte binding protein b
CDS-20a	73	13	89842	89678	microcon		MAL13P1.60 (exon 3) erythrocyte binding antigen-140 crosses splice
CDS-20b	73	13	89879	89714	microcon		MAL13P1.60 (exon 2) erythrocyte binding antigen-140 crosses splice
CDS-21	55	14	2713027	2713076	microcon		PF14_0632 26S proteasome regulatory subunit, putative exon3
CDS-22	92	14	247643	247734	microcon		PF14_0065 (exon 1). Putative glideosome associated protein
INTERGENIC							
INTER-1a	44	4	595499	585524	gel	C	partial GRE; RUF 6
INTER-1b	44	4	580274	580299	gel	C	partial GRE; RUF 6
INTER-1c	44	12	1715853	1715878	gel	C	partial GRE; RUF 6
INTER-2a	132	4	595407	595537	gel	C, M	RUF 6; RNAz ID:3145
INTER-2b	132	4	580182	580312	gel	C, M	RUF 6; RNAz ID:3134
INTER-3	31	14	84611	84641	gel		between PF14_0023 and PF14_0024

Sequence number	length	Chromosome	start	end	RNA extraction method	Matches	notes
CDS-8b*	26	13	2802674	2802699	gel		between MAL13_5.8s and 13P1.461
CDS-8c*	26	7	1144986	1145011	gel		between MAL7_28s and MAL7_28Sa
INTER-4	31	14	982106	982136	gel	O	PF14TR004 non protein coding
INTER-5	100	11	1166810	1166908	flashPAGE	R	PFa SnoRU14a, between PF11_0313 and PF11_0314
INTER-6	83	9	992941	993023	microcon		between PFI1200 and PFI1205
INTER-7	73	14	982097	982169	microcon	C*, R, O	snoR30; SnoR_30; PF14TR004
INTER-8a	31	11	1166876	1166905	rRNAdep/gel		between PF11_0313 and PF11_0314
INTER-8b	31	11	1166651	1166680	rRNAdep/gel		between PF11_0313 and PF11_0314
INTRON							
INTRON-1	76	13	1637300	1637375	microcon	C, R, Mi	SnoR20; PFS17; MAL13P1.209, 60S ribosomal subunit protein L18, 2nd intron,
INTRON-2	70	8	1162332	1162401	flashPAGE	C, R, Mi	Partial snoR05; PFS2; PF08_0019, receptor for activated c kinase
INTRON-3	61	11	391778	391837	rRNA/depletion	C*, R	snoR10; SnoscaR_10; PF11_0105 2nd intron
INTRON-4	59	11	391779	391837	gel	C*, R	snoR10; SnoscaR_10; PF11_0105 2nd intron

Sequence number	length	Chromosome	start	end	RNA extraction method	Matches	notes
INTRON-5	70	11	390440	390509	all	C*, R	snoR01; PF11_0105 5th intron
INTRON-6	51	8	240206	240256	gel	C, R	Partial snoR06; PF8_0124 2nd intron
INTRON-7	72	3	307980	308051	flashPAGE	C, M	Partial snoR03; partial RNAz ID:2814; PFC0290w 2nd intron, 40s ribosomal protein
INTRON-8	35	9	1181268	1181290	microcon		PFI1450c 1st intron
MITOCHONDRIAL GENOME							
MITO-1	67		296	362	rRNAdep/gel		malmito_rna_LSUG:rRNA
MITO-2	99		506	593	gel/flashPAGE		malmito_rna_1:rRNA
MITO-3	109		628	724	flashPAGE		malmito_rna_10:rRNA
MITO-4	71		1638	1696	gel		mal_rna_8:rRNA
MITO-5	90		1830	1909	gel		mal_rna_10:rRNA
MITO-6	134		1916	2030	rRNAdep/gel		mal_rna_11:rRNA
MITO-7	134		1916	2030	rRNAdep/gel		mal_rna_11:rRNA
MITO-8	91		1916	2004	gel		mal_rna_11:rRNA

Sequence number	length	Chromosome	start	end	RNA extraction method	Matches	notes
MITO-9	76		4624	4698	microcon		malmito_rna_RNA4:rRNA
MITO-10	52		4648	4698	flashPAGE		mlamito_rna_RNA4:rRNA
MITO-11	86		4720	4802	gel		malmito_rna_RNA5:rRNA
MITO-12	28		4964	4988	gel		malmito_rna_RNA18:rRNA
MITO-13	49		4997	5027	gel/flashPAGE		mal_rna_13:rRNA
MITO-14	49		4997	5027	flashPAGE		mal_rna_13:rRNA
MITO-15	80		5380	5446	gel		mal_rna_15:rRNA
MITO-16	29		5423	5439	gel/flashPAGE		mal_rna_15:rRNA
MITO-17	72		5452	5507	microcon		malmito_rna_SSUF:rRNA
MITO-18	68		5452	5507	microcon		malmito_rna_SSUF:rRNA
MITO-19	30		5478	5507	microcon		malmito_rna_SSUF:rRNA
MITO-20	52		5513	5546	flashPAGE		malmito_rna_RNA14:rRNA
MITO-21	169		5603	5771	rRNAdep/gel		mal_rna_17:rRNA

Sequence number	length	Chromosome	start	end	RNA extraction method	Matches	notes
RIBOSOMAL RNA (rRNA)							
rRNA-1a	191	5	1289594	1289784	microcon		MAL5_18S
rRNA-1b	191	7	1139137	1139327	microcon		MAL7_18S
rRNA-2a	41	5	1292162	1292202	flashPAGE		MAL5_5.8S
rRNA-2b	41	7	1141706	1141746	flashPAGE		MAL7_5.8S
rRNA-3a	61	5	1295908	1295947	flashPAGE		MAL5_28S
rRNA-3b	61	7	1145449	1145488	flashPAGE		MAL7_28Sa
rRNA4-a	59	5	1292140	1292198	flashPAGE		MAL5_5.8S
rRNA-4b	59	7	1141684	1141742	flashPAGE		MAL7_5.8S
rRNA-5a	74	5	1292470	1292542	rRNAdep/gel		MAL5_28S
rRNA-5b	74	7	1142013	1142085	rRNAdep/gel		MAL7_28S
rRNA-6a	126	14	780916	781034	rRNAdep/gel		MAL14_5S_3
rRNA-6b	126	14	780006	780124	rRNAdep/gel		MAL14_5S_2

Sequence number	length	Chromosome	start	end	RNA extraction method	Matches	notes
rRNA-6c	126	14	779377	779495	rRNAdep/gel		MAL14_5S_1
rRNA-7a	164	5	1292044	1292203	rRNAdep/gel		MAL5_5.8S
rRNA-7b	164	7	1141588	1141747	rRNAdep/gel		MAL7_5.8S
rRNA-8a	160	5	1292044	1292203	rRNAdep/gel		MAL5_5.8S (1 nt different to rRNA-7)
rRNA-8b	160	7	1141588	1141747	rRNAdep/gel		MAL7_5.8S (1 nt different to rRNA-7)
rRNA-9a	49	5	1294699	1294745	gel		MAL5_28S
rRNA-9b	49	7	1144240	1144286	gel		MAL7_28S
rRNA-10a	207	5	1291135	1291328	rRNAdep/gel		MAL5_18S
rRNA-10b	207	7	1140675	1140868	rRNAdep/gel		MAL7_18Sa
rRNA-11a	159	5	1292044	1292202	microcon/gel/flash		MAL5_5.8S
rRNA-11b	159	7	1141588	1141746	microcon/gel/flash		MAL7_5.8S
rRNA-12a	125	14	780916	781032	flashPAGE		MAL14_5S_3
rRNA-12b	125	14	780006	780122	flashPAGE		MAL14_5S_2

Sequence number	length	Chromosome	start	end	RNA extraction method	Matches	notes
rRNA-12c	125	14	779377	779493	flashPAGE		MAL14_5S_1
rRNA-13a	51	7	1144224	1144273	flashPAGE		MAL7_28S
rRNA-13b	51	5	1294683	1294732	flashPAGE		MAL5_28S
rRNA-14a	44	5	1294075	1294118	microcon		MAL5_28S
rRNA-14b	44	7	1143615	1143657	microcon		MAL7_28S
rRNA-14c	44	1	480138	480177	microcon		MAL1_28S
rRNA-15a	61	5	1291147	1291207	microcon		MAL5_18S
rRNA-15b	61	7	1140687	1140747	microcon		MAL7_18Sa
CDS-8d*	26	5	1295445	1295470	gel		MAL5_28S
CDS-8e*	26	11	1932324	1932349	gel		MAL11_rRNA
TRANSFER RNA (tRNA)							
tRNA-1	56	4	524651	524687	microcon	C	MAL4_tRNA_Ala1

Sequence number	length	Chromosome	start	end	RNA extraction method	Matches	notes
tRNA-2	26	11	143355	143380	rRNAdep/gel		1399.pre-tRNA-Gly-1
tRNA-3	72	4	525051	525122	gel	C	MAL4_tRNA_Glu1
tRNA-4	37	7	381486	381521	rRNAdep/gel		MAL7_tRNA_Thr1
ARTEFACT							
A-1	39				gel		*H. sapiens* rRNA 28S
A-2	46				rRNAdep/gel		*H. sapiens* rRNA 28S
A-3	27				rRNAdep/gel		*H. sapiens* rRNA 28S
A-4	25				gel		*H. sapiens* rRNA 28S
A-5	132				rRNAdep/gel		*H. sapiens* haemoglobin alpha 1
A-6	85				flashPAGE		no hits
A-7	77				flashPAGE		no hits
A-8	60				flashPAGE		no hits
A-9	70				flashPAGE		no hits

Sequence number	length	Chromosome	start	end	RNA extraction method	Matches	notes
A-10	21				flashPAGE		no hits
A-11	12				gel		no hits
A-12	27				flashPAGE		no hits
A-13	12				flashPAGE		no hits
A-14	81				gel		no hits
MULTIPLE HITS							
MULTI-1	14				flashPAGE		many not P. falciparum
MULTI-2	13				flashPAGE		many not P. falciparum
MULTI-3	21				flashPAGE		many not P. falciparum
MULTI-4	24				flashPAGE		many not P. falciparum
MULTI-5	23				flashPAGE		many not P. falciparum
MULTI-6	25				flashPAGE		many not P. falciparum

Sequence number	length	Chromosome	start	end	RNA extraction method	Matches	notes
MULTI-7	18				flashPAGE		many not P. falciparum
MULTI-8	36				flashPAGE		many not P. falciparum
MULTI-9	17				flashPAGE		many not P. falciparum
MULTI-10	23				rRNAdep/gel		many not P. falciparum
MULTI-11	18				rRNAdep/gel		many not P. falciparum
MULTI-12	16				rRNAdep/gel		many not P. falciparum
MULTI-13	50				rRNAdep/gel		many not P. falciparum
MULTI-14	15				flashPAGE		many not P. falciparum
MULTI-15	19				gel		many not P. falciparum
MULTI-16	16				gel		many not P. falciparum
MULTI-17	26				gel		many not P. falciparum
MULTI-18	14				gel		many not P. falciparum
TOO SHORT							

Sequence number	length	Chromosome	start	end	RNA extraction method	Matches	notes
SHORT-1	6				rRNAdep/gel		too short
SHORT-2	9				gel		too short
SHORT-3	8				gel		too short
SHORT-4	5				flashPAGE		too short

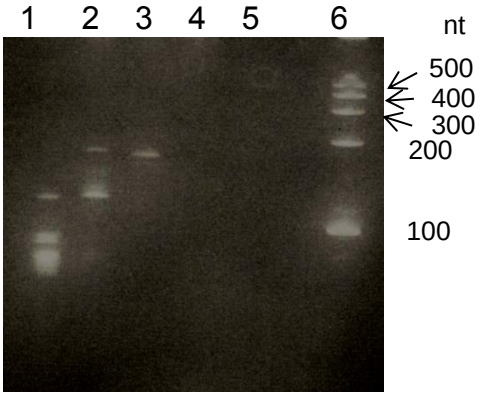


Figure 3.3 Separation of small RNA by FlashPAGE. 1. 0-150 nucleotides (nt) 2. 5.8s 3. 200+ nt 4. 300+ nt 5. 400+ nt 6. Century markers (Ambion). 500 ng of each fraction loaded and run at 200V on 15% denaturing acrylamide gel. Samples 3, 4 and 5 pooled as 200+ fraction for cloning.

To reduce repetition of sequencing of the abundant rRNAs, clones were initially pre-screened with rRNA probes (see methods). A method to deplete rRNA before cloning was developed and applied to the acrylamide gel size selected RNA. This reduced the number of clones containing rRNA from approximately 44% to 2.5% (Table 3.5). Although a higher percentage of colonies would not grow for further plasmid preparation, or gave poor plasmid yield, the percentage of candidate sequence results (excluding rRNA) increased from 9.5% (undepleted) to 14.3% (depleted). No additional classes of ncRNAs were found by the rRNA depletion method (table 3.4).

	RNA Purification Method				
Location	Microcon	Gel	Flash	rRNA depletion/gel	Total[1]
CDS	18	2	1	1	22
Intron*	1	0	0	0	1
Intergenic*	1	3	0	1	5
*not including known ncRNAs					
Known ncRNAs					
rRNA					
5s rRNA	0	0	1	1	2
5.8s rRNA	1	1	3	2	5
18s rRNA	2	0	0	1	3
28s rRNA	1	2	2	1	6
Mitochondrial rRNA					
LSU	0	1	2	1	3
SSU	3	4	1	2	9
RNA1-18	1	4	4	1	9
tRNA	1	1	0	2	4
snoRNA	3	3	4	2	9
RUF1-5	0	0	0	0	0
GRE	0	2	0	0	2
Aretefact/short/multiple sequences	0	10	18	8	36

[1] Where total does not tally with previous columns, sequences were found with more than 1 purification method.

Table 3.4 Analysis of ncRNA clones. Classes of known and novel ncRNAs found by the modified GeneRacer method. The division of results by RNA purification method is also shown.

	No depletion		rRNA depletion	
	n	%	n	%
Small (not seq)[a]	77	23.6	37	31.1
No insert (seq)[b]	43	13.2	25	21
5.8s (hyb)[c]	139	42.6	3	2.5
5.8s (seq)[d]	5	1.5	(3)	0
Good seq[e]	31	9.5	17	14.3
Seq failure[f]	9	2.8	9	7.6
Others[g]	22	6.8	28	23.5
Total	326	100	119	100

Table 3.5 Effect of depleting rRNA on modified GeneRacer cloning.

Comparison of GeneRacer clones from 0-150 and 200+ RNA fractions of FlashPAGE purified RNA, with rRNA depleted GeneRacer clones. To allow direct comparison, the data presented here are from the 5.8s mixed probe only. [a] Considered too small by PCR screening to have an insert, so not sequenced. [b] Sequenced, but had no insert. [c] 5.8s rRNA sequences detected by 5.8s mix probe, not sequenced. [d] 5.8s rRNA sequences not detected by 5.8s mix probe, sequenced. The value in parentheses was detected both by sequencing and hybridization. [e] non 5.8s rRNA sequences, sequenced. [f] Sequencing reactions that did not produce readable sequence, or had mixed inserts. [g] Others are clones which would not grow or gave poor plasmid prep yields.

Discussion

The modified GeneRacer technique was used to successfully clone ncRNAS in *P. falciparum*. The first attempt using Microcon columns to size separate RNA yielded many CDS fragments. Although these could represent a new class of small RNA, these sequences have not been identified by other bioinformatic and experimental screens. They would not be expected to be found by Mourier et al (178), as they filtered coding regions from their screen. Otto et al (150) size selected their cDNA to between 200-250 nucleotides. This would have excluded even the largest

ncRNA of 112 nucleotides, found with the Microcon size separation method. However, Raabe and colleagues did no such CDS filtering and included cDNA covering this size range, but did not identify these sequences (146). We also did not detect these sequences using acrylamide gel based size separation methods. This leads us to conclude that the CDS fragments from the Microcon method are most likely an artefact from the purification technique.

In their first ncRNA report, Upadhayay and colleagues suggested "that combinations of multiple approaches will yield maximum success." (144). Six years on this has proved to be true as no published dataset completely overlaps. Comparing our results to those obtained by Chakrabarti and colleagues (125) shows that many of their verified ncRNAs were not detected here. Of their 31 snoRNAs, only 7 were identified here. A further snoRNA (INTER-5) was detected both here and in the Raabe set (PFa SnoRU14a) (146). Similarly, only 3 of the 44 tRNAs known in the *P. falciparum* genome were detected here, for reasons discussed below. We cloned four novel sequences from intergenic regions and one from an intron, none of which were predicted by the Mourier *et al* RNAz algorithm (178). A comparison of our dataset (table 3.3) against published screens has shown that none of these sequences have been detected by all of the other data sets. We have also identified, but not verified, five novel sequences.

The high proportion of clones containing rRNA meant that many had to be sequenced to identify the few novel ncRNAS. When rRNA sequences were removed by pre-screening, the next most prevalent class of ncRNA identified was mitochondrial RNA. It is possible that

experimental design led to some classes being unable to be cloned, although clones were obtained from several classes of ncRNA. *Plasmodium* DNA is known to be highly unstable when cloned into *E. coli* due to its high A+T content. It is also possible that some of the small RNAs were toxic when cloned into *E. coli*, so would never be found.
The large percentage of clones containing rRNA here was comparable to those found by this method in *D. discoideum* (180). Whilst 44.1% of clones were 5.8S rRNA in *P. falciparum*, the figures for *D. discoideum* were 21.8% for 5.8S and 54.9% for 5S rRNA in their 50-150 nucleotide insert library. This rRNA figure rose to 98% in their 150-500 insert library. The authors did not find any tRNA clones, which they suggest is due to secondary structure. Although we did not find many tRNAs, we find that this method is suitable for tRNA cloning.

Depletion of rRNA by biotinylated oligonucleotides and streptavidin beads led to a significant reduction in the percentage of clones that contained rRNA. Only one round of depletion was used here. More rounds of depletion, either just with the rRNA oligonucleotides, or with other common sequences such as mitochondrial RNA, may increase the proportion of novel species in the cloning reaction. It is interesting to note that reducing the fraction of clones with rRNA inserts led to an increase in the percentage of clones that would not grow or grew poorly. Clearly, rRNA sequences are not toxic to *E.coli*. It is possible however that this rRNA depletion has led to an increase of different sequences within these clones, some of which may well be toxic or unstable.

This work was discontinued due to the amount of screening required to produce an ever decreasing amount of new data. An

automated approach to mass sequencing of clones may have identified more interesting clones. At the same time it was known that the sequencing techniques that were being developed at the Wellcome Trust Sanger Institute (WTSI) which would remove the need for bacterial culturing. The WTSI were then to sequence the RNA transcriptome using the RNAseq method, although their fragment size was limited to 200-250 bp. The experimental approach described herein was one of several used by different laboratories to clone ncRNA in *P. falciparum.* Most of these, though successful, were not complete. Technology has now advanced to the extent that RNAseq is commonplace, alleviating the need for more traditional methods.

Although the screen reported here did not find any novel ncRNAs thought to have the potential to be involved with *var* genes, some were identified by other genome wide screens. Raabe and colleagues described 73 *var* associated antisense ncRNA (146). These were found for all Ups classes except for UpsE and most were specific for a single *var* gene. There was a second group of 17 transcripts that were complementary to more than one *var* gene. The lncRNAs described in chapter 1 are suggested by Broadbent and colleagues to be possible precursors of the small antisense transcripts described by Raabe (103). The only other ncRNA identified with possible connection to *var* was the GRE. This was identified by both Chakrabarti and Upadhyay but was omitted from Upadhyay's final list of candidiates, as it had no homology to *P. yoelii* (125,144). As our data had not identified any other potential *var* related ncRNA, further work in this thesis focuses on the role of the GRE.

Chapter 4

Development of an HB3 *var* gene qPCR primer set

Introduction

In order to focus on the potential role of the GC rich element (GRE) in *var* gene regulation, it was decided to utilise a panel of HB3 clones for a sample of isogenic parasites expressing different *var* genes. Progeny from a genetic cross of 3D7xHB3 were also to be used to study the potential role of GRE in meiotic recombination. To do this it was necessary to develop tools to analyse their *var* repertoire. The publication of qPCR primers for the 3D7 isolate by Salanti and colleagues (56) has enabled *var* gene expression to be studied by qRT-PCR. Previously, this could have been investigated by northern blot techniques. Generic probes to the exon 2 region could be used to follow switching events between different size *var* genes, but this would not identify which genes were involved. Identification of specific genes required hybridization with gene specific probes or RT-PCR. RT-PCR used primers designed for unbiased amplification of DBLα (186) followed by cloning and sequencing of the products (187). With qRT-PCR, the relative level of expression for each *var* gene can be determined in a single experiment and each gene is identifiable immediately.

With standard qRT-PCR, the expression of one gene is examined in many samples at the same time. For this, a standard curve of known template concentration is run concurrently with the samples and the levels of expression can be determined quantitatively. To study *var* gene expression by qPCR, 50-60 different genes are tested in the same experiment. It is both impractical and impossible to run standard curves for each set of primers at the same time. For this reason, relative quantification is used to analyse the results. With relative quantification, results are expressed relative to the expression of a reference gene. The choice of reference gene is crucial as it must have a steady level of expression throughout the cycle. Housekeeping genes that are constitutively expressed are usually used as reference genes. Salanti and colleagues analysed transcription levels of three housekeeping genes in *P. falciparum* and determined that seryl-tRNA synthetase had the most consistent level across the intraerythrocytic lifecycle (56).

To date, qPCR primers are available for the 3D7and IT4 isolates (56,188). However, *var* gene regulation is most often studied in 3D7 or its parent isolate NF54 because of the availability of the genome sequence. More weight would be given to novel findings being *Plasmodium* specific, rather than isolate specific, if they could be reproduced in a separate isolate. Therefore, a set of primers for qPCR from an isolate genetically distinct from 3D7 would be ideal. Although such primers exist for IT4, at the time of starting this project, its genome was not sequenced. The HB3 isolate had been sequenced by the Broad Institute (73). HB3 originally comes from Honduras in Central America (74,161) whereas 3D7 groups with African isolates (74). This is one reason why they were originally

chosen for the 3D7xHB3 cross (161). As the progeny of the cross are a useful genetic tool, a set of var gene primers for HB3 would complement those for 3D7.

With the sequencing of the HB3 genome at the Broad Institute (73) and the subsequent analysis of the *var* gene repertoire (74), the data were available to generate a set of qPCR primers for HB3. The *var* sequences were used to develop a set of primers which were tested on a panel of HB3 clones. The new HB3 primers were designed and tested to have no cross reactivity to 3D7 *var* genes. To complement this, the existing 3D7 *var* primers were tested for cross reactivity to HB3 and redesigned where necessary. The qRT-PCR data from HB3 clones and 3D7xHB3 progeny were verified by northern blot, comparing these techniques directly.

Materials and Methods

Primer design

HB3 *var* exon1 sequences that had originally been downloaded from the Broad Institute (73) for analysis by Kraemer et al (74) were used to design the primers.

Primers were designed using Primer3plus (173), on regions of exon 1 selected using the following criteria: the final kilobase of exon 1 sequence was removed to exclude sterile transcripts from the bi-directional intronic promoter; the 5’ kilobase of exon 1 was removed to exclude the DBLα conserved domain. HB3 primer sequences were then

searched against the 3D7 genome and any that were present in the 3D7 genome were redesigned.

Primer Testing

As described in chapter 2, it was necessary to keep qPCR conditions standard between 3D7 and HB3 primers. For that reason, conditions previously developed for 3D7 were used in this study (56,165).

HB3 primers were initially tested on HB3 gDNA. Amplicons were checked for a single product of the expected size by agarose gel electrophoresis. Melt analysis confirmed single products. HB3 primers were then tested on 3D7 gDNA. Those pairs that cross reacted with 3D7 gDNA were re-designed; the replacement pairs were tested as above. For relative quantification data to be valid, the efficiency of each primer pair must be as close as possible to all other pairs, and to the reference gene. Take Off (T.O.) and Amplification (Amp) data were determined using the Relative Quantification package of the RotorGene 6.0 and RotorGene Q software (172). Following 5 replicates on gDNA, the average T.O. and Amp were calculated. As the theoretical optimal amplification efficiency for PCR is 2.0, primer pairs with efficiency less than 1.85 were redesigned. The median T.O. of the primer set (all pairs) was also calculated and those pairs that were +/- 1.5 cycles from the median were redesigned. Primer pairs for two out of 40 *var* genes consistently failed to meet these criteria so the best pairs were included to cover the full repertoire. The primers tested for the HB3 set are shown in appendix i.

qPCR controls

Previous studies using *var* gene qPCR have included seryl-tRNA-synthetase as the control gene for relative quantification (56). MIQE guidelines on qPCR suggest that at least two housekeeping genes are necessary (189), therefore both actin and fructose-bisphosphate aldolase were tested.

In addition to the housekeeping controls, I included stage specific controls. As *var* genes are expressed at late ring stage, I chose both ring and trophozoite genes to act as controls for correct staging of parasites for RNA purification. Knob associated histidine rich protein (KAHRP) is traditionally used as a control for ring stage (190). By searching for genes that are expressed solely at trophozoite stage, single copy and not part of a gene family, S-adenosylmethionine synthetase (PFI1090w) was selected and primers devised. Appendix i contains all qPCR primer sequences.

cDNA synthesis and qRT-PCR

The primer set was tested on ring stage cDNA from 31 clones prepared by limiting dilution from an HB3 parent.

Following DNase treatment and testing for successful DNA removal, 4 µg RNA was split into two duplicate cDNA syntheses. Each cDNA was prepared in 20 µl, and 160 µl water was added to each reaction before qPCR. Most duplicate cDNA syntheses were tested, but as all were to be confirmed by Northern blots, for some only one cDNA was tested. Complete primer sets were run in duplicate for each cDNA. No-template controls were included for the control primers.

qPCR Data Analysis

Correction Factor

To correct for the slight variation in T.O. values between primer pairs, a correction factor was applied to the average T.O. of the individual pairs. To calculate this, the final set of primers was again run five times on HB3gDNA. The correction factor was calculated thus:

The average T.O. of each pair was calculated (X).

The geometric mean of the set of averages was calculated (Y).

Correction factor = X-Y

Data treatment

From melt curve analysis, those samples that had multiple peaks or shoulders, or those which had no peak, were removed from subsequent analysis. The Relative Quantification package of the RotorGene software was used to analyse the data with seryl-tRNA synthetase as the reference gene. The RNA staging controls and extra housekeeping genes were not included in *var* analysis. The average T.O. and Amp of duplicate samples was determined for each primer pair. To relate each *var* gene (*x*) to the reference gene (*ref*), the following formula was used:

$$\frac{A_{ref}^{TO-CF}}{A_{x}^{TO-CF}}$$

Where A = Amp, TO = Take off and CF = Correction factor.

Northern Blot and Hybridization

Ring stage RNA was resolved on 0.8% agarose gels (Chapter 2), 10 µg RNA per lane. Transfer and hybridization was as described previously. PCR products were labelled by random priming. Details of PCR and

hybridization conditions for each probe are shown in table 4.1. To extend the life of the filters a pool of probes was also used. The pool contained probes 1074-1, 1296-3 and 699-5.

Alterations to the published 3D7 qPCR primer set

The primers in the published 3D7 set (56) were tested on HB3gDNA. Those that amplified HB3gDNA were redesigned using the parameters described above and are listed in appendix i.

3D7xHB3 cross progeny analysis

RNA was prepared from 5 progeny clones of the 3D7xHB3 cross. Each cDNA synthesis was 4µg RNA in a 40µl reaction. The resulting cDNA was stored in 10 µl aliquots and 80 µl of water was added before use in qPCR. Two aliquots were used with the set of HB3 primers and two with HB3 primers.

To aid data analysis, each progeny's gDNA was tested with both sets of qPCR primers. Only those amplicons present in each respective genome were used in subsequent analysis. qPCR on XP8, X10 and X33 was repeated using only those amplicons present in their genomes., Correction factors were not applied to the cross progeny data sets. This was because a separate correction factor would be needed to be calculated for each progeny's *var* repertoire, requiring multiple replicates on each progeny's gDNA. The qPCR results were verified using northern blots. Hybridization probes are shown in table 4.1.

Results

Generation of an HB3 specific primer set

The HB3 primers were tested against HB3 and 3D7 gDNA. The average Take Off and Amplification values of five replicates of HB3, and one replicate of 3D7 are shown in table 4.2. For primers to be HB3 specific it is necessary that there is a significant difference in T.O. values between the different isolates. The range of difference between HB3 and 3D7 was 9.7-23.6 cycles, with the average of 16.2 cycles. An example of primers tested with both samples of gDNA is shown in figure 4.1. Primer pairs, e.g.HB3-1210, were redesigned if there was not a big enough difference between the T.O. values of the two isolates.

As discussed previously, relative quantification analysis of qPCR expression data relies on the efficiency of the primers being as close as possible. The amplification value of a fully efficient reaction should be 2. The range of Amplification values for the set is 1.9-2.1 with an average of 2.0. Inefficient pairs were redesigned e.g. HB3-1210-1zc.

The correction values were determined for the final set and are shown in table 4.2.

Verification of the HB3 primer set for qRT-PCR

The HB3 qPCR primer set was tested on cDNA from 31 clones generated by limiting dilution from an HB3 culture and the parental culture. cDNA was made in duplicate, although not all duplicate samples were analysed by qRT-PCR. Stage specific controls indicated correct staging of parasites in most clones. In some clones, KAHRP expression was absent,

Probe	Forward primer	Reverse primer	PCR mM $MgCl_2$	PCR Anneal temp°C	Hybridi-zation temp°C	Wash conditions °C and SSC (+0.1% SDS)
1074-1	AGTGGTGGTAATAGCGAAAA	CCCTTCTGGAATACCTTCTT	2.5	50	60	65 0.1x
1074-2	ACGTCTGCCCAATAGTAAAA	ACTATTTGGTTGTGAGGGTG	2.5	55	60	65 0.5x
1074-3	CTTGAGAGAGCAATGGAATC	TTTTCCATATTTGGTTGGAC	2.5	55	60	65 0.5x
1296-3	TTTGGTGATAAAGAGCTGGT	TACAAGAATTCGCACAACTG	2.5	55	60	65 0.5x
699-2	GTGGTGGAAAACACAGAAGT	AAAATGCCTCTTTCAGATCA	2.5	50	60	65 0.5x
699-5	CCTGATTTTGTTCTTGAAGG	TTCTTCCACGACTTCTGTTT	2.5	50	60	65 0.5x
1499	TGTGAGTGATAACGGTGAAA	TATAAAACGGTTCCTGCTGT	2.5	50	60	65 0.5x
1107	CTAGTGTCCAACAACCCAGT	TTCCATTGCACTTATTTCCT	2.5	50	60	65 0.1x
1737	ATCTGTTACCGCACCTCTTA	AAACTTTTGTCGTTTTTCCA	3.5	55	60	60 0.5x
1235	CTTATTCAGATACTAGAAGG	GTTTCTTTCATTACGGCCA	2.5	50	55	60 0.5x
1459	GGGCATCAATAACACTTCAT	TTGTCCACGACATTTGTCTA	3.5	55	60	65 0.5x
Var2csa	ATACTATAATACATGGAG	CATTATTAGTGCATGCGTC	2.5	45	50	60 0.5x

Table 4.1 Hybridization probes for verification of *var* gene expression in HB3 clones and 3D7xHB3 cross progeny

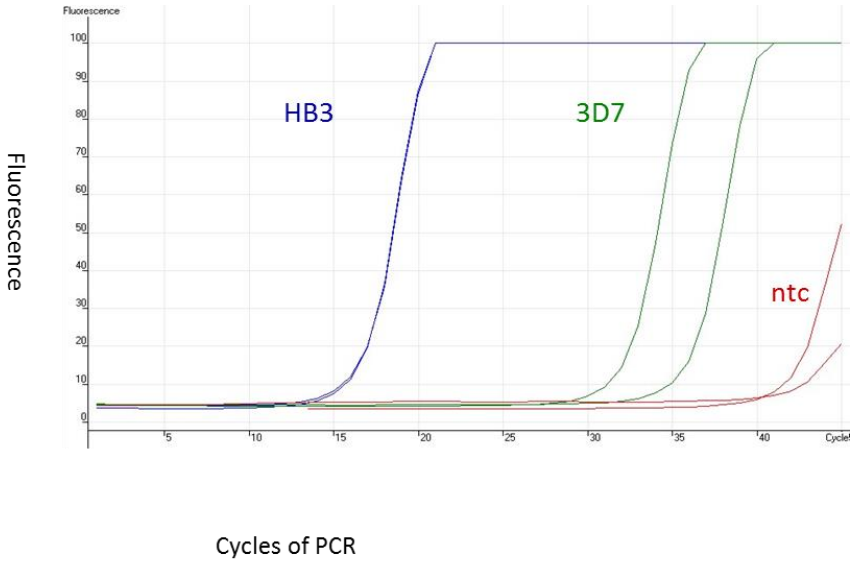


Figure 4.1 HB3 v 3D7 genomic DNA PCR. qPCR trace of a representative successful pair of HB3 qPCR primers tested on HB3 and 3D7 gDNA. The difference in take off values between the genomes is apparent. This is necessary to make the HB3 set specific for use in 3D7xHB3 progeny.

Table 4.2 Primer pairs tested for HB3 *var* gene qPCR. Those pairs with a correction factor were selected for the final set. Where there is no reason given for failure of a pair, they have not failed. In these cases more than one pair were designed and tested at the same time. The correction factor determined for each pair is included and was applied to the expression data.

Primer Pair	Amp average	Take Off average HB3	Take Off average 3D7	Notes	Correction factor
HB3-1000-1	2.018	15.78	29.8		0.21
HB3-1000-2alt	2.062	15.12	28.4		-0.33
HB3-1040	2.016	15.94	29.6		0.05
HB3-1074-1	2.04	15.46	23.7	3D7+	
HB3-1074-1-1zc	2.034	15.16	30.8		-0.67
HB3-1074-1-2zc	1.942	16.52	31.5		
HB3-1074-2alt	2.026	15.8	31.4		0.05
HB3-1074-3	2.018	16.12	27.4		0.37
HB3-1107	2.038	15.84	30.9		-0.01

Primer Pair	Amp average	Take Off average HB3	Take Off average 3D7	Notes	Correction factor
HB3-1108 var1 like	2.006	16.94	30.1		-0.03
HB3-1210	2.048	16.12	20	3D7+	
HB3-1210-1zc	1.828	17.92	34.7	poor amp	
HB3-1210-2zc	1.938	16.42	30.3		-0.07
HB3-1210alt	1.826	38.44	38.5	fail	
HB3-1235	1.998	17.76	31.7		3.23
HB3-1296	2.022	16.1	39.3		0.25
HB3-1296-2	2.078	15.28	32.8		-0.33
HB3-1296-3	2.006	15.94	39.3		-0.09
HB3-1296-3-1zc	1.992	15.64	25.8	3D7+	
HB3-1296-3-2zc	2	15.74	24.6	3D7+	
HB3-1308-1zc	2.004	15.7	30		-0.05
HB3-1308-2zc	2.044	15.14	16.1	3D7+	
HB3-1334	2.012	15.88	36.8		0.21
HB3-1334alt	2.02	39.18	16.2	fail	
HB3-1408	2	16.74	38.9		0.41
HB3-1459	2.048	15.48	18.9	3D7+	
HB3-1459-1zc	2.016	15.46	30.1		-0.59
HB3-1459-2zc	1.938	16.52	32.8		
HB3-1499	2.046	16.08	29.5		0.79
HB3-1514	2.09	15.36	33.9		-0.09
HB3-1514alt	2.014	15.7	31		
HB3-1523	2.08	14.3	23.5	3D7+	
HB3-1523-1zc	1.956	15.04	37		-1.15
HB3-1523-2zc	1.91	16.2	31.3		

Primer Pair	Amp average	Take Off average HB3	Take Off average 3D7	Notes	Correction factor
HB3-1587	2.056	15.64	21.9	3D7+	
HB3-1587-1zc	1.994	15.82	29.3		0.29
HB3-1587-2zc	2.008	15.64	29.7		
HB3-1604-1	2.058	15.4	27.8		0.05
HB3-1604-2	2.052	15.56	28.2		-0.17
HB3-1671	2.04	15.56	39.1		0.01
HB3-1703	1.668	22.94	38.3	Amp fail	
HB3-1703-1zc	2.01	15.66	9.5	3D7+	
HB3-1703-2zc	2.004	15.48	29.6		-0.11
HB3-1727zc	1.964	15.48	32.8		-0.77
HB3-1737	2.054	26.7	30.3	fail	
HB3-1737alt	2.024	35.34	37.2	fail	
HB3-1737hr	1.866	16.34	31.6		-0.89
HB3-1817zc	1.998	14.7	29.3		-1.09
HB3-2007zc	2.028	14.72	29.8		-0.97
HB3-209	1.976	18.14	31.2		3.6
HB3-699-1	2.008	15.6	31.6		-0.27
HB3-699-2	2.054	15.56	30.3		-0.13
HB3-699-3	2.022	16.62	39.4		0.67
HB3-699-4	2.028	15.82	39.1		-0.11
HB3-699-5	2.024	15.46	39.1		-0.41
HB3-752-1	2.032	16.9	31.3		0.31
HB3-752-2	2.006	16.04	21	3D7+	
HB3-752-2 -1zc	2.034	15.52	30.5		-0.13
HB3-752-2 -2zc	1.974	14.74	28.9		
HB3-752-3	2.032	15.44	29.2		-0.31

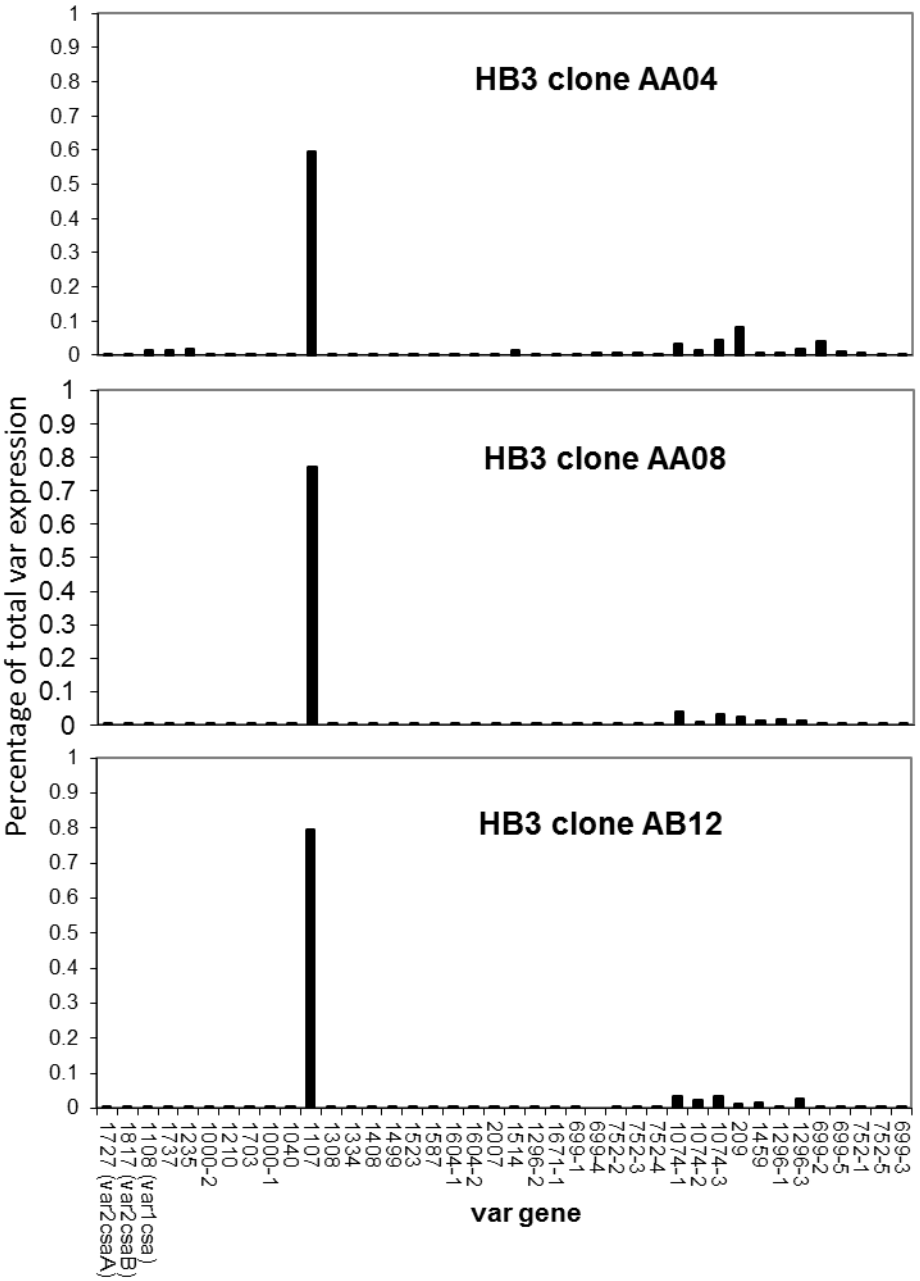
Primer Pair	Amp average	Take Off average HB3	Take Off average 3D7	Notes	Correction factor
HB3-752-4	2.038	15.42	27.4		-0.09
HB3-752-5	2.018	16.1	25.8		-0.07
HB3-752-5alt	1.96	38.44	39	fail	
Controls					
seryl1a	2.012	15.9	14.4		-0.57
Aldo	2.02	16.1	14.6		
PFB0100	2.05	15.4	14.2		
PFI1090	2.028	15.74	14.7		

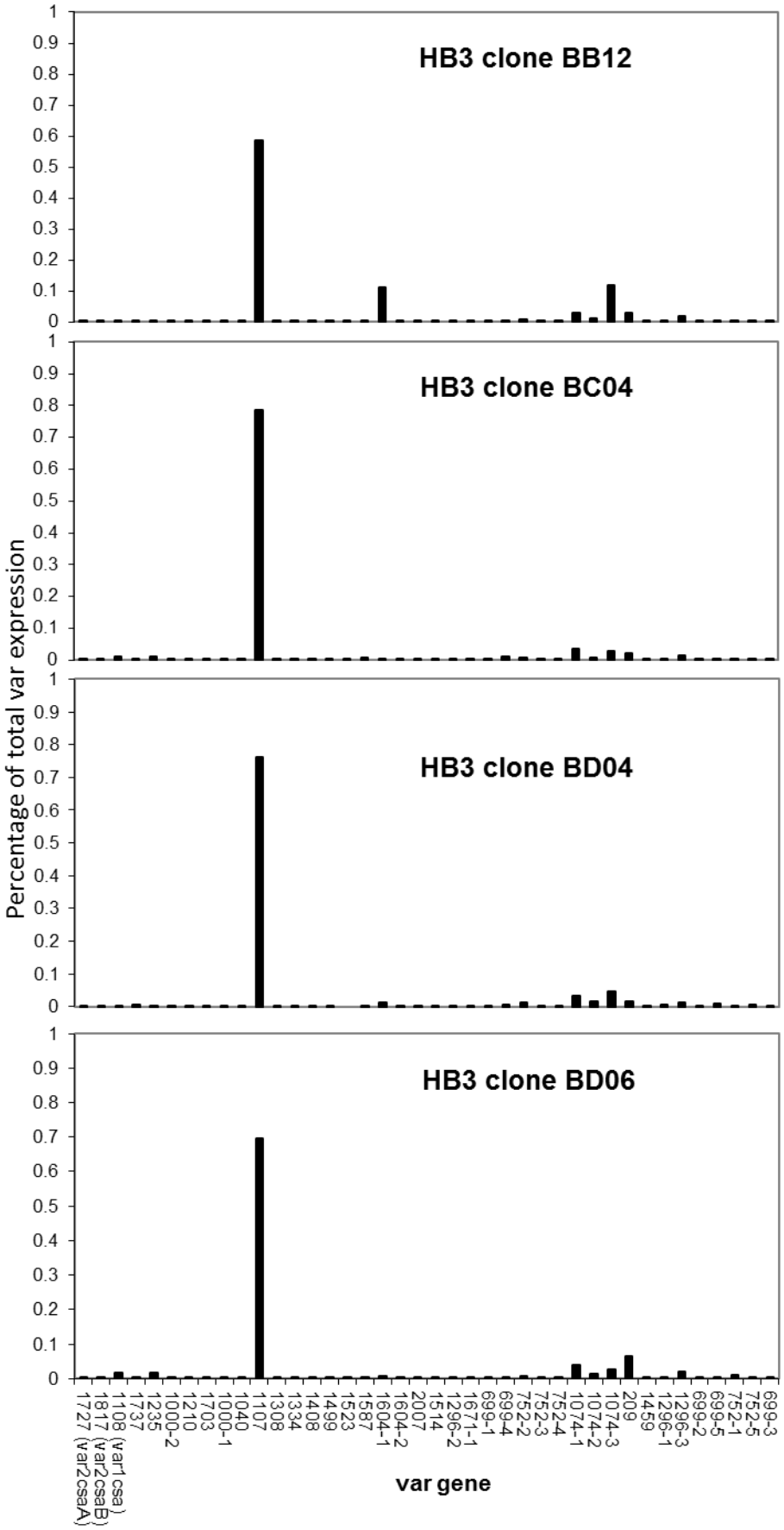
consistent with the common deletion event on chromosome 2 (191). In these cases northern blots confirmed correct staging. The expression profiles of the progeny clones are shown in figure 4.2. Nine different *var* genes were expressed as the dominant transcript in the 31 clones (Figure 4.3a and Table 4.3). The profile of the parent culture is shown in figure 4.3 b. This has been cultured without selection pressure for many generations and shows a similar pattern to the frequency of the dominant *vars* expressed in the clones.

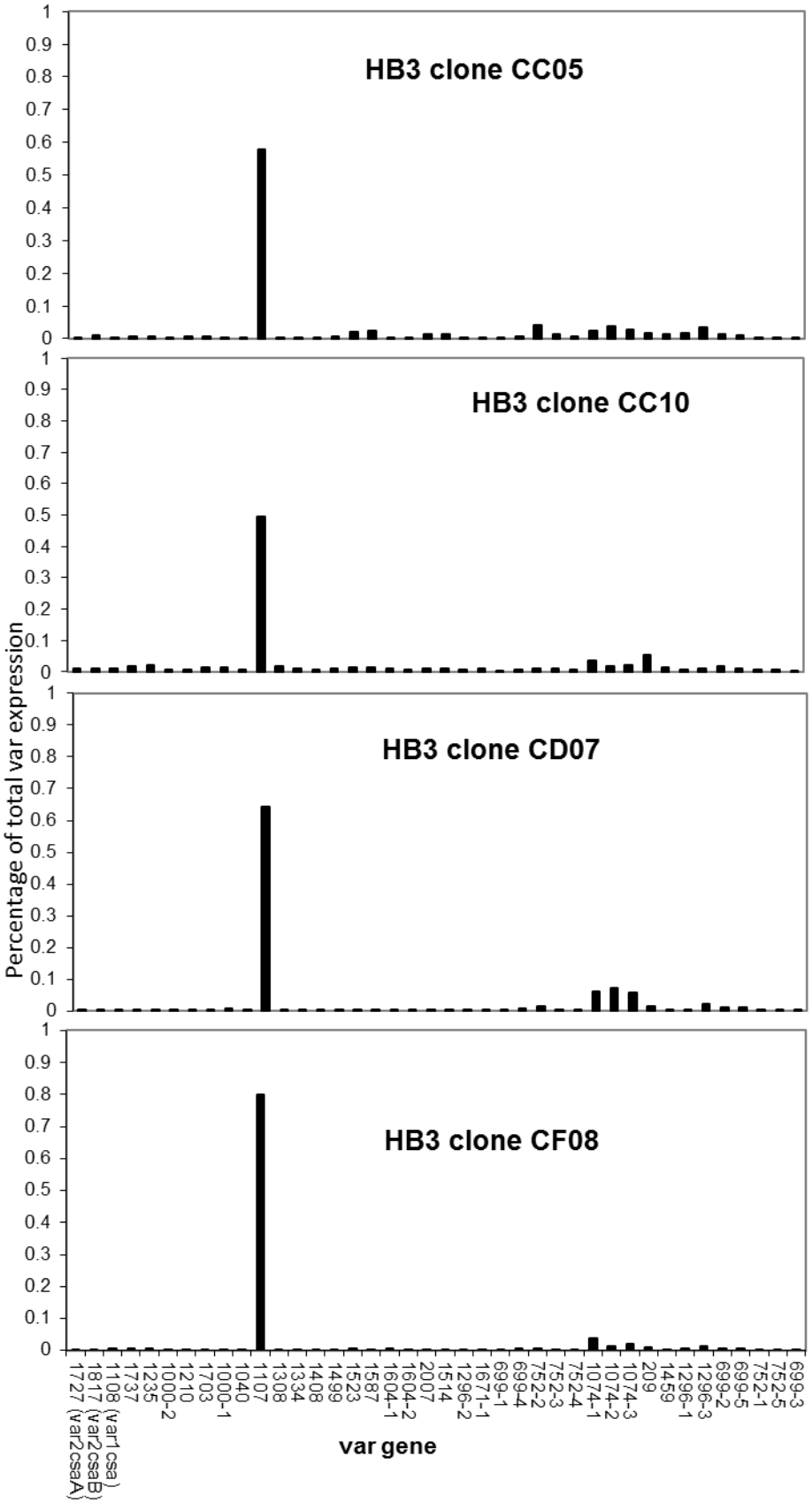
Examples of the 3 major *var* Ups promoter classes, UpsA, UpsB and UpsC were found within the panel. The most prevalent was HB3-1107, an UpsB gene that is predicted to be telomeric but has not been mapped to a chromosome (74). Both subtelomeric and central UpsB classes were expressed as dominant genes. All 3 *var* genes from HB3 contig 1074, a

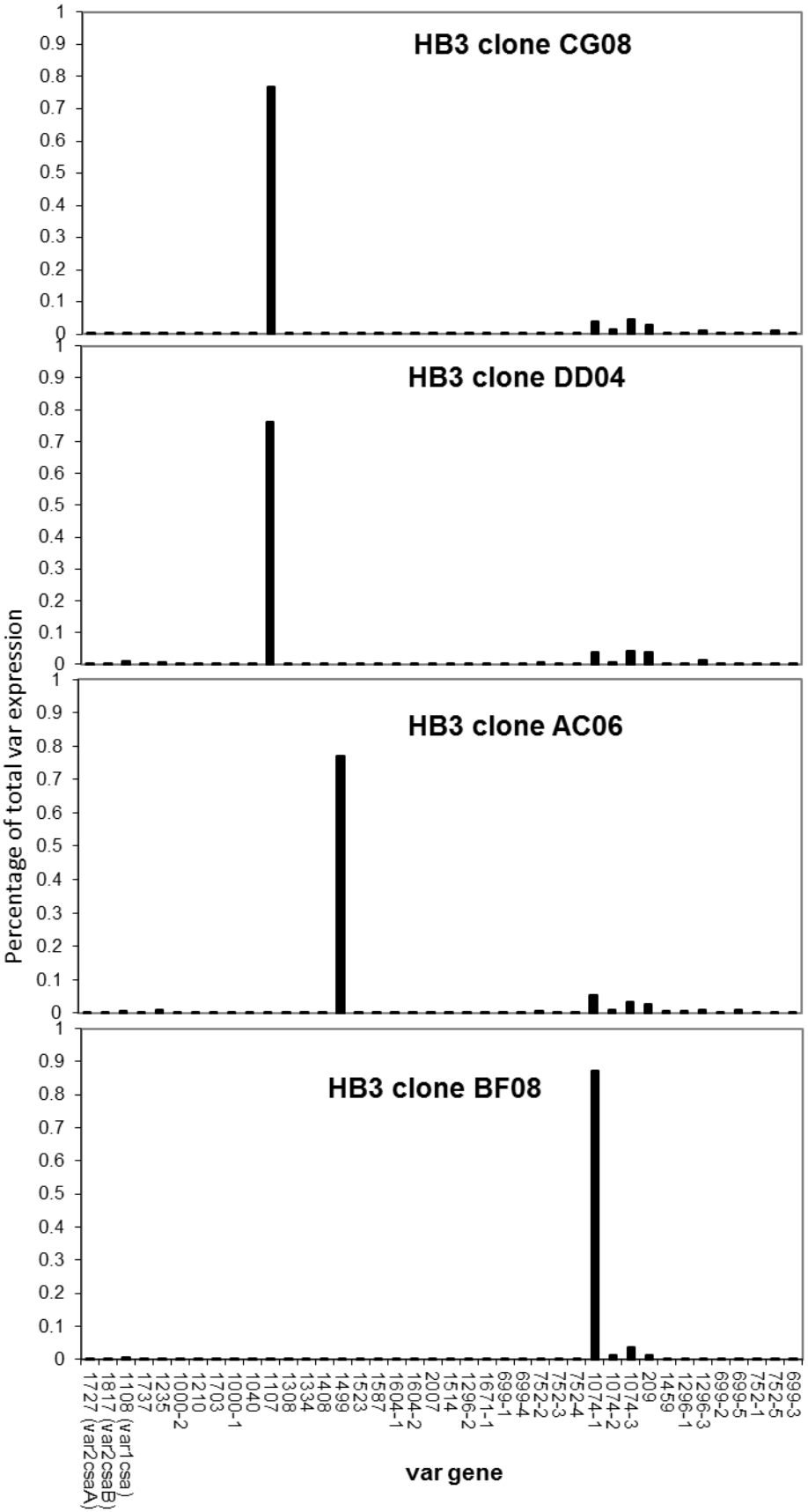
central cluster on chromosome 4 were represented. HB3-1074-1 is an UpsB type and HB3-1074-2 & 3 are UpsC types.

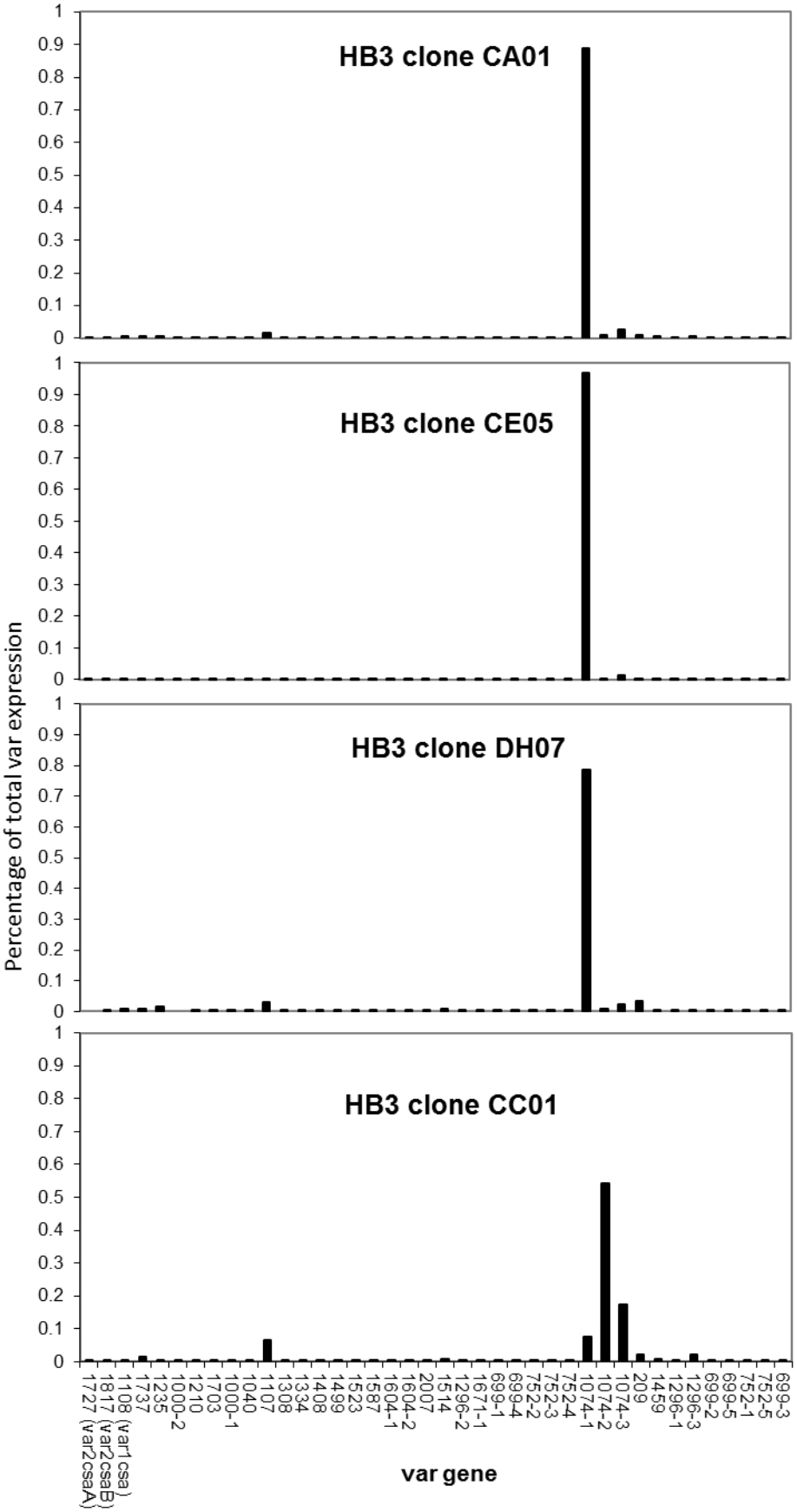
Figure 4.2. *var* gene expression profiles of 31 HB3 clones. Expression is shown as percentage of total *var* gene expression, following relative expression analysis to seryl-tRNA synthetase.

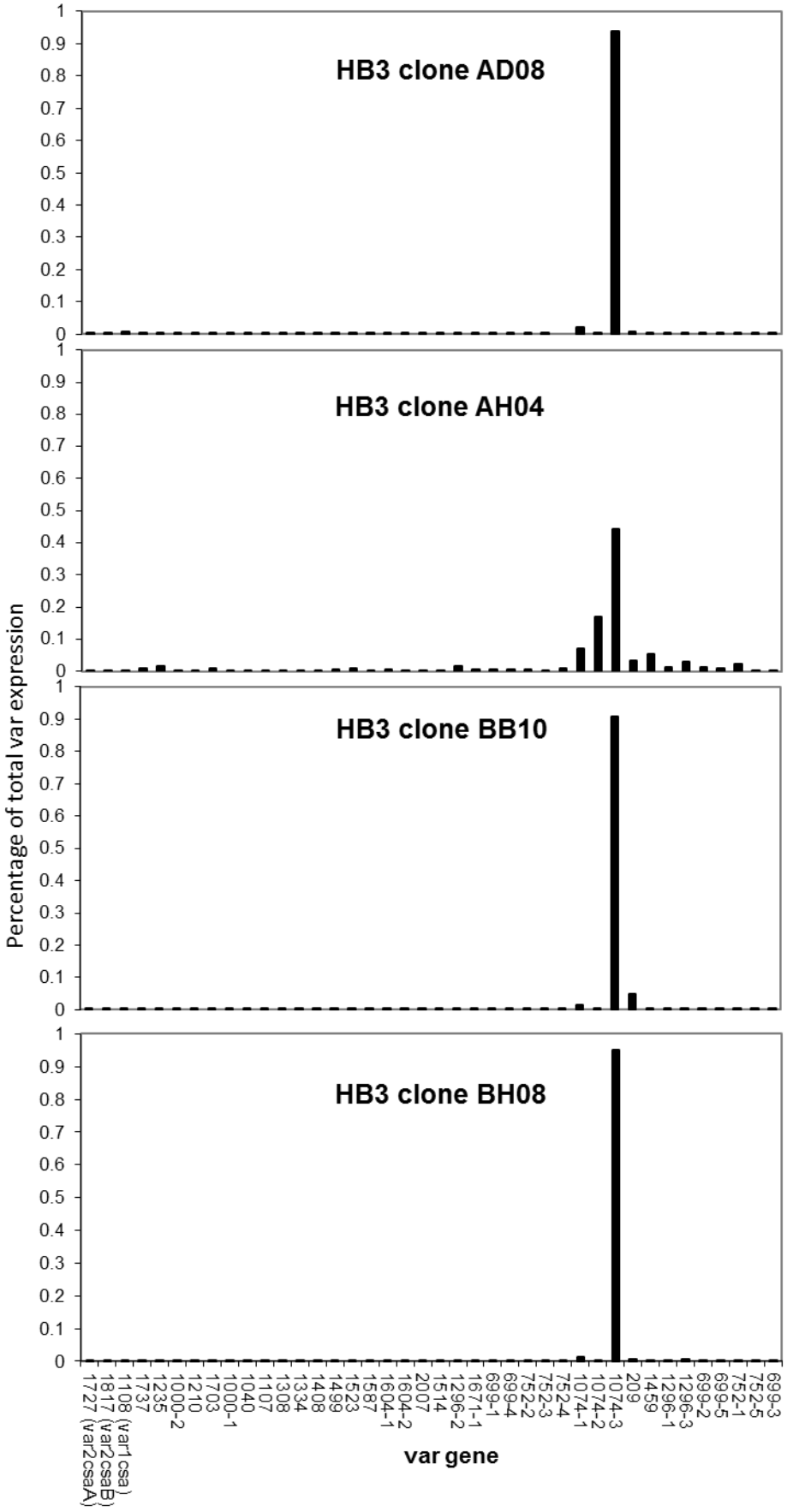


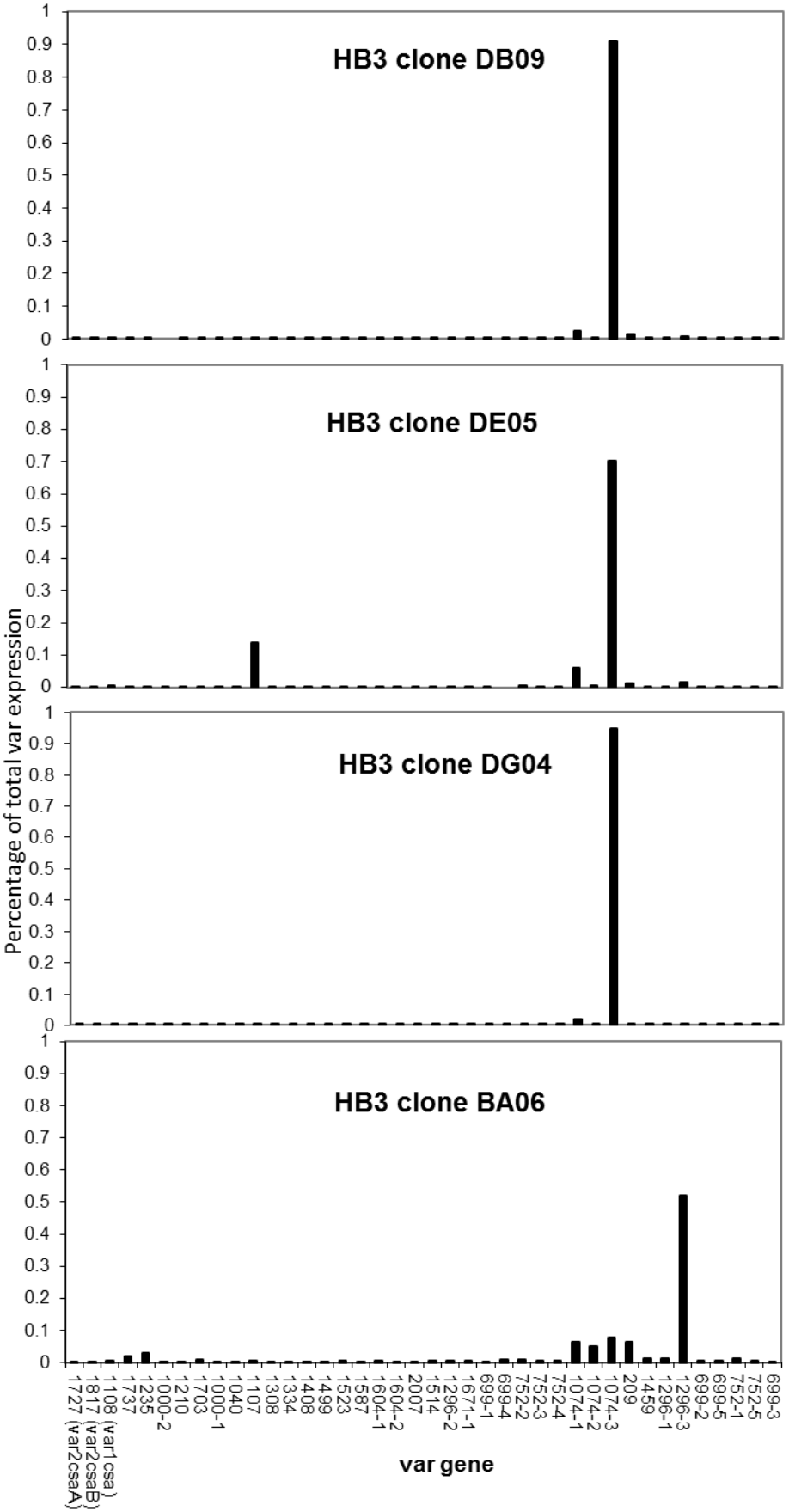

HB3 clone BB12
HB3 clone BC04
HB3 clone BD04
HB3 clone BD06
Percentage of total var expression
var gene
1727 (var2csaA)
1817 (var2csaB)
1108 (var1 csa)
1737
1235
1000-2
1210
1703
1000-1
1040
1107
1308
1334
1408
1499
1523
1587
1604-1
1604-2
2007
1514
1296-2
1671-1
699-1
699-4
752-2
752-3
752-4
1074-1
1074-2
1074-3
209
1459
1296-1
1296-3
699-2
699-5
752-1
752-5
699-3

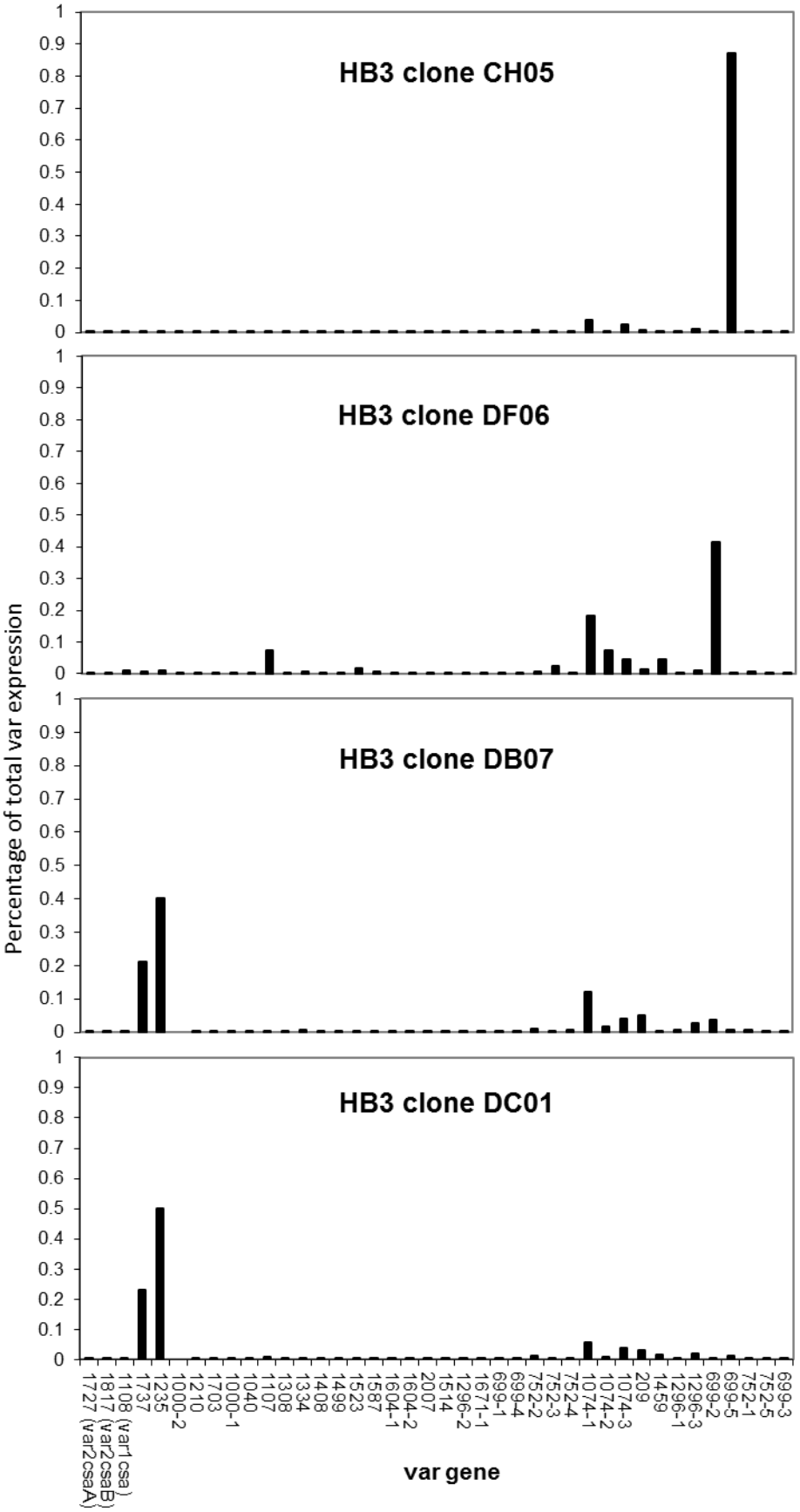

HB3 clone CC05
HB3 clone CC10
HB3 clone CD07
HB3 clone CF08
Percentage of total var expression
var gene
1727 (var2csaA)
1817 (var2csaB)
1108 (var1csa)
1737
1235
1000-2
1210
1703
1000-1
1040
1107
1308
1334
1408
1499
1523
1587
1604-1
1604-2
2007
1514
1296-2
1671-1
699-1
699-4
752-2
752-3
752-4
1074-1
1074-2
1074-3
209
1459
1296-1
1296-3
699-2
699-5
752-1
752-5
699-3


HB3 clone CG08
HB3 clone DD04
HB3 clone AC06
HB3 clone BF08
Percentage of total var expression
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
1727 (var2csaA)
1817 (var2csaB)
1108 (var1csa)
1737
1235
1000-2
1210
1703
1000-1
1040
1107
1308
1334
1408
1499
1523
1587
1604-1
2007
1514
1296-2
1671-1
699-1
699-4
752-2
752-3
752-4
1074-1
1074-2
1074-3
209
1459
1296-1
1296-3
699-2
699-5
752-1
752-5
699-3
var gene


HB3 clone CA01
HB3 clone CE05
HB3 clone DH07
HB3 clone CC01
Percentage of total var expression
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
1727 (var2csaA)
1817 (var2csaB)
1108 (var1csa)
1737
1235
1000-2
1210
1703
1000-1
1040
1107
1308
1334
1408
1499
1523
1587
1604-1
1604-2
2007
1514
1296-2
1671-1
699-1
699-4
752-2
752-3
752-4
1074-1
1074-2
1074-3
209
1459
1296-1
1296-3
699-2
699-5
752-1
752-5
699-3
var gene


HB3 clone AD08
HB3 clone AH04
HB3 clone BB10
HB3 clone BH08
Percentage of total var expression
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
1727 (var2csaA)
1817 (var2csaB)
1108 (var1csa)
1737
1235
1000-2
1210
1703
1000-1
1040
1107
1308
1334
1408
1499
1523
1587
1604-1
1604-2
2007
1514
1296-2
1671-1
699-1
699-4
752-2
752-3
752-4
1074-1
1074-2
1074-3
209
1459
1296-1
1296-3
699-5
699-2
752-1
752-5
699-3
var gene


HB3 clone DB09
HB3 clone DE05
HB3 clone DG04
HB3 clone BA06
Percentage of total var expression
var gene
1727 (var2csaA)
1817 (var2csaB)
1108 (var1csa)
1737
1235
1000-2
1210
1703
1000-1
1040
1107
1308
1334
1408
1499
1523
1587
1604-1
1604-2
2007
1514
1296-2
1671-1
699-1
699-4
752-2
752-3
752-4
1074-1
1074-2
1074-3
209
1459
1296-1
1296-3
699-2
699-5
752-1
752-5
699-3


HB3 clone CH05
HB3 clone DF06
HB3 clone DB07
HB3 clone DC01
Percentage of total var expression
var gene
1727 (var2csaA)
1817 (var2csaB)
1108 (var1csa)
1737
1235
1000-2
1210
1703
1000-1
1040
1107
1308
1334
1408
1499
1523
1587
1604-1
1604-2
2007
1514
1296-2
1671-1
699-1
699-4
752-2
752-3
752-4
1074-1
1074-2
1074-3
209
1459
1296-1
1296-3
699-2
699-5
752-1
752-5
699-3

HB3 *var* gene	Ups type	Chromos-omal location[1]	Chromos-ome[1]	Number of clones	Percentage of total *var* expression
1235	A	Telomeric	5	2	40%-50%
1107	B	Telomeric	Unknown	13	49%-79.9%
1499	B	Telomeric	Unknown	1	77%
1074-1	B	Central	4	4	78.5%-97%
1074-2	C	Central	4	1	54%
1074-3	C	Central	4	7	44%-95%
699-2	C	Central	12	1	41%
699-5	C	Central	12	1	87%
1296-3	C	Central	7	1	52%

[1] Chromosome assignment and location from Kraemer et al (74).
Table 4.3. Summary of *var* gene expression data for 31 HB3 clones.

Confirmation of HB3 qRT-PCR results by northern blot analysis

To verify the results obtained by qRT-PCR, northern blots were prepared with 10 µg RNA. A generic exon 2 probe was used first to confirm the number of transcripts in each clone. Gene specific probes were then sequentially hybridized to the filters. The hybridization results are shown in figure 4.4, and confirm the qPCR data.

Clones DC01 and DB07 appear to simultaneously express two UpsA genes

Clones DC01 and DB07 appeared to express two similar UpsA type genes, HB3-1737 and HB3-1235. To further investigate this possible expression of two var genes in a single clone, the sequences of the two genes were more closely examined. A pseudogene HB3-1278 was also found to have regions of similarity and was included in the investigation.

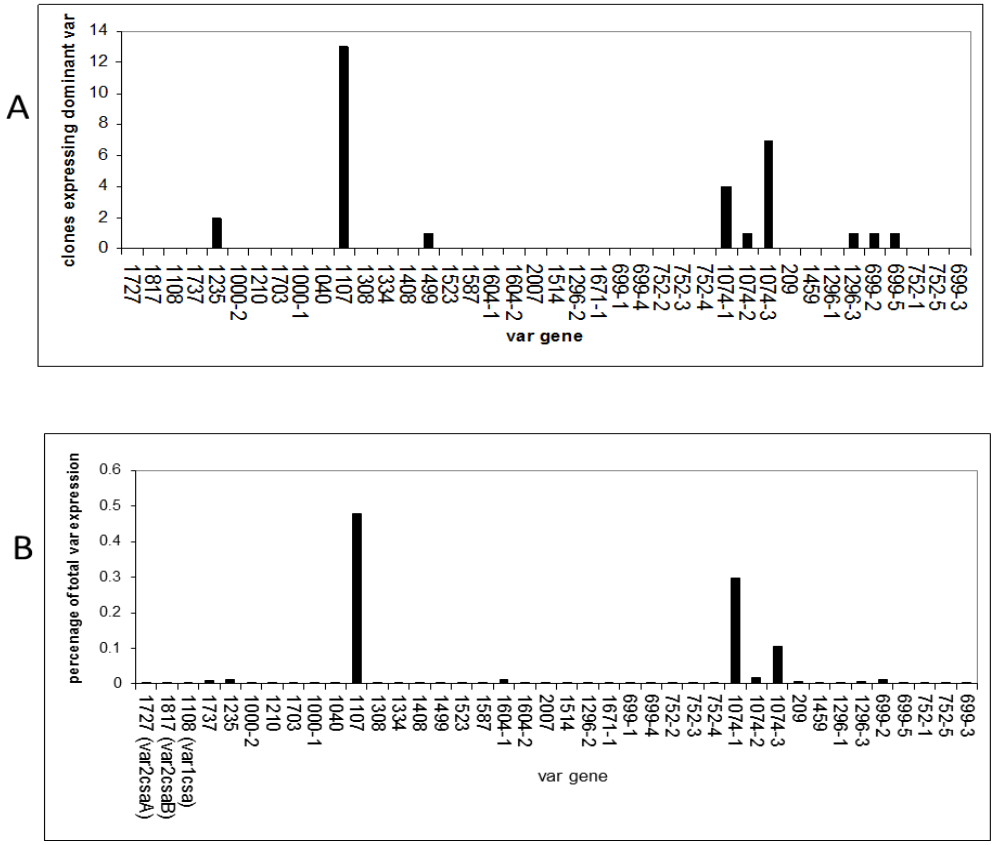


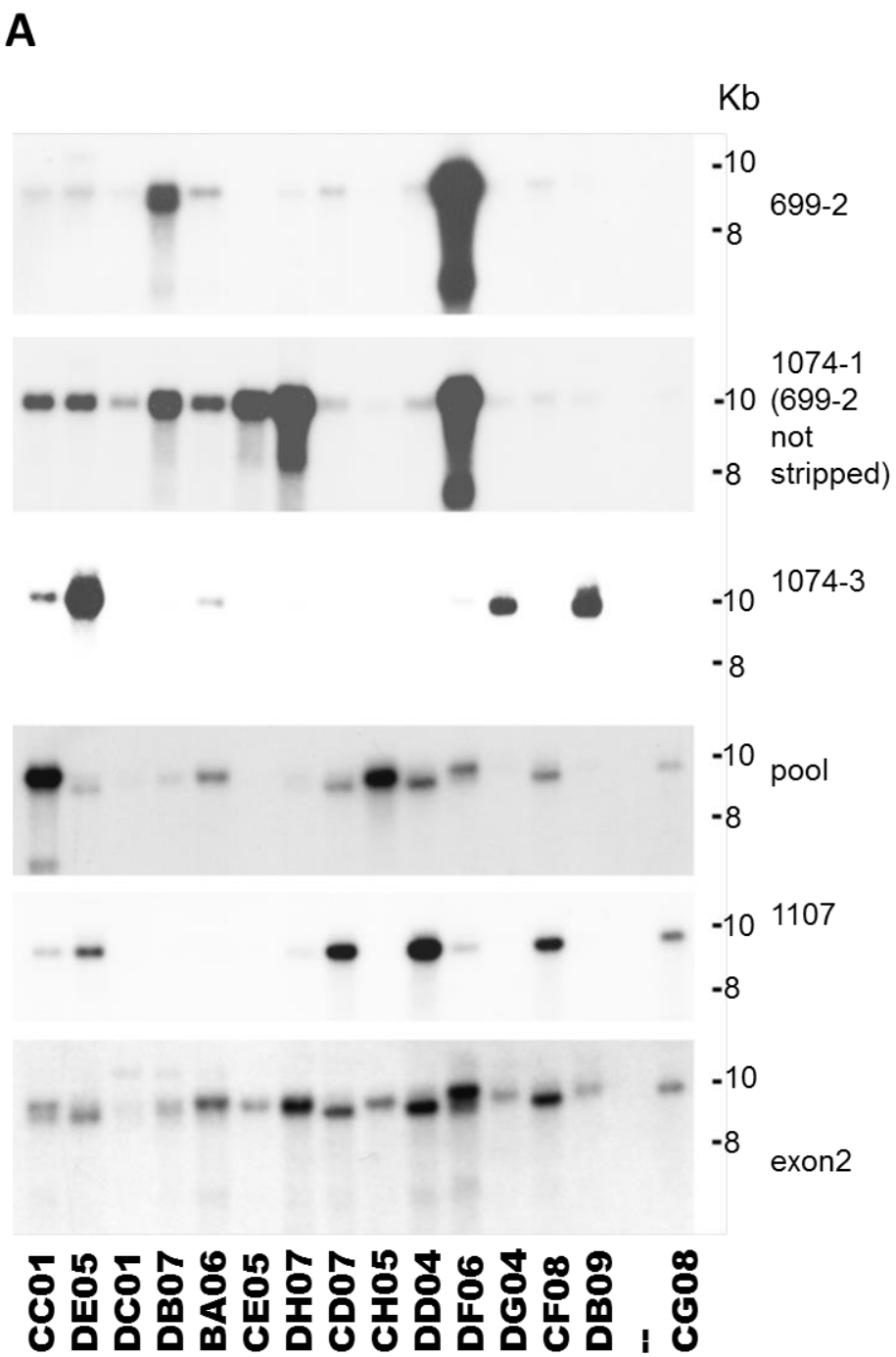
Figure 4.3 *Var* expression in HB3 clones and parent. A. Frequency of dominant *var* gene expression in panel of 31 clones (number of clones expressing each type). B. Expression profile of unselected HB3 parental culture.

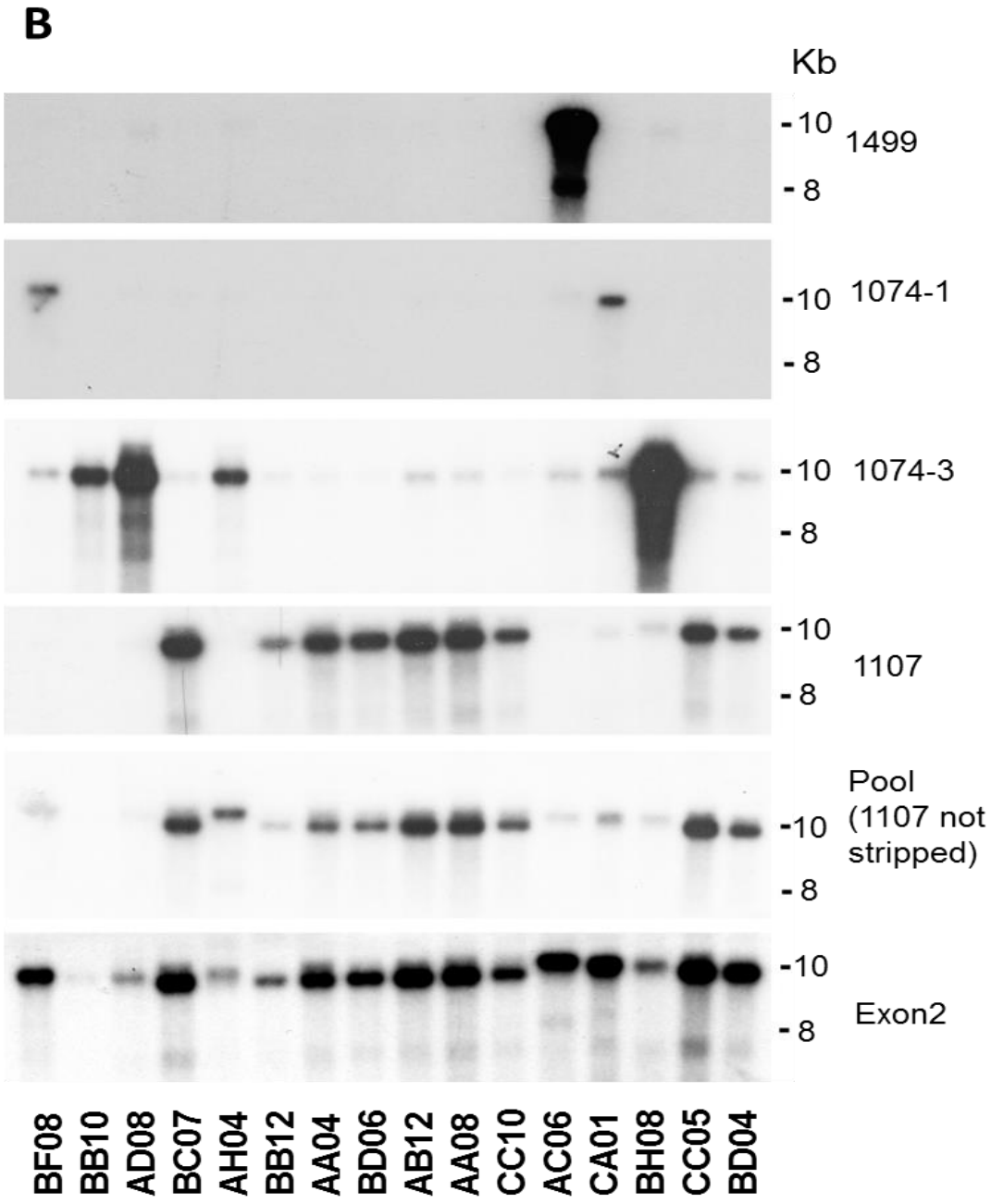
A schematic of the genes is shown in figure 4.5. Clearly, the qPCR primers HB3-1737hr are not gene specific and will also detect HB3-1235 and HB3-1278. Two pairs had been designed for HB3-1737 in the non-homologous sequence region, but neither of these worked. Additionally, the HB3-1235 pair is one of two pairs that did not meet the parameters set out in the methods.

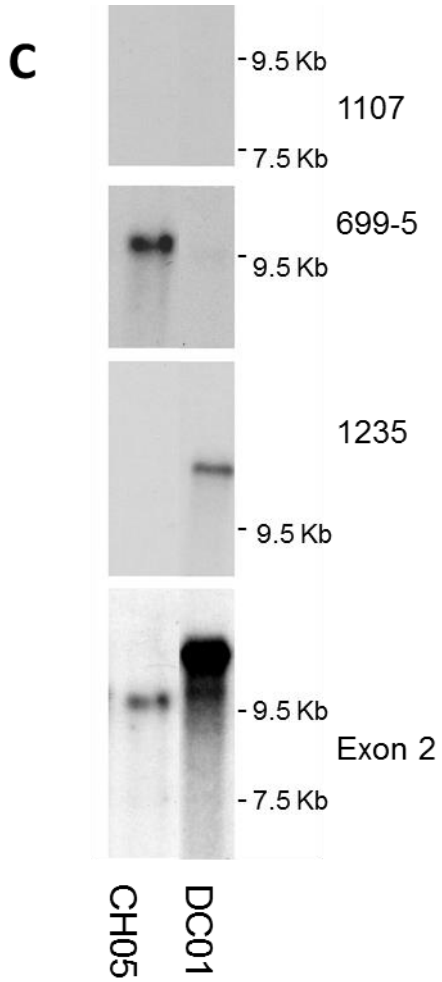
To resolve this, hybridization probes were made. The HB3-1737 probe did not hybridize to the RNA. To ensure that there was not a technical problem, HB3-1737 was hybridised to a Southern blot and did detect the gDNA sequence. The HB3-1235 hybridization result shown in figure 4.5

confirms that clones DC01 and DB07 express HB3-1235, an UpsA gene mapped to chromosome 5 (74).

Figure 4.4. Verification of qPCR data obtained for HB3 clones by northern blot hybridization. **A**. Northern blot hybridized sequentially with gene specific and generic exon2 DNA probes. Some of the probes were not stripped before the next was added, this is indicated. The pool probe was a mixture of HB3-1074-2, HB3-1296-3 and HB3-699-5. **B**. As for A, with different clones. **C**. Separate confirmation of clone CH05 result with HB3-699-5, and DC01 with HB3-1235.




B
Kb
10
8
1499
10
8
1074-1
10
8
1074-3
10
8
1107
10
8
Pool
(1107 not
stripped)
10
8
Exon2
BF08
BB10
AD08
BC07
AH04
BB12
AA04
BD06
AB12
AA08
CC10
AC06
CA01
BH08
CC05
BD04



Alterations to the published 3D7 qPCR primer set

A previous student had already altered some of the 3D7 primer set to increase efficiency. These and the published set were tested against HB3 gDNA. Nineteen new pairs were designed to increase isolate specificity or to increase gene specificity, and are shown in appendix i.

***var* gene expression in progeny of 3D7xHB3 genetic cross**

The HB3 and modified 3D7 primer sets were tested on cDNA from 5 progeny of the 3D7xHB3 cross. Initially both full sets of primers were tested on each cDNA and the data expressed relative to seryl-tRNA synthetase which was included in each run. It was noted that as expected,

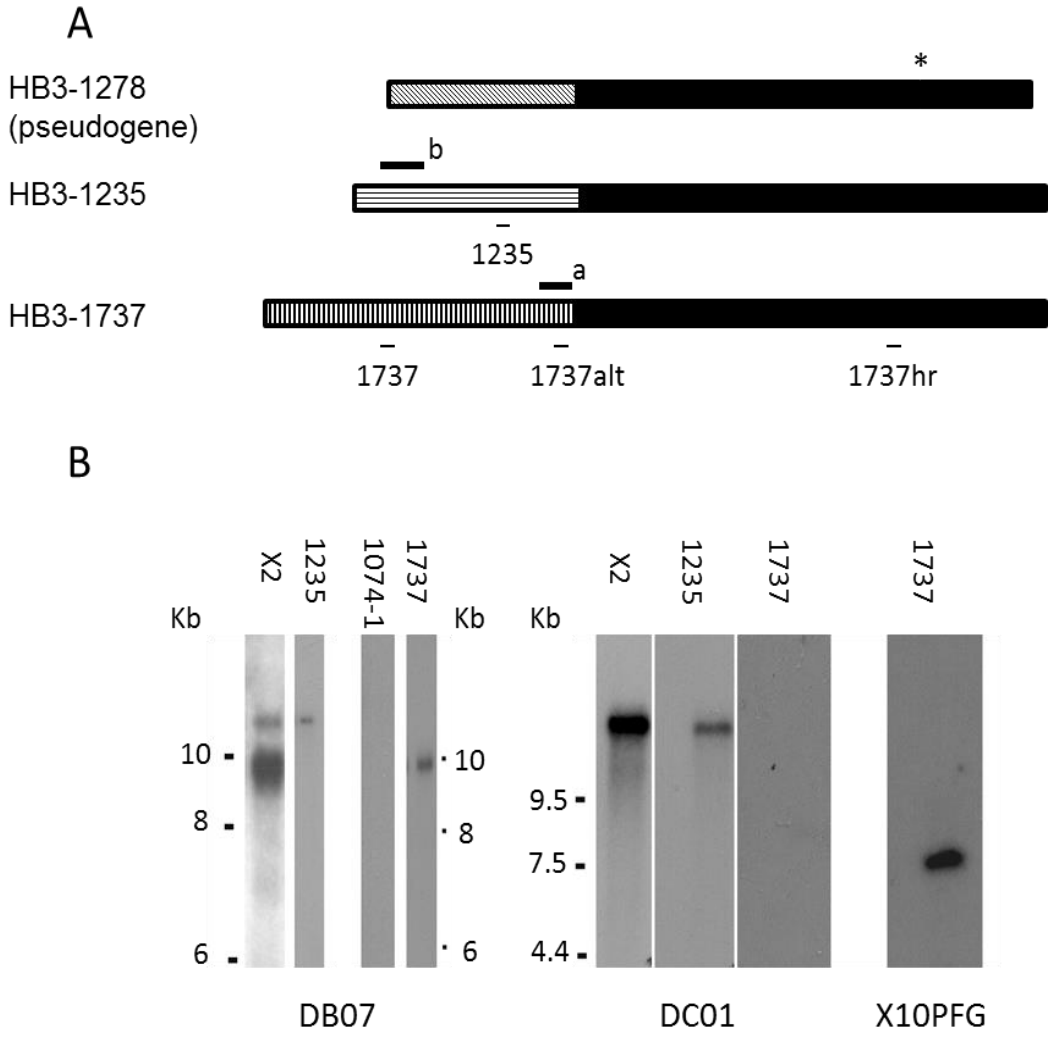


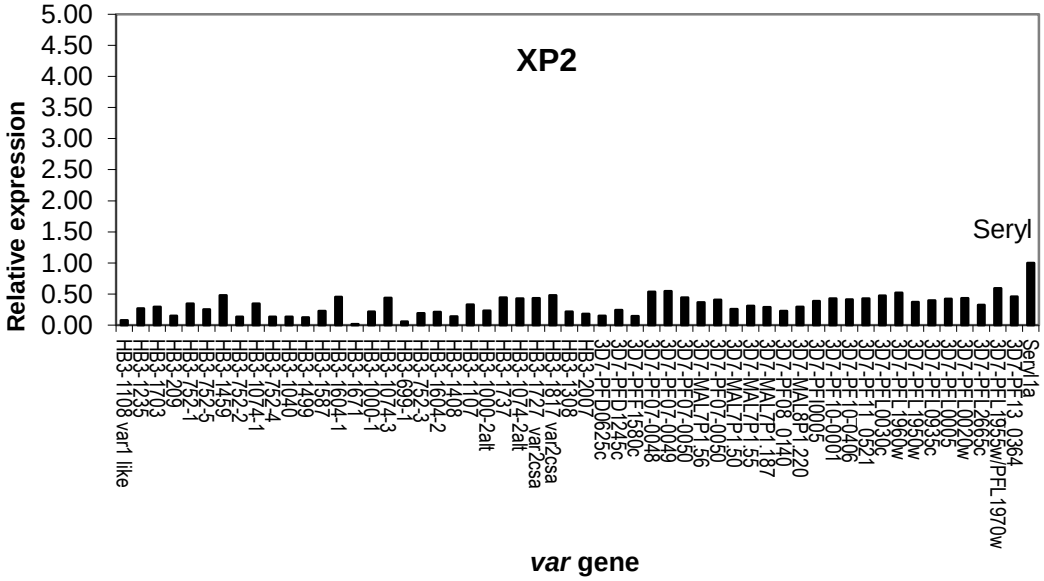
Figure 4.5 Verification of expression in clones expressing similar *var* genes.
A. Schematic of HB3 *var* genes 1737, 1235 and 1278 (pseudogene). The solid areas indicate areas of identical sequence between the 3 genes. * indicates a frameshift in the pseudogene HB3-1278. The qPCR amplicon locations are indicated below each gene. 1737 and 1737alt did not work; 1737hr are used in this study. Clearly the 1737hr primers will detect all three sequences. The bars above the genes show the location of hybridization probes a.HB3-1737, b. HB3-1235. **B**. Northern blot data for clones DB07 and DC01. The HB3-1737 probe is A and HB3-1235 probe is C in the schematic above. A Southern blot pulsed field gel, X10PFG of 3D7xHB3 progeny clone X10 was hybridized as a positive control for probe a.

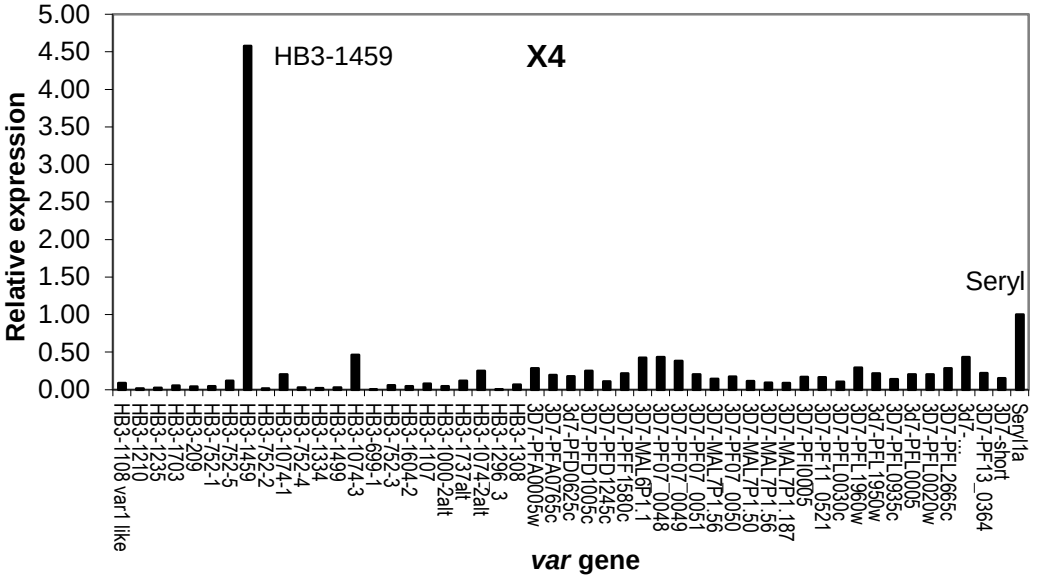
approximately 50% of the *var* genes primer pairs did not produce a product. As each progeny is assumed to be from a meiotic recombination event, it is expected that each will contain a mixture of *var* genes from both 3D7 and HB3, but unlikely a full repertoire of both.These were easy to identify, appearing as a flat-line on the raw data curve. Unfortunately the analysis software converted these to false positives (by taking the second derivative of the curve), so these had to be removed manually. To improve efficiency of analysis, gDNA from each cross progeny was tested with the two primer sets. Only those *var* genes present in a progeny's genome were used in its subsequent data analysis.

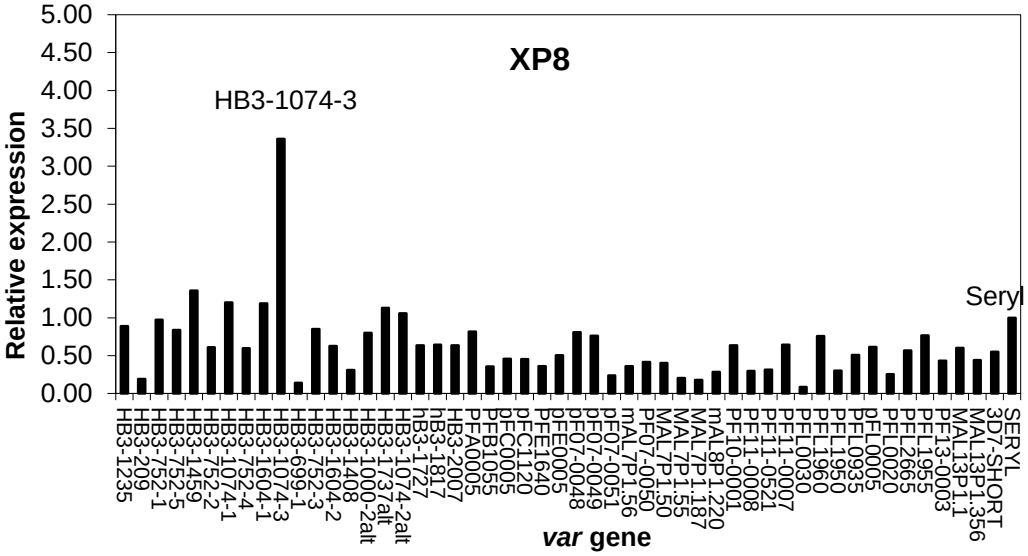
Although the ring stage control gene was expressed more abundantly than the trophozoite control, a northern blot hybridized with the exon 2 probe indicated that the staging of the RNA preparation was incorrect for 3 of the samples, XP8, X10 and X33 (data not shown). These were recultured and the qPCR repeated. For these samples, only the amplicons that were present in the genome were included in the reactions. This meant that a complete genome set could be processed in one run. The data was treated as for the HB3 clones, but a correction factor was not applied to the results (see methods). The *var* gene qRT-PCR expression data are shown in figure 4.6.

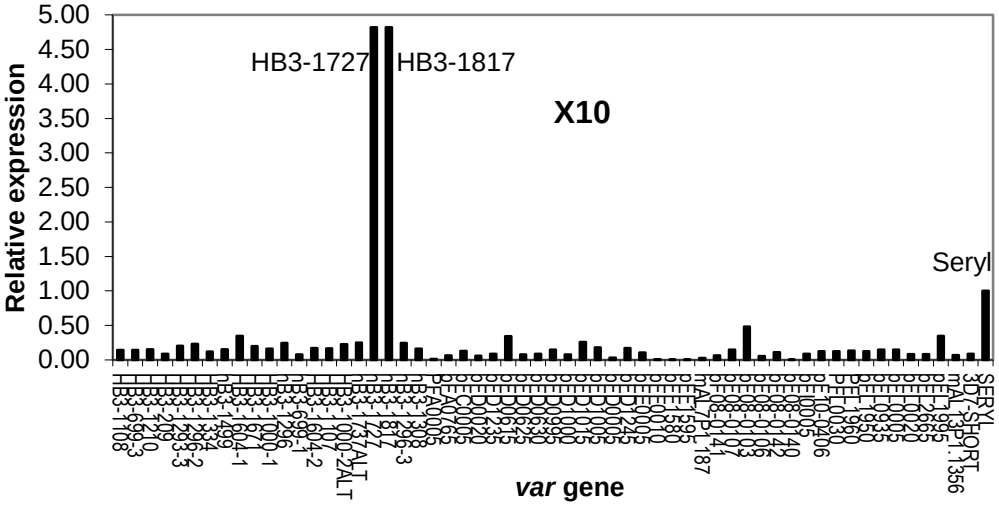
The progeny clones XP2 and X33 do not appear to express a dominant *var* gene. It is common for a population of parasites that have been cultured for many generations not to express a dominant *var* gene. Without selection for a particular adhesion type, switching will occur.

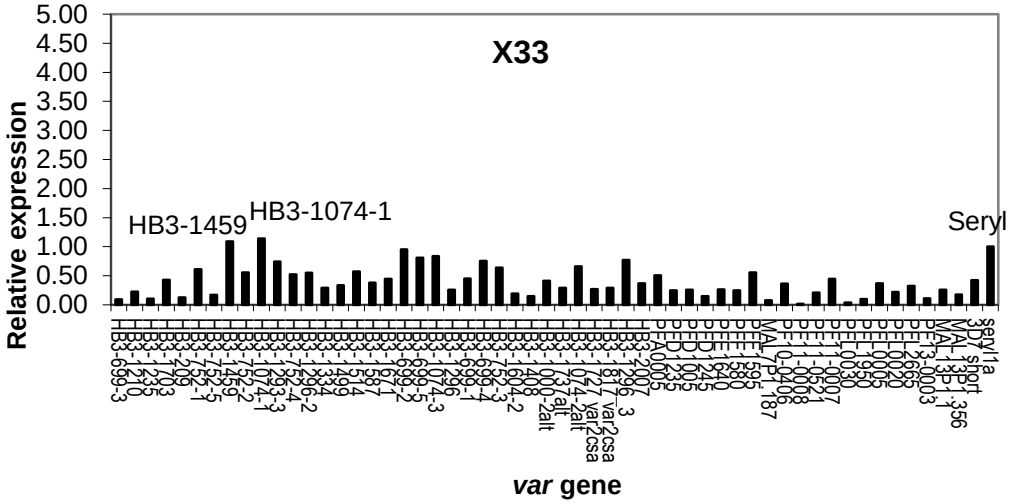
Figure 4.6. *var* gene expression profiles of 3D7x HB3 progeny. Expression is shown relative to seryl-tRNA synthetase. Clones XP2 and X33 show no dominant *var* expression. Clone X4 expresses HB3-1459, clone XP8 expresses HB3-1074-3 and clone X10 expresses HB3-1817*var2csa* and HB3-1727*var2csa.*






XP8
HB3-1074-3
Seryl
Relative expression
5.00
4.50
4.00
3.50
3.00
2.50
2.00
1.50
1.00
0.50
0.00
var gene


X10
HB3-1727
HB3-1817
Seryl
Relative expression
5.00
4.50
4.00
3.50
3.00
2.50
2.00
1.50
1.00
0.50
0.00
var gene


X33
HB3-1459
HB3-1074-1
Seryl
Relative expression
5.00
4.50
4.00
3.50
3.00
2.50
2.00
1.50
1.00
0.50
0.00
var gene

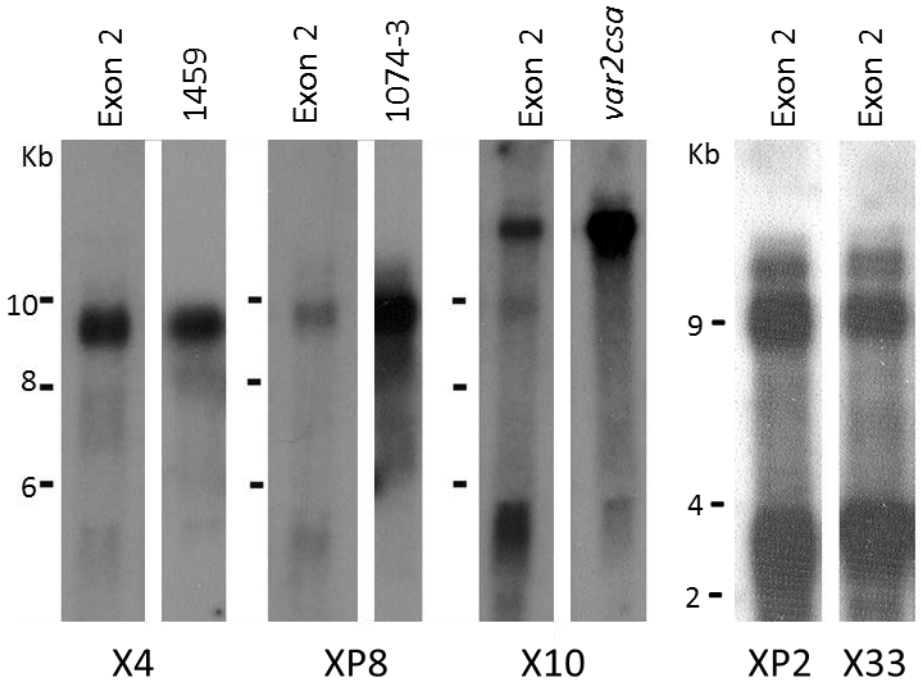


Figure 4.7 Verification of *var* qPCR expression data by northern blot.
X4. Clone X4 has a dominant *var* transcript. qPCR and northern data confirm this to be HB3-1459.
XP8. Clone XP8 has a dominant *var* transcript. qPCR and northern data confirm this to be HB3-1074-3.
X10. Clone X10 has a dominant and a minor *var* transcript. qPCR and northern data confirm the dominant transcript to be *var2csa*.
XP2 and X33. Both of these clones have multiple transcripts.

Exon 2 probe hybridized at 50°C and washed at 50°C in 0.5xSSC/0.1% SDS. All specific *var* probes hybridized at 60°C and washed at 65°C in 0.5xSSC/0.1% SDS. The small transcripts <6kb are sterile exon 2 transcripts present in trophozoite stages and indicate a staging issue for the RNA preparation

The northern blot data support this for clones XP2 and X33, as there are several bands with the exon 2 probe. There is however a dominant band visible. This could suggest that the qPCR primers have not detected this *var* gene, or the dominant band could comprise signals from similar size var genes.

The qRT-PCR results for clone X10 shows that it expresses an HB3 *var2csa* gene. This is the UpsE type, upregulated in CSA binding to IE, of

which there are two copies in HB3. Although our qPCR primers can distinguish between the HB3 and 3D7 *var2csa* genes, I was unable to test for cross-reactivity of the *var2csa* primer pairs within HB3. Therefore I cannot distinguish between the expression of either one of the HB3 *var2csa* genes or both.

Clones X4 and XP8 both express UpsC type *var* genes, predicted to be located on different central *var* clusters on chromosome 4 (74). HB3-1459 expressed in clone X4 is a solitary *var* gene, whereas HB3-1074-3 expressed in clone XP8 is located within a cluster of three *var* genes. These data are summarized in table 4.4.

Use of *var* primer sets to identify meiotic recombination events in cross progeny

We used the information obtained from qPCR of gDNA of the cross progeny to see if we could construct a map of the genomes identifying regions where meiotic recombination had occurred. Using the fully sequenced 3D7 genome and the partial sequence and chromosomal localisation available for HB3, we could identify blocks of genes from the respective parents. Using the 3D7 genome as a scaffold we then aligned the blocks along each chromosome. The *var* gene repertoire of the cross progeny assigned by chromosome is shown in table 4.5. The data are incomplete because eleven of the HB3 *var* genes have yet to be assigned to a chromosome.

Discussion

The set of qPCR primers presented here for the HB3 clone will be of use to the *var* research community. They have been tested on 31 HB3 clones,

and the qRT-PCR results verified by northern blot hybridization. They are based on the sequencing data from the Broad institute (73) and analysed by Kraemer and colleagues (74). If this data set is updated in the future it may be necessary to revisit the sequences and update them accordingly. One issue with *var* qPCR was identified during the validation of the HB3 primer set. Distinguishing between genes with regions of high sequence similarity can be problematic. In this HB3 primer set, there are two pairs that detect two genes each, HB3-1737/HB3-1235 and the two *var2csa* pairs. Although not ideal, this is substantially reduced from the published 3D7 *var* gene primer set, for which seven pairs had two or more target genes (56).

Clone	**XP2**	**X33**	**X10**		**X4**	**XP8**
var expression	mixed	mixed	HB3 *Var2csa*		HB3-1459	HB3-1074-3
Ups type			Ups E		Ups C	Ups C
SuperContig			1.262	1.295	183	100
Contig			1727	1817	1459	1074
Presumed chromosome			12	1	4	4
Location			subtelomeric	unknown	Central	Central
Adjacent *var*			none	unknown	None[1]	1074-1 (UpsB) 1074-2 (UpsC)

Table 4.4 *var* gene details for cross progeny. Summary of *var* gene expression, Ups type*, location* and contig* details for cross progeny. * from (74,192) [1] solitary var in contig.

Table 4.5 Identification of regions of meiotic recombination in 3D7xHB3 progeny using *var* primer sets.

3D7 genes are shown in red and HB3 in blue. The *var* genes are organised in order from chromosome 1-13. Where a column changes from red to blue within a chromosome, a crossover is assumed to have occurred in the region between the 2 *var* genes. Genes shown in black are in unresolved areas. In these cases, it is possible that recombination may have led to multiple *var* genes in these locations, or because not all of the HB3 *var* gene locations have been assigned. [1] The short primers amplify 3 genes, so this anomaly could be explained by the short primers amplifying a different gene. [2] The 1737 primers also amplify 1235, which is present in these clones, so this could be a false positive.

Var genes		Var repertoire of Cross progeny				
3D7 var	HB3 var	XP2	X4	XP8	X10	X33
Chromosome 1						
PFA0005w			PFA0005w	PFA0005w		PFA0005w
PFA0015c(short)			short	short	short	short
PFA0765c	1817	1817	PFA0765c	1817	1817	1817
Chromosome 2						
PFB0010w						
PFB1055c				PFB1055c		
Chromosome 3						
PFC0005w				PFC0005w	PFC0005w	

PFC1120c /PFC0005w				PFC1120c		
Chromosome 4						
PFD0005w					PFD0005w	
PFD0020c					PFD0020c	
PFD0615c	1074-1	1074-1	1074-1	1074-1	PFD0615c	1074-1
PFD0625c	1074-2	1074-2	1074-2	1074-2	PFD0625c	1074-2
PFD0630c /PFD0635c	1074-3	1074-3	1074-3	1074-3	PFD0630c	1074-3
PFD0635c	1459	1459	1459	1459		1459
PFD0995c					PFD0995c	
PFD1000c					PFD1000c	
PFD1005c			PFD1005c		PFD1005c	PFD1005c
PFD1015c					PFD1015c	
PFD1235w /MAL7P1.1					PFD1235w	PFD1235w
PFD1245c	1161				PFD1245c	PFD1245c
Chromosome 5						
PFE0005w	1235	1235	1235	PFE0005w/1235	PFE0005w	1235
PFE1640w				PFE1640w		PFE1640w

Chromosome 6

PFF0010w					PFF0010w	
PFF0020(short)			short	short	short	short
PFF0845c	209	209	209	209	209	209
PFF1580c			PFF1580c			PFF1580c
PFF1595c			PFF1595c		PFF1595c	PFF1595c

Chromosome 7

MAL7P1.212	1210		1210		1210	1210
PF07_0048	1278	PF07_0048	PF07_0048	PF07_0048		
MAL7P1.50	1296-1	MAL7P1.50	MAL7P1.50	MAL7P1.50	1296	1296-1
PF07_0049	1296-2	PF07_0049	PF07_0049	PF07_0049	1296-2	1296-2
PF07_0050	1296-3	PF07_0050	PF07_0050	PF07_0050	1296-3	1296-3
PF07_0051		PF07_0051	PF07_0051	PF07_0051		
MAL7P1.55	1671-2	MAL7P1.55	MAL7P1.55	MAL7P1.55	1671-2	1671-2
MAL7P1.56	1671	MAL7P1.56/1671	MAL7P1.56	MAL7P1.56	1671	1671
MAL7P1.187		MAL7P1.187	MAL7P1.187	MAL7P1.187	MAL7P1.187	MAL7P1.187

Chromosome 8

PF8_0142					PF8_0142	
PF08_0141					PF8_0141	
PF08_0140		PF8_0140			PF8_0140	
	752-1	752-1	752-1	752-1		752-1
PF08_0103	752-2	752-2	752-2	752-2	PF8_0103	752-2
PF08_0106	752-3	752-3	752-3	752-3	PF8_0106	752-3
PF08_0107	752-4	752-4	752-4	752-4	PF8_0107	752-4
MAL81.207	752-5	752-5	752-5	752-5		752-5
MAL8P1.220				MAL8P1.220		

Chromosome 9

PFI0005w	1040	PFI0005w/1040	PFI0005w		PFI0005w	
PFI1820w(short)	1000-2	1000-2	short/1000-2[1]	short/1000-2[1]	short/1000-2[1]	short/1000-2[1]
PFI1830c	1000-1	1000-1			1000-1	

Chromosome 10

PF10_0001	2007	PF10_0001 /2007		PF10_0001/2007		2007
PF10_0406	1308	PF10_0406 /1308	1308		PF10_0406 /1308	PF10_0406

Chromosome 11

PF11_0007	1737	1737/1235	1737/1235	PF11_0007/1737[2]	1737/1235	PF11_0007 /1737[2]
PF11_0008	1523			PF11_0008		PF11_0008
PF11_0521		PF11_0521	PF11_0521	PF11_0521		PF11_0521
Chromosome 12						
PFL0005w	1727	PFL0005/1727	PFL0005w	PFL0005/1727	PFL0005/1727	PF0005/1727
PFL0020w	1514	PFL0020w	PFL0020w	PFL0020w	PFL0020w	PFLL0020 /1514
PFL0030c		PFL0030c	PFL0030c	PFL0030c	PFL0030c	PFL0030c
	699-1	699-1	699-1	699-1	699-1	699-1
PFL0935c	699-2	PFL0935c	PFL0935c	PFL0935c	PFL0935c	699-2
PFL1950w	699-3	PFL1950w	PFL1950w	PFL1950w	PFL1950w/ 699-3	699-3
PFL1955w/PFL1970w	699-4	PFL1955w/ PFL1970w	PFL1955w/ PFL1970w	PFL1955w/ PFL1970w	PFL1955w	699-4
PFL1960w	699-5	PFL1960w	PFL1960w	PFL1960w	PFL1960w	699-5
PFL2665c	1181	PFL2665c	PFL2665c	PFL2665c	PFL2665c	PFL2665c
Chromosome 13						
PF13_0001	1587	1587		PF13_0001		PF13_0001 /1587

PF13_0003				PF13_0003		PF13_0003
PF13_0364		PF13_0364	PF13_0364	PF13_0364	PF13_0364	PF13_0364
	unassigned					
	1678					
	1604-1	1604-1		1604-1	1604-1	
	1604-2	1604-2	1604-2	1604-2	1604-2	1604-2
	1107	1107	1107		1107	
	1426					
	1499	1499	1499		1499	1499
	1334		1334		1334	1334
	1408			1408		1408
	1703	1703	1703			1703
	1108	1108	1108		1108	

The extensive sequence homology between the genes HB3-1737, HB3-1235 and pseudogene HB3-1278, made it difficult to design sequence specific primers. The HB3-1737 pair detects all three genes and the HB3-1235 pair fall outside the optimal parameters (56). The raw qRT-PCR data for DC01 and DB07 indicated HB3-1737 expression but not HB3-1235, supporting HB3-1737 expression. However, hybridization with HB3-1737 probes did not confirm the qRT-PCR result. We suspected a problem with quantification of these highly similar genes. When data were re-analysed including a correction factor for the primer efficiency, both HB3-1737 and HB3-1235 were expressed. Hybridization with an HB3-1235 probe confirmed its sole expression in these clones. Ideally, to improve the qPCR primer set, a new pair of primers would have to be developed for HB3-1737 that are sequence specific. As will be shown in chapter 6, the primer pair HB3-1235 has been improved. It will probably be necessary in the future to resolve the expression from either gene by hybridization.

Controls were included to confirm correct staging of parasites. KAHRP was chosen for the ring stage. Results from some of the HB3 clones showed no expression of this gene, although northern blot data with exon 2 confirmed correct staging. This is consistent with the well-documented observation that the end of chromosome 2 containing KAHRP is frequently deleted (193). In three of the cross progeny KAHRP was the most abundantly expressed gene, yet northern data indicated incorrect staging. This could be due to KAHRP expression continuing into early trophozoite forms. It may be better to design an alternative ring stage control.

The expression result obtained for the progeny clone X10 indicated that one or both *var2csa* genes were being expressed. It would be a very interesting result if both were expressed, as it would show non- mutually exclusive expression. On closer examination of the primer sequences in light of this result, it was found that there is only one base difference between forward primers for the two genes, and none in the reverse. The two copies of *var2csa* have 89.6% sequence similarity (192). Having attempted to design specific primers and failing, it was felt that it would be difficult to improve these within the design parameters.

A study in 2009 by Brolin and colleagues (192), described simultaneous transcription from both HB3 *var2csa* loci. They identified a region to design specific qPCR primers and used these to study transcription in single cells isolated by micromanipulation. In CSA selected HB3, they reported cells that expressed either one or other gene, and some that expressed both. They confirmed this by RNA-FISH, where 78% of cells expressed both genes, and co-localised to the same perinuclear location. Although these data show that it is possible to design specific primers, they also indicate that it may not be necessary. If CSA selected parasites do express both copies of *var2csa,* it would be unlikely that gene specific primers would be needed.

The qRT-PCR data for two clones, XP2 and X33 showed no dominant *var* gene expression. With a generic exon 2 probe XP2 has a dominant band plus one other band, and X33 has a dominant band plus two others. When clones are grown in continuous culture without selection it is likely that different cells within the culture will switch to different *var*

genes. Over many generations this will lead to a culture that simultaneously expresses many genes, due to sub-populations of cells each expressing a different *var* gene.

For clone XP2 although a dominant band was detected on a northern blot, no dominant *var* gene was detected by qRT-PCR. As the full sets of both 3D7 and HB3 primers were used for qRT-PCR on this clone, it cannot be due to a mis-assignment of genes chosen for testing the clone-specific repertoire. Although a genome specific set of primers were applied to X33, the *var* repertoire had been assigned by PCR on gDNA (table 4.5). A correction factor was not applied to the cross progeny results. This was due to the cost and time involved in running replicate sets of qPCR for each clone's gDNA, as each clone's primer set would have its own correction factor. Also, any result was to be confirmed by northen blot. As was seen with DC01 and DB07, the correction factor was crucial in identifying the dominant *var* gene. With these HB3 clones there was some expression of the dominant *var* (HB3-1235) above reference levels in the raw data. No expression above reference levels was detected with the raw data for XP2. It is therefore unlikely that a correction factor would alter the lack of a dominant qRT-PCR result. One possible explanation of the XP2 and X33 expression results is that meiotic recombination may have occurred within *var* genes. Two groups have reported recombination events within *var* genes in the progeny of the 3D7xHB3 cross (94,194). More recently, Frank and colleagues used the 3D7 qPCR primers on gDNA from a parasite E5, derived by limiting dilution cloning from NF54 (195). They found that 23 genes were not detected in the E5 genome, including several telomeric clusters. This was suggested to be due to gene

conversion events. Either of these recombination events would lead to the XP2 and X33 results, a dominant northern signal that cannot be detected by qRT-PCR. Sequence data from the WTSI shows no evidence of recombination within *var* genes in these progeny (C. Newbold personal communication). As recombination has not occurred, the most likely explanation of the northern signal is that the dominant band is comprised of signals from multiple *var* transcripts of similar sized genes.

The use of the primer set to analyse cross progeny gDNA for precise sites of meiotic recombination was limited for two main reasons. Firstly, as several of the HB3 genes are not assigned to a chromosome, the data set is incomplete. Secondly, *var* gene data only give an indication that a crossover event has occurred. Unless this falls within a *var* gene cluster, it may only narrow the region down to several megabases.

The HB3 primer set described and validated here is a useful addition for investigating *var* gene expression in a third standard laboratory isolate. Their use in distinguishing highly similar sequences such as those of HB3-1235 and HB3-1737 was resolved by applying a correction factor to the entire dataset. The HB3 primer set was also optimised to be used on progeny of the 3D7xHB3 cross. To further enhance *var* gene analysis on the 3D7xHB3 progeny, the published 3D7 set (56) was altered to ensure isolate specificity. Despite this, some *var* genes may not be detected due to recombination within *var* genes. Therefore, to use qRT-PCR on the progeny of a genetic cross successfully, each progeny clone must be treated as a new genome and a set of clone specific primers designed.

Chapter 5

GRE analysis techniques

Introduction

Having developed reagents to follow *var* gene expression in HB3 and the cross progeny, we needed to develop methods to analyse their GRE components. In order to investigate this we would like to have a reproducible, quick method for GRE sequence determination. This needs to be able to distinguish individual GRE sequences obtained from cDNA. Preliminary investigations into GRE expression utilised basic cloning and sequencing of RT-PCR products. This method is reliable, if enough clones are sequenced, but it is lengthy and time consuming. Other methods were tested to see if they were suitable for GRE analysis.

The first is High Resolution Melt PCR (HRM) a technique first described by Wittwer and colleagues (196). HRM uses "3rd generation" dyes, which are intercalating dyes that fully saturate double stranded DNA (dsDNA) and fluoresce when the DNA is double stranded. Following a PCR reaction, the temperature is slowly increased. As this happens, the dsDNA disassociates to single stranded DNA (ssDNA) which does not fluoresce. Every DNA species has a characteristic melt profile which is influenced by the amplicon length, sequence and GC content. This is now established

as a genotyping technique to differentiate single nucleotide polymorphisms (SNP) as an alteration at a single base can be detected.

The use of HRM was first reported in *P. falciparum* by Andriantsoanirina and co-workers in 2009 (197). They looked at point mutations in genes associated with drug resistance. Although most of the SNPs were tested successfully, they noted that SNPs that were close together were not. This was in agreement with Kristensen who had previously reported that efficient genotyping relies on only one SNP per amplicon (198). Their study on genes involved in methyl metabolism also suggested that primers should not cover a SNP. Due to the polymorphic nature of the GRE, it may be unsuitable for HRM.

The second method to be tested for GRE expression profiling is Mismatch PCR, originally described as Amplification Refractory Mutation System (ARMS) by Newton and colleagues, and PCR Amplification of Specific Alleles (PASA) by Botteria both in 1989 (199,200). The theory behind this technique is that if the 3' nucleotide of the primer does not pair with the template sequence, *Taq* polymerase cannot extend the product (199,200). Work by Qiang and colleagues showed that a single mutation in a primer is not always sufficient to abrogate the action of *Taq,* and they introduced another alteration 3 nucleotides in from the 3' end (201). Primers are designed that match the SNP sequence, but do not match the genomic sequence at the 3' end. Lack of the test PCR product implies that the genomic sequence is present. As lack of a product is the diagnostic tool for system, an internal control PCR is required to ensure that the reactions have worked. To apply this technique to GRE, it will be

necessary to have primers for each polymorphism within the GRE genomic sequence.

Work in this chapter will attempt to apply the techniques described above to GRE analysis. It will also describe alterations to basic PCR methods required for accurate GRE expression data.

Material and methods

Identification of HB3 GRE sequences

The GREs in the HB3 genome were identified by BLAST searches of the Broad Institute's HB3 database using a 3D7 GRE as a query sequence (73,175). To compare the RNA structures, each HB3 GRE was analysed using the RNA folding programme, Mfold (126). Structural prediction comparisons of multiple sequences used the RNAalifold programme (202).

GRE expression analysis

High Resolution Melt PCR

To study GRE expression, high resolution melt PCR (HRM) was attempted. The genomic GRE sequences from HB3 and 3D7 isolates were compared and primers designed by the author. The primers are shown on a pileup of the sequences as figure 5.1 and in appendix i. To test for PCR artefacts, each T primer was paired with each B primer in 10 µl reactions (0.2 µM each primer, 1 µl 10xPCR buffer ll (Applied Biosystems), 3 mM $MgCl_2$, 0.2 mM dNTPs, 0.05 units AmpliTaq (Applied Biosystems)). PCR conditions were as follows: 95^oC for 5 minutes; followed by 35 cycles of 95^oC for 15 seconds, 60^oC for 15 seconds and 72^oC for 10 seconds. HRM PCR cycles were as above, but with 10 minute hold and 40 cycles. HRM

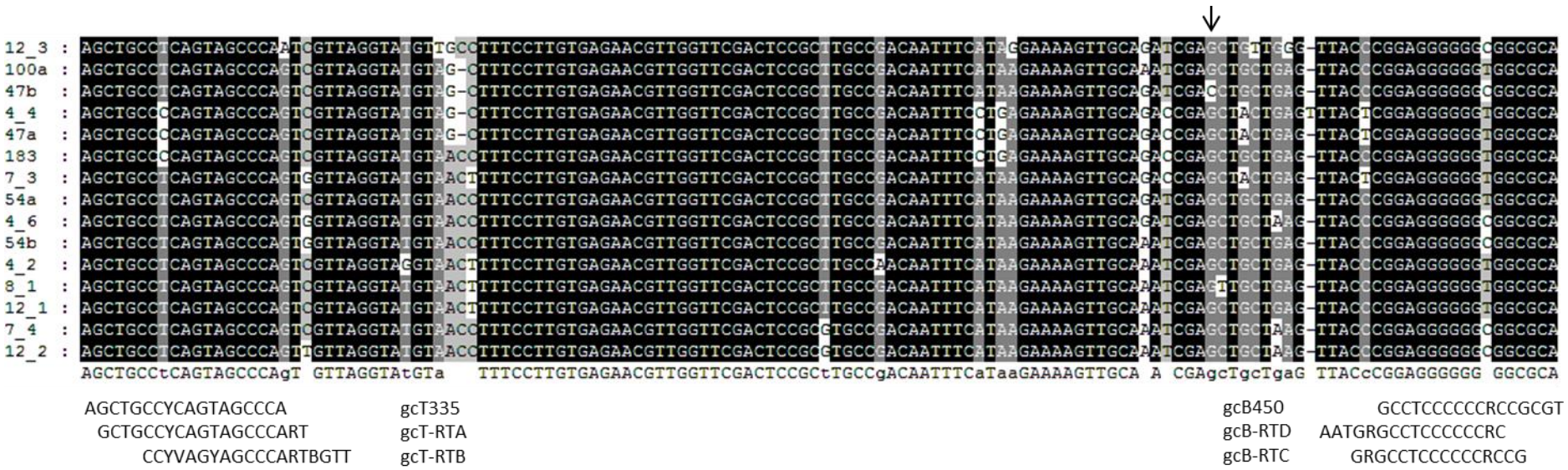


Figure 5.1 Location of oligonucleotides tested in HRM PCR
The 6 primers tested for HRM PCR are shown under a pileup of the GRE sequences found in HB3 and 3D7 genomes. Loci with identical sequences have been removed for simplicity. In HRM testing each gcT oligo is paired with each gcB oligo. The novel SNP found only in 47b is marked with an arrow.

was 65^{o}C – 95^{o}C (172). 10 µl reactions were 5 µl Sensimix (Quantace), 0.4 µl EvaGreen (Quantace), 0.2 µM primer gcT-RTB, 0.2 µM primer gcB-RTC and 1µl template.

A panel of DNAs representing known GRE sequences from the HB3 and 3D7 genomes was prepared. This was assembled from plasmid DNAs prepared using Wizard *Plus* SV Minipreps (Promega) (chapter 2). Each DNA represents a bacterial colony obtained from cloning RT-PCR products from preliminary experiments studying GRE expression in 3D7 and A4, a clone of the IT04 isolate (46). Novel sequences that had been identified in these experiments which did not match gDNA were also included in the panel. The details of the panel are shown in Table 5.1. Plasmid prep DNA was tested at 5 ng and 0.005 ng per reaction. cDNA was prepared using both random hexamers and the GRE primer gcB450 in standard 20 µl reactions. To increase the concentration of template in the reaction, 1 µl cDNA was added per reaction.

Mis-match PCR

Mis-match PCR was attempted as a method for analysing GRE expression. As the 3D7 genome sequence is complete, this was used as the test case. Primers were designed with the 3’ nucleotide matching the SNP sequence and to pair with previously designed primers gcB450 and gcT335. To cover the complete range of SNPs within the GRE genomic sequence it was necessary to design primers to both ends of the GRE. Control primers designed within the conserved portion of the GRE were also designed to amplify with gcB450 and gcT335. Primer sequences are shown in figure 5.2 and appendix i.

Figure 5.2 Primers designed for Mis-match PCR. SNP specific primers are shown in purple, with the diagnostic nucleotide enlarged. The control primers are as follows; Green = gcT control, Blue = gcB450, Red = gcT335, Orange = gcB control. A PCR reaction contains the specific primer paired with either, both gcT335 and gcT control or both gcB450 and gcB control.

Clone	3D7 GRE	HB3 GRE
NR32-1	4_2/4_3	
NR32-7	8_1	
NR13-5	12_1	
NR13-6	12_2	
NR13-8	7_4	
NR26-1	4_1/7_1/7_2	
3D716-3	4_6	
NR13-19	12_3	47c
A42B2-21		100a/100b
A42B2-11		54b
3D716-6		
A42B2-9		
3D716-2		
3D716-1		
NR32-6		
NR13-1		
A42B2-27		

Table 5.1. Panel of clones for HRM testing. Details of DNA samples used as a panel to test HRM. Those that are neither HB3 nor 3D7 sequences contained novel sequences. There are three 3D7 and three HB3 genomic GRE sequences missing that had not been found in the *E.coli* clones sequenced.

A pilot study was performed using four pairs of primers and plasmid DNA templates of known sequence (table 5.1). 20 µl pilot reactions contained 1 µM of each primer (gcB450, control T and the test T primer); 1x AmpliTaq buffer; 2.5mM $MgCl_2$; 200 µM dNTPs and 0.25 units AmpliTaq (Applied Biosystems). Cycling conditions were as follows: 94°C for 5 minutes,

followed by 35 cycles of 94°C for 15 seconds, 58°C for 15 seconds and 72°C for 15 seconds. Alterations were made to the dNTP concentration, with 20 µM and 2 µM tested, and the control primer concentration.

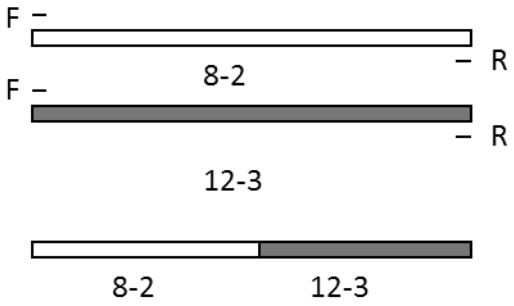
Cloning and Sequencing

Analysis of GRE expression and gDNA sequences was performed using the PCR, KAPA HiFi PCR (proofreading), cloning, wizard prep plasmid preps and sequencing techniques outlined in chapter 2. The routine gcT335/gcB450 PCR requires 3.5 mM $MgCl_2$, 45°C annealing temperature and 20 second extension.

Analysis of Potential Chimeras

To determine if novel sequences detected by KAPA HiFi PCR were due to PCR artefact, they were analysed by PCR to surrounding intergenic regions. If the novel sequences are in the genome, it should be possible to retrieve them from gDNA by PCR. If however the sequence is a PCR chimera, it will not be present in the gDNA. As the GRE sequences are so similar, it is necessary to move primers to intergenic regions. Intergenic sequences encompassing GRE involved in potential chimeras (4-4 and 7-3; 8-2 and 12-3; 12-2, 4-6 and 7-4) were obtained from PlasmoDB and compared by Clustalw (156,203). PCR primers to these potential chimeras were designed to be sequence specific with a Tm of 58-60°C (Appendix i). Diagrammatic details are shown in Figure 5.3. Regular PCR reaction conditions were optimised for each set per sequence being tested. To test for chimera NR13-F-18, for example, the following reactions were performed: 12-3F v 12-3R; 8-2F v 8-2R; 12-3F v 8-2R; and 8-2F v 12-3R. Analysis of PCR products was by agarose gel.

A



B Predicted results

PCR primers	8-2F/8-2R	12-3F/12-3R	8-2F/12-3R
PCR chimera	+	+	-
Mitotic event	+	+	+

Figure 5.3 Experimental Principles of testing for PCR chimeras.
A. Diagram of GRE sequences 8-2 and 12-3 showing F and R primers. Sequencing results predict the 8-2/12-3 hybrid.
B. Expected results of PCR with the primer combinations. If the hybrid sequence is as a result of a mitotic recombination event within the genome, a PCR product should be generated using a F and R primer from the separate sequences. If it is a PCR chimera, it will not be present within the genome. Note that both 8-2F/12-3R and 12-3F/8-2R will be tested for a mitotic event**.**

Results

HB3 GC rich elements

BLAST analysis of the HB3 genome revealed nine GRE. These are shown aligned with the 3D7 GRE in figure 5.1. To compare the structures with the hairpin structure predicted for all 3D7 GRE, each HB3 GRE was analysed by Mfold (126). Of the nine sequences, five have a single prediction of a hairpin as 3D7, figure 5.4. The identical sequences HB3-100a and b, and

another sequence HB3-54a have a primary prediction in common with 3D7, and 2 further predictions. These other predictions are still hairpin structures with a large bulge, either internally or at one end. The final GRE, HB3-47b has a hairpin prediction with a large bulge in the stem. The second prediction is similar to those of 3D7. Although the mfold predictions are similar structures, analysis with RNAalifold predicted a difference between 3D7 and HB3 GRE at a structural level, figure 5.5 (126,202). Removing the HB3-100a and b, and 47b sequences from the analysis altered the prediction so that it was an identical prediction to 3D7. There are no SNPs in HB3-100a or b that do not match those in other genomic sequences. There is a novel SNP in HB3-47b marked with an arrow in figure 5.1.

HRM PCR

In order to assign GRE sequences with HRM, it was necessary to assemble a panel of known sequences for comparison with test results. Although the panel in Table 5.1 is incomplete in terms of both 3D7 and HB3 genomic sequences, the sequences that had been found in cDNA in preliminary experiments were included. The primers in figure 5.1 were tested and following analysis on agarose gels (figure 5.6) the gcT-RTB/gcB-RTC combination was chosen for HRM.

Figure 5.7 A & B show results obtained with plasmid DNA containing a single GRE sequence. It was hoped that each GRE sequence would give a single peak, ideally with different melt temperatures. Although some of the plasmids gave single peaks, many of them had multiple peaks and a lot of these overlapped.

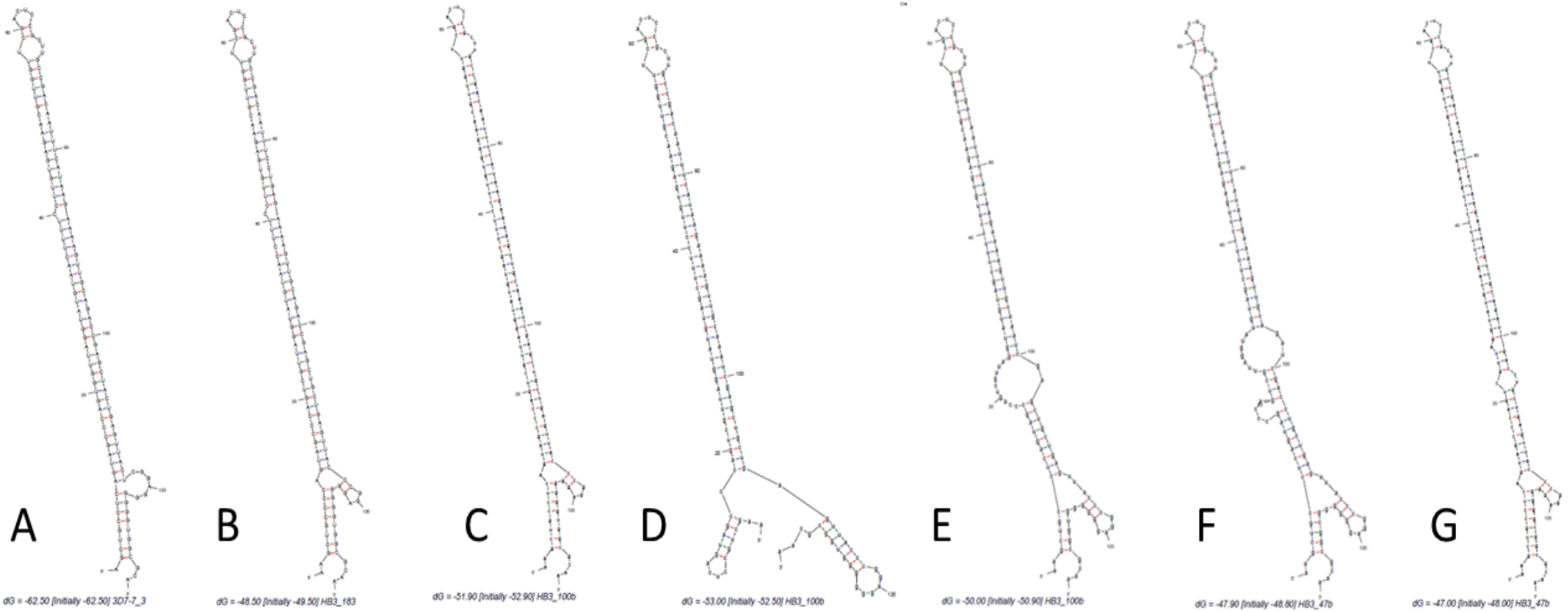


Figure 5.4. Mfold structure predictions of HB3 GRE.
Representative structures of GRE as predicted by the RNA folding programme mfold (126). A, 3D7-7_3. B, 3D7-4_1, 7_1, 7_2 and HB3-183. A and B represent the common prediction. C, D and E, HB3-100a and b, C is the strongest prediction and shows the common structure. D and E are other predicted structures. F and G, HB3-47b, F is the strongest prediction, but the common structure G is also predicted.

Initial experiments with random hexamer derived cDNA were unsuccessful. These were then repeated to compare cDNA generated with either random hexamers or gcB450, a GRE primer. Figure 5.7 C & D clearly indicate that the specific primer is required to generate cDNA for

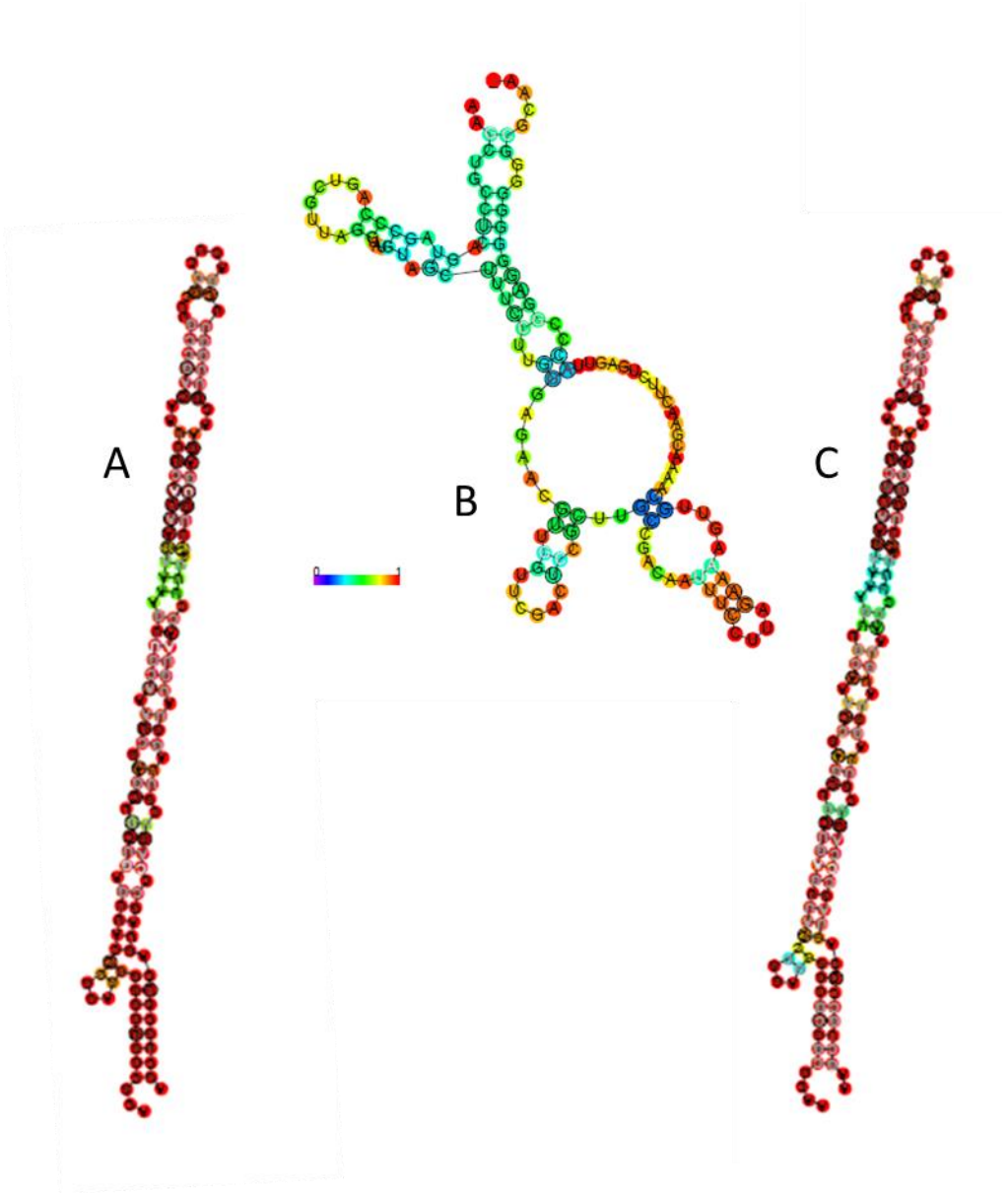


Figure 5.5. RNA alifold base pair probability predictions for GRE. Structure and base pair probability predictions for GRE sequences using RNA alifold (202). A, prediction for all 3D7 sequences. This predicts the structures in figure 5.5. B, prediction for all HB3 sequences. The structure of this prediction is significantly different to those predicted by mfold (126). C, if the sequences representing HB3-100a and b, and HB3-47b are removed, the prediction reverts to the common structure.

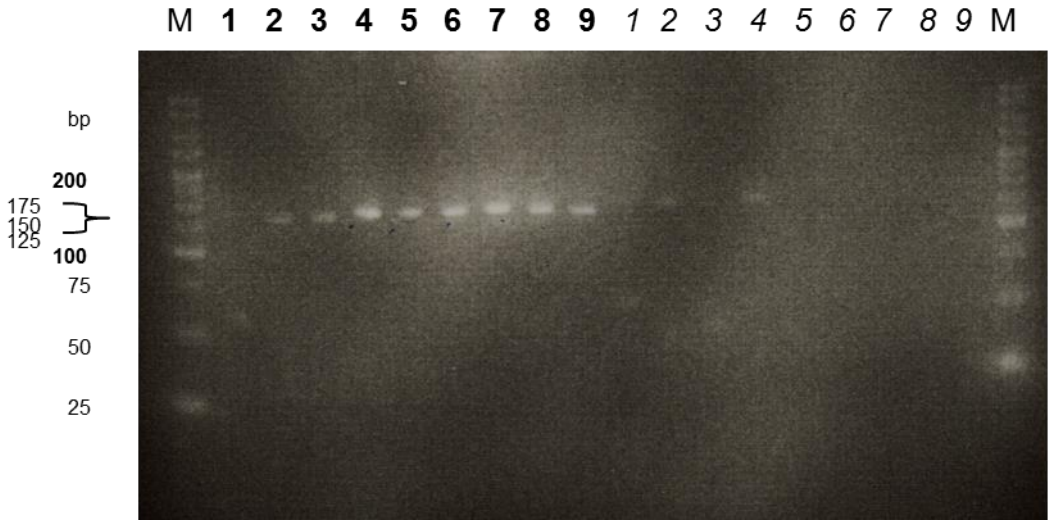


Figure 5.6 HRM primer testing.
Primers were paired as follows; 1 gcT335/gcB450, 2 gcT335/gcB-RTC, 3 gcT335/gcB-RTD, 4 gcT-RTA/gcB450, 5 gcT-RTA/gcB-RTC, 6 gcT-RTA/gcB-RTD, 7 gcT-RTB/gcB-450, 8 gcT-RTB/gcB-RTC, 9 gcT-RTB/gcB-RTD. Lanes in bold type are with gDNA and lanes in italics are no template controls. 1 has primer-dimer, 2, 3 and 4 have background in controls. 5-9 are suitable for further use.
PCR conditions were as described in text. Products were separated on 2% metaphor agarose gel.

HRM. From these two results, it was impossible to determine the composition of GRE in the cDNA. To further test the method, cDNAs were made from RNA whose GRE composition had been studied previously by cloning and sequencing. NR32 and NR26 were chosen as they each expressed a single predominant GRE. The results of HRM on these cDNAs are shown in Figure 5.8. It is clear that the melt temperature of 82.5^{o}C for NR26 did not match that of 82.2 ^{o}C of NR26-1, which 93% of the cloned sequences contained. For NR32 the Melt temperature was similar, but the NR32-1 DNA has a shoulder on the curve which was not present in the cDNA with matching melt temperature. The duplicate cDNA does have a similar pattern to the curve, but the melt temperature is lower, at 80.7 ^{o}C instead of 80.87 ^{o}C. Other DNAs also had similar melt temperatures.

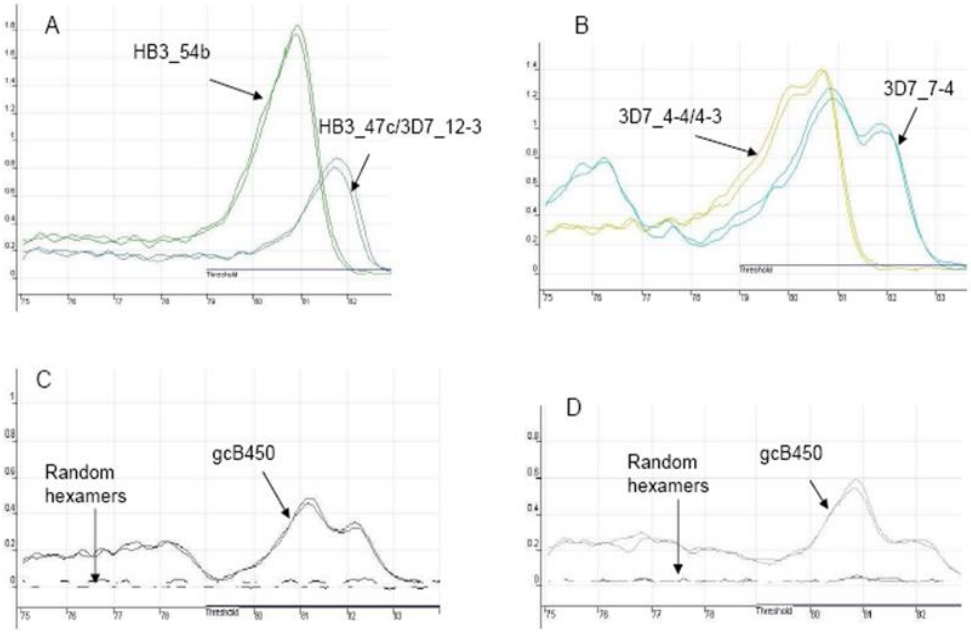


Figure 5.7. HRM results
Melt data from HRM experiments. A, GRE plasmid DNA amplicons showing single peaks. B, GRE plasmid DNA amplicons showing multiple peaks. C (X4cDNA) and D (3D7cDNA) comparing cDNA from random hexamers and GRE specific primers.

The HRM software (172) has the facility to genotype samples, by comparison to reference samples in the experiment. The software was unable to assign genotypes to the cDNAs studied. As single copy sequences in plasmid DNA did not all give discrete melt peaks and HRM could not assign genotypes to cDNA, it was necessary to investigate alternative techniques.

Mis-match PCR

As with HRM, mis-match PCR was tested on plasmid DNA templates shown in table 5.1. The results of a pilot experiment using four templates and corresponding templates is shown in Figure 5.9. Although not perfect, it was promising enough to move to a complete set. Each diagnostic

primer was tested with the relevant gcTcontrol or gcBcontrol primer. These were tested with 20 µM and 2 µM dNTPs. The results, shown in Figure 5.10, confirm that each primer pair works on 3D7gDNA, but only at 20 µM dNTPs. When the positive control primers were added to the reactions, 9/11 reactions gave only the control product, as shown in figure 5.10b.

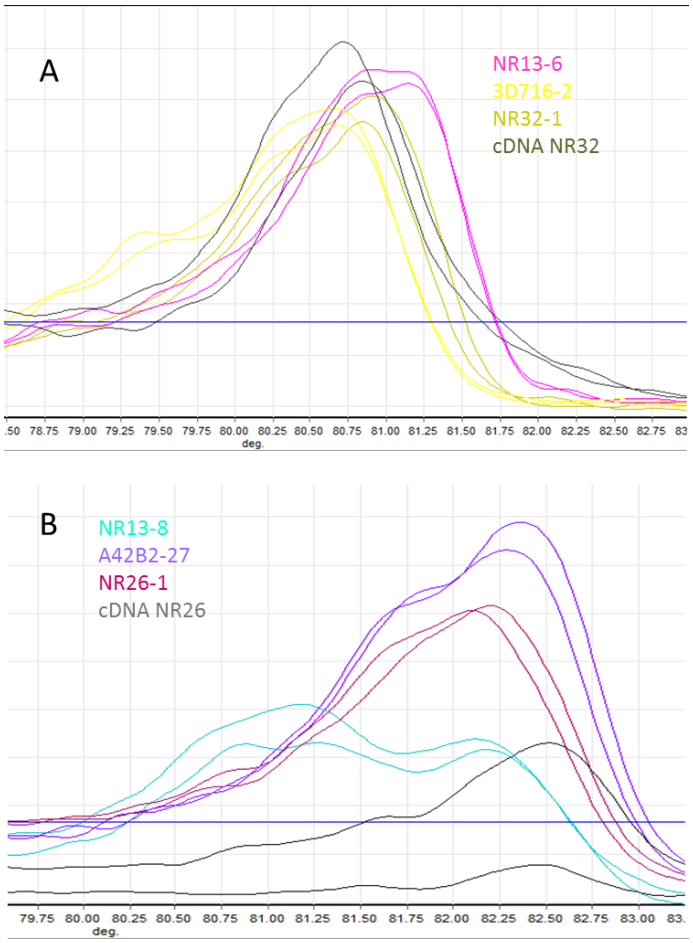


Figure 5.8. HRM testing of cDNAs with known GRE reportoire.
A. cDNA NR32. 10/13 clones sequenced previously had NR32-1 sequence. HRM is unable to distinguish between NR13-6, 3D716-2 and NR32-1.
B. cDNA NR26. 15/16 clones sequenced previously had NR26-1 sequence. HRM is unable to assign this as NR26-1, it more closely resembles A42B2-27.

Doubling the concentration of the gcB45 or gcT335 primers, to give enough to work with both test and control primers, did not work. Increasing the dNTP concentration to ensure that it was not limiting the reaction was also unsuccessful (in both cases, data not shown).

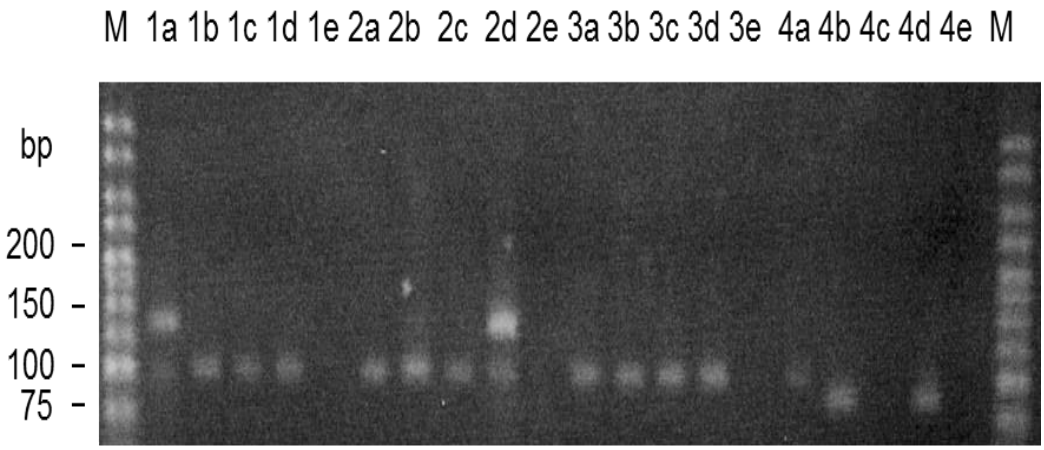


Figure 5.9 Mis-match PCR Pilot study.
The 97 bp product is the positive control and should be in all lanes a-d. Templates are plasmid DNAs containing the following GRE sequences: a = 3D7-12_3; b = 3D7-12_2; c = 3D7-4_6/7_3; d = 3D7-7_4; e = no template control. All reactions contain primers gcB450 and control T. GRE specific primers are as follows; 1 = 12-3T (135bp), 2 = 12-2T (133bp), 3 = 4-6T (133bp), 4 = 7-4T (85bp).
The reaction in 1a-e show the expected result. Only the DNA containing the GRE sequence is detected. Reaction set 4 have the expected test result, as the 7-4T primer should detect 7_4 and 12_2. The positive control reaction has failed in 4b-d. Primer 4-6T has not worked in set 3. Closer inspection of the 12-2T sequence indicated that they should detect 12-2 and 7-4, they have detected 7-4.

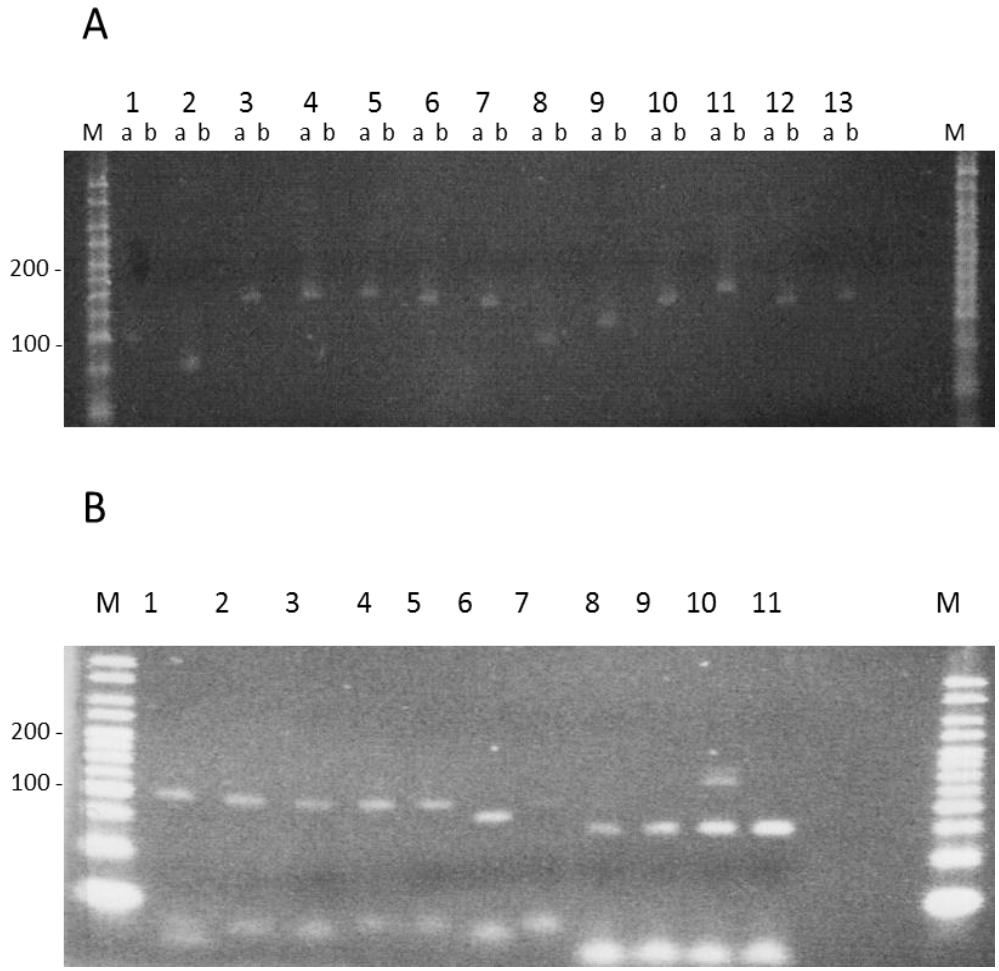


Figure 5.10 Mis-match PCR
A. Testing of each primer pair on gDNA, to ensure they work and give the anticipated size product. a = 20 µM dNTPs, b = 2 µM dNTPs. The reactions only work with 20 µM dNTPs. 1 gcTcontrol/gcB450, 97bp. 2 gcBcontrol/gcT335, 67bp. 3 gc12-3F/gcB450, 135bp. 4 gc12-2T/gcB450,133bp. 5 gc4-6/7-3T/gcB450, 133bp. 6 gc4-2/4-3T/gcB450, 124bp. 7 gc4-4/4-5T/gcB450, 119bp. 8 gc7-4/12-2T/gcB450, 85bp. 9 gc4-1/7-1B/gcT335, 100bp. 10 gc8-1B/gcT335, 120bp. 11 gc7-2B/gcT335, 135bp. 12 gc8-2B/gcT335, 116bp. 13 gc12-1T/gcB450, 118bp.
B. Testing of each pair and control primer in the same reaction. 1-7 have both gcTcontrol and gcB450 primers plus the following specific primers: 1 gc12-3T, 2 gc12-2T, 3 gc4-6/7-3T, 4 gc4-2/4-3T, 5 gc4-4/4-5T, 6 gc7-4/12-2T, 7 gc12-1T. 8-11 have both gcBcontrol and gcT335 primers plus the following specific primers: 8 gc4-1/7-1/7-2B, 9 gc8-1B, 10 gc7-3B, 11 gc8-2B. Reactions 1-5, 7, 8, 9 & 11 only have a product generated from the control reaction. 6 has the test product only. 10 has both the positive control and test product.

GRE expression in 3D7xHB3 progeny

As neither HRM nor mismatch PCR gave interpretable results, the following analyses were done by the original cloning and sequencing method. Sequence data for GRE expression was obtained from cDNA for five cross progeny as described in chapter 2. The sequences derived from 13-16 colonies from each cross progeny are shown in Table 5.2. These data show that, disregarding obvious deletions, there are 18 novel sequences found in the progeny of the cross. The novel sequences are shown in figure 5.11. Fifteen of these have nucleotide changes that match polymorphic sites in the genomic sequences, but X4-L, XP2-T and XP8-L have SNPs at novel sites. It is possible that these novel sequences could be a result of recombination, but it is also possible that they could be due to polymerase error. The significance of the GRE expression data will be discussed in relation to *var* gene expression data in chapter 6.

Comparison of Polymerases with and without Proof-reading

To investigate if the new GRE sequences were due to polymerase error, the PCR was performed using either Kapa HiFi Polymerase (HiFi) or AmpliTaq. The HiFi enzyme has proof-reading, whereas AmpliTaq does not. To rule out novel sequence generation by meiotic recombination, parental HB3 and 3D7 gDNA was used as template. The results were compared to the sequences in the PlasmoDB and Broad databases (73,156). A pile up of the new sequences that did not match the databases is shown in figure 5.12. For 3D7, all of the sequences obtained with HiFi were in the genome, but 5 novel sequences were found with AmpliTaq. With HB3, one new sequence was detected with HiFi, and three others

3D7xHB3 cross progeny (*var* expression)					
	XP2 (mixed)	X33 (mixed)	X10 (*var2csa*)	X4 (1459)	XP8 (1074-3)
GRE Locus					
HB3_47c 3D7 12_3			4 (0.06)	2 (0.29)	1 (0.27)
HB3_54b				1 (0.38)	2 (0.19)
HB3_100a & b	1 (0.29)				7 (0.0016)
3D7 4_2/4_3		2 (0.29)	4 (0.06)	1 (0.27)	
3D7 12_1	1 (0.38)				1 (0.38)
3D7 8_1					
3D7 12_2	1 (0.38)				2 (0.19)
3D7 7_4	1 (0.38)			1 (0.38)	1 (0.38)
HB3_183 3D7 4_1/7_1/7_2	3 (0.23)	1 (0.08)		4 (0.23)	
3D7 4_4/4_5			2 (0.29)		2 (0.29)
HB3_54a 3D7 8_2	3 (0.18)			2 (0.29)	
Novel* singles	2	7	3	2	
Novel* double within clone		2 1x2			
Novel* single (but present in 2 progeny)	1		1		
Deletions (42-96bp)	2	2		3	
N=	15	14	14	16	16

Table 5.2 GRE sequences from cDNA of the progeny of the 3D7/HB3 cross. Numbers in parentheses are the probability of the number of bacterial clones containing this GRE out of the total number sequenced for each HB3 clone being significant and were calculated by binomial probability (204). * Novel sequences, not present in 3D7 or HB3 databases. Deletions appeared in some sequences, but may be bacterial cloning artefacts.

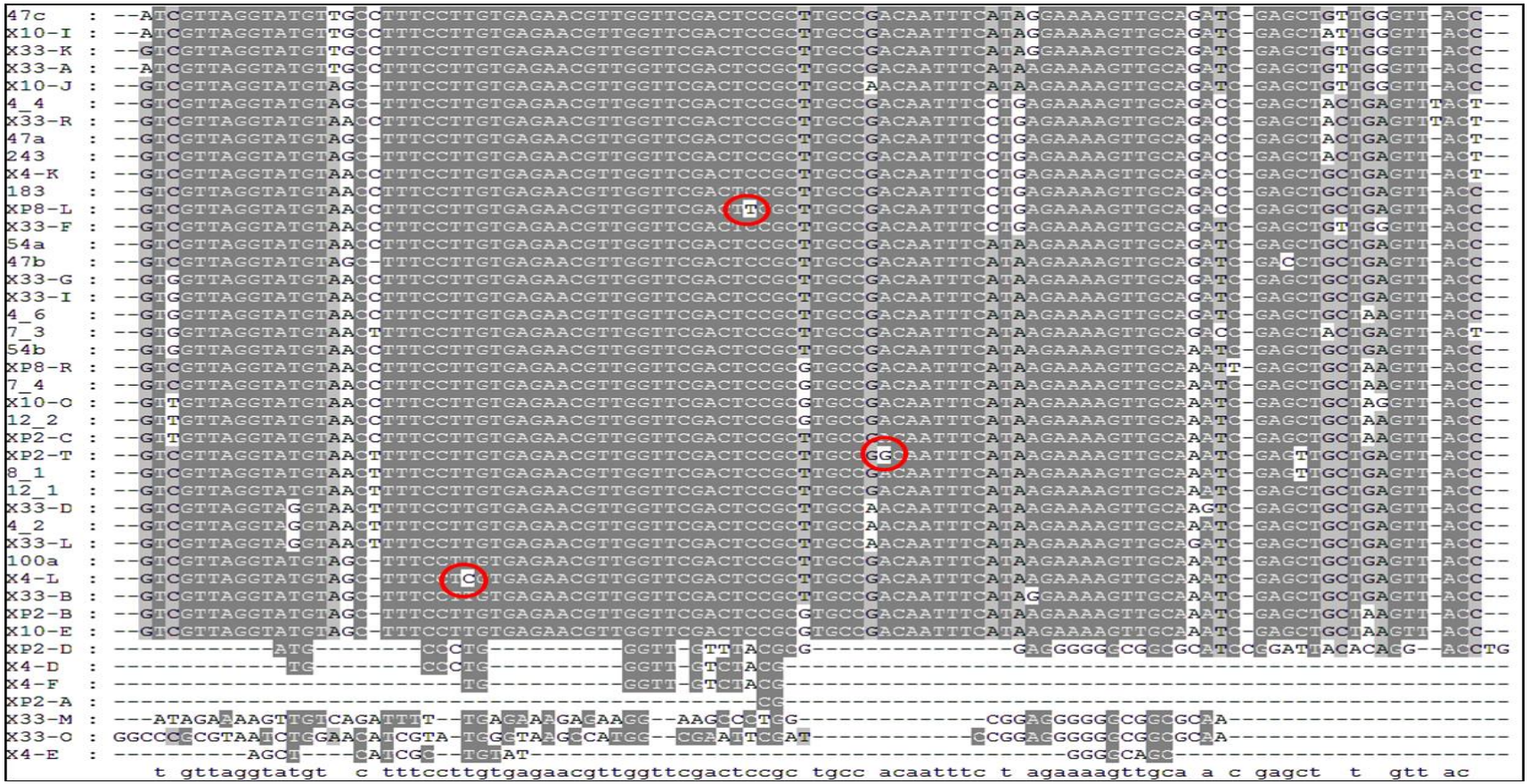


Figure 5.11 Novel GRE sequences in Cross Progeny. Alignment of novel GRE sequences found in progeny of 3D7xHB3 cross. Genomic sequences are included for comparison. The degenerate oligo sequences and all multiple sequences have been removed. Novel base changes are circled in red.

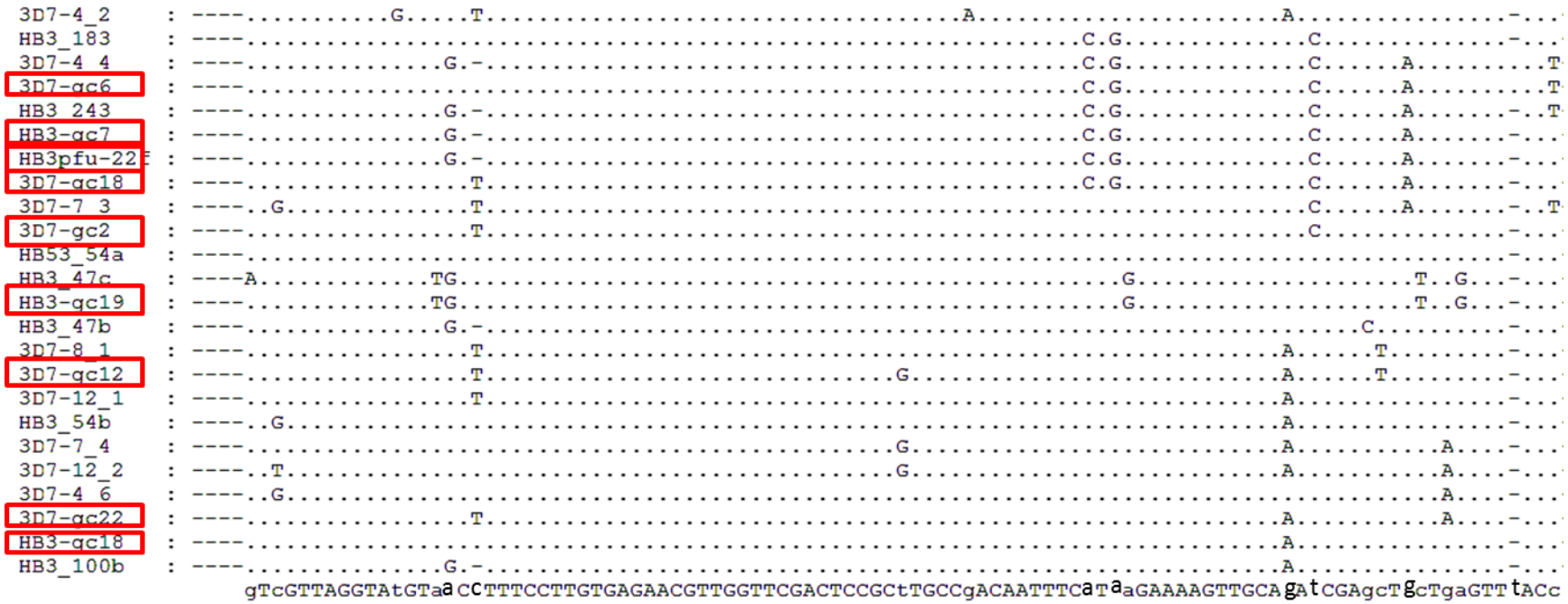


Figure 5.12 Novel GRE sequences generated from genomic DNA
Alignment of novel GRE sequences generated by both amplitaq and PFU PCR on HB3 and 3D7 gDNA. Novel sequences in red boxes. GRE sequences from Plasmodb and Broad databases are included. Multiples of sequences and degenarate oligo sequences have been removed.

with AmpliTaq. As the HB3 genome is not completely assembled, it is possible that the new sequence detected with HiFi may be a valid result. To further test the possibility of Amplitaq error, the PFU PCR reaction was repeated on NR13 gDNA. NR13 is a clone from NF54, so has the same parent as 3D7, therefore any results obtained can be compared to the 3D7 sequence. NR13, a clone selected from a rosetting phenotype was tested soon after cloning (NR13-0) and after being cultured for 100 cycles without selection (NR13-F). With the exception of new sequences that may potentially occur as a result of mitotic recombination, the genomes should be identical. The results, comparing NR13 sequences to 3D7 genomic sequences, are shown in Table 5.3. Although they share the same parent as 3D7, there are three novel sequences. A pileup of the novel sequences and those from 3D7 are shown in Figure 5.13. The differences in the new sequences are all at known polymorphic sites.

For the mutations to be due to polymerase error, it would be expected that any new sites would be randomly distributed within the GRE sequence. A further possible explanation could be PCR chimeras, which was tested using the rationale described in Figure 5.4. The results in Figure 5.14 show that NR13F-18 and NR13F-23 are PCR artefacts. Despite repeated attempts the origin of NR13-0-23 could not be resolved. The significance of the sequences in studies of mitotic recombination will be discussed in chapter 7.

Discussion

The use of recently described techniques to study the GRE repertoire of *P. falciparum* was unsuccessful. To test HRM a panel of

plasmid DNAs, each containing a single GRE was developed. To test all of the GRE sequences within one amplicon, the primer pair gcT-RTB/gcB-RTC was chosen. Kristensen in 2008 demonstrated that primers should not overlay a SNP as this may lead to differential amplification, before the HRM step (198). Primer gcT-RTB contains three degenerate bases and gcB-RTC contains two. Sequencing of RT-PCR products has shown that GREs are frequently multiply expressed. A mixture of cDNAs differentially amplified with 5 different potential polymorphisms in the primers could lead to a broad range of templates for HRM. To cover the entire repertoire of GRE with one amplicon cannot be done without using degenerate primers. One possibility would be to use two amplicons covering either end of the GRE. One of each pair of primers could be designed in the conserved central region, but this would still require that the second primer be degenerate. In the same paper, Kristensen and colleagues stated that only one SNP should be covered by an amplicon. More recently, in a study of SNP detection in lactose persistence, Janukonyte and colleagues were able to differentiate samples with up to 4 potential SNP differences in an amplicon (205). This genotyping was done using gDNA alleviating any potential variables from cDNA synthesis. The GRE amplicon has 24 potential polymorphic sites within it. As HRM is a relatively new technique, it is possible that improvements will be made to allow for more SNP detection.

GRE Locus	NR13-0	NR13-F
3D7 4-2/4-3	6	3
3D7 12-2	2	0
3D7 8-2	1	2
3D7 4-4/4-5	5	2
3D7 12-1	2	1
3D7 8-1	1	2
3D7 7-4	0	2
3D7 4-6	0	2
3D7 12-3	2	3
3D7 4-1/7-2/7-1	1	4
3D7 7-3	3	0
Novel sequence NR13-0-23	1	0
Novel sequence NR13-F-18	0	1
Novel sequence NR13-F-23	0	1
N=	24	23

Table 5.3 Comparison of gDNA GRE sequences by PFU PCR
Number of bacterial clones containing the GRE sequence, after cloning and sequencing PFU PCR products. NR13-0 was selected as a rosetting phenotype at cycle 0. NR13-F is clone NR13-0 cultured without selection for 100 cycles

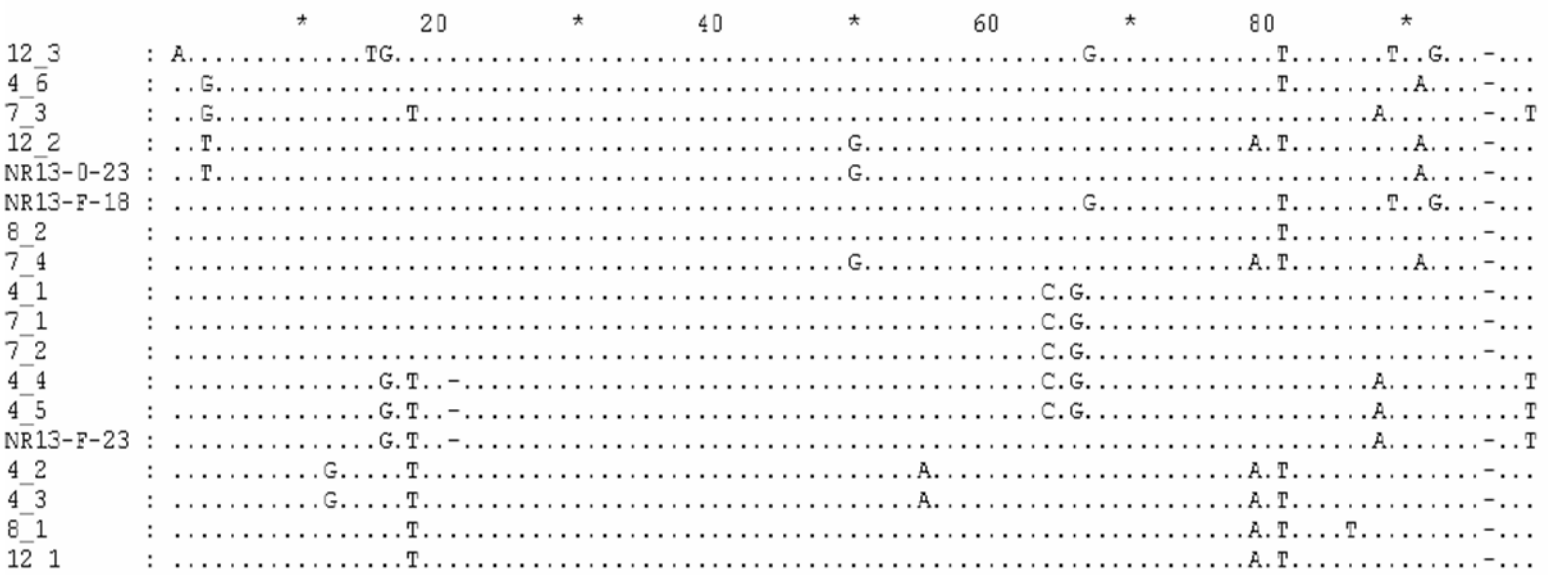


Figure 5.13 Alignment of novel sequences from NR13-0, NR13-F and 3D7 genomic database.

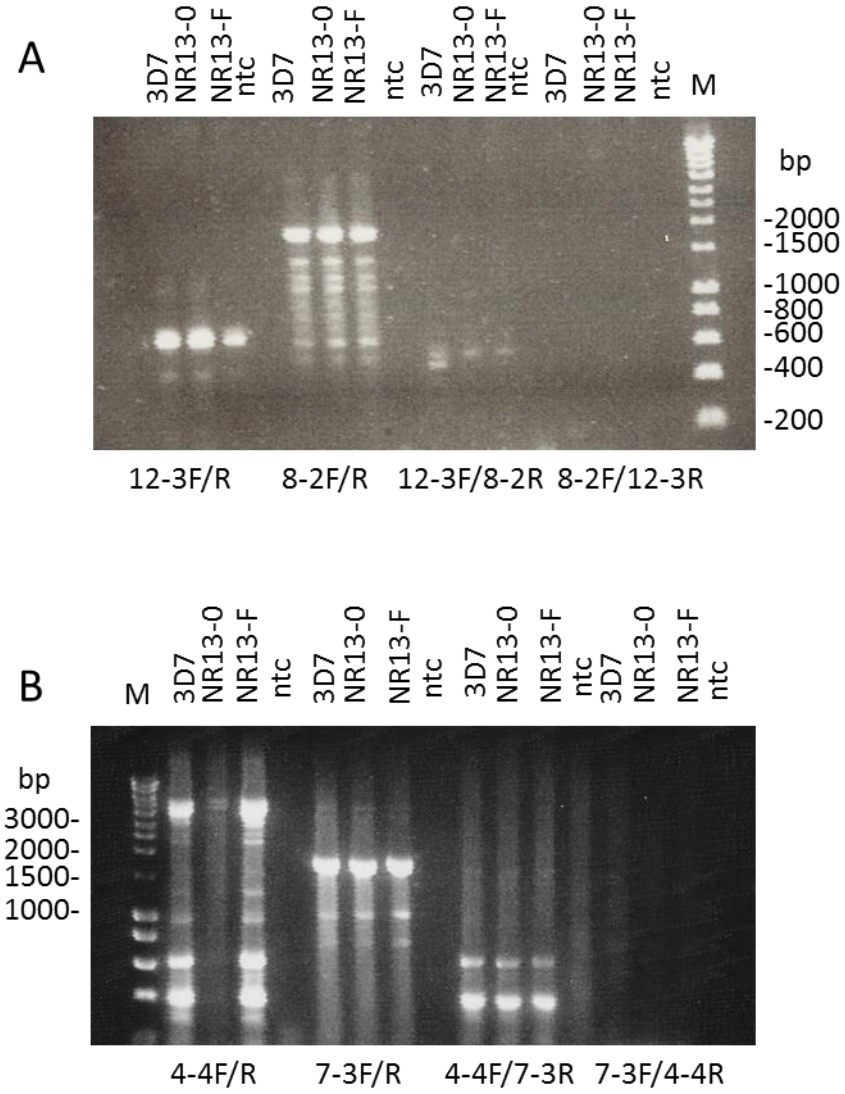


Figure 5.14 NR13-F-18 and NR13-F-23 are PCR chimeras.
PCR testing of gDNA with PCR primers for specific GRE. If the novel sequences detected in NR13-F-18 and NR13-F-23 are mitotic recombination events they would be detected in the NR13-F genome, and not in NR13-0 or 3D7.
A. NR13-F-18. The predicted product sizes are as follows: 12-3F/R 709 bp; 8-2F/R 2086 bp; 12-3F/8-2R 881 bp; 8-2F/12-3R 1619 bp. The products found with 12-3f/8-2R are too small and most likely PCR artefacts.
B. NR13-F-23. The predicted product sizes are as follows: 4-4F/R 3.6 Kb; 7-3F/R 1.7 Kb; 4-4F/7-3R 3.8 Kb; 7-3F/4-4R 1.2 Kb. The products found with 4-4F/7-3R are too small and most likely PCR artefacts.

The limit of detection for HRM is 0.2^{o}C (172). Although there were greater differences between the DNAs shown in Table 5.1, some had very similar melt temperatures. The differences between duplicate samples were greater than 0.2 ^{o}C in many cases. Many of the curves had shoulders on them, thought to be due the relaxation of secondary structure during the melt stage. The high GC content, length of the amplicon and the number of polymorphisms in combination with the number of cDNAs present in a sample made it difficult to assign a sample to a known template. The RotorGene software was unable to do this even when cDNA was used from a sample known to be expressing 93% of a single GRE (NR26-1) (Figure 5.5). To continue with HRM would require dissecting the GRE into smaller amplicons and testing these separately. This would lead to multiple reactions being necessary for each sample. It is still likely that with multiple GRE sequences being expressed in many samples that analysis of the products would not be conclusive.

The second method tested for GRE detection was mis-match PCR. This technique requires many different successful PCR reactions per sample but would lend itself to being automated. The basic PCR reactions worked. However, in the presence of positive control primers, designed to confirm the PCR reactions worked, most reactions gave only the control product. As lack of test product in the test reactions is part of the diagnostics, the positive control is necessary. With the knowledge that many cDNAs would contain multiple sequences it was felt that the results would be difficult or impossible to interpret. Although more testing was done to improve the conditions, it was decided due to time constraints to return to cloning and sequencing.

An initial comparison of the HB3 and 3D7 genomic GRE sequences revealed three HB3 sequences that altered the RNAalifold hairpin prediction (figure 5.1). These sequences were not found in the screen of HB3gDNA using a proof reading enzyme. Initially the concern was whether these sequences were genuine however, this is most likely due to not screening enough colonies. Data to be presented in chapter 6, studying GRE expression by proofreading RT-PCR, shows that the sequences are present in at least five different HB3 clones. We conclude from this that the HB3-100a, 100b, and 47b sequences are genuine. Eighteen novel GRE sequences were detected in cDNA from progeny of the 3D7xHB3 cross. Interestingly, three of these had base changes at previously non-polymorphic sites. These sequences were derived from AmpliTaq RT-PCR reactions. To address the possibility that these novel bases came from AmpliTaq error, gDNA from both HB3 and 3D7 was tested with both AmpliTaq and a proofreading enzyme. The KAPA HiFi DNA polymerase chosen as an enzyme with proofreading is sold as having an error rate 100 times lower than regular Taq DNA polymerase (206). For 3D7, all of the HiFi sequences were present in the genomic sequence. For HB3 one HiFi sequence was not present in the Broad database. As this is not fully assembled, it is possible that this is a valid result. With AmpliTaq, five novel sequences were found in 3D7 and 3 in HB3. None of these sequences generated new SNPs. If the novel sequences were purely due to AmpliTaq error, we would expect any nucleotide changes to be random throughout the sequence.

Having established that proofreading DNA polymerase is necessary for GRE sequence analysis, it was decided to further test this. Two

different gDNA samples from the same clone, separated by 100 generations were examined. After 100 generations two novel sequences were found. Further testing of these showed them to be PCR chimeras. Generation of new sequences during PCR that cannot then be detected in the genome was reported previously in *P. falciparum* (194). This study, using generic primers to study *var* gene DBLα, used AmpliTaq. Although proofreading enzymes may help reduce the likelihood of novel sequences due to polymerase error, they clearly do not stop the production of chimeric sequences. DNA recombination frequencies during PCR have been reported as 1-5% (207). Such recombination is thought to be due to incomplete extension from one primer, then annealing of that product to other templates, and extension in following cycles (208). This incomplete extension could be due to premature termination or pausing of Taq on the template (207). As the GRE RNA is assumed to adopt the hairpin secondary structure, it could be imagined that during the annealing stage of the PCR reaction, the single stranded DNA could begin to adopt a secondary structure too. This could then lead to the stalling or premature termination of Taq. However, a study of PCR recombination in HIV1 *tat* gene sequences reported that the PCR recombination regions did not coincide with secondary structures (207).

Although not tested, it would seem likely that many of the new sequences found using AmpliTaq on HB3 and 3D7gDNA are due to PCR chimeras. As 15/18 of the novel GRE in the cross progeny also did not generate new SNPs, some of these could be due either to PCR or meiotic recombination. This potential recombination rate of 18% in the cross progeny is much higher than previously reported for PCR recombination

(207). It is more likely therefore, that as these were generated with AmpliTaq, they are due to a combination of recombination and AmpliTaq error. GRE sequences generated using proofreading enzymes have a recombination/error rate of 9%, significantly less than the 18% for AmpliTaq.

In conclusion, for reasons most likely due to the highly polymorphic nature of the GRE sequence, and the lack of a single expressed transcript, HRM was unsuccessful. Similarly, multiple sequences in cDNA, meant that the results of mis-match PCR would be difficult to interpret. Cloning and sequencing of PCR products was the most reliable method tested. However it is essential that the enzyme chose for the initial PCR has proofreading abilities. The findings discussed herein can now be applied to GRE analysis in relation to *var* gene expression in chapter 6 and with respect to recombination in chapter 7.

Chapter 6

The potential role of GRE in *var* gene regulation

Introduction

The primary aim of this thesis is to investigate the potential role that the GRE may play in *var* gene regulation. As described in chapter 1, the GRE are only located within, or adjacent to the central *var* gene clusters on chromosomes 4, 7, 8 and 12. As they are only found at central clusters, it was postulated that they may have a role in central *var* gene regulation. At the time of their initial discovery in 2002 (58), little was understood about the regulation of *var* gene expression. In the past decade however, our knowledge, particularly of epigenetic factors, has increased considerably (reviewed in Chapter 1). This increase can be attributed to several things including the genome sequencing projects, advances in bioinformatics, gene expression analysis by qPCR, and technical advances in transfection studies in *P. falciparum*.

ncRNA and *var* genes

In one of the original description of *var* genes, Su and colleagues identified 1.8-2.4 Kb “sterile” transcripts that hybridized to the exon 2 region (20). This was the first description of ncRNA with relation to *var* genes. Kyes and colleagues showed that these sterile transcripts are only detectable between 24-48 hours post invasion (190). In 2009, Epp and co-

workers demonstrated that these transcripts derived from a bi-directional promoter in the *var* gene intron (64). These RNAs were found to be non-polyadenylated and remained within the nucleus, possibly associated with the *var* gene clusters. A noncoding antisense RNA of exon 1, originating from the intron promoter, was also described in the same publication.

Recently, a family of 22 long non-coding RNAs (lncRNAs) was described (103). Transcribed at 40 hours post invasion from TARE 2-3 of the Telomere Associated Repeat Region, they are called lncRNA-TARE. The lncRNA-TAREs contain the SPE2 binding sites for the ApiAP2 transcription factor PfSIP2. SPE2 binding sites are also found in the promoters of UpsB *var* genes, and intriguingly, each lncRNA-TARE was located adjacent to an UPsB *var*. The authors propose 4 models for the action of lncRNA-TARE, including one that involves lncRNA-TARE recruiting chromatin remodelling factors to provide silencing memory marks at UpsB *var* genes.

var class specific factors

That lncRNA-TARE has an effect on UpsB *var* genes is only a model at present. This study however follows on well from a previous study looking at PfSIP2. Using concatemerised SPE2 elements to isolate SPE2 binding proteins, Fluek et al identified PfSIP2 (102). Genome wide ChIP analysis revealed that PfSIP2 binding only occurs at TARE 2-3, and a region with a cluster of SPE2 elements upstream of UpsB *var* genes. No binding was found at chromosome-internal sites. Taken together, the data presented by Broadbent and Fluek suggest that PfSIP2 has a role in regulation of UpsB *var* genes, but not centrally located ones.

Further evidence for differential regulation of *var* gene groups was demonstrated by Tonkin and colleagues looking at Sir2 homologues (92). Using 3D7 parasites containing either a disrupted PfSIR2A or PfSIR2B gene, they performed global transcription analysis. Looking specifically at *var* genes, they found a correlation between the disrupted PfSIR2 gene, and the direction of transcription and location of the activated gene. UpsA, UpsE and UpsC genes were activated upon disruption of PfSIR2A, and UpsB genes by disruption of PfSIR2B. They concluded that PfSIR2B has a role in silencing UpsB *var* genes, and PfSIR2A in silencing UpsA, UpsC and UpsE. Although the modes of action are as yet unclear, it is apparent that different factors act on different classes of *var* genes. One of these factors has been shown to be a ncRNA. In this chapter I will investigate the hypothesis that the GRE are a new group of Ups class specific factors.

Patterns of *var* gene switching

Whilst the mechanisms underlying *var* gene switching and expression are still unclear, patterns of switching are becoming clearer. Studies of switching patterns involve two main elements, switch rates and gene to gene switching probabilities. These experimental studies also combine with mathematics to produce models of gene switching.

Switch rates of surface expressed antigen genes were first described in 1992, by Roberts et al (46). Using a clonal line A4, derived from the IT isolate, they subcloned the A4 parasites. Using agglutination assays to determine that a different surface molecule had been expressed, they determined that new antigenic types appeared at 2.4% per generation. A further publication in 1995 correlated these results with

expression of individual members of the newly described *var* gene family (19). The same A4 clone was also used to measure both on and off rates of *var* genes by Horrocks and colleagues (187). Using new clonal populations derived from A4, they determined that individual *var* genes have different switch rates.

More recently, since the classification of *var* genes by their upstream sequence (66), switch rates can be correlated to the class of *var* gene involved. Using two patient derived clones, Peters and colleagues looked at levels of expression from different classes of *var* genes (209). They found that over 6-10 days following culturing from a patient sample, Ups classes A, D and E had a larger reduction in levels of transcription than UpsB or UpsC. Frank and co-workers followed expression in NF54 derived clones for up to 28 weeks (210). They found that all of the clones that were expressing a sub-telomeric *var* gene at t=0 had switched expression by 16 weeks. In contrast, those clones that were originally expressing a central *var* gene were expressing the same gene at 16 weeks. They determined off rates for central genes to be 0-0.3% per generation. To determine off rates for sub-telomeric genes they used transgenic parasites whose expression could be controlled by drug selection pressure. The off rates were determined as 2.5-2.7%. Data produced for a model of switching (211) also followed several parasite clones for 60 generations. Two of their IT derived clones stably expressed central *var* genes, one an UpsB/C, the other an UpsC gene.

The studies above all used established laboratory isolates. Two publications in 2011 looked at clinical isolates whilst they were being

adapted as laboratory cultures. Zhang and colleagues observed a culture stably expressing an UpsB/C for 60 days (212). Bachmann followed two *ex-vivo* isolates that were expressing UpsA type genes. After 12 generations, they were both expressing genes from all three classes (213). With all of the studies, once in *in vitro* culture the parasites are not subject to immune pressures, therefore experimental results may not necessarily replicate the outcome of switching events in a clinical infection.

Transfection in *P. falciparum*

As detailed above, the study by Frank and colleagues utilized transgenic parasite lines to study switch rates (210). Many of the recent advances in our knowledge of *var* gene regulation can be attributed to improvements in transfection techniques. There were two major obstacles to the development of the technique. Firstly, the instability of A-T rich DNA when cloned into *E. coli* was, and is still, a problem in generating the plasmid constructs used to transfect. Secondly, in order to deliver externally produced DNA to the parasite nucleus, it must cross four membranes; the red blood cell, parasitophorous vacuole, parasite plasma and nuclear membranes, yet remain intact.

The first successful report of transfection in *Plasmodium* species was in *P. gallinaceum*, an avian parasite (214). To circumvent the issue of crossing four membranes, the authors used the extracellular female gamete and fertilized zygote stages. They then cloned the firefly luciferase gene on to a plasmid containing the pgs28 gene, which was known to be expressed in gamete and zygote stages. Following transfection, luciferase activity was detected after 24 hours. This experiment showed that DNA

could be stably cloned in *E. coli*, and started to define the control elements that were required in a construct. In 1995, Wu and colleagues successfully transfected *P. falciparum (215)*. This work further refined the requirements for 5' and 3' flanking control elements, using a series of constructs containing a chloramphenicol acetyltransferase (CAT) gene. More significantly, the constructs were successfully electroporated into intraerythrocytic ring and schizont parasites.

In the early studies using transfected parasites, only transient transfections were studied by following a reporter gene, such as CAT or luciferase, for a short time. In 1996 two groups reported stable transfection with transfected DNA integrated into *P. falciparum* chromosomes (216,217). Unlike transient constructs, plasmids for stable transfections require a selectable drug resistance gene. Application of the drug pressure maintains the DNA within the cell. Both groups used dihydrofolate reductase-thymidylate synthase (*dhfr-ts*) that confers pyrimethamine resistance. Following 4 or 5 weeks of culturing with added pyrimethamine, the transfected DNA was maintained within the parasites, but episomally. Although the groups had differing drug regimens, they both used a period of growth without pyrimethamine followed by culturing back in the presence of drug. After this, both groups reported integration of the plasmid DNA into *P. falciparum* chromosomal DNA. Wu and colleagues found integration into three separate locations, each corresponding to the chromosomal locus of an element in the construct (216). Production of stably integrated transfected *P. falciparum* is a lengthy process, taking approximately 3 months.

In early stable transfectants, the integration event was a single crossover. This inevitably meant the plasmid backbone was also incorporated into the chromosome. In 2002, the first report of a double crossover integration was published (218). Using a construct with two selectable markers, one of which was outside the target gene, the authors applied a negative selection protocol. This selected against cells containing the plasmid episomally. The ability to replace or disrupt target genes opened up the field to knock-out studies.

There have been many developments in transfection technology since the initial reports. New selectable markers have been developed. These include human DHFR, selected for by WR99210 (219) and blasticidin S deaminase, conferring resistance to Blasticidin (220). A new DNA delivery method was described by Deitsch et al (221). Plasmid DNA is electroporated into uninfected erythrocytes, and then parasites are cultured in these cells which spontaneously take up the plasmid DNA. This is less traumatic to the parasites as they are not electroporated. It also enables greater parasite loading, as newly loaded RBC can be added to the culture.

RNA polymerase lll transcripts

We have previously established that the GRE is a RNA polymerase lll (Pol lll) transcript. The distinguishing feature of most Pol lll promoters is the location of motifs within the transcribed region (reviewed in (222)). Consequently, when the transcription complex binds, it covers the gene. There are three different classes of Pol lll promoters, types l, ll and lll. Type l is unique to 5SrRNA, has internal A and C blocks and is bound by

the transcription factor TFIIIA. Type I is the most common; it is found in tRNAs and other species, including GRE. The sequence motifs are the A and B box (figure 1.10) and these are bound by TFIIIC (222). Of the six subunits of TFIIIC, some are conserved between yeast and humans, whereas there is only limited homology between the yeast subunit TFC7 and its human counterpart TFIIIC35 (223). To date only a potential homologue to TFIIIC5 has been identified in *P. falciparum* (79,156). The type III promoters do not have sequence motifs common to all members, and these elements are not within the transcription sites. Extra TFIIIC sites (ETC) have been identified which bind TFIIIC, but do not recruit Pol III (224). The 21 nucleotide ETC consensus sequence in *S. cerevisiae* is less conserved at the B box than in transcribed genes (224). Only a conserved B box, not an ETC sequence is present in the GRE. The Pol III transcription system and ETC sites have been implicated in other nuclear processes, some of which are described below.

It has been observed that transcription from tRNA genes can inhibit transcription from adjacent Pol II promoters (225,226). Termed tRNA gene-mediated silencing (225), Hull and colleagues demonstrated that it is not dependent on orientation of the tRNA or on a location at a specific distance upstream of the Pol II promoter (227), but does require localisation in the nucleolus (225,228). Whilst *S. cerevisiae* ETC sites are located in the nuclear periphery, they do not localise to either the nucleolus or the telomeric foci (229). However, the Pol III transcription factor TFIIIC mediates this localization to the nuclear periphery.

Another aspect of tRNA involvement in gene regulation is as a boundary element. There are two types of boundary elements that prevent gene regulatory elements from acting on genes other than those they are intended for. Barriers interrupt the spread of heterochromatin, whereas insulators prevent enhancers from activating transcription of adjacent genes (230). tRNA sequences have been shown to act as either (231-233).

In this chapter, I will investigate the role of GRE in relation to *var* gene expression and switching. A longitudinal study of clonal HB3 cultures will be undertaken and the relative level of all *var* and GRE expression examined in each. Several techniques previously used in this thesis will be employed. *Var* gene expression will be followed by qPCR and northern blot. GRE expression will be investigated by cloning and sequencing. To gain more conclusive evidence of the effect of GRE, we will attempt to make transfection constructs containing either UpsB or UpsC promoter regions and the GRE. Transfection studies will allow us to determine if the GRE has a specific effect on either class of *var* genes.

Material and Methods

GRE expression

GRE expression was analysed using RT-PCR, cloning and sequencing as described previously.

Var gene expression

The *var* gene expression for 31 HB3 clones and HB3x3D7 cross progeny had been determined in chapter 4. Identical methods were used to follow *var* gene expression in this chapter.

Var gene switching

To follow *var* gene switching longitudinally, eleven of the 31 HB3 clones were chosen along with the original, non-clonal parent culture. The clones were chosen to cover a range of Ups types. Table 6.1 describes the eleven clones and a simplified numbering system for this experiment.

Culturing methods, RNA isolation, qPCR and data analysis have been described previously. The clones were cultured for up to 90 cycles. RNA was prepared as soon as possible after thawing the parasites, this was then referred to as t=0. RNA was then prepared at approximately 10-cycle intervals for qPCR analysis.

To make cDNA, 5 µg RNA was DNase treated in a 20 µl reaction. Following confirmation of removal of gDNA, 6 µl RNA was used as template in a 20 µl cDNA synthesis reaction. cDNA was made in triplicate from each RNA, and 1 µl was removed before storage to be used for GRE analysis. For qPCR, 121 µl of water was added to each cDNA aliquot. Each cDNA sample was then run in duplicate with the HB3 primer set. Two new pairs of qPCR primers, HB3-209b and HB3-1235c were designed and tested to replace the pairs in the original HB3 set. Their details are in appendix i.

Var gene switch rates were determined using the method of Horrocks et al (187). $r_{off} = 1-p^{1/n}$ where p = proportion of population with phenotype at generation n and n = number of generations since cloning.

Clone number	HB3 clone	Var gene expression*	Ups type	location
P		1107	B	telomeric
1	AC06	1499	B	telomeric
2	AA08	1107	B	telomeric
3	BD04	1107	B	telomeric
4	DC01	1235	A	telomeric
5	BF08	1074-1	B	central
6	DH07	1074-1	B	central
7	CC01	1074-2	C	central
8	BH08	1074-3	C	central
9	DF06	699-2	C	central
10	CH05	699-5	C	central
11	BA06	1296-3	C	central

**var* gene expression as determined in chapter 4.

Table 6.1 HB3 clones used in switching experiment

Transfection Constructs

A range of constructs was required comprising either a GRE element, or tRNA as a control and either an UpsB or UpsC *var* promoter region. All constructs also required a luciferase reporter gene and 3' control region from *hrp2* (215). Each of the resulting constructs was to be tested with and

without an intron promoter. All were constructed into pGEM-T (Promega) as the vector backbone, figure 6.1. A schematic of the constructs is shown in figure 6.2. All primer sequences are in Appendix i.

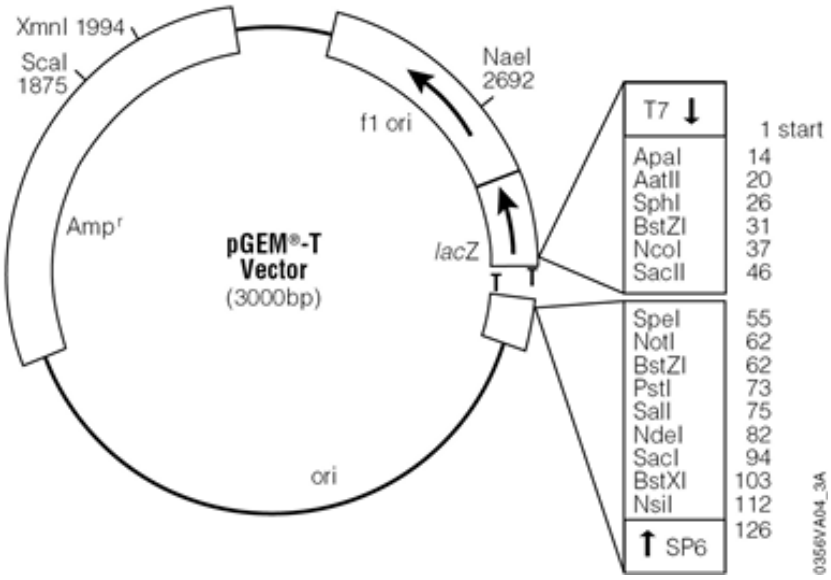


Figure 6.1 pGEM-T vector.
Reproduced with permission from Promega Corporation.

GRE and tRNA elements

A GRE sequence on 3D7 chromosome 4, Pf3D7_04_v3:557,457..557,592 (156) was chosen, also known in this thesis as 4-1 and GC-10 (chapter 8). Whilst the data in table 6.7 show that there is no consensus as to the location of a transcribed GRE with respect to the expressed *var* gene, a distance of 3 Kb upstream was chosen as such for the constructs. The other possible permutations in table 6.7 would make the size of resulting constructs impractical. Primers were designed to a region of chromosome 4, starting 3.4 Kb upstream of the GRE and ending 0.15 Kb downstream.

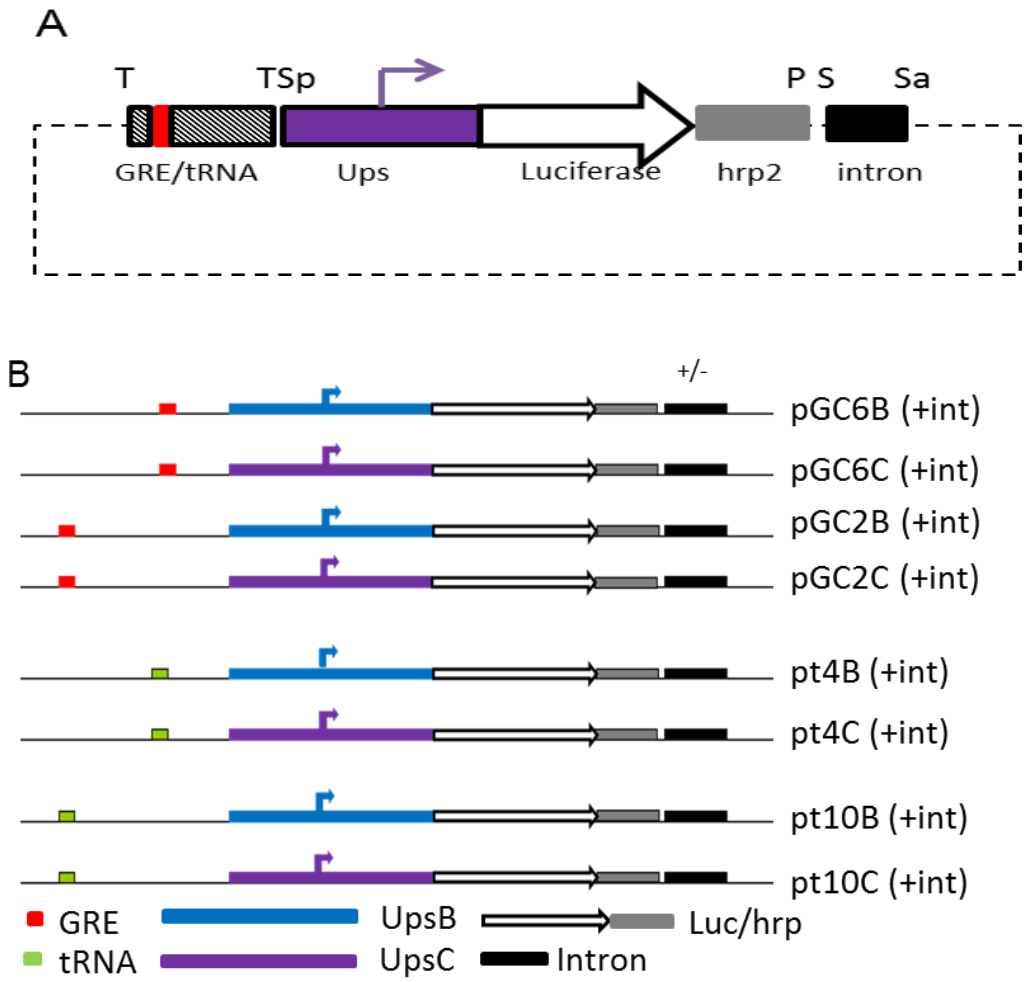


Figure 6.2 Design of constructs for GRE transfection studies. A. Basic cloning strategy showing location of elements within the pGEM-T backbone (dashed line). T. PCR cloning site, Sp. Spel, P. Pstl, S. Sall, Sa. Sacl. B. Full list of required constructs e.g. pGC6B and pGC6B+int are the same construct with +int having an intron sequence cloned into the Sall and Sacl sites.

tRNA was chosen as a control element to compare to GRE containing constructs. Primers were designed to a region surrounding tRNA MAL_5_tRNA_Leu1 (chromosome 5, 451342-451435). To mimic GRE as closely as possible, given intergenic constraints, 2.9 Kb of upstream and 0.6 Kb of downstream sequence was included. KAPA HiFi PCR conditions, using 3D7gDNA as template, were identical for each primer pair. Cycling conditions were 95^{o}C for 5 minutes, followed by 30

cycles of 98^{o}C for 15 seconds, 55^{o}C for 10 seconds and 68^{o}C for 1minute 45 seconds. KAPA HiFi GC buffer was used, and 3.5mM $MgCl_2$.The GC products were checked by DraI digest, and tRNA by HinfI digest prior to cloning. Both PCR products were cloned into the PCR cloning site of pGEM-T and transformed into JM109 cells (Promega) using manufacturer's instructions.

Plasmid DNA was prepared from colonies as in chapter 2. To confirm the orientation of the insert, GRE clones were digested with Nde1 and tRNA clones with EcoRV. Following sequencing, four clones were chosen to accept the other components of the constructs. A representative containing either GRE or tRNA approximately 3 Kb upstream of the 5' cloning site, and ones with the element in the opposite orientation were selected.

Luciferase and hrp2 3' control sequence

The luciferase coding sequence joined to hrp2 3' control sequence was amplified from pHLH-1 (221,234). Restriction digest analysis revealed that the laboratory copy of pHLH differs from the published sequence. A second set of primers, LUC-2, specific for our copy of pHLH were designed and tested. Eventually, because of concerns over differences from the published sequence, it was decided to abandon pHLH and use a commercial luciferase gene. Primers were designed to the luciferase gene from pGL2-Basic (Promega). A colleague generated this PCR product and following gel purification and digestion with HindIII, ligated it to the 3' hrp2 control sequence isolated by restriction digestion with HindIII and PstI from pHLH. This was then used as template to generate a PCR product.

The luciferase/hrp2 3' PCR product was A tailed as described in chapter 2. After resuspending in 5 µl water, 2 µl was ligated into pCR-TOPO-4 (see chapter 2). Plasmid DNA from luc-positive colonies was sequenced.

Ups sequences

PCR primers were designed to the UpsC sequence from HB3-1074-3 (HB3-UpsCF & R), and HB3-1074-1for UpsB (HB3-UpsBcenF & R). The HB3-UpsC reaction used 100ng HB3 gDNA as template in a standard 50 µl KAPA HiFi reaction supplemented with 1.25 mM $MgCl_2$. Cycling conditions were 95^{o}C for 5 minutes followed by 30 cycles of 98^{o}C for 20 seconds, 65^{o}C for 15 seconds and 68^{o}C for 1 minute 45 seconds. The HB3-UpsB reaction used standard 50 µl Phusion PCR reactions supplemented with 1 mM $MgCl_2$ and 100 ng HB3 gDNA as template. Cycling conditions were 98^{o}C for 1 minute followed by 25 cycles of 98^{o}C for 15 seconds, 63^{o}C for 10 seconds and 68^{o}C for 1 minute 15 seconds. A final 10 minute extension at 68^{o}C was included. Both products were A tailed as described previously and 2 µl of each was directly ligated into pCR4-TOPO. Plasmid DNA and sequencing was as described in chapter 2. Following unsuccessful attempts at generating a fusion product (see later section) a new UpsC-F3 primer was designed. This was used in a standard KAPA HiFi reaction with HB3-UpsCR. The only difference from the HB3-UpsCF & R reaction was that the annealing temperature was reduced to 60^{o}C.

Primers to two alternative UpsB sequences were also designed. These were HB3-699-1 and HB3-752-4. The HB3-752-4 primers were used in a standard Phusion reaction at 3.5 mM $MgCl_2$, with the cycles as for HB3-

UpsB except that the 63^{o}C anneal was replaced by 55^{o}C anneal. The HB3-699-1 reactions were unsuccessful. All of the products from Ups PCRs were gel purified.

Intron

The intron fragment was designed to be cloned into the Sall and Sacl sites in the multiple cloning site of pGEM-T. The initial PCR using INTRON F & R was a standard Phusion reaction supplemented with 2 mM $MgCl_2$, using 64^{o}C annealing temperature, 1 minute extension and 3D7gDNA as template. PCR products were checked by restriction digest with Dral. A secondary PCR to add restriction sites used INTF+Sal and INTR+Sac used identical conditions to the primary reaction, but replaced 3D7gDNA with primary PCR product as template. The resulting product was gel purified.

Fusion of Ups and Luciferase/hrp2-3' products

Two methods were attempted to join the Ups PCR products to the Luciferase/Hrp2-3' product.

PCR fusion

The methodology behind PCR fusion is shown in figure 6.3 (235,236). Three primers sets with 16 nucleotide overhangs were designed. The overhangs were added to the Ups and Luc/Hrp2-3' primary PCR products, in reactions which also added the restriction fragments to the opposite ends. When it was realised that the exonuclease activity of both Phusion and HiFi were removing the overhangs, we moved to standard PCR reactions. Both AmpliTaq and BIOTAQ (biolines) were used. The primers

pairs used, and their annealing temperatures are shown in table 6.2, extension time was 2 minutes and 30 seconds. Products were gel purified.

Several different attempts were made to assemble the fusion products. The successful method used approximately equal amounts of Ups and Luc/Hrp2-3' as template in 25 µl primary reactions. Each contained 1x PCR buffer (Bioline), 2.5 mM $MgCl_2$, 200 nM each dNTP, 2.5 units BIOTAQ and each template DNA in the range 130-500 ng. Several reactions were assembled, using a range of template concentrations.

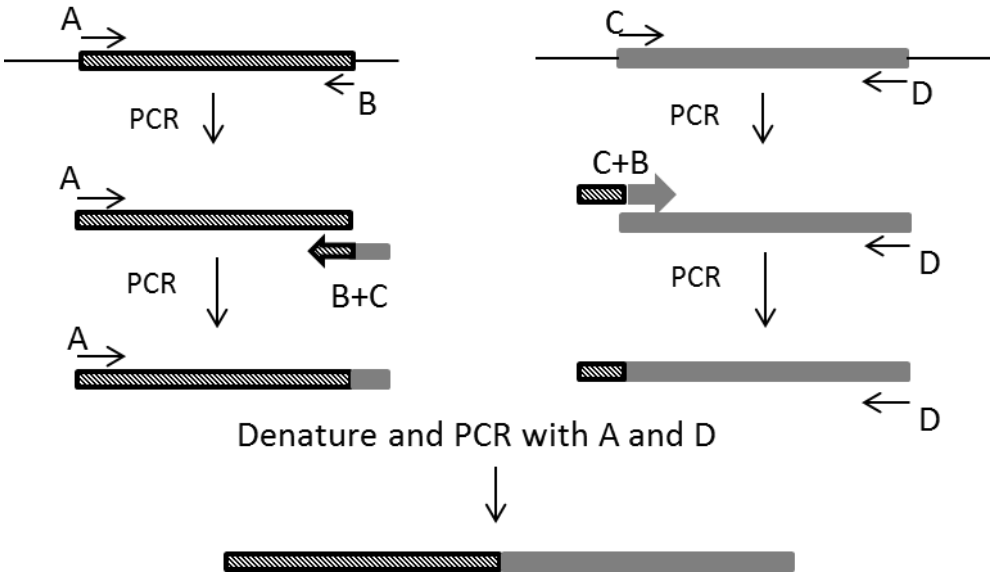


Figure 6.3 PCR fusion technique. To join PCR products A/B and C/D (first line in figure), each product is effectively extended by 16 bp, matching the desired, joined sequence (two separate PCR reactions, second line). In the final PCR reaction using A and D (third line in figure), the denaturing and annealing step in the PCR should produce some templates with the overhanging oligo sequences annealed to the other template. These spliced templates can then act as templates for primers A and D.

F primer	R primer	Final $MgCl_2$ concentration	Anneal temperature
Luc-2+UpsR	Hrpterm R2	2.5 mM	55^{o}C
UpsCF	Ups+Luc2F	3.25 mM	68^{o}C
UpsBF	Ups+Luc2F	2.5 mM	68^{o}C
UpsBF+Spe	pGL2+UpsR	3.5 mM	63^{o}C
UpsC3F+Spe	pGL2+UpsR	2.5 mM	60^{o}C
UpspGL2F	HrptermR2+Pst	2.5 mM	55^{o}C

Table 6.2 Reaction conditions for primer pairs used for PCR fusion.

Cycles were 94^{o}C for 3 minutes followed by 29 cycles of 94^{o}C for 30 seconds, 52^{o}C for 10 seconds, 48^{o}C for 10 seconds and 60^{o}C for 3 minutes. All products of approximately 5 Kb were gel purified, then used as template in 50 µl secondary reactions. The component concentrations were as the primary reactions with the addition of 1 µM of the respective UpsF+Spe and HRP2R+Pst primers. Again several reactions were assembled using 2.5-10 µl of primary template. Cycling conditions were as the primary but with the extension time increased to 5 minutes. Again PCR products were gel purified.

Blunt end ligation

To generate PCR products to ligate, the primary Ups and Luc/hrp2-3' PCR products or plasmid DNA prepared from clones containing these products were used as template. For Ups reactions, the respective F

oligo+ SpeI site and the UpsR oligo were used. For the Luc/hrp2-3' the respective F oligo and the hrpR2+pstI site were used. Standard AmpliTaq reactions were used; the primer pairs, $MgCl_2$ concentration and annealing temperatures are shown in table 6.3. Gel purified PCR products were treated with Klenow and T4 PNK (chapter 2). After ethanol precipitation, the DNA was resuspended in 44 µl water. The Ups samples were digested with SpeI (10 units) and the Luc/hrp2-3' with PstI (20 units) in 50 µl reactions containing 1x buffer and 5 µg BSA. Following heat inactivation, the small digest fragments were removed by Qiagen PCR clean up columns.

Concentrations were determined by spectrophotometry and equal amounts of Ups and Luc/hrp2-3' were ethanol precipitated together (amount of single component ranged from 330 ng to 11 µg). Ligation reactions of 20 µl contained DNA, 1x T4 DNA ligase buffer and 400 units T4 DNA ligase and were incubated at room temperature for 2 hours or overnight at 16^{o}C. Ligations in the ng range were used directly; those in the µg range were gel purified first.

F primer	R primer	Final $MgCl_2$ concentration	Annealing temperature
LucF2	HrptermR2+Pst	2.5 mM	55^{o}C
UpsBF+Spe	UpsR	3.5 mM	65^{o}C
UpsCF+Spe	UpsR	2.5 mM	65^{o}C
UpsC3F+Spe	UpsR	2.5 mM	60^{o}C
pGL2LucF	HrptermR2+Pst	2 mM	65^{o}C

Table 6.3 Reaction conditions for primers used in blunt end ligation

Cloning of Ups/Luc/hrp2-3'

Each of Ups/Luc/hrp2-3' products were initially ligated directly into the four pGEM-T plasmids containing tRNA or GRE fragments that had been digested with SpeI and PstI. As this was unsuccessful, it was decided to clone the Ups/Luc/hrp2-3' into pGEM-T. To achieve this, the fusion PCR products were A-tailed (see above) then ethanol precipitated and cloned into pCR-TOPO-4. The sequence of the UpsC/luc/hrp2-3' clone C4 was confirmed using the following oligos: M13F & R, 1074-3seq4, 1074-3seq11, 10743-seq12, pGL2Lucseq13 and HRPseq14.

Cloning of Ups/Luc/hrp2-3' into GC and tRNA vectors

To prepare vectors containing GC or tRNA to ligate to the Ups/Luc/hrp2-3' fragment, clones that had previously had the Ups/luc/hrp into the GC and tRNA vectors were used. Although these had deletions within the Ups/Luc/hrp2-3' sequence, the vector backbone was intact. By using a double digest to excise a fragment, it confirms that each enzyme was successful. Each of the GC and tRNA vectors was digested in 100µl reactions for 3 hours at 37^{o}C with SpeI and PstI. The reactions contained 1x NEB buffer 2, 10 µg BSA, 60 units SpeI, 120 units PstI and 4-13 µg DNA. The 6.5Kb band was gel purified. The UpsC/Luc/hrp2-3' fragment was excised from the C4 clone first by digestion with PstI in a 100 µl reaction containing 1x NEB buffer 2, 10 µg BSA, 60 units PstI and 10 µg DNA for 1 hour at 37^{o}C. Following gel confirmation of digestion, 30 units of SpeI and 60 units of XmnI were added and digested for a further 2 hours at 37^{o}C. The 4.6 Kb fragment was gel purified.

50 ng vector (GC or tRNA) and 70 ng insert (UpsC/Luc/hrp2-3') were ligated in a 20 µl T4 ligase reaction for 2 hours at room temperature. Transformation of 2 µl of the ligation reaction into JM109 cells followed manufacturer's instructions.

Colony screening

To reduce the number of plasmid preps, bacterial colonies potentially containing Ups and luciferase components were screened by PCR. These used Luc-screen and Ups-screen primers. Both reactions used 0.625 units of MangoTaq (Bioline) in a 10 µl reaction also containing 1x buffer, 2 mM $MgCl_2$, 1 µM of each primer and 200 µM each dNTP. Cycling was as a standard PCR with 20 second extension and annealing temperatures of 50^{o}C for Ups-screen and 55^{o}C for Luc-screen. Products were visualised on 2% agarose gels.

Transfections

The plasmid pHLH (hrp5'+ luciferase/hrp2-3') was used to test transfection protocols. Large scale plasmid preps were made using NucleoBond Extra Maxi kits, following the manufacturer's protocol. A check digest was made to confirm no obvious rearrangements before use. RBC loading and transfections followed the method of Deitsch et al (221). Briefly, 300 µl washed RBC were washed in 5 ml incomplete cytomix. Plasmid DNA was made up to 100 µl with cytomix and added to the RBC. The sample was electroporated, on ice, using a BioRad Gene Pulser at 0.31 kV and 960 µF.

Transfection was by invasion of parasites into pre-loaded RBC. Using 0.5 ml trophozoites at 5% and 4.5 ml culture media with DNA

loaded RBC at 5% haematocrit. Parasites were cultured by standard techniques for 72 hours.

Luciferase assay

Cells were harvested by centrifugation and washed with PBS. The cell pellet was lysed with an equal volume of 1x Cell culture Lysis Reagent (CCLR) from the Promega Luciferase Assay System at room temperature for 5 minutes. To assay, 20 µl lysate was added to 100 µl Luciferase assay reagent and the sample measured in a TD20/20 luminometer (Turner Designs). Saponin lysis prior to addition of CCLR supplemented with 0.1% saponin was tested.

Results

Var gene and GRE expression in cross progeny

In chapter 4 the *var* gene expression of the cross progeny was determined whilst verifying the HB3 and altered 3D7 primer set. The GRE expression of the same progeny clones was determined in chapter 5. The results of these experiments are combined in table 6.4. We were interested to determine whether the expression of any of the GRE elements was linked to the expression of var genes within the same internal cluster. Four of the five clones have no dominant GRE expression. The results for clone XP8 are potentially more informative. XP8 expresses HB3-1074-3, an UpsC type *var* located in central cluster on chromosome 4. A diagram of the contig is shown in figure 6.4a. This clone also expresses a single predominant GRE from the same contig. This is the second example of GRE being expressed from within the same central cluster as the dominant *var* gene.

GRE expression in HB3 clones

The GRE expression of a selection of the 31 HB3 clones was determined by cloning and sequencing RT-PCR products. HB3 clones representing all 3 *var* Ups classes found were chosen, including both telomeric and central UpsB genes. The sequences of the GRE *E. coli* clones were compared to the known HB3 genomic sequences using Clustal analysis (203). The results are shown in table 6.5. To see if there was a relationship between Ups type and GRE expression, the data from table 6.5 is summarized in Table 6.6. As the HB3 supercontigs are not assembled into chromosomes, the telomeric *vars,* and the central *vars* plus GRE are on different supercontigs. Consequently, we cannot get a telomeric *var* expressed from the same supercontig as a GRE. Comparison of the central *vars* showed no GRE expression from the same supercontig with an UpsB type. Four of the seven UpsC type *vars* did express a GRE from the same supercontig. Two of these are *var* genes from supercontig 1.100, shown in Figure 6.3. The other two are from supercontig 1.47 on chromosome 12. This contig, shown in figure 6.4b, contains 5 *var* genes, including the only published case of an UpsA *var* gene in a central location (74). The clone BA06 expresses *var* 1296-3 from a supercontig, 1.142, that does not contain any GRE copies. This supercontig is predicted to form a central cluster of *var* genes with supercontig 1.243 (74). Supercontig 1.243 has one GRE. The presumed central *var* cluster on HB3 chromosome 7 is shown in figure 6.4c.

Clone	XP2	X33	X10		X4	XP8
var expression	mixed	mixed	HB3 *var2csa*		HB3-1459	HB3-1074-3
Ups type			Ups E		Ups C	Ups C
SuperContig			1.262	1.295	183	100
Contig			1727	1817	1459	1074
Presumed chromosome			12	unknown	4	4
Location			subtelomeric	unknown	Central	Central
Adjacent *var*			none	unknown	none*	1074-1 (UpsB) 1074-2 (UpsC)
Adjacent GRE					183	100a & 100b
Dominant GRE expression	none	none	none	none	none	100a & 100b

*This is the only *var* gene in this contig.

Table 6.4 *Var* gene and GRE expression details for cross progeny. Summary of *var* gene expression, Ups type, location, contig and GRE expression details for the HB3x3D7 cross progeny.

Many of the clones had GRE sequences not present in the HB3 genome database (73). A novel sequence in AH04 was found in 9/16 bacterial clones. This same sequence was also present in 7/16 bacterial clones from AC06. BLAST searches using this novel sequence failed to recover any matching hits in the *P. falciparum* HB3 genome. None of the novel sequences matched those found in sequencing the GRE repertoire of HB3 with either Pfu or AmpliTaq (data not shown). Whilst this could be a novel sequence as yet unannotated in the genome, it could also be a result of PCR error. If the polymerase were to make an error in an early round of PCR, this would then be amplified in following cycles, and represent a significant proportion of the final product.

Analysis of the relationship between GRE and *var* expression

There are now several cases where GRE and UpsC type *var* gene expression occur from the same cluster. Could the GRE be acting as a *cis*-regulatory element on UpsC type genes? To investigate this, I examined the spatial relationship between the expressed GRE and *var* genes. Unfortunately, the two copies of GRE in supercontig 1.100 are identical, so no conclusions can be drawn from them. Both possible permutations are included in the data shown in Table 6.7 to see if there could be a common theme. The only potential common factor is that 3 of the 4 cases could have a GRE expressed that is approximately 3Kb upstream of the start of exon 1.

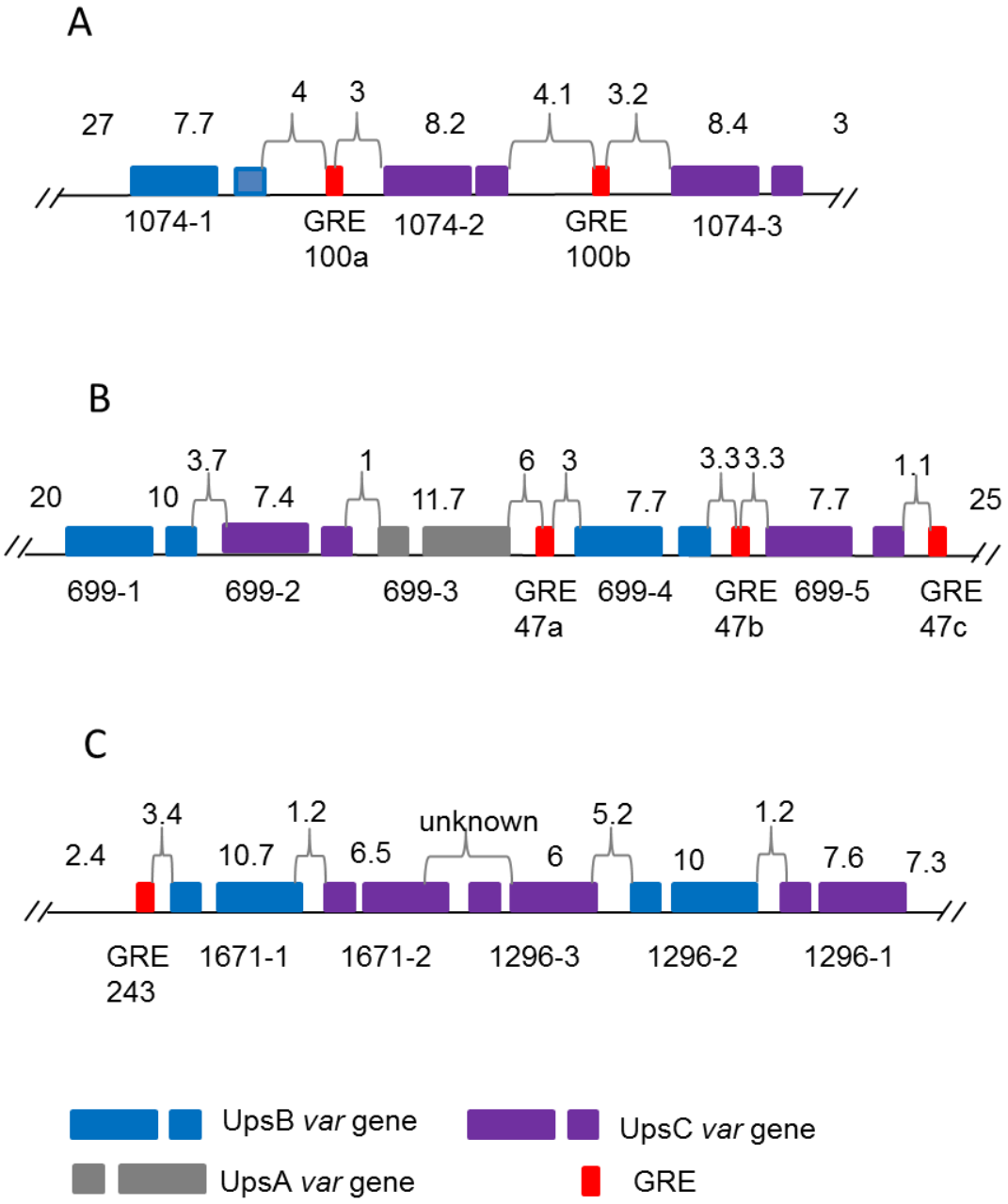


Figure 6.4 Central *var* clusters in HB3. Diagram of some central var clusters in the HB3 genome. The *var* gene and GRE are shown beneath the diagram, distances in Kb above. The distance at the end of each, is to the end of the Ssupercontig. A. Supercontig 1.100, one of the central clusters on chromosome 4. B. Supercontig 47, one of the central clusters on chromosome 12. C. Supercontigs 243 and 142, forming the central cluster on chromosome 7 (74).

.

Table 6.5 GRE expression data in HB3 clones
Data from sequencing RT-PCR products cloned into *E. coli.* The Ups types are shown, for Ups B their chromosomal location is also given, t= telomeric, c=central. *Var* gene expression was determined in chapter 4. Numbers in parentheses are the probability of the number of bacterial clones containing this GRE out of the total number sequenced for each HB3 clone being significant and were calculated by binomial probability.

	Clone							
	DC01	**AC06**	**BD04**	**BB12**	**CE05**	**BF08**	**DH07**	**CC01**
Ups type	A	Bt	Bt	Bt	Bc	Bc	Bc	C
Var gene	1235	1499	1107	1107	1074-1	1074-1	1074-1	1074-2
Chr**	5	u	u	u	4	4	4	4
GRE Locus	Number of bacterial clones containing each HB3 GRE sequence							
47a/ 243	0	2 (0.2)	0	0	0	2 (0.176)	2 (0.102)	0
47b	3 (0.12)	1 (0.32)	2 (0.27)	0	5 (0.05)	0	3 (0.21)	0
47c	6 (0.001)	2 (0.28)	8 ($3.4x10^{-5}$)	4 (0.057)	10 ($2.7x10^{-5}$)	13 ($1.5x10^{-10}$)	4 (0.11)	9 (0.001)
100a& b	3 (0.25)	2 (0.2)	2 (0.22)	8 (0.007)	4 (0.2)	0	3 (0.175)	16 ($1.5x10^{-5}$)
54a	0	0	0	1 (0.32)	0	0	2 (0.28)	0
54b	0	1 (0.32)	0	0	1 (0.22)	0	1 (0.24)	1 (0.15)
183	0	0	0	1 (0.32)	0	0	2 (0.28)	0
Novel	1	7[1]	2	1	1	1	3	0
Dele-tions	0	0	0	0	0	0	0	0
N=	13	15	14	15	21	16	20	26

	Clone						
	DB09	**BH08**	**DE05**	**AH04**	**BA06**	**DF06**	**CH05**
Ups type	**C**	**C**	**C**	**C**	**C**	**C**	**C**
Var gene	1074-3	1074-3	1074-3	1074-3	1296-3	699-2	699-5
Chr**	4	4	4	4	7	12	12
GRE Locus	Number of bacterial clones containing each HB3 GRE sequence						
47a/ 243	0	0	0	0	0	0	0
47b	0	0	0	1 (0.3)	2 (0.29)	0	12 ($1.7x10^{-7}$)
47c	4 (0.11)	6 (0.008)	10 ($9.1x10^{-6}$)	1 (0.3)	8 ($1x10^{-4}$)	11 ($9.3x10^{-7}$)	0
100a&b	10 (0.004)	7 (0.05)	8 (0.027)	1 (0.08)	3 (0.23)	6 (0.12)	0
54a	0	0	0	1 (0.3)	0	0	0
54b	0	0	1 (0.25)	0	2 (0.29)	1 (0.25)	0
183	3 (0.21)	0	0	2 (0.29)	0	0	0
Novel	3	3	0	10*	1	1	0
Dele-tions	0	2	0	0	0	0	8
N=	20	18	19	16	16	19	20

Table 6.5 continued

Ups type	A	B telomeric	B central	C
GRE expression on same supercontig as *var*	0	0	0	4
GRE expression on different supercontig from *var*	1	2	2	3
Mixed GRE expression	0	1	1	1

Table 6.6 Summary of data comparing GRE to *var* expression

Parasite clone	GRE	*var*	Distance upstream to start of *var* Kb	Distance upstream to end of *var* Kb	Distance downstream to start of *var* Kb	Distance downstream to end of *var* Kb	Number of *var* genes between GRE and exon1
CC01*	100a	1074-2	3	11.2			0
CC01*	100b	1074-2			12.3	4.1	0
DB09*	100a	1074-3	18.5	27			1
DB09*	100b	1074-3	3.2	11.7			0
CH05	47b	699-5	3.3	11			0
DF06	47c	699-2			52	45	3

Table 6.7 Distance between expressed GRE and expressed *var* gene. The distances between the expressed GRE and expressed *var* gene in a supercontig are shown in Kb. As GRE 100a and 100b are identical*, both possible situations for clones CC01 and DB09 are described

GRE involvement in *var* gene switching

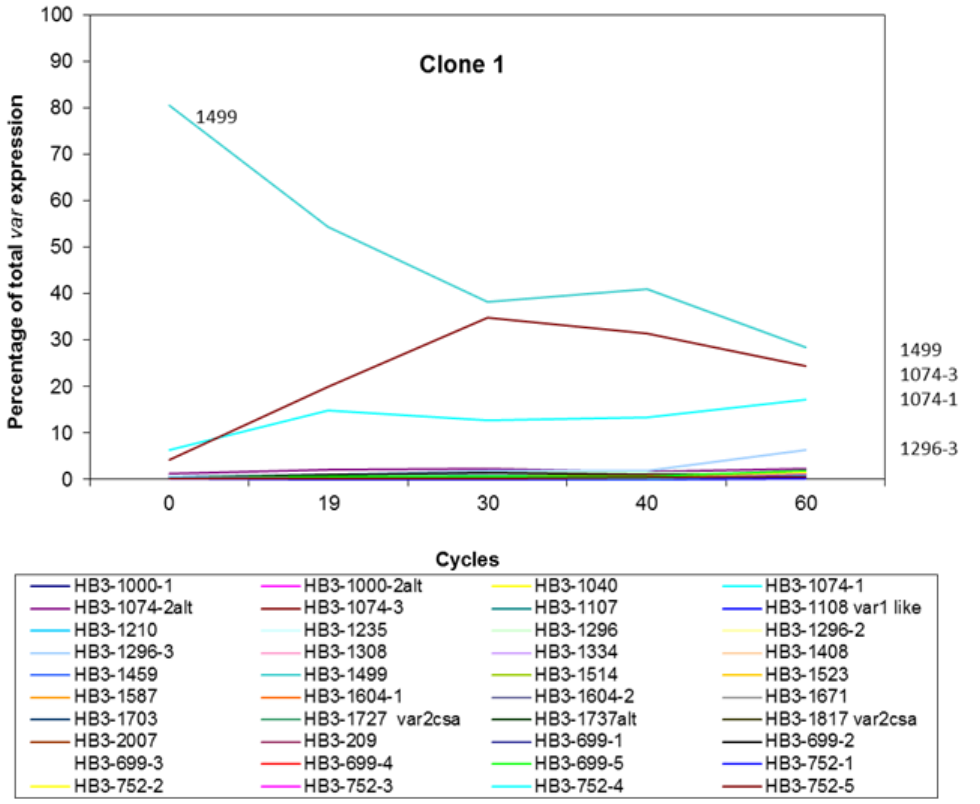
One of the major areas of research involving *var* genes is into switching mechanisms. The potential involvement of GRE in *var* gene switching was studied by following clonal cultures over 50-90 cycles. Eleven of the 31 HB3 cultures were chosen to cover a range of Ups types, along with the non-clonal parent culture.

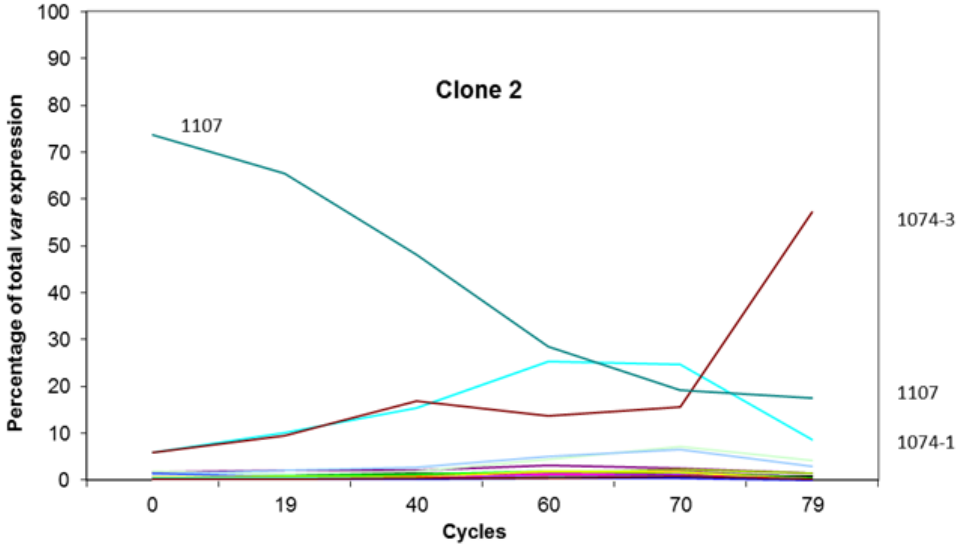
Var gene switch rates

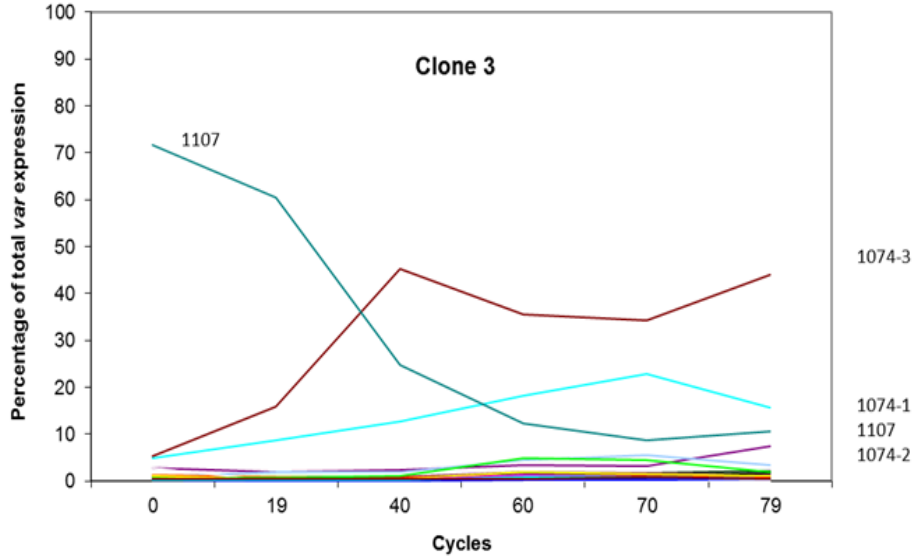
Figure 6.5 shows the *var* gene expression profiles for the 11 clonal cultures over time. It should be noted that for clone 4, the data for HB3-1737 has been omitted from the data set. As detailed in chapter 4, these primers amplify both HB3-1235 and HB3-1737, but northern blot data confirmed only HB3-1235 expression. Using these data, we calculated the off rates of the *var* genes that were initially expressed. These data are presented in table 6.8 and figure 6.6. As n is the number of cycles since cloning, the switch cycle in table 6.8 does not match the cycles in Figure 6.5. Those in figure 6.5 are cycles since the start of the longitudinal study.

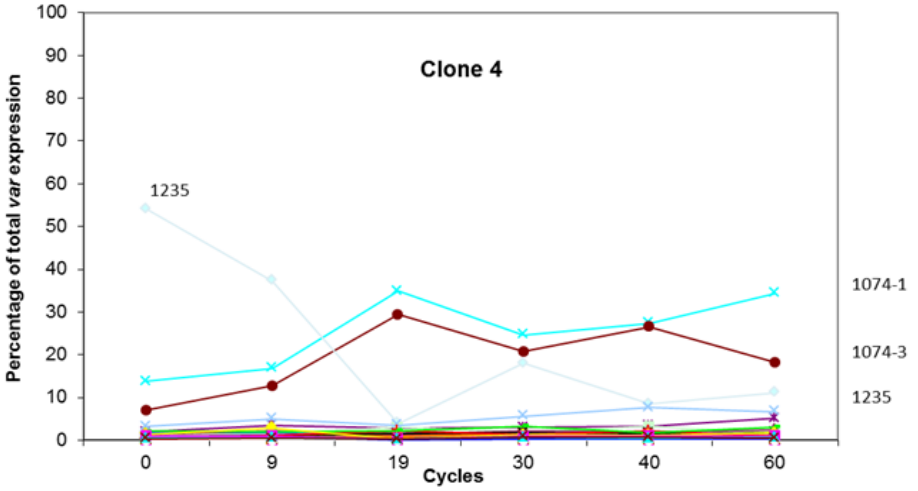
Subtelomeric *var* genes have off rates between 1.42-2.8% per generation. Central *var* genes have a wide range of off rates from 0.3-2.9% per generation. We do not find that, overall, subtelomeric *var* genes have higher off rates than central genes. However, in the case of the UpsB genes, the central UpsB genes have lower off rates (0.63-0.71% per generation) compared to the subtelomeric UpsB genes (1.42-1.98% per generation).

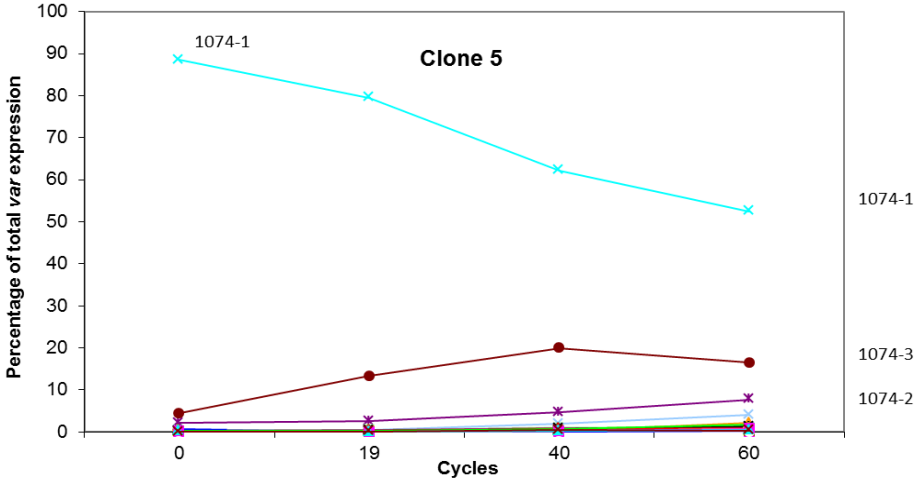
Figure 6.5 *Var* gene expression profiles of HB3 clonal cultures. Data for each *var* gene is expressed as a percentage of the total *var* gene expression at each time point.

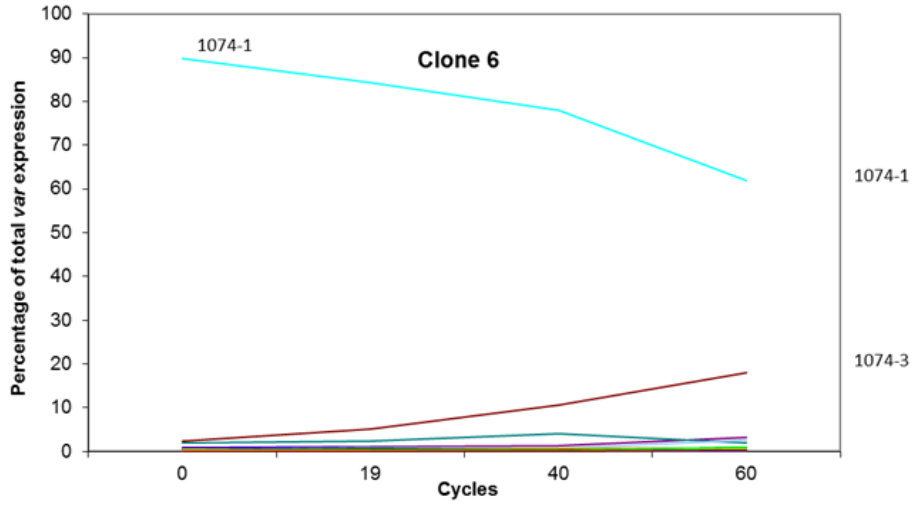


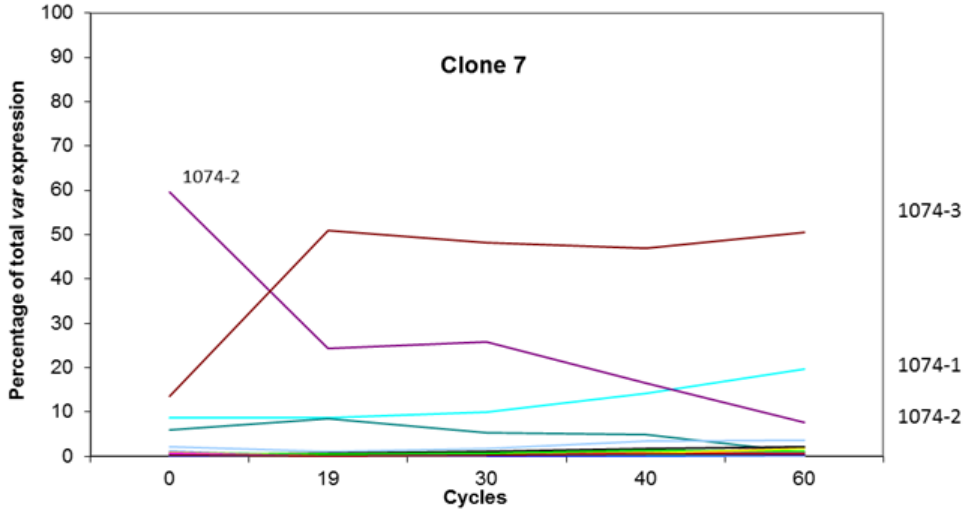


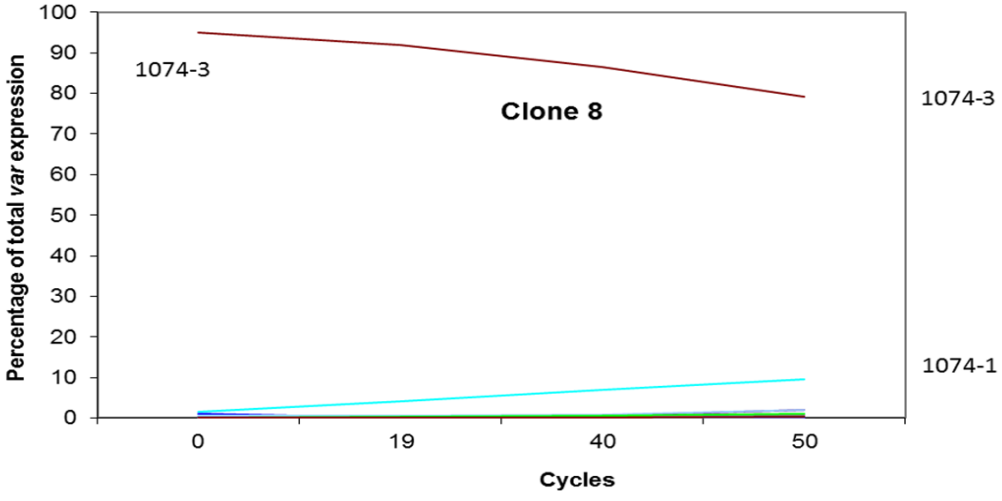

Clone 3
Percentage of total var expression
Cycles
1107
1074-3
1074-1
1107
1074-2
0
19
40
60
70
79

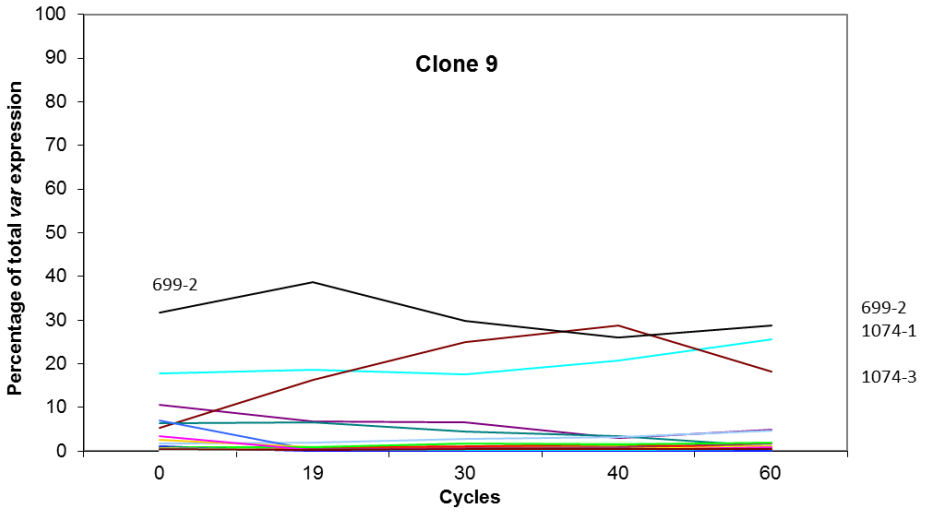

Clone 4
Percentage of total var expression
Cycles
1235
1074-1
1074-3
1235
0
9
19
30
40
60

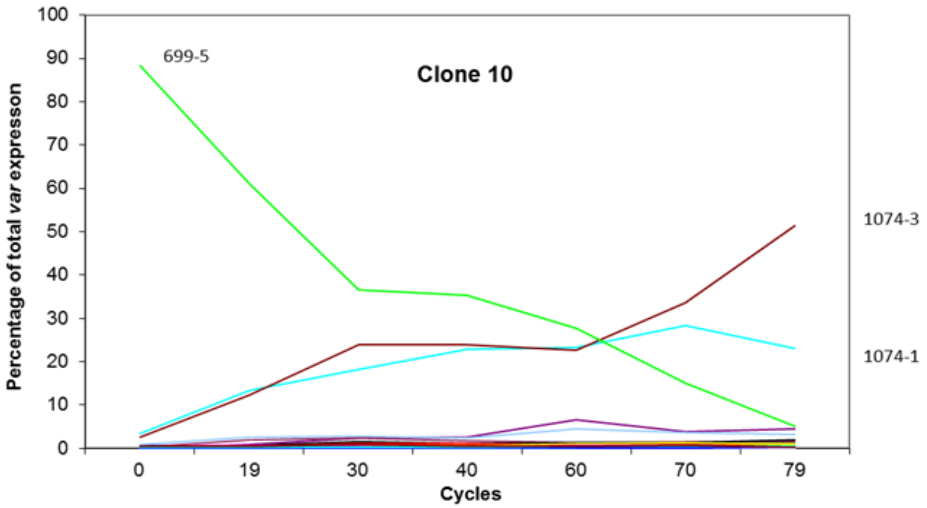

Clone 5
Percentage of total var expression
Cycles
1074-1
1074-1
1074-3
1074-2
0
19
40
60

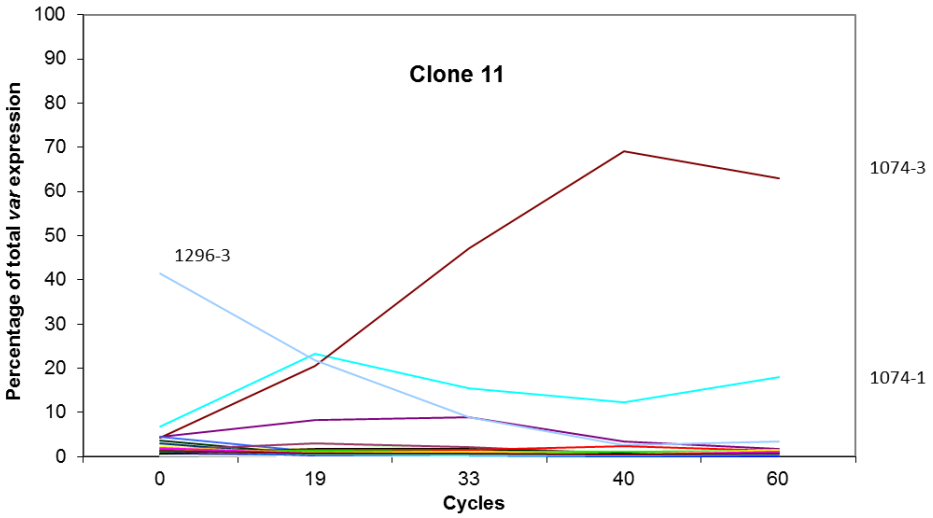

Clone 6
1074-1
Percentage of total var expression
Cycles
0
19
40
60
1074-1
1074-3


Clone 7
1074-2
Percentage of total var expression
Cycles
0
19
30
40
60
1074-3
1074-1
1074-2


Clone 8
1074-3
Percentage of total var expression
Cycles
0
19
40
50
1074-3
1074-1


Clone 9
Percentage of total var expression
100
90
80
70
60
50
40
30
20
10
0
699-2
699-2
1074-1
1074-3
0
19
30
40
60
Cycles


Clone 10
Percentage of total var expresson
100
90
80
70
60
50
40
30
20
10
0
699-5
1074-3
1074-1
0
19
30
40
60
70
79
Cycles


Clone 11
Percentage of total var expression
100
90
80
70
60
50
40
30
20
10
0
1296-3
1074-3
1074-1
0
19
33
40
60
Cycles

Clone number	Initial var gene	Ups type , location and chromosome (if known)	Switch cycle (since cloning)	Final dominant var gene (Bold if switched)	Ups type , location and chromosome (if known)	Percentage of culture expressing final Dominant *var* at switch cycle.	Off rate %
1	1499	B tel	N/A (cultured for 89 cycles)	1499	B tel	28.4	1.42
2	1107	B tel	99	**1074-3**	C cen 4	19.2	1.65
3	1107	B tel	70	**1074-3**	C cen4	24.7	1.98
4	1235	A tel 5	60	**1074-1**	B cen 4	17.9	2.8
5	1074-1	B cen 4	N/A (cultured for 90 cycles)	1074-1	B cen 4	52.5	0.71
6	1074-1	B cen 4	N/A (cultured for 89 cycles)	1074-1	B cen 4	56.6	0.63
7	1074-2	C cen 4	63	**1074-3**	C cen 4	25.8	1.2
8	1074-3	C cen 4	N/A (cultured for 80 cycles)	1074-3	C cen 4	79.1	0.3
9	699-2	C cen 12	69	**1074-3**	C cen 4	26.0	1.79
10	699-5	C cen 12	99	**1074-3**	C cen 4	15.1	1.89
11	1296-3	C cen 7	51	**1074-3**	C cen 4	21.8	2.9

Table 6.8 var gene switching data and off rates for HB3 clone

We noticed that specific *var* on rates varied depending on the identity of the preceding dominant transcript. Specifically, for clones 5 and 6, the on rate of HB3-1074-3 is very low, leading to fairly stable expression of HB3-1074-1. In contrast, for clone 11, the on rate of HB3-1074-3 is high, leading to rapid switching from HB3-1296-3. HB3-1074-1 is an UpsB central *var* gene. Overall, all Ups types, with the exception of UpsB centrals, were capable of switching. Interestingly, all of the clones that switched expression switched to dominant *var* genes on the same central cluster on chromosome 4. This is the same as supercontig 1.100, found with the cross progeny XP8 (figure 6.4a).

We also followed the expression of the unselected parent HB3 culture; these data are shown in figure 6.7. At cycle 0, the parent was expressing central *var* genes, but also approximately 25% of the expression was from HB3-1107, a telomeric UpsB type. By cycle 60, the HB3-1107 expression had decreased to 7% and increased at the UpsC gene HB3-1074-3.

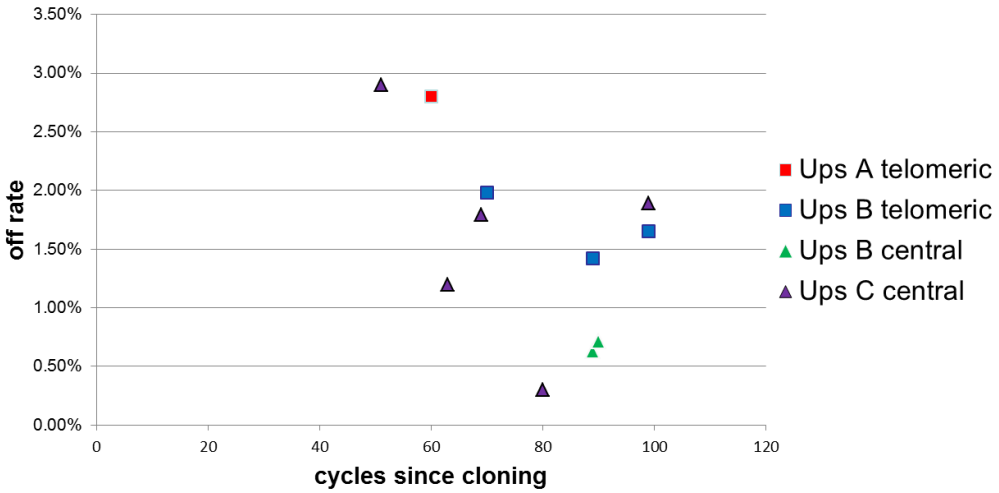


Figure 6.6 Off rates of different Ups classes of *var* genes.

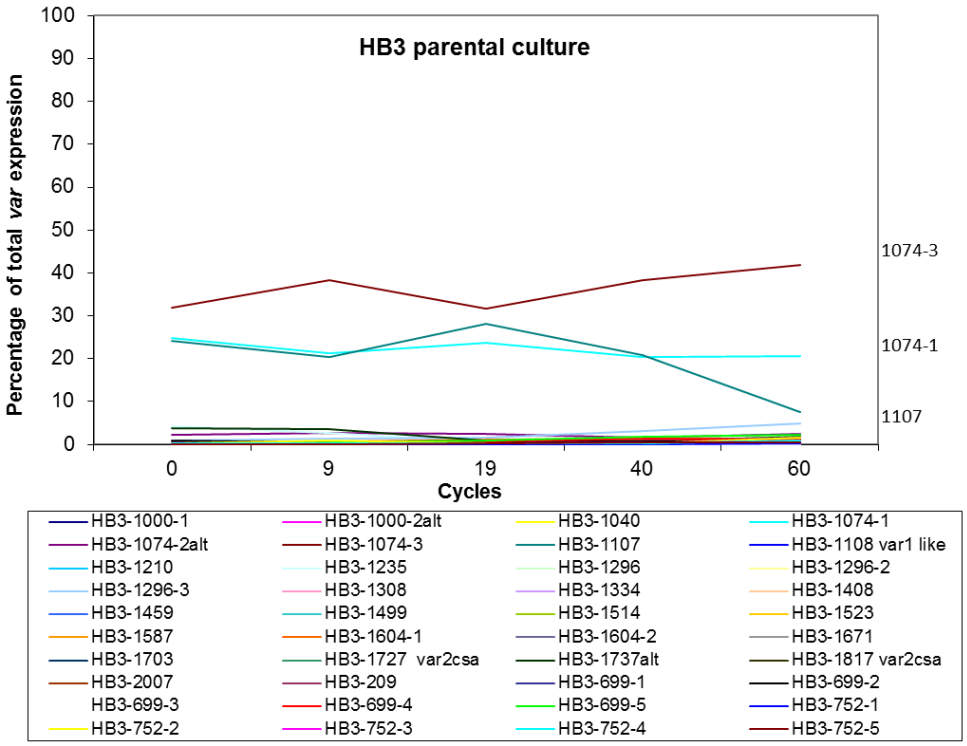


Figure 6.7 *var* gene expression profile of the HB3 parental culture.

GRE and var gene switching

Expression of GRE was studied by cloning and sequencing RT-PCR products from a variety of parasite clones, at a number of time points. Samples for investigation were selected on the basis of the raw qPCR data before correction factors were applied. Those analysed were: clone 1, t=0 and t=40; clone 4, t=0 and t=40; clone 7, t=60; clone 8, t=19 and t=50; clone 10, t=79 and clone 11, t=40. These data are presented in tables 6.9 and 6.10. When an UpsB gene is the dominantly expressed *var* gene, the expressed GRE is not from the same central cluster. However, when an UpsC is the dominant *var* gene, a GRE is expressed from the same cluster. The only exception is clone 9 at cycle 40. At this time point

	Clone number (cycle)										
	4 (0)	**4 (0)**	**5 (0*)**	**6 (0*)**	**7 (0*)**	**7 (60)**	**8 (0*)**	**8 (19)**	**8 (50)**	**9 (0*)**	**9 (40)**
GRE Locus											
100a/b	12 (**0.0002**)	0	0	3	16 (**$1.5x10^{-5}$**)	6 (**0.005**)	7 (**0.01**)	19 (**$4.8x10^{-11}$**)	14 (**$1.4x10^{-5}$**)	6	2
47a/243	0	0	2	2	0	0	0	0	1	0	0
47b	1	0	0	3	0	1	0	1	1	0	1
47c	4	16 (**$8x10^{-15}$**)	13 (**$3.2x10^{-11}$**)	4 (**0.08**)	9 (**0.001**)	2	6 (**0.01**)	1	5	11 (**$4.4x10^{-7}$**)	8 (**$1.6x10^{-5}$**)
54a	2	0	0	2	0	0	0	0	0	0	0
54b	1	1		1	1	0	0	0	0	1	2
183	0	0	0	2	0	0	0	0	0	0	0
Deletions	1	4	0	0	0	1	2	0	0	0	3
Novel	2	0	1	3	0	6	3	0	0	1	0
n=	23	21	16	20	26	16	18	21	21	19	16

Table 6.9 GRE expression during *var* gene switching.
The numbers of bacterial colonies containing the respective GRE are shown.
Numbers in parentheses are the probability of the number of bacterial clones containing this GRE out of the total number sequenced for each HB3 clone being significant and were calculated by binomial probability. The cycles marked * were data from the original assessment of GRE in these clones shown in table 6.5

Table 6.9 continued

	clone number (cycle)			
	10 (0*)	**10 (79)**	**11 (0*)**	**11 (40)**
GRE Locus				
100a/b	0	12 **(0.0002)**	3	12 **($5.9x10^{-5}$)**
47a/243	0	1	0	0
47b	12 **($3.5x10^{-12}$)**	2	2	0
47c	0	4	8 **($6.5x10^{-5}$)**	6
54a	0	0	0	0
54b	0	1	2	0
183	0	0	0	0
Deletions	8	0	0	1
Novel	0	0	1	0
n=	**20**	**20**	**16**	**19**

clone number	**4**	**4**	**5**	**6**	**7**	**7**	**8**	**8**	**8**	**9**	**9**	**10**	**10**	**11**	**11**
cycle	0	40	0*	0*	0*	60	0*	19	50	0*	40	0*	79	0*	40
Dominant GRE	100a/b	47c	47c	47c	100a/b	100a/b	100a/b	100a/b	100a/b	47c	47c	47b	100a/b	47c	100a/b
Dominant var	1235	1074-1	1074-1	1074-1	1074-2	1074-3	1074-3	1074-3	1074-3	699-2	1074-3	699-5	1074-3	1296-3	1074-3
% dominant var expression of total var expression	54	27	87*	78*	54*	50	97*	92	79	40*	28[1]	87*	51	52*	69
Ups type	At	Bc	Bc	Bc	Cc	Cc	Cc	Cc	Cc	Cc	Cc	Cc	Cc	Cc	Cc
GRE and var expression from same contig	na	n	n	n	y	y	y	y	y	y	n	y	y	n	y

Table 6.10 Relationship between GRE and *var* expression in switching experiment. The cycles marked * were data from the original assessment of GRE in these clones shown in table 6.5**.** [1] This clone has 28% expression of 1074-3 but 26% expression of 699-2 at this time point. The dominant GRE, 47c, is from the 699-2 contig.

the dominant expressed gene is HB3-1074-3. The expression of HB3-1074-3 is only 28% of the total *var*, whereas the expression of HB3-699-2 is 26% of the total, figure 6.3. This is also the only time point when HB3-1074-3 is expressed above HB3-699-2 levels. GRE expression is from the same cluster as *var* gene HB3-699-2.

With the exception of clone 9 at cycle 40, the data presented in table 6.10 shows that GRE and *var* expression are from the same central cluster, only when an UpsC type *var* gene is expressed. This provides circumstantial evidence towards, but does not prove, the hypothesis that GRE has a role in *var* gene regulation. It was decided however that this was enough evidence to proceed to transfection studies.

Transfection constructs

To prove the hypothesis that GRE has a role in regulation of central *var* gene expression, it was decided to make transfection constructs. By placing a GRE upstream of a reporter gene with either an UpsB or UpsC promoter, it was hoped to demonstrate that the GRE would drive expression from the UpsC promoter, but not the UpsB type. As a negative control, an alternative Pol lll transcript, tRNA was used in place of GRE. Because of the time required to select for integrated transfection constructs, it was decided to use transient transfections in the first instance.

The initial stage of assembly, PCR cloning of two regions of gDNA, one containing GRE and the other tRNA into pGEM-T was successful. As described in the methods, generating the luciferase/hrp2 3' control region by PCR from pHLH was problematic. To circumvent this we cloned the

luciferase gene from the pGL2-Basic vector (Promega). Cloning and sequencing of the pGL-2 luciferase/hrp product revealed two amino acid alterations. The first is a substitution of Asp by Thr at position 6, and the second a substitution of Phe by Ser at amino acid 14, figure 6.8.

Both UpsB and UpsC PCR products, 1074-1 and 1074-3 were cloned into pCR4-TOPO and confirmed by sequencing. Several attempts at PCR fusion and blunt end ligation to join them to the luciferase/hrp were unsuccessful. Eventually a new UpsC product generated from UpsC-3F was joined by PCR fusion to the luciferase/hrp and cloned into pCR4-TOPO. The insert of this clone, C4 was fully sequenced to confirm that there were no rearrangements or deletions. The UpsC/luciferase/hrp was isolated by restriction digest. This fragment was then cloned using SpeI and PstI into the multiple cloning sites of the pGEM-T vectors that contained GRE or tRNA. Many clones were positive by PCR screening for luciferase and Ups, and all of these were checked by restriction mapping. The few that fulfilled the screening criteria were sequenced at the cloning sites. Clones pGC6C, pGC2C and pt4C (figure 6.2) were successfully assembled

Despite repeated attempts using three different UpsB sequences, HB3-1074-1, HB3-752-4 and HB3-699-1, and both joining methods, we were unable to isolate any clones that had UpsB joined to luciferase/hrp.

.

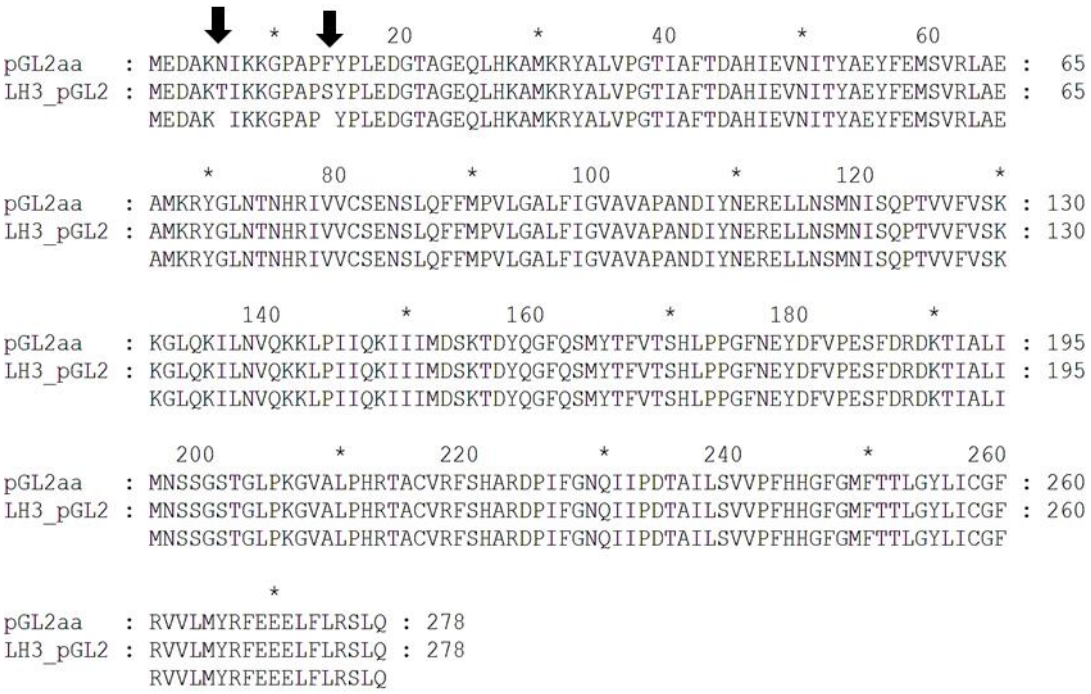


Figure 6.8 Differences in amino acid composition from pGL2-Basic luciferase and when cloned into LH3_pGL2.

Transfections

As we had no UpsB constructs we were unable to complete the transfection experiments. However, it was decided to trial the experimental procedures that would be required should the constructs become available. An initial trial of transfections using plasmid pHLH (hrp5'+ luciferase/hrp2-3') compared saponin lysis to Promega's cell culture lysis reagent (CCLR) supplemented with 0.1% saponin. The results, using DNA electroporated into RBC at 0-75 µg, are shown in figure 6.9. One sample was removed from the 75 µg DNA cells prior to saponin lysis and treated by lysis directly into CCLR + saponin. In the luciferase assay, the average for this sample was 0.498 units, whereas the equivalent sample saponin

lysed was 0.844 units. Direct lysis into CCLR + saponin reduced luminescence by 40%.

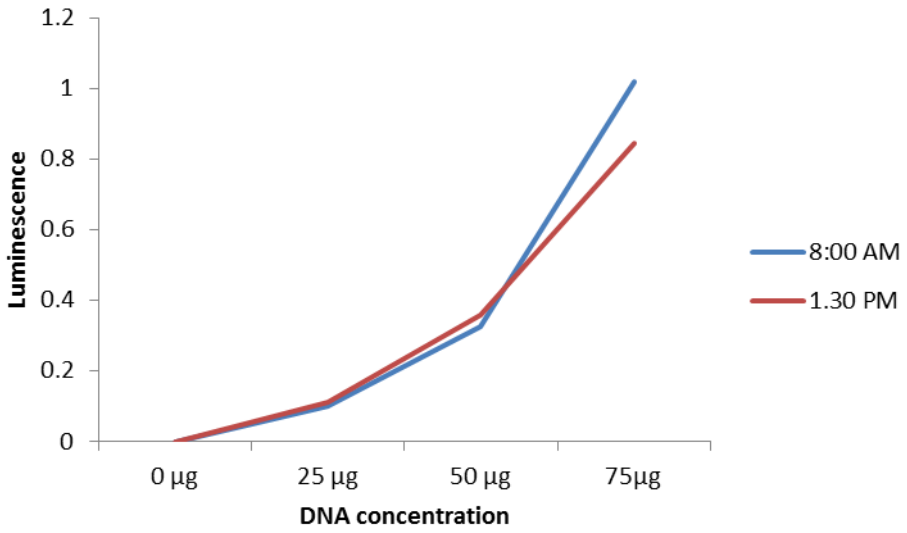


Figure 6.9 Effect of DNA concentration in pre-loaded RBC on transfections. Samples were taken at two time points to ensure parasites were at late ring stage, and treated by saponin lysis prior to the luciferase assay. Each time point is an average of two samples.

Discussion

Prior to the start of this study, we had found only one case of a GRE and *var* expression from the same central cluster. This was from a clone of 3D7 that was expressing PFL1955w and the GRE 12_2 located 15 Kb downstream from the *var* gene on chromosome 12 (our unpublished data). The data presented in tables 6.4 and 6.5 show five more cases with expression of both from the same cluster. In all cases of co-expression, the *var* gene is of the UpsC type. The association is not exclusive, as there are also five cases where there is no link between GRE and *var* gene expression. Whilst HB3-1459, expressed in X4, is documented as being on chromosome 4, it is the solitary *var* gene in the contig, alongside

a single GRE. Whilst solitary internal *var* genes exist, in all other cases, the GRE are found in or at the end of clusters. It is possible therefore that this contig awaits assembly with another, to form a larger central cluster. The HB3 clone BA06 expresses a *var* gene from a supercontig which has no GRE within it; however Kraemer et al link the supercontig to another, forming the central cluster on chromosome 7 (74). If this is the case, the minimum distance between the two elements is 22 Kb. For clone BH08, and to a lesser extent DE05, the presence of duplicate GRE sequences can explain the result. BH08 had more colonies positive for GRE100a & b than 47c, but as GRE 1000a & b are identical, the result is less significant. It is entirely possible that expression is from only one of these loci, but this cannot be determined. Interestingly, AH04 expressed a novel GRE sequence. Two explanations are possible. Firstly, that this is indeed a novel sequence that has not been assembled. The other explanation is that the sequence arose from a PCR error. As this sequence was not found in the PCR of HB3 gDNA the second explanation is most likely.

Long term culturing studies showed some clear differences in switching rates between different classes of *var* genes. Whilst sub-telomeric UpsB genes have switch rates of 1.42%-1.98% per generation, central UpsB genes tend to be stable. Central UpsB genes have very low off rates, and were the only Ups class that did not switch. We found that UpsC *var* genes were able to switch expression, but only to other central *var* genes. Overall, without selection pressure, HB3 cultures favour switching expression to *var* genes on a central cluster on chromosome 4. Our findings are generally in agreement with those of Frank and colleagues who studied switching in the NF54 lineage (210). The specific

off rates differ however, due to differences in the formulae used to calculate them. Frank et al calculate the difference in transcript levels between two time points, divided by the number of cycles between the two time points. Frank's method, unlike the one used in this study, does not take into account the total number of cycles since cloning. Their data from NF54 clones also show that clones expressing central *var* genes, preferentially expressed those from a central cluster on chromosome 4.

The natural progression from analysing *var* gene switching, in the context of this work, was to ask if GRE expression altered in tandem with *var* switching. To this end, cDNA from selected clones and time points was analysed by RT-PCR. In all samples except for two, dominant GRE expression was from the same contig as dominant *var* gene expression. One of these cases is t=0 for clone 11, this is the same sample as BA06 discussed earlier. The other sample is clone 9 a t=40. Whilst it is expressing HB3-1074-3, it is only at 28% of the total *var* expression. It is also expressing 26% of HB3-699-2 which is from the same contig as the dominant GRE, 47c, and is the *var* gene it has switched from.

Whilst supporting the hypothesis that the GRE has a role in *var* gene expression, the switching data is not proof of such. To provide more conclusive evidence that there is an involvement, transfection constructs were assembled. There are now several cases linking GRE HB3-100a and b with the UpsC HB3-1074-3, but none with the UpsB gene HB3-1074-1 from the same central cluster. For this reason these two Ups sequences were originally chosen as the Ups elements for the constructs. The inability to clone UpsB/luciferase/hrp2 fusion inserts is most likely

technical, as other researchers have made constructs containing the UpsB sequence (65,237). There are two main possible points of failure, although as UpsC was cloned these should not be insurmountable. First, overlap PCR was originally described in fusion of four small PCR products to assemble a final insert of 971 bp (235), but fusion products up to 10.8 Kb have since been reported (238). As our product was 4.7 Kb it should be achievable, and was indeed visible, albeit faintly, on a gel. As there was so little product, it was impossible to quantify and optimise the ligation reactions, and therefore very few colonies resulted. As with the majority of cloning involving *Plasmodium* DNA, the AT richness makes it unstable in *E. coli*. It was hoped this would not be a major problem as the luciferase gene, which accounts for 50% of the insert, comes from *Photinus pyralis.*

In the original description of the GRE (58), no correlation was found with respect to location of GRE and adjacent *var* genes. Looking for common factors where a GRE was expressed at the same time as an adjacent *var* gene was difficult. This was because the sequences of the two GRE elements on supercontig 1.100 are identical. However, there are three possible cases where the GRE is 3-3.3 Kb upstream from the start of exon1. For this reason this distance was chosen for the GRE element in the constructs. As a control for the distance of the GRE from the start of the *var* gene being crucial, the PCR product was cloned in both directions. To ensure that any effect seen was due to the GRE, similar constructs were made using a section of DNA containing a tRNA. This was chosen as, like the GRE, it is a small RNA polymerase lll transcript.

Owing to the inability to isolate a clone with the UpsB fused to the luciferase/hrp sequence, the transfection work could not be completed. As this experiment comparing the effect of GRE on UpsB and UpsC is essential in proving a role of GRE on *var* expression the work will continue. We will continue with other UpsB sequences to get a fusion clone. Although we have only used HB3 sequences to date, it may be necessary to try with 3D7 sequences as well. Once this has been successful we can move onto transfection studies. We have trialled transfections, testing both saponin lysed samples and cclr plus saponin lysis as reported recently (239). We also find that luminescence is reduced with cclr plus saponin lysis, most likely due to the presence of haemoglobin which has been shown to quench luciferase luminescence (240). However, although our number of replicates was lower, we did not find saponin lysis prior to addition of lysis buffer to be inconsistent.

Although we have only circumstantial evidence of GRE involvement in *var* gene regulation, it is worthwhile postulating on the role of GRE within the cell. It has already been shown that central *var* genes localise to the nuclear periphery (95). A model that requires the recruitment of TFIIIC and direct localization as a site positioning mechanism is appealing. However, localization at the nuclear periphery, has only been demonstrated from DNA elements of ETC sites (229). As the GRE contain both A and B boxes, and are actively transcribed, any model would be expected to include transcription within it. The data from Kendall et al show that mutation of the CBF5 RNP in yeast reduces nearby pol II silencing and disassociates tRNA from the nucleolus (225). Here, the transcript is involved in localization of the DNA to the nucleolus. Although

the ETC and tRNA involve localization to different areas of the nucleus, these studies both use yeast as a model. Could the GRE transcript which is a Pol lll transcript, like tRNA, be involved in localization to the nuclear periphery in a similar way?

The data do not support GRE being directly involved in silencing Pol ll transcription from adjacent *var* gene loci. In the supercontig 47 on chromosome 12, there are three *var* genes between the dominant *var* gene and the transcribed GRE. As the orientation of the Pol lll transcript is not critical (227), the data from the other examples on chromosome 12 and 4 could support direct silencing. However, because of the identical GRE sequences on the chromosome 4 cluster, it is impossible to dissect which GRE is being transcribed.

One likely potential role of GRE is as a boundary element. tRNA genes have been identified as barrier elements, in *S. cerevisiae* and *S. pombe* (241,242). At *S. pombe* centromeres, tRNA genes mark the transition boundaries between silent and active chromatin. At the *S. cerevisiae* Hidden Mat Right (HMR) locus involved in mating type switching, the boundary has also been mapped to a tRNA gene (233). Mutations in the A and B boxes that were known to reduce transcription resulted in loss of barrier activity, as did mutations in the TFlllC and TFlllB. More recent work at the same locus has shown that the silent domain is enriched in Sir3p, but this disappears at the barrier. The region flanking the opposite side of the barrier is enriched in acetylated histones, but free of Sir proteins (243). The barrier tRNA region was depleted of histones. Whilst deletion of either the tRNA or mutations in acetyltransferases did

not lead to a spread of silenced chromatin, the combined effect of both did. Whilst all of this has been demonstrated at a telomeric locus in yeast, Sir2 paralogues in *P. falciparum* have been shown to act on *var* genes throughout the genome (92). PfSir2A was shown to have a role in silencing UpsC genes, with PfSir2B acting on UpsB types. Although the GRE is not a tRNA gene, it does have all of the regulatory sequence elements of one. One possibility is that the GRE sequence acts as the tRNA barrier does in yeast, preventing the spread of the silencing factor PfSir2B from adjacent UpsB *var* genes. The Spe2 motif that is recognised by PfSip2 is found in duplicate copies adjacent to UpsB *var* genes, but the postulated PfSip2 silencing only acts on subtelomeric UpsB genes (102). Could the GRE be an alternative element that acts only with internal UpsB *var* genes? This would be valid for co-transcription of GRE47b and HB3-699-5. However, the inability to differentiate between GRE100a and b renders us unable to draw conclusions from the HB3-1074 cluster. Alternatively, the GRE could act as a barrier preventing the spread of PfSir2A from adjacent UpsC genes. If either case were correct, there must be other co-operating interactions involved in central *var* gene regulation, as there are not GRE located adjacent to each central *var* gene.

In conclusion, we now have a substantial amount of data showing that GRE and UpsC var genes are frequently co-expressed from the same cluster. This is likely to be class specific, as there is no co-expression of GRE and central UpsB genes. In addition, when *var* gene switching occurs from one central cluster to another, dominant GRE expression switches accordingly. We were unable to assemble a construct containing GRE, UpsB, and reporter gene sequences, which was necessary as a control for

comparison to similar constructs containing UpsC sequences. Because of the lack of a control construct, we were unable to prove a role for GRE in the regulation of UpsC *var* genes. While this study is ongoing, in the following chapter we will investigate possible roles for GRE in recombination.

Chapter 7

The potential role of GRE in recombination

Introduction

The identification of a GC rich element (GRE) by workers in the malaria genome project in 2002 started investigations into its possible role within the malaria parasite (58). Interest originally centred on the observation that the GREs are only located in close proximity to central clusters of *var* genes. It is possible however that they may fulfil other roles in the cell. As they were found to be transcribed in all isolates studied except for one, it was felt unlikely that cells would put so many resources into their generation if there was no benefit to be gained.

One possible role for the GRE is a recombination hotspot. Initial evidence for this came when the central *var* clusters were identified as synteny breakpoints between Plasmodium species (159). Comparing regions of synteny between species is one method of evolutionary analysis. Syntenic blocks contain contiguous genes on a chromosome. Where this diverges from one block to another a rearrangement has occurred (244). The *Plasmodium* study identified the central *var* clusters

as synteny breakpoints, with recombination potentially occurring anywhere within the cluster. The authors also looked in more depth at intergenic regions for repetitive sequences that might act as recombination points (159). They identified a family of genes they termed *var internal cluster associated repeat (vicar)* genes consisting of 7 genes and 8 pseudogenes. Five of the genes were reported to have a transmembrane domain, and five to have a signal peptide, creating open reading frames. These *vicar* genes are the GRE but our data (chapter 1) support a RNA polymerase lll transcript with the classical internal A and B box promoter elements (245). Although they are repetitive elements, there is little likelihood in their direct involvement in generating the synteny break points as they are not present in the rodent malaria genomes. Also, it is now known that in comparisons of the *P. reichenowi* and *P. falciparum* genomes the GRE and *var* genes are co-linear with no evidence of cross-overs at the GRE (C. Newbold, personal communication). The *vicar* gene MAL7P1.39 no longer exists and in recent versions of PlasmoDB the GRE are now annotated as ncRNA (156).

Recombination occurs primarily during meiosis, generating genetic diversity between progeny. During meiosis, duplicated chromosomes pair with homologous chromosomes to form a bivalent. Double strand breaks then lead to recombination between non-sister chromatids at chiasma. As meiosis 1 progresses, the crossovers are resolved and the homologs separate (246). This exchange of DNA within chromosomes results in daughter cell chromosomes being a combination of DNA from both parents.

Genetic recombination within *Plasmodium* was confirmed experimentally by studies on the progeny of a genetic cross between two strains of *P. berghei* (247) . The first description in *P. falciparum* was of recombination in the *Msp-1* gene on chromosome 9 (248). In 1999 a genetic map was constructed from the HB3xDd2 cross, and a rate of one cross-over per $1.67x10^6$ base pairs was reported (249). Recombination between *var* genes was demonstrated in progeny of the HB3xDd2 cross by Freitas-Junior and colleagues (94). Mitotic recombination has also been reported in *var* genes, with the recombination event occurring within the intron (250).

Meiotic recombination hotspots

Several pieces of evidence from other species have led to the hypothesis that GRE may have a role in meiotic recombination. In yeast, three classes of hotspots have been described (reviewed in (158)). The α class require binding of transcription factors (251,252); the β class are created by sequences that prevent nucleosome formation (252) and the γ class are associated with DNA containing a high GC content (158,253,254). Although we are primarily interested in the ncRNA, the template DNA also has a high GC content. Finally, some recombination hotspots in both mouse and fisson yeast are associated with ncRNAs (255,256).

γ hotspots have elevated GC content

There is a significant body of evidence linking γ hotspots with elevated levels of GC. In analysing the *Saccharomyces cerevisae* genome Gerton and colleagues reported a striking correlation between recombination hotspots and high GC content (253). A separate study looked at correlates

of sequence variables and recombination rates in mouse, rat and humans (254). Although other associations were also reported, there was a correlation in both mouse and humans between GC content and recombination rate. More evidence for GC rich sequences being associated with hotspots came with the reporting of a 13mer C rich sequence motif at 40% of human hotspots (257). Petes suggested two possible mechanisms for the γ class hotspots (158). The first was that the increase in GC content may represent a DNA sequence that causes replication fork blockage, leading to local histone modification. This may then be a signal to the recombination machinery. The alternative theory was that the GC sequence itself is recognised by proteins that bind high GC DNA, and these then recruit the recombination machinery. These increases in GC content have been reported in genomes that already have a high GC content with respect to *Plasmodium* genomes. The GC content within the GRE is more striking in the background of an AT rich genome.

ncRNA association with hotspots

In 2004, a 2.4Kb ncRNA was identified that is associated with a recombination hotspot in the mouse (255). Wahls and colleagues investigated the *prl* class of ncRNAs, 300-500 nucleotide polyadenylated ncRNAs that are only produced during meiosis (256). They proposed a mechanism whereby ncRNAs help to position recombination protein complexes at hotspots. Although these 2.4 Kb and *prl* transcripts have no similarity to the GRE, they have no similarity to each other. It is therefore possible that *P. falciparum* has different species of ncRNA that are involved in meiotic recombination.

A final piece of evidence for ncRNA in recombination comes from the report that there is increased recombination between tRNA genes when they are actively transcribed (258). As both tRNA and GRE are transcribed by RNA polymerase lll, and share the same promoter elements, could this be a common feature between RNA polymerase lll genes?

Mitotic recombination

Mitotic recombination is a much rarer and less widely studied event. It is thought to occur as the consequence of DNA damage, such as double strand breaks, mismatches or replication defects. Areas with elevated levels of transcription are also implicated (259). Mitotic hotspots have been identified, for example, an 80kb region between the *lysA* and *hisH* genes in *Aspergillus niger* (260). No genetic element was found in this study, but it is a region of high transcription with tracts of purine bases in non-coding DNA. In *S. cerevisae*, long tracts of (GAA)n(TTC)n increase the mitotic recombination rate 10,000 fold, but have no effect on the meiotic recombination rate (261). This trinucleotide expansion is the same sequence that forms a DNA triplex structure causing Freidrich's ataxia in humans (261,262). Lee and colleagues produced a fine resolution map of mitotic crossover events in a 120kb region of *S. cerevisae* and determined that they were non-randomly distributed (263).

Although there is no evidence to suggest a role for GRE in recombination during mitosis, mitotic recombination is known to occur in *var* genes (250). It is possible that the ncRNA may have an as yet unknown role in mitotic recombination.

Taken together, the data discussed above regarding recombination, give enough pointers to investigate the GRE as a possible marker of meiotic recombination. The progeny of genetic crosses between different *P. falciparum* strains are the only populations available in which to study meiotic recombination in *Plasmodium*. As seen in chapter 4, we have 5 progeny of the 3D7xHB3 cross available (161). We also have two clones of NF54 available to study mitotic recombination. NR13 is a rosetting clone which was examined close to cloning at t=0 (NR13-0) and after 100 generations (NR13-F). Any recombination found in NR13-F that is not in NR13-0 may be due to mitotic recombination.

Material and methods

GRE analysis

Following the data obtained in comparison of enzymes with and without proofreading (chapter 5), GRE sequences were obtained using Kapa HiFi, cloning and sequencing (chapters 2 and 5).

Southern-blot based techniques for recombination analysis

Pulsed Field Gel Electrophoresis

To test for meiotic recombination events around the GRE, a Southern blot was prepared using pulsed field gel electrophoresis (PFGE). Genomic DNA embedded in agarose blocks was digested with EcoRI. Blocks were washed 4 times in TE for 30 minutes at 37°C. Blocks were then equilibrated in 100 µl Restriction mix, 60 minutes at 4°C. Restriction mix contains 1x restriction buffer, 2.5 mM spermidine, 0.1 $mgml^{-1}$ BSA. This was then replaced with 100 µl Restriction mix and EcoRI (20 units per

block, NEB) at 4°C overnight. A further 20 units EcoRI was added then incubated at 37°C for 5 hours.

Blocks were loaded into 1% agarose, 0.5xTBE gels and run on CHEF-DRIII system (BioRad) at 18°C in 0.5x TBE. Electrophoresis conditions were 5.8v/cm and pulse was start 2 seconds, finish 8 seconds, for 18 hours. Staining and visualisation was as chapter 2.

To aid transfer of large fragments, the gel was depurinated in 0.25 M HCL for 25 minutes. It was rinsed in deionised water then blotted by capillary transfer (Chapter 2).

Liquid digests

For analysis of smaller restriction fragments, 2 µg liquid gDNA was digested in a 15 µl reaction containing 1x restriction buffer and 10 units of HinfI, DraI or both DraI and EcoRI (NEB). Digestion was at 37°C for 2 hours. Digested DNA was separated on 1% agarose gel in 1x TBE at 110v for 3 hours. The gel was blotted by capillary transfer.

Hybridization

Hybridization used 25 ng plasmid prep DNA from a clone containing GRE HB3-47c. Hybridization was at 55°C and washes were at 60°C in 1x SSC.

Results

GRE potential for Mitotic Recombination

Whilst assessing PCR error with AmpliTaq and PFU PCR in chapter 5, the PFU PCR was tested on NR13gDNA. NR13, a clone selected from a rosetting phenotype was tested soon after cloning (NR13-0) and after

being cultured for 100 cycles without selection (NR13-F). Although it was chosen for PCR enzyme comparison, it was also chosen to see if mitotic recombination had occurred at the GRE. If mitotic recombination between GRE is responsible for generating novel sequences, it is likely this could occur during the 100 cycles between sampling. The results, in table 5.3 and figure 5.13 show three novel sequences, two after 100 cycles, and one in the original clone. Further investigations revealed that the two novel sequences in NF13-F were PCR artefacts and the origin of the novel NR13-0 could not be resolved. From these data, we have no evidence of mitotic recombination at the GRE.

GRE as a meiotic recombination hotspot

Original attempts to investigate recombination at the GRE in the 3D7xHB3 cross progeny used PCR. These were derived by AmpliTaq, and yielded 18 novel sequences from five clones, figure 5.11. When using AmpliTaq on 3D7 and HB3 gDNA, eight novel sequences were detected. It is therefore likely that most if not all of the novel GRE sequences could be due to polymerase error, or PCR chimeras.

As an alternative method to determine if a recombination event had occurred at a GRE in the cross progeny, restriction digestions were done on gDNA. Initial digests with HinfI, or DraI using liquid DNA were uninformative. Liquid gDNA preparation methods will only reliably resolve DNA fragments up to 12Kb. Larger than this, the DNA shears during the preparation. To move to larger size fragments it was necessary to analyze gDNA embedded in agarose blocks by PFGE. EcoRI digested DNA was resolved by PFGE, transferred, and hybridized with plasmid DNA

containing the HB3-47c GRE sequence. Low resolution mapping with EcoRl produces 0-100 Kb fragments whereas mapping with Dral produces fragments are in the 0-10 Kb range. A recombination event would be expected to result in a new fragment that is not present in either parent. The results of the Dral and EcoRl digests in figures 7.1 and 7.2 show that there have been no recombination events involving fragments containing GRE in the cross progeny.

Discussion

Having established in chapter 5 that proofreading DNA polymerase is necessary for GRE sequence analysis, it was decided to further test this whilst investigating mitotic recombination. Two different gDNA samples from the same clone, separated by 100 generations were examined. After 100 generations two novel sequences were found. Further testing of these showed them to be PCR chimeras. From these data, we have no evidence of mitotic recombination within the GRE in this NR13 clone. Whilst we found no mitotic recombination events, the number of clones sampled was too small to dismiss the possibility. Unless a recombination event happened immediately after cloning, it is unlikely to have occurred in every parasite within the population. Not enough clones were sampled to identify each known genomic GRE sequence. Therefore there was little likelihood of identifying a novel sequence that would not have been present in each parasite in the population.

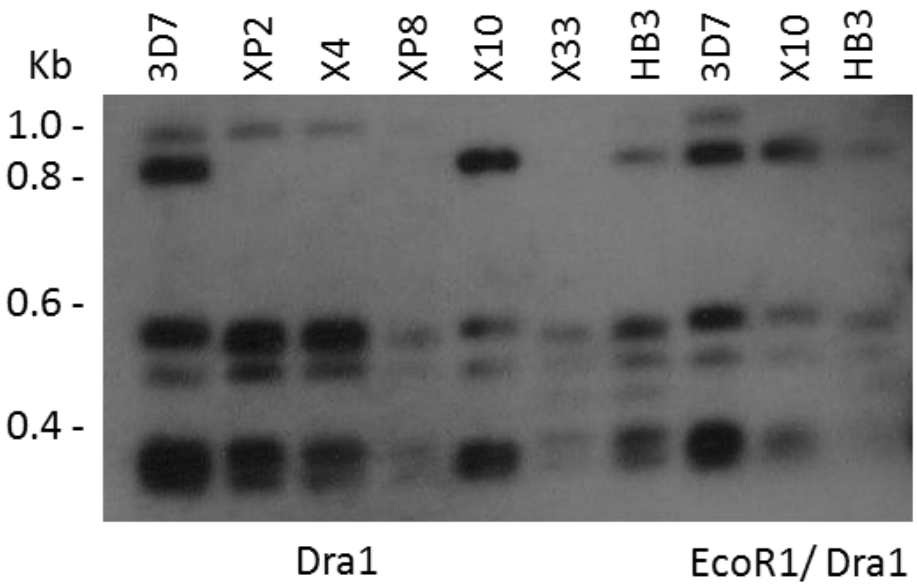


Figure 7.1 Dra1 restriction digests are uninformative in meiotic recombination analysis of cross progeny. Southern blot of EcoR1/Dra1 digest of 3D7, X10 and HB3 and Dra1 digest of 3D7, HB3 and cross progeny.Only parental bands are detected when hybridized with HB3-47c GRE.

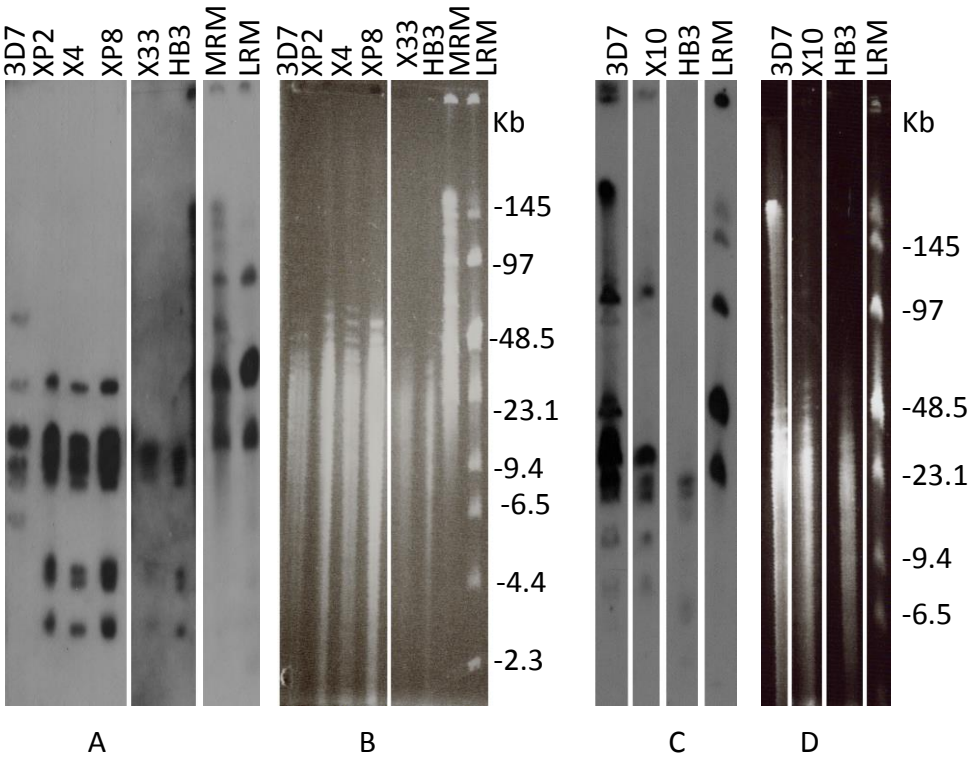


Figure 7.2 EcoR1 restriction digests are uninformative in meiotic recombination analysis of cross progeny.

A, Autoradiograph and B, gel photograph of PFGE of EcoR1 digested gDNA from 3D7, HB3 and cross progeny. C, Autoradiograph and D, gel photograph of PFGE of EcoR1 digested gDNA from 3D7, HB3 and X10. A and C are hybridized with GRE sequence HB3-47c. All bands in the cross progeny match those in either parental genome and therefore have no rearrangement within EcoR1 fragments that contain GRE. MRM, mid range markers. LRM, low range markers (both NEB).

PFGE investigations into GRE as a meiotic recombination hotspot in progeny of the 3D7xHB3 cross found no evidence of such. Recent work has pointed to the DBL regions being associated with recombination events in *var* genes (62,195). Regions with frequent recombination events were found in the variant antigen families of two laboratory clones. Closer analysis identified a gene conversion within DBLα of a *var* gene (195). Further evidence for recombination occurring in the DBLs of *var* genes came from a detailed bioinformatics analysis of *var* gene sequences which identified an intra DBL breakpoint (62). GREs are located in intergenic regions, but a recent publication found that the majority of hotspots in *P. falciparum* occur in protein coding regions (264). Jiang and colleagues also identified five sequence motifs that are enriched in hotspot sequences (264). One is similar to the TARE-6 sequence, rep20. The other four are 12 base pair GC rich sequences, three of which have G every three base pairs. GC rich DNA is common in hotspots in other species. None of these motifs are present in the GRE sequences.

Although we found no evidence of GRE as a recombination hotspot, we had access to only 5 progeny clones from the cross. To make definitive statements from such a small sample size is not possible. Although the *var* gene data presented in table 4.5 shows a recombination event in X10 between HB3-1296-1 and MAL7P1.187 this does not involve sequences that contain a GRE. This was confirmed by SNP data from the Wellcome Trust Sanger Institute, which shows a recombination event between the central cluster and the sub-telomeric *var*, but far beyond the central *var* cluster (S. Asafa, personal communication). A sequencing project is currently underway, sequencing all of the progeny of the

3D7xHB3 cross. Analysis of this data, when available, will determine if recombination occurs at GRE.

Recently, a deletion event resulting in the removal of a GRE from 3G8, a clone derived from the IT lineage, has been described (J. Lemieux, personal communication). From a central var cluster on chromosome 4, the deleted region also includes part of a *var* gene and a pseudo *var* gene. A similar event has been seen at the same region in another clone derived from IT (A. Rowe, personal communication). Deitsch and colleagues first reported a spontaneous 18 Kb deletion of a *var* gene from a central cluster on chromosome 12 in 1999 (265). Looking at the sequences involved, a GRE is also within this deletion event. Deletions at regions of homology are widely reported (258,266). For the GRE to be involved as site for this would not be surprising, as other RNA polymerase lll elements including tRNA and Alu are known to be sites (258,267). In the majority of cases of deletion and insertion at repetitive DNA sequences, there are multiple copies e.g Alu has approximately 1 million copies in the human genome (268), yet the maximum number of GRE identified so far in *Plasmodium* genomes is 15. Also, as GREs are not adjacent to each other, but separated by at least one *var* gene, any deletion event would result in the permanent loss of that gene from the repertoire. In the deletion events mentioned above, sequence data is not available to determine if other GREs are involved. Closer inspection of the sequences at either side of the deletion in 3G8, reveal large regions of homology. It would seem likely therefore, that it is the homologous sequences within the *var* gene, rather than the GRE, responsible for these mitotic rearrangements

In summary, we have found no evidence for the GRE being the site of either a mitotic or meiotic recombination hotspot in *P. falciparum*. With some correlation between UpsC *var* and nearby GRE expression, and no solid evidence for an association with recombination, we reasoned that the GRE RNA itself may have a *var* specific function. This is investigated in the following chapter, with a focus on identification of GRE RNA binding proteins.

Chapter 8

Structural and binding properties of GC rich elements

Introduction

In the preceding chapters of this thesis, experiments were designed to determine a role for the GC rich element (GRE) within *P. falciparum*. In this chapter, an alternative approach is used, aiming to complement the other studies. By investigating possible structural features and binding of GRE to proteins and identifying any such proteins, it is hoped to narrow down the mode of action of the GRE.

Hairpin RNAs

The predicted structure of the GRE is that of a hairpin. Hairpin RNAs are the most commonly predicted secondary structure in RNA and are abundant throughout nature (269). Although they differ in size, they share the same structural features; a double stranded stem which exhibits primarily Watson:Crick base pairing; a single stranded loop; and internal bulges where base pairing does not occur. RNA folding diagrams show the structures as 2-dimensional, although they must be thought of as 3-dimensional structures, (figure 8.1). In the simplest form, the stem is a 3-dimensional helix. A common structure is a pseudoknot; these result from single stranded regions forming stems with other single stranded regions e.g. loops. This folding brings non-contiguous stretches of the molecule

together, resulting in a complex 3-dimensional structure.

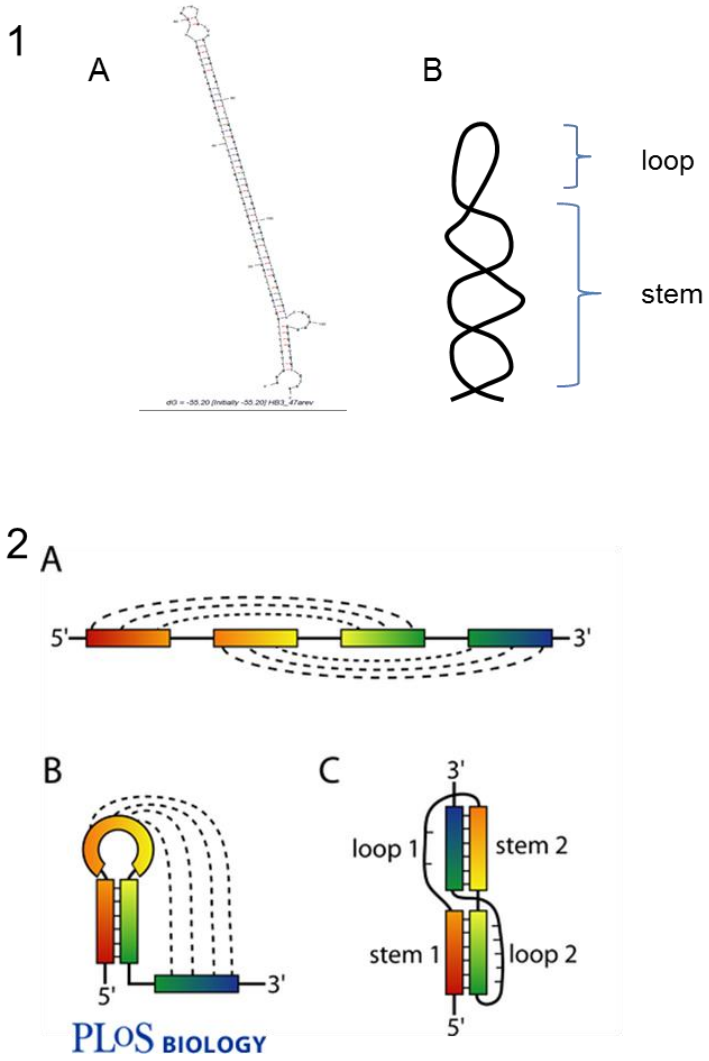


Figure 8.1 RNA secondary structure.
RNA molecules form complex secondary structures. 1A. A common form is the hairpin, represented by a GRE. 1B. A hairpin is not a 2 dimensional structure, it forms a helix in the stem. A tetraloop contains 4 nucleotides in the loop. 2. At another layer of complexity, a pseudoknot is formed. 2A. The linear RNA molecule. 2B. A hairpin structure forms. 2C. A pseudoknot is formed when a region at the 3' end folds back to form a stem with the initial loop.

2 adapted from (270)
http://www.plosbiology.org/article/info:doi/10.1371/journal.pbio.0030213

Hairpin RNAs have many roles within cells with both *cis* and *trans* acting

examples. The widely studied miRNA and siRNA, described in Chapter 1,

are examples of *trans* acting hairpins. There are many cases of *cis* acting hairpin RNAs, some of which will be described below.

Cis acting hairpin RNAs

Regulation of bacterial operons

Regulation of some bacterial operons involves the generation of hairpins within the leader mRNA transcript. The secondary structure was first described for the attenuator at the tryptophan operon in *E. coli* and *Salmonella typhimurium* and a model subsequently proposed (271,272). This coordinated control of both transcription and translation can occur in the single cytoplasmic compartment in prokaryotes. However it would not be expected in *P. falciparum*, as in eukaryotes, transcription occurs in the nucleus and translation in the cytoplasm, so are spatially distinct.

Direct tRNA interaction

tRNA was shown to act as a terminator in *Bacillus subtilis.* The tyrosyl-tRNA synthetase transcript has a hairpin terminator at the end. In conditions of excess tyrosine, the charged tRNA cannot bind directly to the mRNA, leading to termination. When tyrosine levels are low, uncharged tyrosyl-tRNA is able to bind directly to a bulge in the mRNA, forming an alternative secondary structure at the terminator region. This allows read through of the terminator (273).

Riboswitches

The term riboswitch was first used to describe the effect of direct binding of coenzyme B_{12} to *btuB* mRNA (274). Binding of the substrate or metabolite to the mRNA produces a conformational change in the mRNA leading to inhibition of translation.

IRES

The Internal Ribosome Entry Site is a feature of viral mRNA transcription first discovered in picornaviruses (275,276). Eukaryotic translation begins at the 5’ cap structure of mRNA, with recruitment of proteins. Many viruses bypass the cap structure by formation of a secondary hairpin structure on the viral RNA.

RNA binding proteins

We have seen in chapter 1 that there are many ncRNA species involved in regulatory roles within cells. These RNAs do not function in isolation however, and their actions are usually in conjunction with proteins. Collectively these are termed RNA binding proteins (RBP). RNA recognition occurs at RNA binding domains (RBD), and most RBPs contain multiple RBD. This modular structure gives the protein multiple binding sites allowing binding of two or more RNAs, or multiple domains recognising a long RNA. This increases the specificity and affinity of the binding. They can also combine with enzymatic domains with the RNA bringing substrate and enzyme together (277). Functionally, the RBP act on, or with RNA in many ways including mRNA biogenesis, splicing, transport, cellular localization and post-translational gene silencing (278).

RBPs are classified by the structure of their RBD. Although there are more than 40 different classes of RBD, some of which are species specific, there are several common domains (278). A brief description of the common classes and *P. falciparum* examples follow.

RNA-recognition motif

The RNA-recognition motif (RRM) is the commonest and most widely studied RNA binding module in eukaryotes. An individual domain recognises 2 to 8 nucleotides of ssRNA, but multiple RRM domains are usually needed, as one is not enough to define specificity (277,279). RRMs are approximately 90 aa long, containing 2 conserved sequences, RNP1 (K/R-G-F/Y-G/A-F/Y-V/I/L-X-F/Y) (279-281) and RNP2 (V/I/L-F/Y-V/I/L-X-N/L) (279,282). Structurally, they form a four stranded antiparallel β-sheet with two α-helices (283). In *P. falciparum*, 72 genes are predicted to contain RRM within their proteins, although only four of these have confirmed protein assignments (156).

PUF

A family of RNA binding proteins is termed PUF, based on its original description in Drosophila as Pumilio proteins, and in *Caenorhabditis elegans* as FBF (284,285). Found only in eukaryotes, PUF proteins have eight tandem repeats of a decapeptide consensus sequence (285). Each repeat recognises a single nucleotide on single-stranded RNA (277), and this modular nature allows for RNA sequence specificity. PUF proteins act by repressing translation of their target mRNAs (285).Two genes coding for PUF proteins, *PfPuf1* and *PfPuf2*, have been identified in *P. falciparum*, with orthologues in other *Plasmodium* species (286). PfPuf2 has dual roles within parasites. One role is regulation of sexual development, by repressing gametocytogenesis (287). The other role controls sporozoite conversion by inhibiting premature sporozoite transformation (288).

Zinc Fingers

The first description of a protein with finger like repeat structures, each containing a zinc molecule was in the transcription factor TFIIIA purified from *Xenopus laevis* (289). TFIIIA interacts with both RNA and DNA (290,291). It has nine zinc fingers, 4-6 interacting with 5S RNA, 5 with DNA and RNA, and 1-3, 5, 7-9 with DNA (292,293). Nucleotide specific binding to either DNA or RNA occurs via residues at the tip of the “finger”. The difference between them being the finger points towards RNA, whereas with DNA it fits in the major groove (294). As with the PUF proteins, the specificity of each finger and multiple fingers brings the sequence specificity to the molecule. The TFIIIA protein is an example of the classical CCHH zinc finger motif which account for 3% of genes in the human genome (295). There are other classes of zinc finger proteins which were originally classified by the residues around the zinc e.g. CCCH. The CCCH class of zinc finger proteins was found to be more abundant in the *P. falciparum* genome, when compared to other eukaryotic genomes, in a search for transcription associated proteins (296). One function of CCCH zinc fingers is in mRNA stability, where they promote the breakdown of mRNA, so have a role in posttranscriptional regulation (297). To date 272 putative zinc finger proteins have been annotated in the *P. falciparum* 3D7 genome (156).

K-homology domain

The K-homology (KH) domain motif, first described in the K protein of the heterogeneous nuclear ribonucleoprotein complex, is (I/L/V)-I-G-X-X-G-X-X-(I/L/V) (298). KH proteins are found in eukaryotes, bacteria and archaebacteria (298), and are involved in transcriptional regulation and

control, and splicing (278). In *P. falciparum*, one protein containing a KH domain has been annotated, PF3D7_1469300, suggested to be involved in rRNA processing (79).

Double-stranded RNA Binding Domain

Whilst the previous examples of RBP all recognise ssRNA, as its name suggests, the double-stranded RNA binding (dsRBD or dsRBM) domain interacts with dsRNA. The domain is 65-70 amino acids long, and although there is a degree of conservation at some residues, there is not a consensus motif (299). The dsRBD is found in eukaryotes, bacteria and viruses. Unlike other RBD that interact with nucleotides, the dsRBD interacts primarily with the phosphodiester backbone and the 2' hydroxyls of the ribose sugar on one side of the helix, covering 16 nucleotides (300).

DsRBDs are not thought to recognise specific nucleotide sequences, so other modes of recognition were determined for the dsRBD containing enzyme, RNase lll (301). This family of enzymes process RNAs with commonly known members, Dicer and Drosha, both having roles in RNAi (Chapter 1) (302). One role of the *S. cerevisiae* Rnt1 enzyme is in processing pre-rRNAs. Rnt1 binds to the tetraloop of the hairpin (figure 8.1B), and cleavage efficiency is regulated by sequences in two regions of the stem (301). Three ribosomal proteins with dsRBDs are annotated in PlasmoDB (156).

S1 domain

The S1 domain was shown to be required for the interaction between the 30S ribosomes and mRNA in *E.coli (303)*. Pfam describes homologues in archaea, bacteria and eukaryotes (304). The motif has since been shown

to be in several proteins including translation initiation factors IF-1 and eIF2-α (305). A putative S1 RNA binding protein is on *P. falciparum* chromosome 11 (PF3D7_1131000) and homologues of IF-1 (PF3D7_1469000) and eIF2-α (PF3D7_1438000 and PF3D7_1019000) are also present (79).

Electrophoretic Mobility Shift Assays (EMSA)

Knowing that the GRE is transcribed into ncRNA, and that some *P. falciparum* proteins are predicted to contain many of the common RBD, would it be possible to identify any GRE binding proteins? The initial step to identifying any RNA binding proteins would be to confirm that binding exists. Garner and Revzin described a polyacrylamide gel based method to study the binding of catabolite activator protein to the *E. coli* lac promoter (306). After mixing and allowing to bind, protein:DNA complexes were resolved on native polyacrylamide gels. The mobility of the complex, seen as a band, decreased when catabolite activator protein was incubated with the DNA, due to the increased size of complex compared to free DNA (306). Fried and Crothers increased the sensitivity of the technique by using radiolabelled DNA (307).This widely used technique is now called the electrophoretic mobility shift assay (EMSA) or gel retardation assay.

Although originally designed to study DNA:protein interactions, EMSAs have since been used to study RNA:protein interactions (308). The technique can also be used as a first step in identifying the protein components (309). Having visualised an RNA:protein complex by autoradiography, the complex is excised from the dried gel. The gel slice

is then subjected to standard mass spectrometry methods. EMSA protein identification is useful only as a preliminary step, because the gel slice will not only contain proteins involved in the complex, but also other proteins that migrate at the same rate as the complex.

The EMSA has been used successfully to identify both DNA and RNA binding proteins in *P. falciparum*. The SPE2 motif was identified by EMSAs using different plasmids to narrow down the region which gave binding (101). Studies looking for proteins regulated by genes containing AP2 domains in *P. falciparum* confirmed bioinformatically derived candidates by DNA EMSA (310). Recently, both DNA and RNA EMSAs have been used in the characterization of a family of proteins that bind both TARE-6 DNA repeats and single-stranded RNA. They have been called the PfAlba protein family (311,312).

Signal recognition particle

Identification of GRE:RNA binding proteins is a step into the unknown. If no binding is achieved, could this be because there is no such complex, or could it be due to technical issues? As RNA EMSA was a new technique in our laboratory we decided to address technical problems first by demonstrating binding to a known transcript. The signal recognition particle (SRP) was chosen as it is a ubiquitous multi protein:RNA complex that targets proteins to the endoplasmic reticulum (313). The SRP RNA in *P. falciparum* was originally identified by bioinformatics comparisons, and homologues of 5/6 mammalian proteins were found (313). Further analysis has identified more components of the pathway (314). Demonstrating SRP

RNA:protein binding allows us to assess for example, the quality of protein extracts, and RNA degradation issues.

Structural digests

RNA folding programmes such as mfold (202) are useful in predicting the secondary structure of an RNA transcript. Predicted features can be confirmed experimentally by RNase digestion (reviewed in (315)), using a variety of enzymes that cleave either single stranded or double stranded regions, with or without sequence specificity (Table 8.1).

Endonuclease	Double or single stranded regions	Sequence specificity
RNase A	Single stranded	3’ of C or U
RNase T1	Single stranded	3’ of G
RNase VI	Double stranded	None

Table 8.1 Enzymes available for structural digest of RNA. Details from Ambion (316) . Although other enzymes are reported (315), only those commercially available from Ambion are included.

In this chapter, I discuss my attempt to confirm the hairpin structure of the GRE, as predicted by computer models, with RNase digestion assays. I confirm GRE:protein binding, *in vitro,* by EMSA. Having established binding conditions, I use these to isolate potential GRE binding proteins and identify them using an in-house mass spectrometry service. The identification of any GRE binding proteins might give an indication of the role of the GRE within the cell.

Materials and Methods

All of the experiments described in this chapter use RNA transcripts generated *in vitro* as their starting material. The following transcripts were made for the work described herein.

1. **GCF**. For EMSA and pull-down studies, the GCF=GC. GC matches the genomic GRE sequences 3D7 4-1/7-1/7-2

2. **GCR**. A reverse transcript GCR was made using a T7+GCR primer.

3. **tRNA-glu**. tRNA have a conserved structure and a tRNA transcript was made using the GCF conditions. Three primers were designed, tRNAgluF, tRNAgluR and T7tRNAgluF.

4. **GCS**. Protein binding studies required a transcript that contained the same nucleotide composition, but in a random order. The GC sequence was randomised and called GCScrambled (GCS) (appendix i).

5. **SRP.** Initial attempts to determine reaction conditions for EMSAs used an SRP transcript. Three oligos were designed for this transcript. SRPF, SRPR and T7SRPF.

Technical details of specific transcripts are in the following section. All primer sequences are in appendix i.

Transcript generation

Transcripts were generated using the MEGAshortscript Kit (Ambion). Primers specific for the required GRE were designed and used on a previously cloned GRE+plasmid template, to ensure generation of a single

RNA species. The primers designed for GC, are in appendix i. These primers were also used to generate a different GRE transcript, NR32-1. Although the primer sequences were identical, the GRE sequence was different and had been cloned in a different plasmid template.

The primary GCF/GCR PCR was a standard reaction (chapter 2) at 2.5 mM $MgCl_2$ with the following cycling conditions: 94^{o}C for 4 minutes, followed by 35 cycles of 94 ^{o}C for 30 seconds, 65 ^{o}C for 30 seconds and 72 ^{o}C for 30 seconds. The secondary PCR, which adds a T7 promoter, was as the primary, using 1 µl primary product in a 50 µl reaction with T7+GCF/GCR primers. The secondary PCR product was run on a 2% agarose gel and gel purified using the Qiagen Gel extraction kit.

The MEGAshortscript reaction was performed following manufacturer's instructions. Briefly each reaction contained 2 µl T7 10x buffer, 2 µl each of ATP, CTP, GTP and UTP solutions (75 mM) and 2µl T7 enzyme mix, 40 pmol template PCR product and water to 20 µl. The reaction was incubated at 37^{o}C for 2-3 hours. Template DNA was then removed by addition of 1 µl DNase, 37^{o}C for 15 minutes. The reactions were terminated and enzymes removed by acid phenol/chloroform extraction, then ethanol precipitated with 3 volumes ethanol. Quantification was by spectrophotometry (chapter 2).

Two primers were designed for GCS. T7+GCscrambledF contained the T7 sequence and the 65 5' nucleotides of the GCS sequence. GCscrambledR was the reverse complement of the last 90 nucleotides. This gave an overlap of 20 nucleotides. The GCS PCR product was generated in a standard reaction containing 1 µM of each primer and 2.5

mM $MgCl_2$. The annealing together of the two primers acted as template. The PCR cycling was 96^{o}C for 5 minutes, followed by 35 cycles of 96^{o}C for 30 seconds and 72^{o}C for 30 seconds. The resulting product was gel purified as above before use as template in a MEGAshortscript reaction.

The primary PCR using SRPF and SRPR was in a standard reaction at 2.5 mM $MgCl_2$ with 3D7gDNA as template. The secondary PCR used T7SPRF and SPRR. Cycling conditions for both reactions were as follows: 95^{o}C for 5 minutes, followed by 35 cycles of 94^{o}C for 20 seconds, 52^{o}C for 20 seconds and 65^{o}C for 20 seconds.

Labelling of transcripts

Several methods were used to label transcripts.

5' end labelling

It was necessary to dephosphorylate the transcript with Shrimp Alkaline Phospatase (SAP) before kinase labelling. Following ethanol precipitation, the resulting pellet was resuspended in 17 µl water. 2 µl 10x SAP buffer and 1µl SAP (Roche) were added and incubated at 37^{o}C for 1 hour. The reaction was then extracted with acid phenol/chloroform and ethanol precipitated. The RNA pellet was resuspended in 16 µl water, 2 µl 10x buffer, 1 µl T4 polynucleotide kinase (T4 PNK) and 1 µl ^{32}P γATP (specific activity 370Mbq/ml) added. The labelling reaction was incubated at 37^{o}C for 1 hour.

3' end labelling by ligation

Due to the secondary structure of tRNA, 5' end labelling is inefficient. It is necessary to label by a ligation reaction of cytidine-3', 5'-bis-phosphate

(pCp) onto the 3' end of the RNA. The RNA pellet was resuspended in 26 µl water, 4 µl 10x T4 RNA ligase buffer, 3 µl T4 RNA ligase (New England Biolabs), 3 µl 10 mM ATP, 1 µl RNase Out (Invitrogen) and 3 µl ^{32}P pCp (specific activity 370Mbq/ml) were added. The reaction was incubated overnight at 4^{o}C.

Biotin labelling

Transcripts for pull down experiments were biotinylated at the 3' end. A labelling reaction contained 1000 pmols RNA; 10 µl 5x buffer, 2 µl terminal deoxynucleotidyl transferase (TdT) (Invitrogen); 2 µl biotin-16-ddUTP (Roche), and water to 50 µl. The reaction was incubated for 1 hour at 37^{o}C. For pilot EMSA studies, 10% of a MEGAshortscript reaction was biotinylated in a 20 µl reaction with 1 µl each of TdT and biotin-16-ddUTP.

Removal of excess label

Excess radioactive label was removed using G-50 columns (GE Healthcare), and excess biotin using the MEGAclear kit (Ambion). Both methods followed manufacturer's instructions. Radioactively labelled transcripts were stored at -20^{o}C. Biotinylated transcripts were stored at -80^{o}C for short term use, or ethanol precipitated at -80^{o}C for longer storage.

Structural digests

Digests to determine the secondary structure of RNA species followed the method described by Knapp (315). All reagents for these assays were sourced from Ambion. To be successful this technique requires precise size markers. Firstly, size markers included Decade markers, which gave a marker every 10 nucleotides. Secondly, a sequencing ladder using site

specific RNases were used as a control for permitted restriction sites in a fully denatured transcript. Finally, a set of structural digests was performed on the re-conformed RNA, using the RNases in table 8.1.

Transcript preparation

Transcripts were labelled by ligation of ^{32}P pCp onto the 3' end. To ensure a single species, transcripts were further purified by running on the FlashPAGE system as described in Chapter 3. GRE and tRNA transcripts were eluted in the fraction running off with the third dye loading. This could be monitored by the radioactivity of the sample. These fractions were then ethanol precipitated. Samples were resuspended in water for 5 minutes at 65^{o}C.

Alkaline hydrolysis ladder

To make an RNA ladder either the transcript under examination or Decade markers were used as template. For reactions containing transcript, 1 µg transcript, 10 µg yeast tRNA and 1x alkaline hydrolysis buffer to 30 µl were incubated at 95^{o}C and 5 µl aliquots were removed after 3, 5, 6, 7, 8, and 9 minutes. Each aliquot had 10 µl Gel buffer ll added, then put on ice before 3 µl was loaded per track. For Decade markers 3 µl RNA, 10 µg yeast tRNA and 11 µl 1x alkaline hydrolysis buffer were incubated as above, with aliquots removed at 3, 5 and 7 minutes. After addition of 10 µl Gel buffer ll, 7.5 µl was loaded per track.

Sequencing ladder and Structural Digests

The preparation of these two components is similar. It starts with a mix containing multiples of the following 18 µl reaction: 2 µl glycogen (20 mg/ml) (Sigma); 2 µl 10x Structural buffer (Ambion); 0.4 µl (2 µg) yeast

tRNA; 4 µl (1 µg) transcript and 9.6 µl water. For the sequencing ladders, half the RNA mix was denatured at 95^{o}C for 5 minutes, and then placed on ice. Digests were assembled on ice, and then incubated at 50^{o}C for 15 minutes. For the structural digests, the RNA in the other half of the mix was allowed to refold at room temperature for 30 minutes. Digests were set up at room temperature and incubated at 18^{o}C for 15 minutes

For RNaseA digests a dilution series of 1/50, 1/100, 1/500, 1/1000 1/5000 and 1/10000 of the enzyme stock were made. For RNase T1 the dilutions were 1/50, 1/100, 1/1000, 1/5000 and 1/10000. With RNase V1, they were 1/10, 1/100, 1/1000 and 1/10000

Following digestion, 20 µl Gel Buffer ll was added to each reaction. Samples were denatured at 95^{o}C for 5 minutes and 15 µl loaded per track onto denaturing acrylamide gels. Gel mixes contained 6-10% acrylamide, 42% w/v urea and 1x TBE. Gels were run at 1500V in 1x TBE for 2 hours. Gels were dried onto Whatman paper and exposed on autoradiography film.

Structural analysis by GeneRacer

As an alternative to gel analysis of structural digests, digest products were also analysed by GeneRacer cloning (Invitrogen), as described in Chapter 3. Both GC and NR32-1 were tested with RNaseA and RNaseT1 at 1/1000 and 1/5000 dilutions. The two dilutions were chosen as the most informative from the gel based method. By pooling them before cloning, it was hoped to ensure that most cut sites would be included, by one or other dilution. Each 20 µl digest contained 1 µg RNA, 2 µl 10x structural buffer, 2 µl glycogen, 13 µl water, and 2 µl RNase to a final concentration

of 1/1000 or 1/5000. After incubation at 18^{o}C for 15 minutes the 1/1000 and 1/5000 digests were pooled, phenol/chloroform extracted and ethanol precipitated. A 5' phosphate group is required for the ligation of the GeneRacer RNA oligo; this was added in a 50 µl kinase reaction. The RNA was resuspended in 45 µl water and 5 µl 10x buffer, 0.5 µl 10mM ATP and 1 µl T4 PNK (Fermentas) were added. The reaction was then incubated at 37^{o}C for 30 minutes. Following inactivation at 65^{o}C for 20 minutes, the sample was ethanol precipitated. C tail addition, RNA oligo ligation, cDNA synthesis and RT-PCR were as described in chapter 3. The primary RT-PCR products were cloned into pCR4-TOPO and sequenced (chapter 2).

Protein Extracts

Protein extractions were prepared by the method of Voss and colleagues (317) with minor alterations. Asynchronous 3D7 cultures were pelleted as described in chapter 2. Infected red blood cells were lysed with 0.1% saponin in PBS at room temperature for 3 minutes. Parasites were pelleted at 5,000 rcf, 4^{o}C for 4 minutes. The supernatant was removed and the pellet washed in PBS at 4^{o}C, before pelleting again. The washes were repeated 3 times, with the pellets being pooled together. Finally the pellet was resuspended in 1 pellet volume of lysis buffer (20 mM HEPES pH 7.8; 10 mM KCl; 1 mM EDTA; 1 mM DTT, 0.65% IGEPAL and 1x Complete EDTA-free Protease Inhibitors (PI) (Roche)). This was inverted 3 times to mix, then incubated on ice for 5 minutes. At this point, this is a total protein extract.

To separate nuclear and cytoplasmic fractions, the total extract was pelleted in a microfuge at 2,500 rcf, 4^{o}C for 5 minutes. The supernatant, removed to a fresh tube, is the cytoplasmic fraction. The pellet was resuspended in 1 pellet volume of extraction buffer (20 mM HEPES pH 7.8, 800 mM KCl, 1 mM EDTA, 1 mM DTT and 1x PI). The tube was vortexed, then shaken in a Mixmate shaker (Eppendorf) for 30 minutes at 4^{o}C. The tube was then spun in a microfuge at 15,700 rcf at 4^{o}C for 30 minutes. The supernatant is the nuclear fraction and was removed to a fresh tube and mixed 1:1 with storage buffer (20 mM HEPES pH 7.8, 1 mM EDTA, 1 mM DTT and 30% glycerol). All protein extracts were stored in single use aliquots at -80^{o}C.

For some experiments the total protein was “cleared” to remove particulates. To do this the extract was spun in a microfuge at 13.2 Krpm, for 10 minutes at 4^{o}C.

Protein Controls

Two control samples were made. For the uninfected RBC control (URBC), 1 pellet volume RBC was resuspended in 1 volume lysis buffer. For uninfected RBC ghosts (ghosts), the RBC were saponin lysed, and washed twice in PBS. They were then resuspended in 1 volume lysis buffer.

Protein concentration determination

To ensure a standard amount of protein was added to experiments, the concentration of the extracts was determined using a Coomassie Plus Bradford Assay kit (Pierce), following manufacturer’s instructions.

Protein Gels

Proteins were visualized by SDS-PAGE using a modification of the method of Laemmli (318). The quality of extracts was assessed by running 5 µg samples on both 6% and 12% gels. The products from pull-down experiments were visualized on 10% or 12% gels. The composition of the gel mixes are shown in Table 8.2. Solution A comprises 0.48 M HCl, 3 M Tris and 20 mM TEMED. Solution B is 0.48 M HCl, 0.5 M Tris and 41 mM TEMED.

	6%	10%	12%	Stacker gel
Solution A	1.25 ml	1.25 ml	1.25 ml	
Solution B				0.47 ml
Water	2.675 ml	2 ml	1.675 ml	5.62 ml
Protogel 29:1	1 ml	1.66 ml	2 ml	0.31 ml
20% SDS	25 µl	25 µl	25 µl	18.5 µl
10% APS	50 µl	50 µl	50 µl	37 µl

Table 8.2 SDS-PAGE composition

Gel mixes were poured to 4.5 cm depth then covered with overlay solution (1 ml solution A, 3 ml water, 20 µl 20% SDS). When polymerized, the overlay was removed and the stacker gel poured. Samples were mixed 1:1 with gel sample buffer (0.1 M Tris pH 6.8, 1% SDS, 10% glycerol, 0.01% Bromophenol blue, 10% β-mercaptoethanol). Samples were denatured at 100^{o}C for 5 minutes before loading. Gels were run in Glycine

running buffer (25 mM Tris, 0.1% SDS, 192 mM glycine) at 100V for 90 minutes.

Coomassie staining

SDS-PAGE gels were fixed and stained in fixation solution (20% v/v methanol, 10% v/v glacial acetic acid) containing 0.05% Coomassie Brilliant Blue G for 60 minutes then destained overnight in several changes of fixation solution.

Silver staining

For pull-down experiments, a more sensitive staining method that was compatible with mass spectrometry was required. These gels were silver stained with the SilverQuest kit (Invitrogen) following manufacturer's instructions.

EMSAs

Although EMSAs used different buffer conditions, different transcripts and either chemiluminescent or radioactive detection, they all followed the same experimental technique. Transcripts were heat denatured in 1x binding buffer (20 mM HEPES, 1 mM $MgCl_2$, 0.2 mM EDTA, 100 mM KCl, 20% glycerol, with 0.5 mM DTT added before use) then allowed to re-anneal overnight. Unless otherwise indicated all EMSA reactions were at room temperature. To prevent non-specific RNA binding molecules sequestering the GRE transcript, a pre-binding reaction of protein and non-specific competitor yeast tRNA (Ambion) was incubated in 1x binding buffer for 20 minutes. Transcript was then added for a binding reaction, and incubated for a further 20 minutes. Heparin sulphate, another non-specific competitor, was then added to a final concentration of 2.5 μg/μl,

and incubated for 10 minutes. Samples were then diluted 1:1 with EMSA sample buffer (16.6 mM HEPES pH 7.8, 13% glycerol and 0.01% BPB) and placed on ice for 5 minutes before loading onto a gel.

Transcript stability

The stability of an RNA transcript during re-conformation overnight at room temperature was assessed. Biotinylated transcripts, 1% of the labelling reaction per sample, were resuspended in 10 µl Binding buffer. The following conditions were compared; 1. 70^{o}C 5 minutes in a dry heating block, then cool gradually in the turned-off dry block overnight. 2. as 1, but with addition of 0.5 µl RNaseOut (Invitrogen). 3. 70^{o}C 5 minutes, then cool in the dry block for 3 hours. Samples were analysed on a 15% denaturing acrylamide gel.

Biotinylated transcript detection

RNA was capillary blotted onto Hybond N^{+} in 0.75 mM sodium hydroxide. Detection was using the Chemiluminescent nucleic acid detection module kit (Pierce), following manufacturer's instructions.

SRP EMSAs

As a pilot study, SRP was chosen for RNA EMSA work. Biotinylated transcripts were used initially. Unincorporated biotin was removed by phenol/chloroform extraction and ethanol precipitation with 3 volumes ethanol. Each reaction contained 1/200 of a transcript labelling reaction. Transcript denaturing was at 70^{o}C for 5 minutes. Pre-binding reactions of 5 µl contained 5 µg proteins, 5 µg tRNA and 1x binding buffer. Binding reactions had 5 µl transcript added. Heparin sulfate (25 µg) was then added to reduce non-specific binding. Samples were analysed on 2%

agarose gels with addition of 5% glycerol,in 0.5x TBE. Gels were run on ice at 50V for 3-4 hours then capillary blotted overnight.

GRE EMSAs

Although Black et al suggest 50 fmols of transcript per reaction (319), no binding was visible with this (data not shown). It was decided to increase this to 1000 fmols SAP treated, kinase labelled GRE transcripts, per reaction. GRE transcripts were denatured at 95^{o}C, before re-annealing. Many different additions to the binding conditions were investigated, as suggested by Black et al [in ref (319)] . A standard 10 µl pre-binding reaction contained 40 µg protein extract, 5 µg tRNA, 5 mM $MgCl_2$ and 1x binding buffer. For binding reactions, 10 µl of transcript was added. Finally, 50 µg heparin sulphate was added.

Samples were analysed on 14x16 cm gels containing 6% acrylamide, 5 mM $MgCl_2$, 5% glycerol, 0.001% TEMED, 1.6% APS and 0.5x TBE. Gels were run in 0.5x TBE, 5 mM $MgCl_2$ at 4^{o}C for 250-300V for 2-3 hours. Gels were dried for 2 hours on a BioRad gel dryer, prior to autoradiography.

Determination of GRE location

To determine the location of the GRE within the cell it was decided to make protein extracts of the cellular fractions, and then perform an RNA extraction with Trizol on them. The rationale was that if GRE was binding to proteins, it would be located in the same compartment as the proteins. This experiment was also designed to check if pre-clearing the total protein extract was removing the GRE. It was vital to keep the cell numbers equivalent at each Trizol pellet. Each pellet was 50 µl, to which was added 500 µl of Trizol and 1 µl glycogen (20 mg/ml). RNA extraction

and analysis by northern blot on a 2% agarose gel continued as in chapter 2. The hybridization probe was a ^{32}P end labelled oligonucleotide, GCB450. Hybridization was at 37^{o}C overnight, with washes at 37^{o}C for 1 hour in 0.5x SSC. A flow chart of the experimental design is in Figure 8.2.

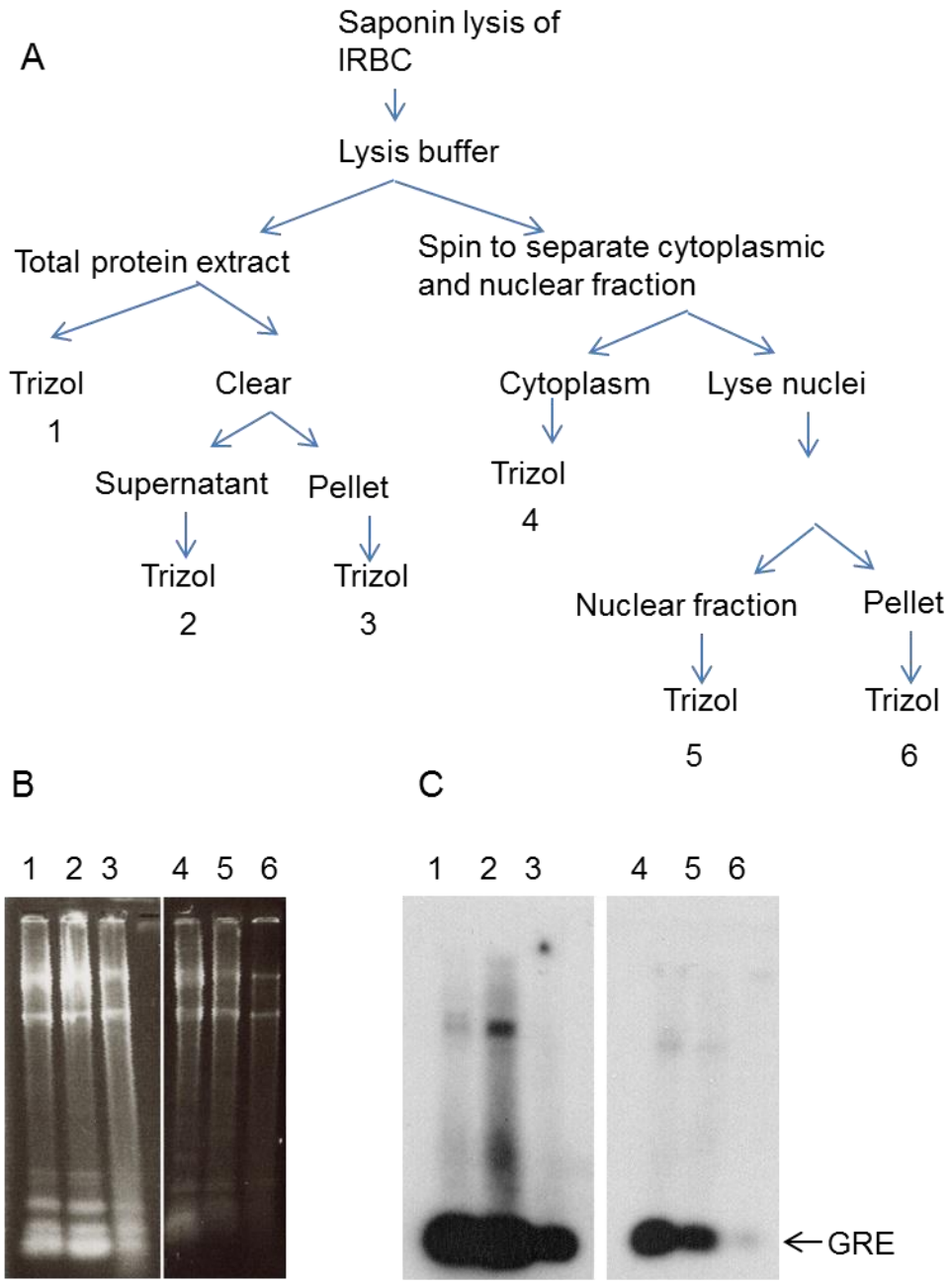


Figure 8.2 GRE RNA locations.
GRE RNA is located mainly in the cytoplasm, with some remaining in the nucleus. A. Experimental design, explaining location of RNA samples run on gel shown in B. B. Northern of samples shown in A, stained with ethidium bromide. C. Autoradiograph of B, hybridized with GCB450. Samples 1-6 were from equivalent cell pellet volumes.

PULL-DOWNS

Two methods were used to perform protein pull-down experiments. The first compared the protein bands obtained from three different reactions; beads plus either GCF, GCS, or GCF plus RNase. The second method was a sequential pull down. Here, protein extracts were pre-cleared with beads, then with beads plus GCS and finally, proteins bound to beads plus GCF were recovered. All incubations and washes involving Dynabeads were performed with gentle rotation.

Comparative pull-down.

This method was performed with cleared total parasite lysate. A control sample of lysate was treated with RNase A (Invitrogen) at 1 µg/µl, for 10 minutes at room temperature. Binding conditions were those which had been previously determined by EMSA. A 250 µl reaction contained 1 mg protein extract, 12.5 µl tRNA (10 mg/ml), 5mM $MgCl_2$, 5 µl RNaseOUT (Invitrogen), 1x PI and 1x binding buffer. Pre-binding was at room temperature for 20 minutes. Re-natured biotinylated transcript was added at 8.8 µg per reaction and incubated for 30 minutes at room temperature. Finally 12 µl heparin sulphate (50 mg/ml) was added for 10 minutes at room temperature. To bind the transcripts to magnetic beads, 1mg M-280 Streptavidin Dynabeads (Beads) (Invitrogen) per reaction were treated for RNA work as per manufacturer's instructions. Beads and binding reactions were incubated with gentle shaking at 4^{o} C for 1 hour. The beads were then washed 4 times for 5 minutes at 4^{o}C in 1x binding buffer. Protein was eluted off the beads in 20 µl SDS sample buffer at 95^{o}C for 5 minutes. Samples were visualised on 10% polyacrylamide gels as described above.

Sequential pull-down

Cytoplasmic and nuclear extracts were compared with this method. The reactions were as described for comparative pull downs. Beads were treated for RNA work as per manufacturers instruction then blocked with yeast tRNA for 1 hour at 4^{o}C (10 mg beads, 1x binding buffer, 500 µg tRNA in 500 µl). Beads were washed once in 1x binding buffer. Biotinylated transcripts were bound to beads in 1x binding buffer at 4^{o}C for 30 minutes, and then washed twice with 1x binding buffer. Each binding reaction was incubated sequentially with the following: 1mg beads without transcript at 4^{o}C, 1 mg beads plus GCS at 4^{o}C, 1 mg beads plus GCS at room temperature, then finally 2 mg beads plus GCF at room temperature. Each binding stage was for 30 minutes, and then following bead removal by magnet, the supernatant was added to the next set of beads. The supernatant was discarded after binding to GCF. The beads were then washed 6 times for 5 minutes at 4^{o}C in 1x binding buffer. Sample elution and visualisation was as for comparative pull downs.

Proteomic analysis on excised gel bands was performed by the Oxford University Central Proteomics Facility using their in-gel digestion protocol, with samples analysed on a nano-LC-MS/MS system (Waters Q-Tof) (320).

Results

Structural Digests

Initially, the structural digests on the GRE were attempted with 5’ labelled transcripts. As these results were impossible to interpret, it was decided to pilot the method on an RNA species with a known structure. To this end it

was decided to use tRNA-glu as this RNA had been isolated and cloned in chapter 3. The structure of tRNA-glu as detailed on the Genomic tRNA database is shown in Figure 8.3A (321). Knapp (315) recommends 3' labelling with pCp for tRNA structures. As this was an alteration to the method, both GRE transcripts GC and NR32-1 were also repeated using pCp labelling. The results of RNase T1 and A digests on tRNA-glu are shown in figure 8.3b. The results for RNase V1 have been omitted as no sequence or structural digest bands were visible with this enzyme. It is assumed that this is due to inactive enzyme, as digest sites would be expected with tRNA of a known structure. RNase A cuts 3' of single stranded C or U and RNase T1, 3' of single stranded G. The potential cut sites that should be visible in the sequencing digests can be determined from figure 8.3A. Although the Decade markers are shown, estimation of band sizes was difficult due to the "smile". At least four predicted cut sites are not present in the sequence digest with RNase A, and three with RNase T1. Conversely, we find three unpredicted sites, one in the sequencing and two in the structural digest, with RNase T1. Although not conclusive, the combined data support the cloverleaf tRNA structure. The G rich loop at 15-18 is predicted by the RNase T1 data, and the C/U rich loops at 32-38, and 53-59 are predicted by that of RNaseA.

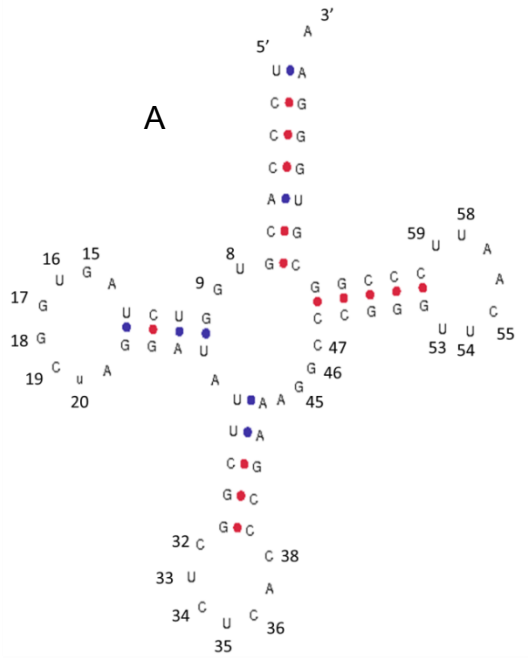
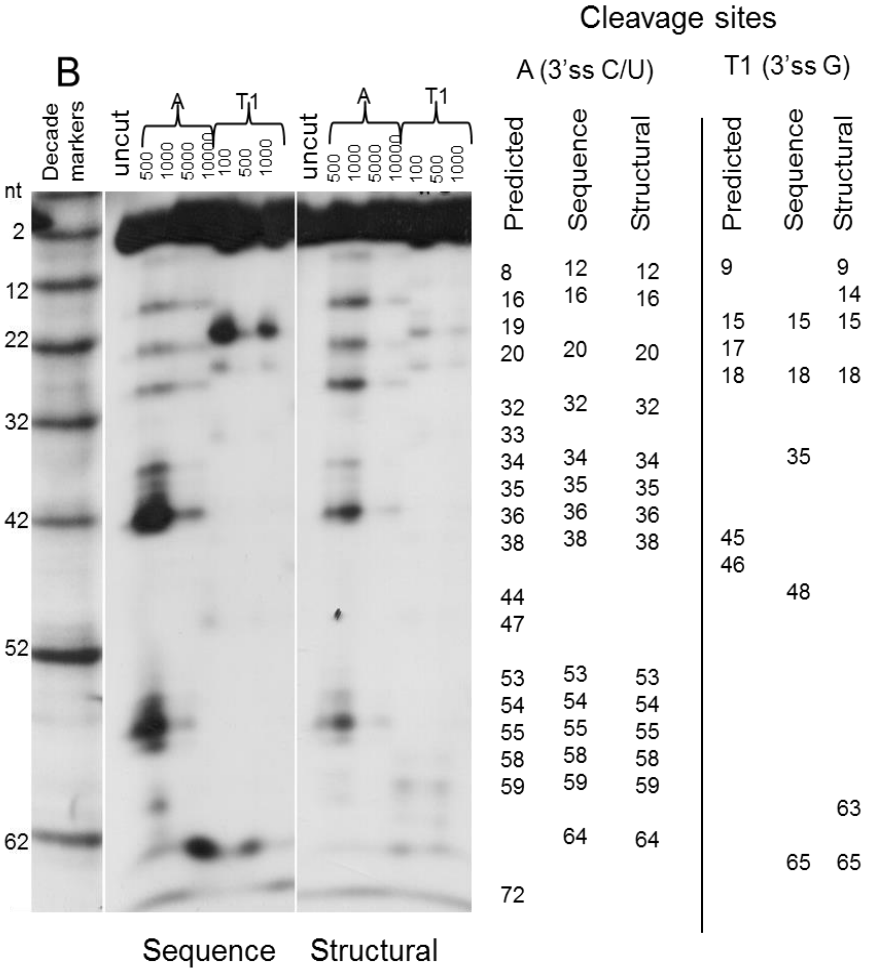


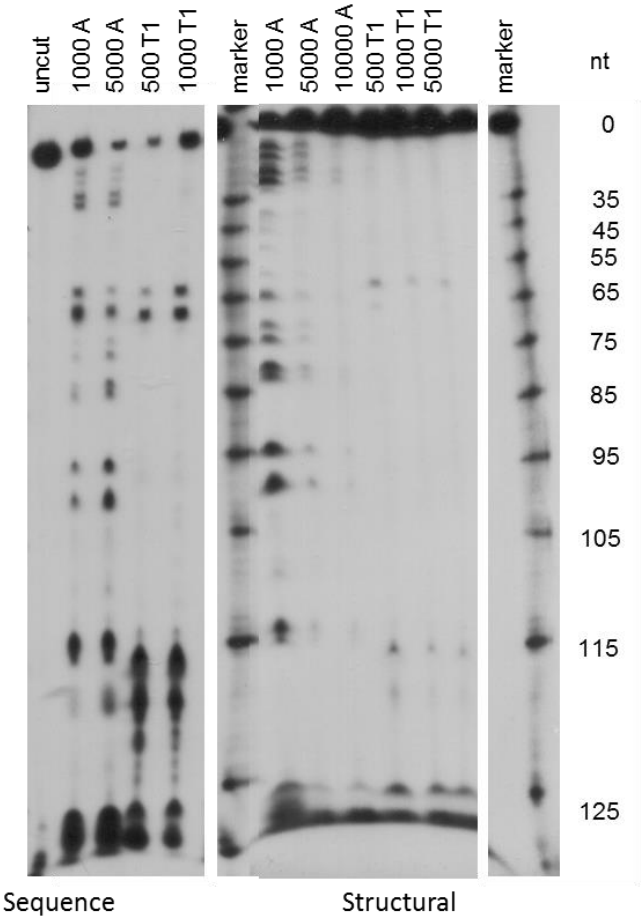
Figure 8.3 tRNA-Glu structural studies. Results of RNase digestion on 3' labelled tRNA-Glu. A. Predicted structure of tRNA-Glu . Sites of predicted structural cuts are numbered. B. Sequence and structural digest with RNase A and T1.

Fragments separated on 8% acrylamide gel. For 3' labelled RNA, cut sites closest to 3' end are shortest therefore at the lower end of the gel. Predicted sites are from A. Sequence sites are digested on linear RNA. Structural sites are from digestion of re-conformed transcript.



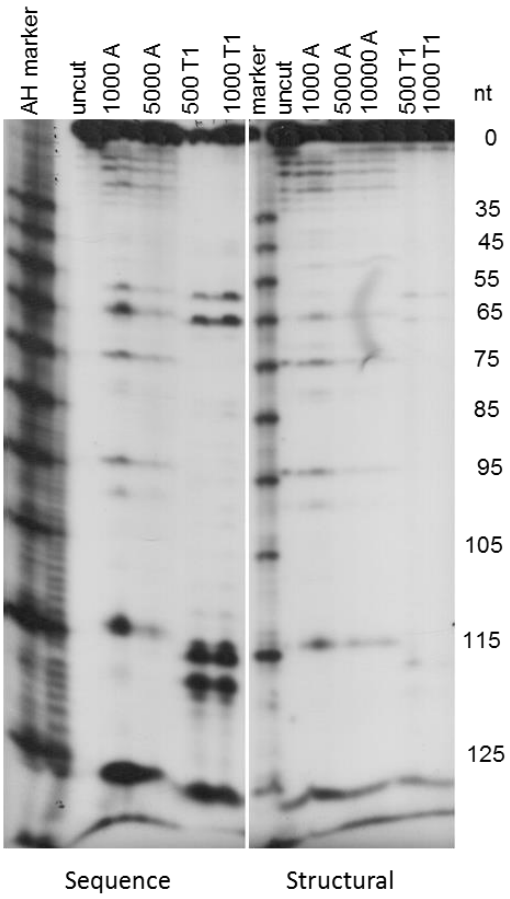
The data for the GRE transcripts NR32-1 and GCFoot are presented in Figures 8.4 and 8.5. Obviously, the sequencing reactions have not been successful. We only see 20%-26% of the predicted cut sites in the GRE transcripts with RNase T1 and 20%- 33% with RNase A. The sequencing digests rely on the transcript remaining single stranded during digestion to cut correctly. It is possible with both the GRE and tRNA that the transcript snaps back into its folded state as soon as it is removed from ice, thus preventing full digestion in a single stranded form. The structural digestion sites obtained with RNases A and T1 are placed on the mfold predictions of the GRE structures in figure 8.6. RNase V1 cuts at double stranded nucleotides so would indicate which parts were paired and therefore not available for digestion with the other enzymes. Without results for this enzyme, it is impossible to conclusively determine a structure for the GRE RNAs. However, we find that with both structures the two GC pairings in a small stem region directly below the tetraloop are digested with enzymes that only cut single stranded nucleotides.

At this point it was decided to try a modification of the GeneRacer technique that had been successful in cloning small ncRNA. Alignments of the sequences of the resulting bacterial clones after digests of both GC and NR32-1 with both RNase A and RNaseT1 are shown in figure 8.7a-d. RT-PCR with both 5’ and 3’ GeneRacer oligos on the NR32-1 cDNA was unsuccessful, so the data presented in figure 8.7a and b are from RT-PCR reactions using the 5’ GeneRacer oligo with gcB450 and the 3’ GeneRacer oligo with gcT335. Although this does not identify RNA with cut sites at both ends of the transcripts, it does identify those with digestion sites at



Known G residues from 5' end	3 6 12 15 20 23 27 28 31 32 45 47 49 53 54 56 57 61 67 71 87 92 95 102 104 107 110 112 119 120 122 123 124 125 126 127 129 130 132
G cut sites in sequencing digests	64 66 115 116 117 119 120 121 122 123 124 125 129
G cut sites in structural digests	64 66 115 117 119 126
Known c/u residues from 5' end	4 5 7 8 9 10 13 16 17 18 21 22 24 25 29 33 36 37 38 39 40 41 42 43 44 46 52 54 55 58 59 60 63 64 65 66 68 69 70 72 73 76 79 80 81 82 84 93 94 96 100 101 105 106 108 109 113 114 116 117 118 128 130 132
C/U sites in sequencing digests	(4 in 0-35 size cannot be determined) 63 65 73 75 81 82 95 103 114 115 118 126 128
C/U sites in structural digest	(4 in 0-35 size cannot be determined) 40 45 50 60 65 73 75 81 82 95 103 114 115 126 128

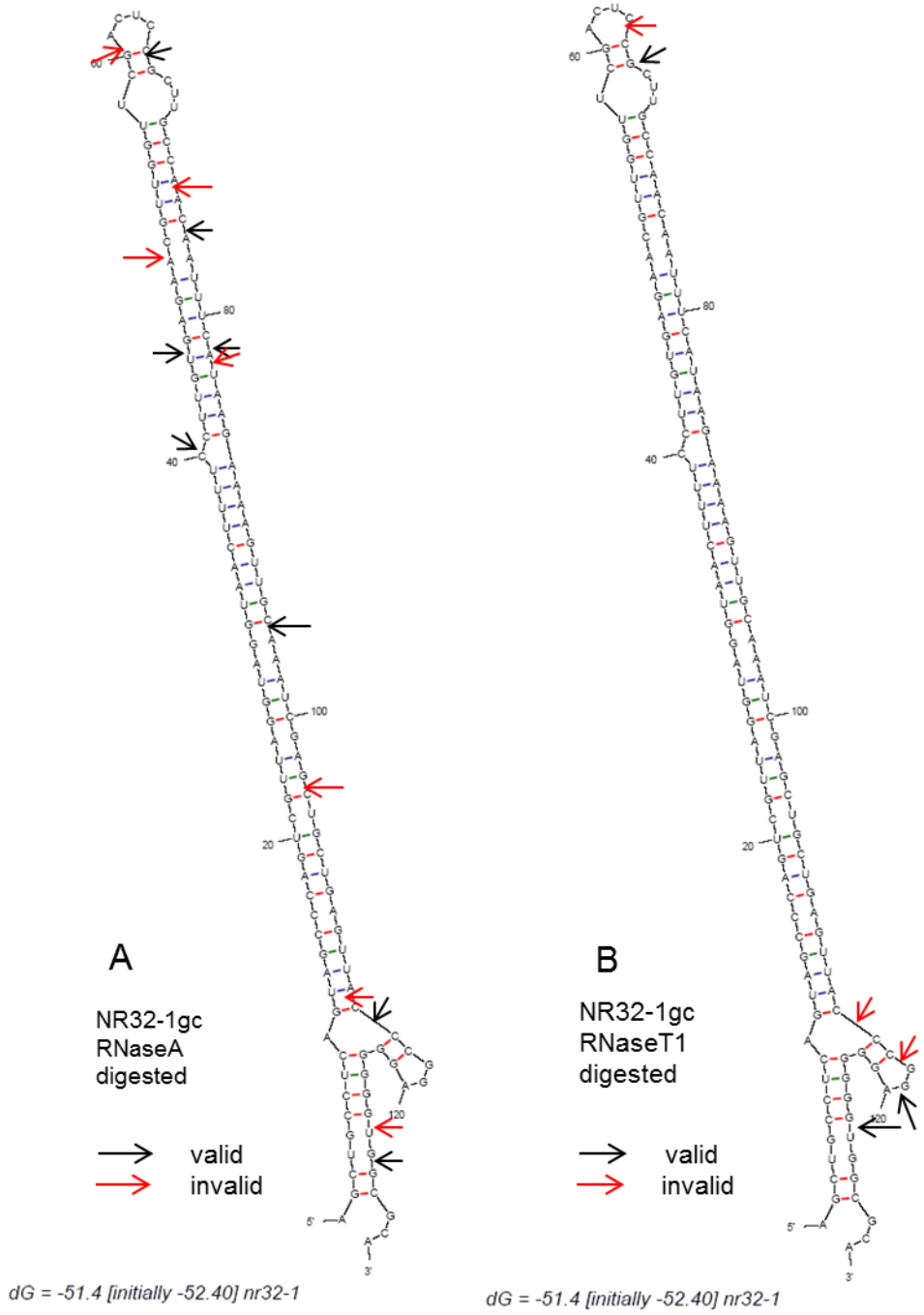
Figure 8.4 Structural studies of GRE NR32-1
Structural and sequencing digests of GRE NR32-1. Results are shown for RNAse A, T1 and V1. The anticipated cut sites as determined from the sequence are shown in the table, along with the sizes of the sequence and structural digests. Fragments are separated on 8% acrylamide gel. Markers are Decade markers (Ambion).

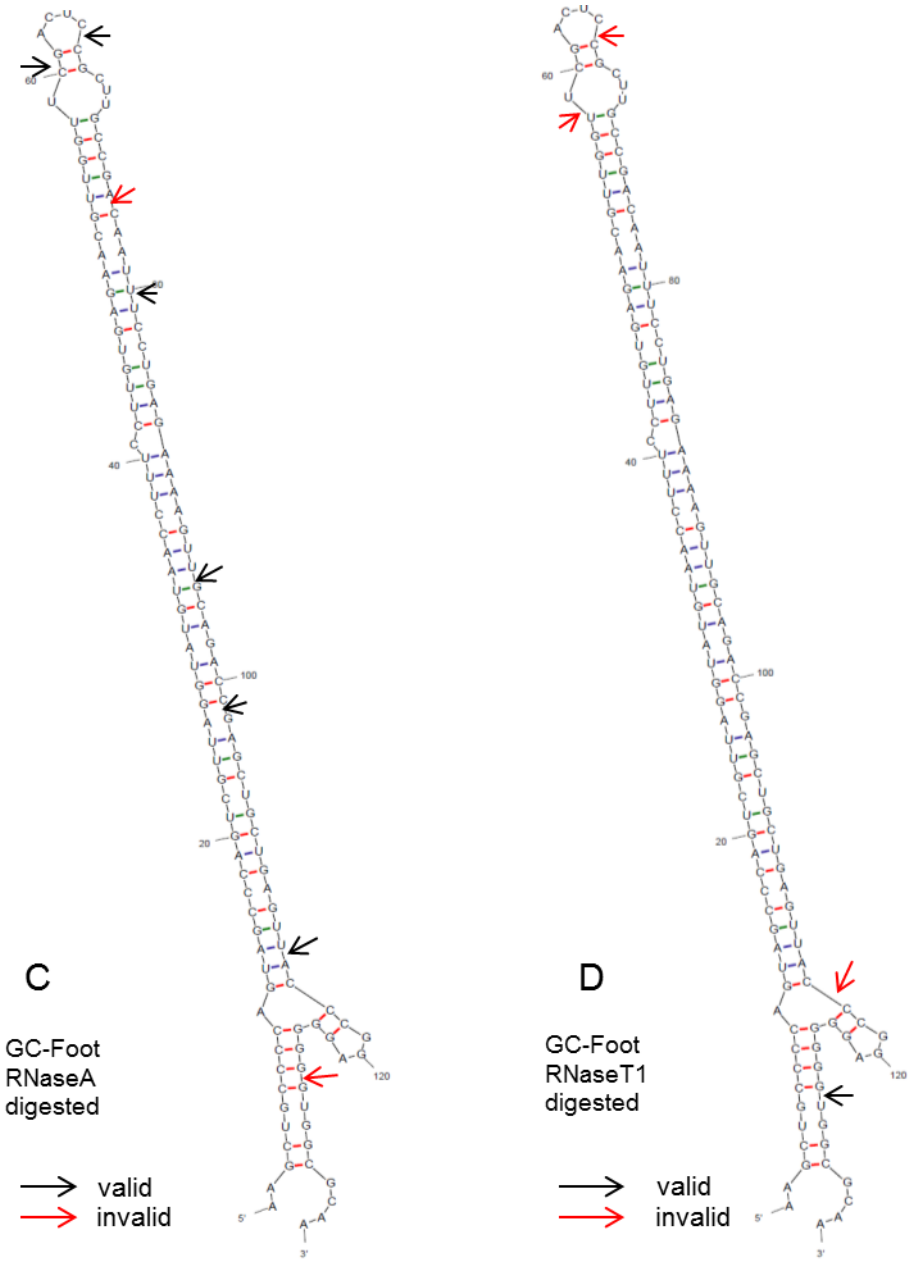


Known G residues from 5' end	3 6 12 15 20 23 27 28 32 45 47 49 53 56 57 61 67 71 74 85 87 92 95 98 102 104 107 110 112 119 120 122 123 124 125 126 127 129 130 132
G cut sites in sequencing digests	58 65 117 118 120 121 122 127
G cut sites in structural digests	(1 in 0-35 size cannot be determined) 58 65 116 127
Known c/u residues from 5' end	4 5 7 8 9 10 13 16 17 18 21 22 24 25 29 31 33 36 37 38 39 40 41 42 43 44 46 52 54 55 58 59 60 63 64 65 66 68 69 70 72 73 76 79 80 81 82 84 93 94 96 100 101 105 106 108 109 113 114 116 117 118 128 131 133
C/U sites in sequencing digests	(6 in 0-35 size cannot be determined) 60 65 75 94 101 114 126
C/U sites in structural digest	(6 in 0-35 size cannot be determined) 60 65 75 80 94 101 114 126

Figure 8.5 Structural studies of GRE GC.
Structural and sequencing digests of GRE GC. Results are shown for RNAse A and T1. The anticipated cut sites as determined from the sequence are shown in the table, along with the sizes of the sequence and structural digests. Fragments are separated on 8% acrylamide gel. Markers are Decade markers (Ambion) and AH markers are an alkaline hydrolysis reaction of Decade markers.

Figure 8.6 RNase digestion sites of GRE. The digestion sites obtained using either RNaseA (cuts 3' of single stranded C and U) or RNaseT1 (cuts 3' of single stranded G) have been indicated on the mfold predictions of GRE structure. A. NR32-1 digested with RNaseA. B. NR32-1 digested with RNaseT1. C. GC digested with RNaseA. D. GC digested with RNaseT1. Valid cut sites that follow the published rules for the enzymes are shown in black. Invalid cuts i.e. after residues that the enzyme is not expected to cut at, are shown in red.





only one end of the transcript. Theoretically, by examining enough clones, both methods should reveal the same digestion sites.

```
                *       20        *       40        *       60        *       80        *      100        *      120        *
2A-5'    : ----------------------------------------------------------------------------------ATAAGAAAAGTTGCAAATCGAGCTGCTGAGTTACCCGGAGGGGGGCGGCGCAA :  53
3A-5'    : --------------------------------------------------------------------------------AATTTCATAAGAAAAGTTGCAAATCGAGCTGCTGAGTTACCCGGAGGGGGGTGGCGCAA :  59
4A-5'    : ------------------------------------------------------------------CCGCTTGCCAACAATTTCATAAGAAAAGTTGCAAATCGAGCTGCTGAGTTACCCGGAGGGGGGCGGCGCAA :  71
7A-5'    : ----------------------------------------------------------GTTCGACTCCGCTTGCCAACAATTTCATAAGAAAAGTTGCAAATCGAGCTGCTGAGTTACCCGGAGGGGGGCGGCGCAA :  79
9A-5'    : ----------------------------------------------------------GGTTCGACTCCGCTTGCCAACAATTTCATAAGAAAAGTTGCAAATCGAGCTGCTGAGTTACCCGGAGGGGGGCGGCGCAA :  80
6A-5'    : ----------------------CGTTAGGTAGGTAACTTTTCCTTGTGAGAACGTTGGTTCGACTCCGCTTGCCAACAATTTCATAAGAAAAGTTGCAAATCGAGCTGCTGAGTTACCCGGAGGGGGGTGGCGCAA : 114
12A-5'   : ------GCCCCAGTAGTCCAGTCGTTAGGTAGGTAACTTTTCCTTGTGAGAACGTTGGTTCGACTCCGCTTGCCAACAATTTCATAAGAAAAGTTGCAAATCGAGCTGCTGAGTTACCCGGAGGGGGGCGGCGCAA : 130
NR32-1gc : --AGCTGCCTCAGTAGCCCAGTCGTTAGGTAGGTAACTTTTCCTTGTGAGAACGTTGGTTCGACTCCGCTTGCCAACAATTTCATAAGAAAAGTTGCAAATCGAGCTGCTGAGTTACCCGGAGGGGGGTGGCGCA- : 133
4A-3'    : --AGCTGCCCCAGTAGCCCAGTCGTTAGGTAGGTAACTTTTCCTTGTGAGAACGTTGGTTCGACTCCGCTTGCCAACAATTTCATAAGAAAAGTTGCAAATCGAGCTGCTGAGTTA------------------- : 114
16xA-3'  : --AGCTGCCTCAGTAGCCCAGTCGTTAGGTAGGTAACTTTTCCTTGTGAGAACGTTGGTTCGACTCCGCTTGCCAACAATTTCATAAGAAAAGTTGCAAAAACA------------------------------ : 102
9xA-3'   : --AGCTGCCTCAGTAGCCCAGTCGTTAGGTAGGTAACTTTTCCTTGTGAGAACGTTGGTTCGACTCCGCTTGCCAACAATTACATAAGAAAAACCAAG---------------------------------- :  96
18A-3'   : ---GCTGCCTCAGTAGCCCAGTCGTTAGGTAGGTAACTTTTCCTTGTGAGAACGTTGGTTCGA---------------------------------------------------------------- :  60
33A-3'   : --AGCTGCCCCAGTAGCCCAGTCGTTAGGTAGGTAACTTTTCCTTGTGAGAACGTTGGTTCGT---------------------------------------------------------------- :  61
15A-3'   : --AGCTGCCCCAGTAGCCCAGTCGTTAGGTAGGTAACTTTTCCTTGTGAGAACGTTGGTTC------------------------------------------------------------------ :  59
12xA-3'  : --AGCTGCCCCAGTAGCCCAGTCGTTAGGTAGGTAA-------------------------------------------------------------------------------------------- :  34
20A-3'   : --AGCTGCCCCAGTAGCCCAGTCGTTAGGTAGGTAAA------------------------------------------------------------------------------------------- :  35
```

Figure 8.7a GRE NR32-1 digested with RNaseA. Each sequence in the alignment was obtained from a single bacterial clone following modified GeneRacer cloning of the NR32-1 transcript digested with RNaseA. RT-PCR used the following pairs of oligos, GeneRacer 5'/gcB450 (5') and GeneRacer 3'/gcT335 (3'). RNaseA cuts 3' of single stranded C or U (T in DNA sequence) residues.

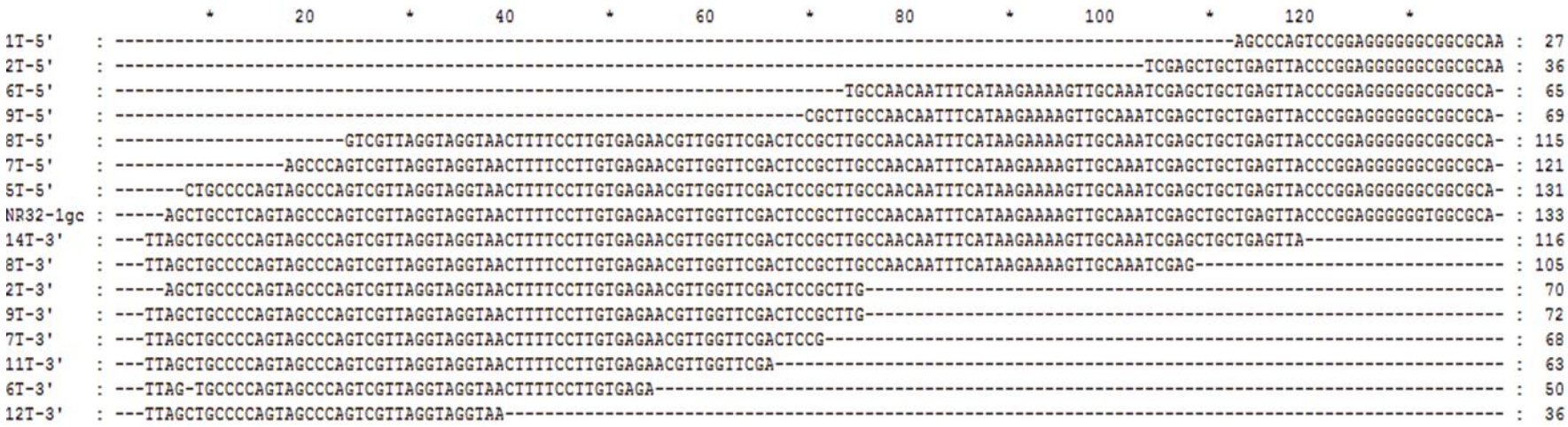


Figure 8.7b GRE NR32-1 digested with RNaseT1.
Each sequence in the alignment was obtained from a single bacterial clone following modified GeneRacer cloning of the NR32-1 transcript digested with RNaseT1. RT-PCR used the following pairs of oligos, GeneRacer 5'/gcB450 (5') and GeneRacer 3'/gcT335 (3'). RNaseT1 cuts 3' of single stranded G residues.

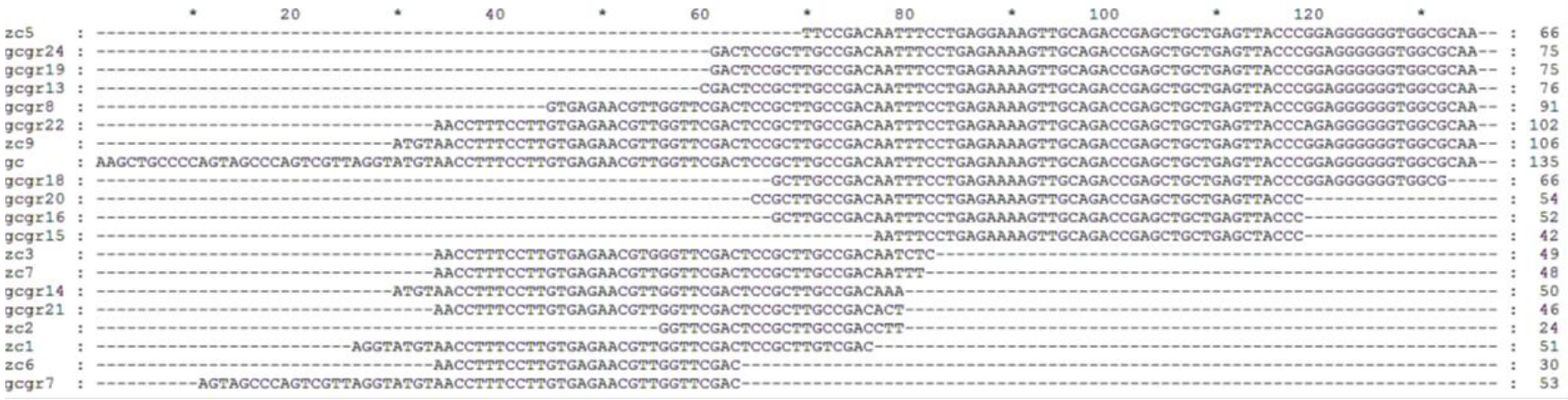


Figure 8.7c GRE GC digested with RNaseA.
Each sequence in the alignment was obtained from a single bacterial clone following modified GeneRacer cloning of the GC transcript digested with RNaseA. RNaseA cuts 3' of single stranded C or U (T in DNA sequence) residues.

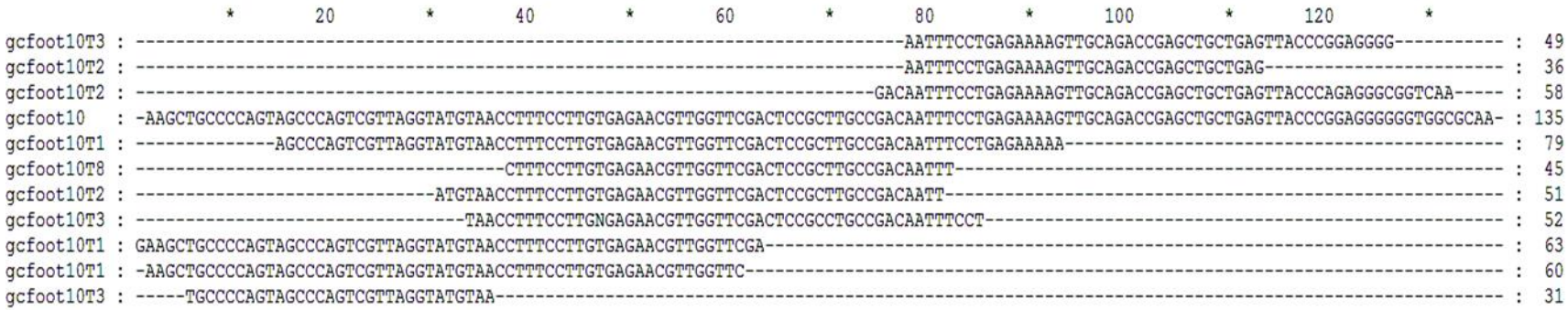


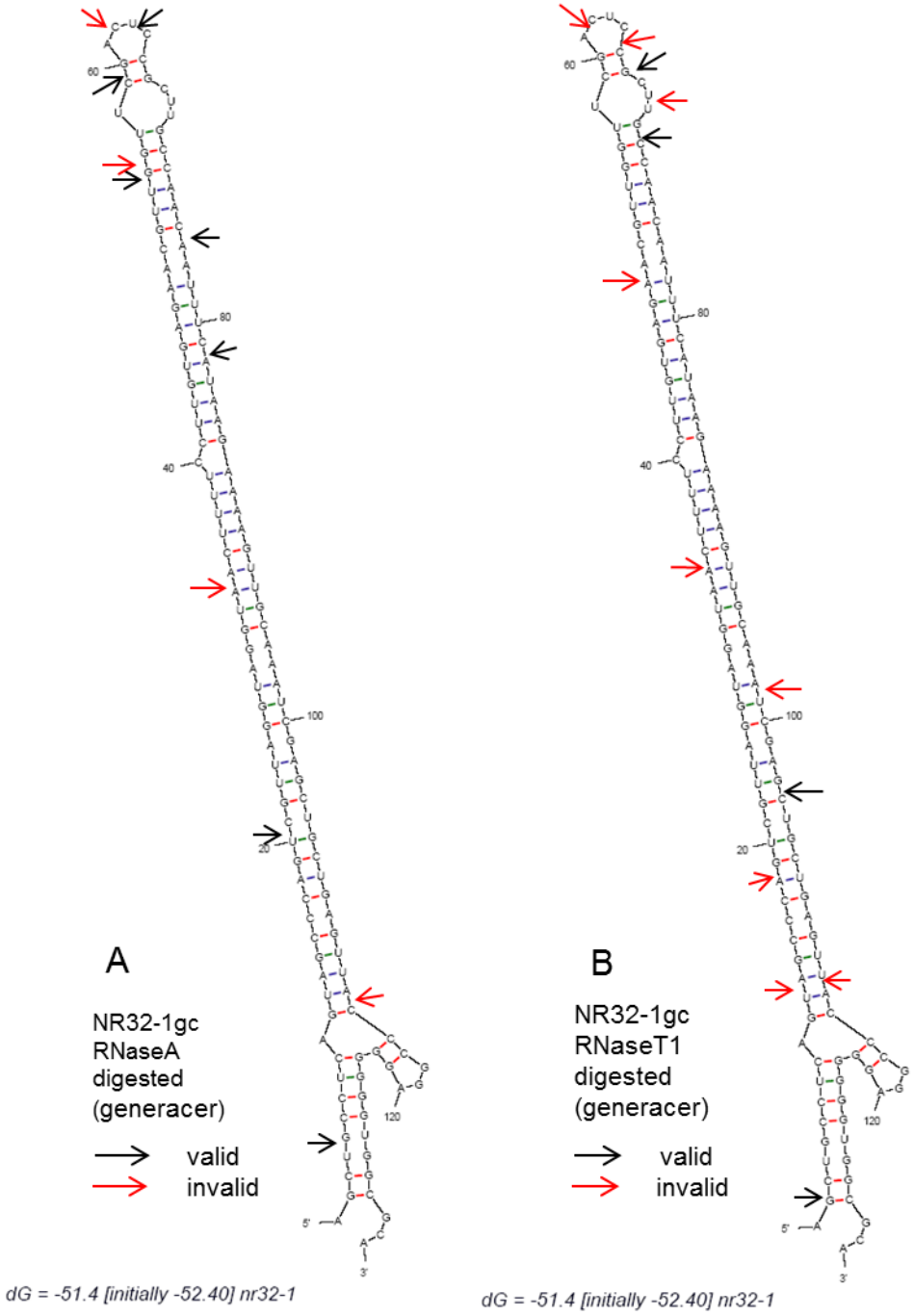
Figure 8.7d GRE GC digested with RNaseT1.
Each sequence in the alignment was obtained from a single bacterial clone following modified GeneRacer cloning of the GC transcript digested with RNaseT1. RNaseT1 cuts 3' of single stranded G residues.

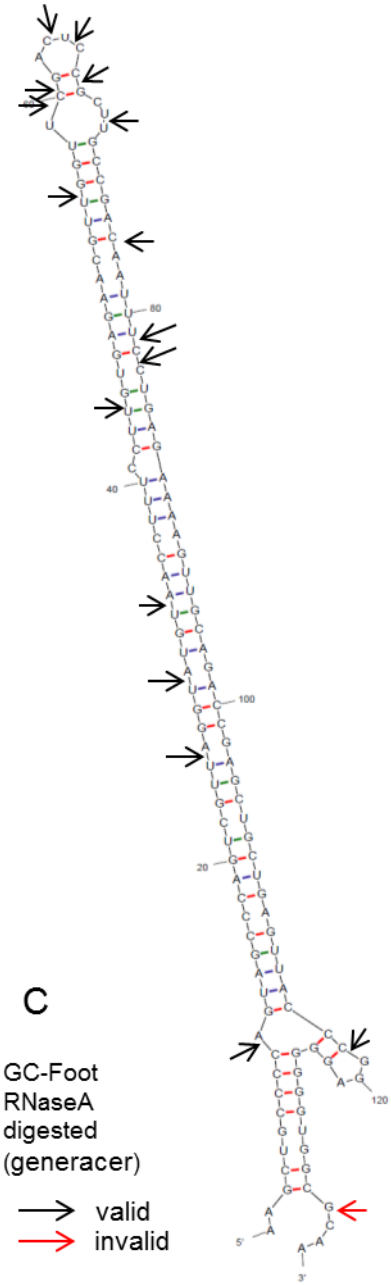
Placement of the digestion sites on mfold figures of the transcripts are shown in in figure 8.8. Whilst some of the cut sites do not appear to be at single stranded cut sites, this is not a concern, especially when they are non-Watson/Crick base pairs. The mfold structure is only a prediction, which we wish to either confirm or alter, so these cut sites could be valid. The data do support a much larger loop at the hairpin as determined from the electrophoresis analysis. Placing the constraints of single stranded nucleotides from 59-70 into an mfold prediction of GC, results in a hairpin with a larger loop, figure 8.9. The placement of other digestion sites as constraints in mfold, lead to a range of predicted structures. A major issue is the number of cut sites, particularly with RNaseT1, that do not follow the published digest rules, it is difficult therefore to have confidence in the data. As with the gel electrophoresis analysis, RNaseV1 information would determine which regions are double stranded. This information would allow us to confirm that the digestion sites we see with RNases A and T1 are correct.

EMSAs

Having looked at structure, we considered what potential functions the GRE could have within the cell. Could we demonstrate that the GRE transcript forms a complex with cellular proteins? To do this I used the EMSA technique. Protein extracts were made and incubated with an *in vitro* labelled GRE transcript. The reaction mixes were then run on a native acrylamide gel using conditions designed to keep any RNA:protein complex associated during the electrophoresis.

Figure 8.8 RNase digestion sites of GRE determined by GeneRacer Cloning. The digestion sites obtained using either RNaseA (cuts 3' of single stranded C and U) or RNaseT1 (cuts 3' of single stranded G) have been indicated on the mfold predictions of GRE structure. A. NR32-1 digested with RNaseA. B. NR32-1 digested with RNaseT1. C. GC digested with RNaseA. D. GC digested with RNaseT1. Valid cut sites that follow the published rules for the enzymes are shown in black. Invalid cuts i.e. after residues that the enzyme is not expected to cut at, are shown in red.




C
GC-Foot
RNaseA
digested
(generacer)
valid
invalid

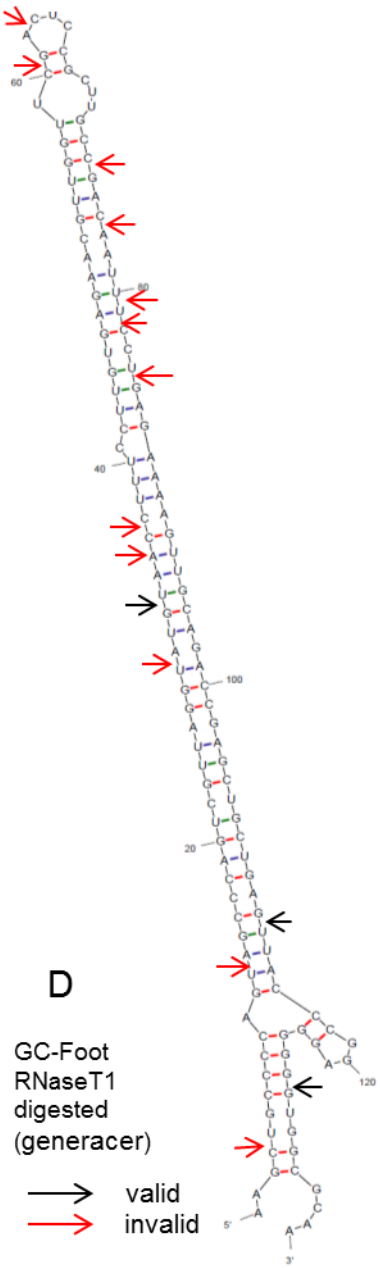
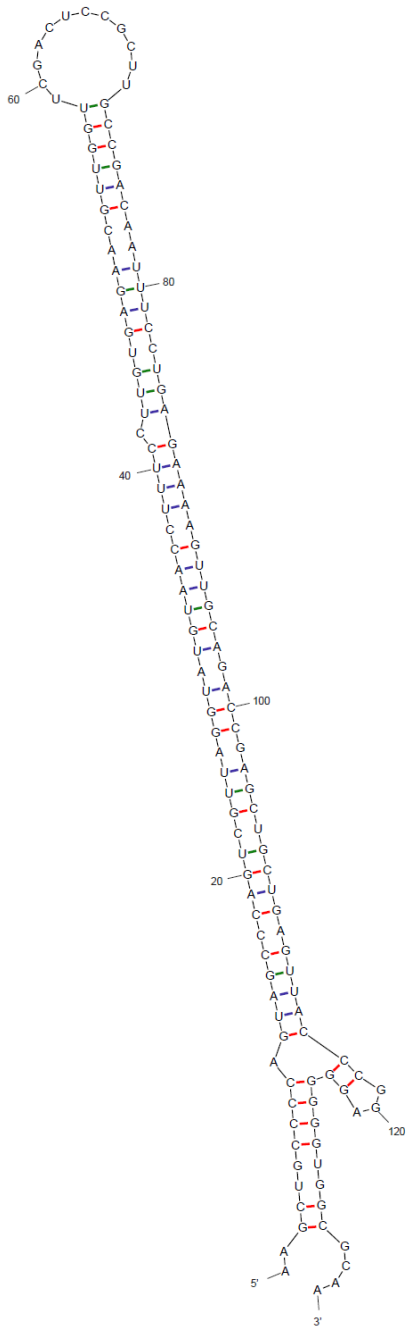

D
GC-Foot
RNaseT1
digested
(generacer)
valid
invalid

Figure 8.9 Mfold structure prediction of GC with constraints at nucleotides 59-70 results in a larger loop



Transcript Stability

EMSA methods require a denaturing of the DNA or RNA, then a slow cooling to allow reannealing. The stability of RNA transcripts overnight whilst cooling to room temperature was assessed and is shown in Figure

8.10. This shows that little or no degradation of the transcript occurred, and RNase inhibitors were unnecessary.

SRP EMSAs

SRP was chosen as the transcript for pilot studies using RNA EMSAs as the protein components of the SRP RNP complex have been identified (313,314). With known binding, a shift should be readily detectable. The EMSA for SRP, showing a shift in the cytoplasm protein fraction, is in Figure 8.11. Owing to the size of the SRP transcript, the EMSA was run in agarose gels.

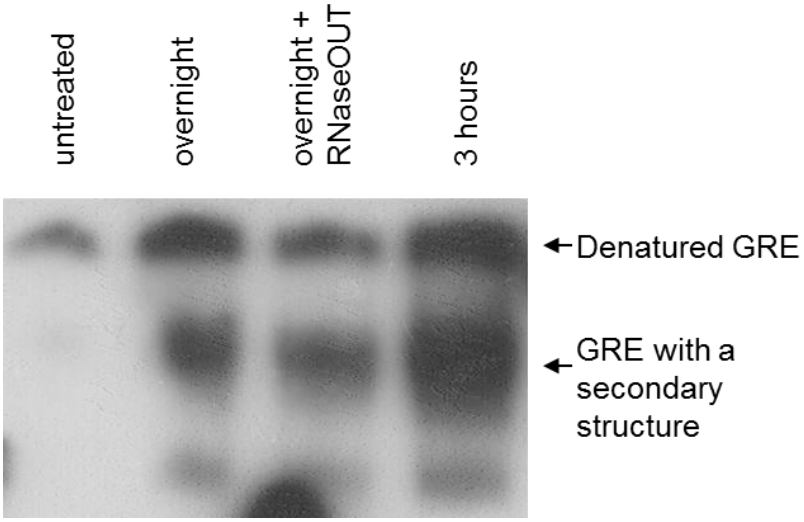


Figure 8.10 Transcript stability**.**
The untreated sample was denatured before loading and shows the full length GRE transcript. The other samples were denatured at 70°C for 5 minutes before being left to cool slowly for the times indicated. This leads to the formation of secondary conformation as indicated. No degradation of the transcript was observed.
Biotinylated transcripts were analysed on 15% denaturing acrylamide gel. RNA was then capillary blotted and detected with a chemiluminescent nucleic acid detection kit.

GRE EMSAs

Several attempts were made to visualize a gel shift, using a biotinylated GRE transcript. GRE binding reactions were run on 6% acrylamide/5% glycerol mini gels. GRE EMSAs and SRP EMSAs were run alongside, using identical protein extracts. It was only possible to detect a gel shift with the SRP transcript (data not shown). This excluded general experimental conditions and the protein extracts as the cause of lack of binding observed for GRE. It was therefore assumed that either the GRE did not have a protein binding capability, or sensitivity of the assay must be preventing detection. It was therefore decided to move to using radiolabelled transcripts.

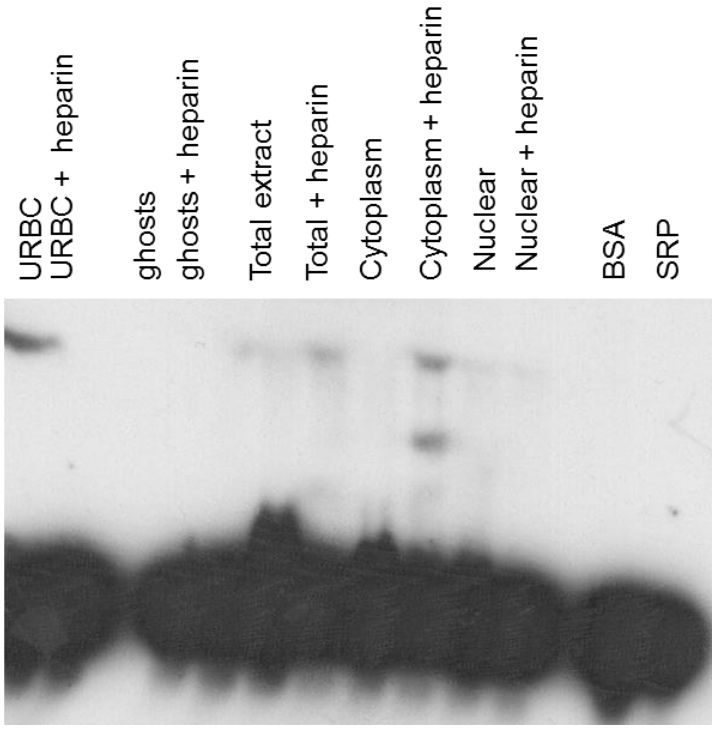


Figure 8.11 SRP EMSA
EMSA gel showing band shift of SRP transcript due to binding of cytoplasmic protein. For experimental conditions see text. BSA was included as a negative control for binding.

Initial experiments with radioactively labelled transcripts were designed to investigate the amount of protein added to the pre-binding reactions. SRP EMSAs had 10 µg extract added per reaction, but Black et al recommends 10-40 µg, so this range was tested (319). The resulting EMSA in Figure 8.12 shows that little binding is detectable with only 10 µg total protein extract. Increasing the protein concentration leads to increased binding, and from here on 40 µg was used as the standard concentration per GRE EMSA. Figure 8.12 shows that there is no binding of GRE in either URBC, or RBC ghosts; this indicates that the GRE binding is solely due to parasite proteins. Addition of heparin has no effect on binding in this EMSA, but as it was not detrimental, it was continued. It was suggested that bromophenol blue may inhibit some protein binding reactions (319). Although bromophenol blue did not abrogate binding, it did reduce it. Bromophenol blue was omitted from EMSA sample buffer in further experiments.

Having established GRE specific protein binding, further experiments were conducted to optimise binding conditions. The effect of 0-50 mM $MgCl_2$ concentration on GRE:protein binding is shown in figure 8.13. As 5 mM $MgCl_2$ gave the maximum number of bands, this was used for further studies. In this experiment, omission of heparin has given less overall binding. The effect of salts was investigated, figure 8.14. Addition of 5 mM $MgCl_2$ and100 mM KCl gives the same main band; however, with tRNA added, the banding pattern alters. It is likely that the common band is due to non-specific binding, and that addition of tRNA reduces this. A complex that has been retained in the wells is visible in the EMSA figures. This large complex was visible in EMSAs with GCF, GCR or GCS, figure

8.15a, and also when GCF was incubated with total, nuclear or cytoplasmic extract (data not shown). To further investigate the origins of the complex, DNase and/or RNase was added to the binding reaction, figure 8.15b. Addition of DNase alone had no effect on the complex, whereas addition of RNase reduced the amount of the complex considerably. As it is present regardless of the RNA transcript, it is likely due to nonspecific interactions with RNA binding proteins.

Protein Pull Downs

The EMSA demonstrated binding between the GRE and parasite proteins *in vitro*; this does not however identify the proteins involved. To identify the proteins it was decided to enrich them by pulling the bound complex out of the cellular lysate. To this end, the transcript was labelled with biotin, and then incubated with cellular lysate. Following binding, the GRE:protein complex was “pulled down” from the lysate using streptavidin coated beads. Preliminary pull-down experiments using Coomassie stained gels did not reveal any protein bands (data not shown). It was then necessary to move to the more sensitive silver staining technique.

The first successful pull down compared proteins obtained in three separate reactions, figure 8.16. These contained GCF RNA, GCS RNA or GCF RNA with addition of RNAse to the protein. Addition of RNase to the protein gave a background control of proteins that came through the pull down, but were not involved in GRE binding. Proteomic analysis of the excised bands, figure 8.16 produced the data set presented in appendix iii. Proteins present in both the GCS and GCF data sets are not specific

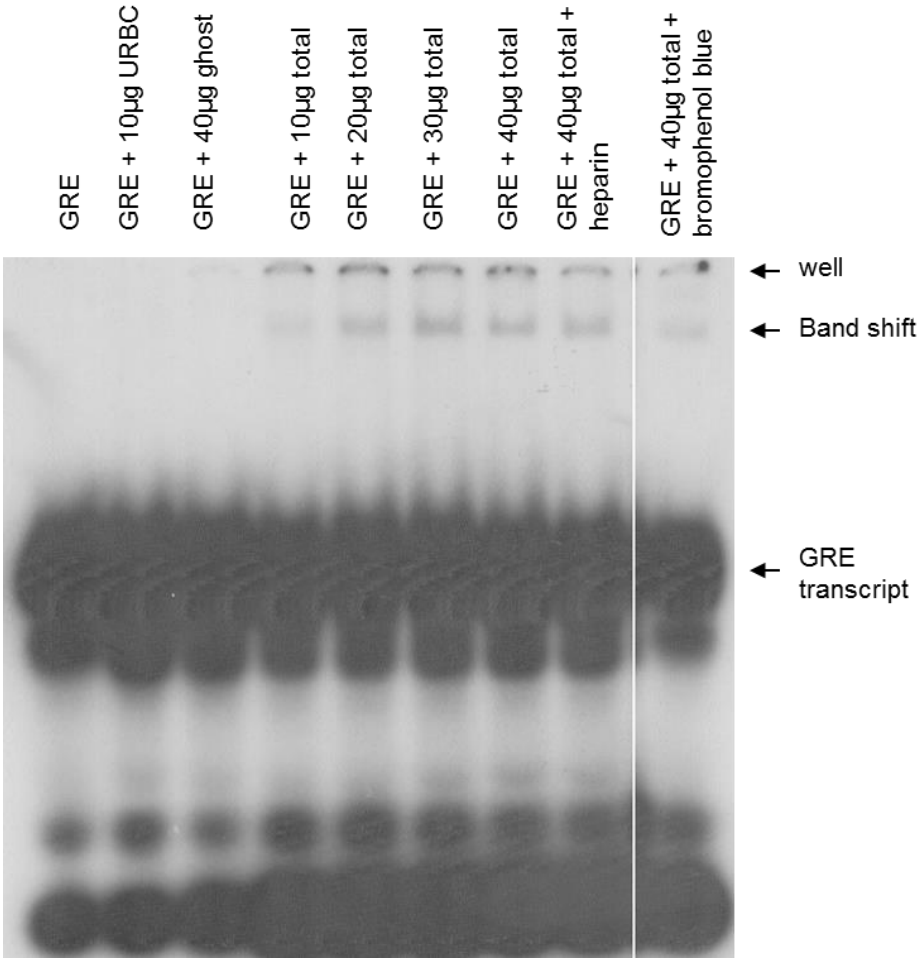


Figure 8.12 GRE protein binding is due to parasite proteins.

Increasing concentrations of total parasite protein extract (total) results in increasing binding of GRE to protein. Lack of a band shift with both URBC and ghost extracts confirms that the binding is due to parasite proteins. Addition of bromophenol blue leads to a reduction in binding.

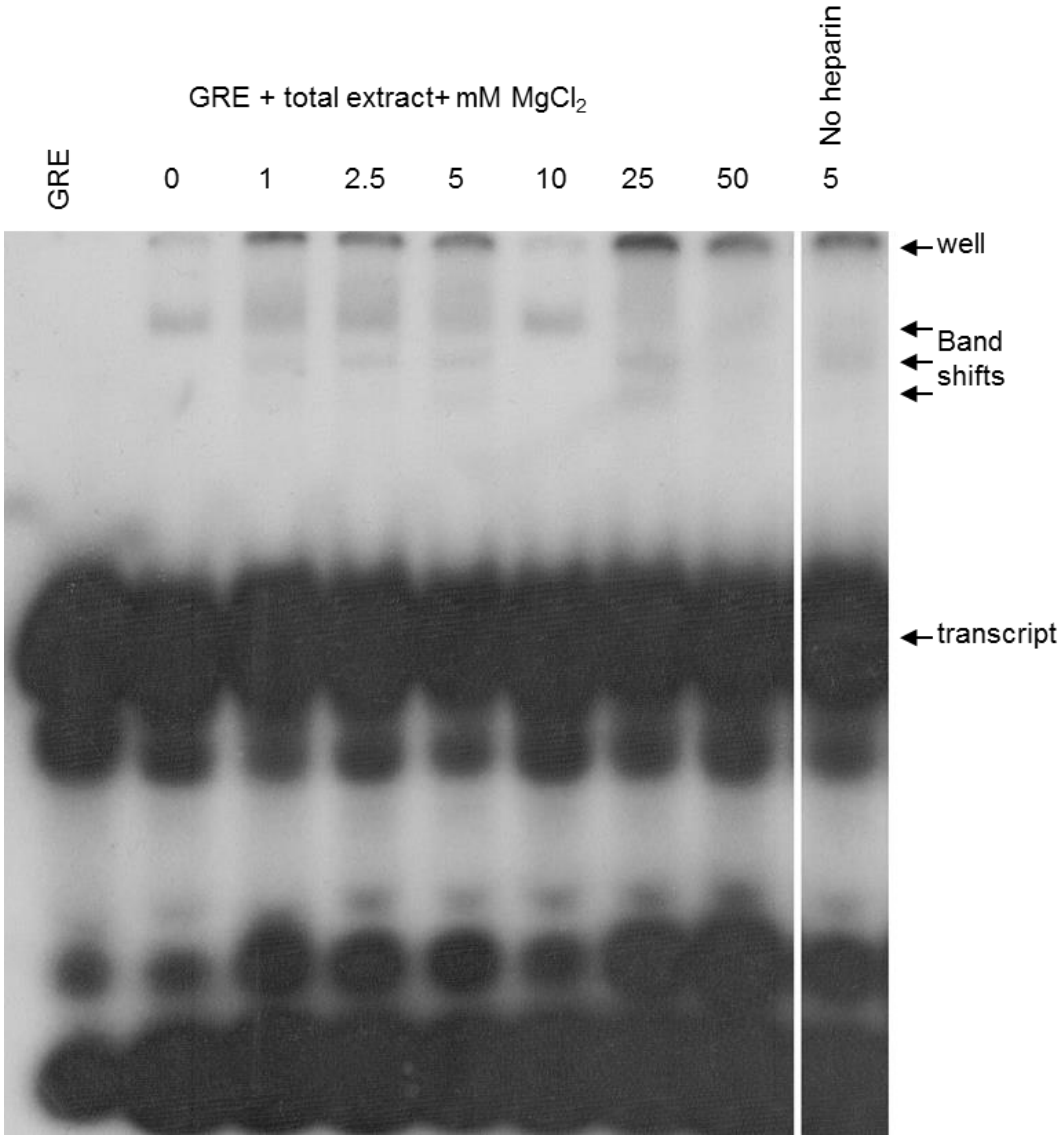


Figure 8.13 Effect of $MgCl_2$ on GRE:protein binding.

$MgCl_2$ has an effect on protein binding with the GRE transcript. 0 and 10 mM $MgCl_2$ gives a single band shift.1-5 mM has 3 bands. 25 mM gives 2 bands. Binding is abrogated at 50 mM. Reactions contain GRE, 40 ug total extract and heparin (for details see text). The final lane contains 5 mM $MgCl_2$ and no heparin, showing that the removal of heparin reduces specific binding.

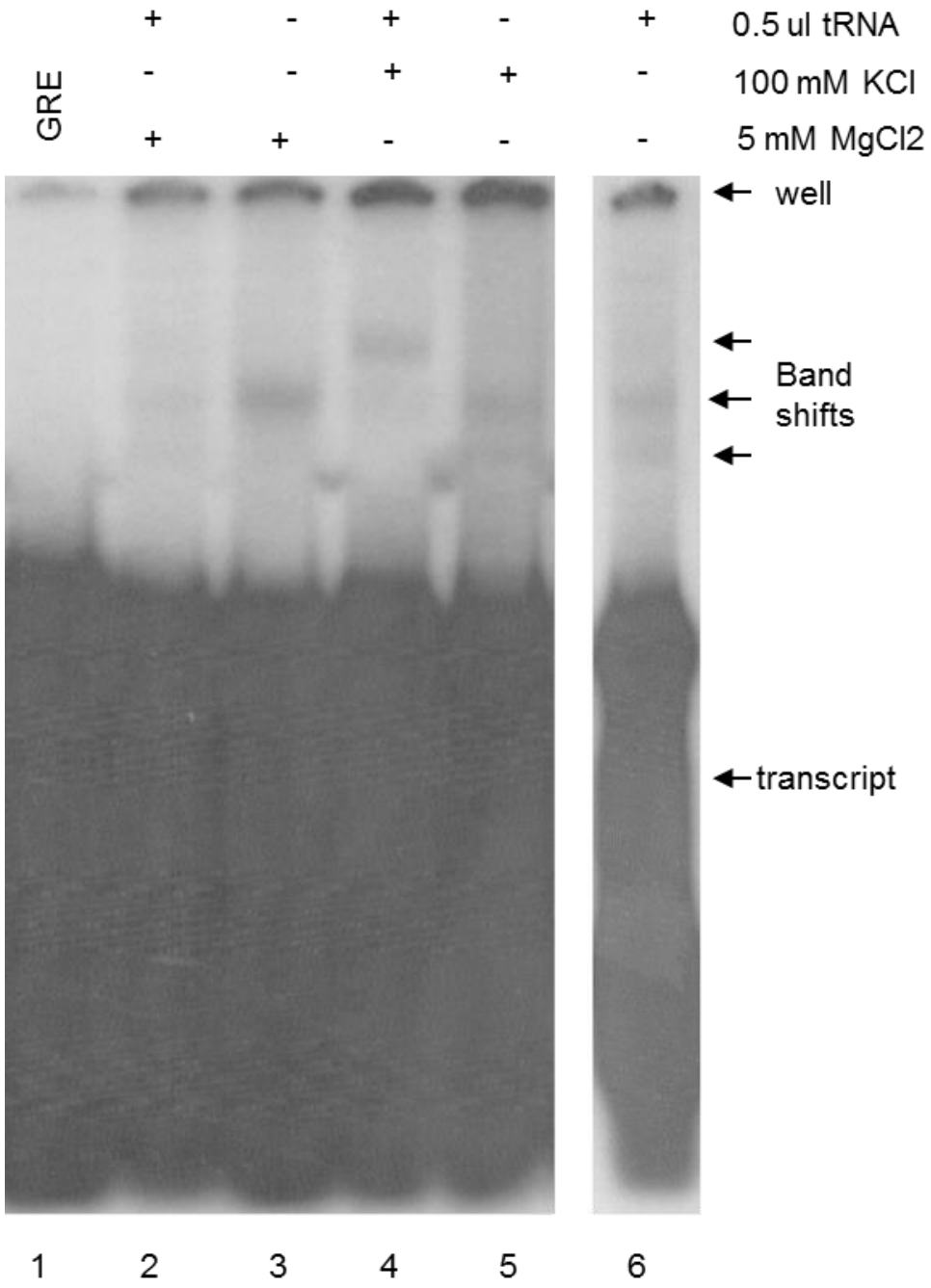


Figure 8.14 Effect of salt on RNA/protein interaction.

Addition of either 5 mM $MgCl_2$ or 100 mM KCl makes no difference to the GRE:protein complex. Addition of tRNA does alter the binding pattern, lanes 2 and 4.

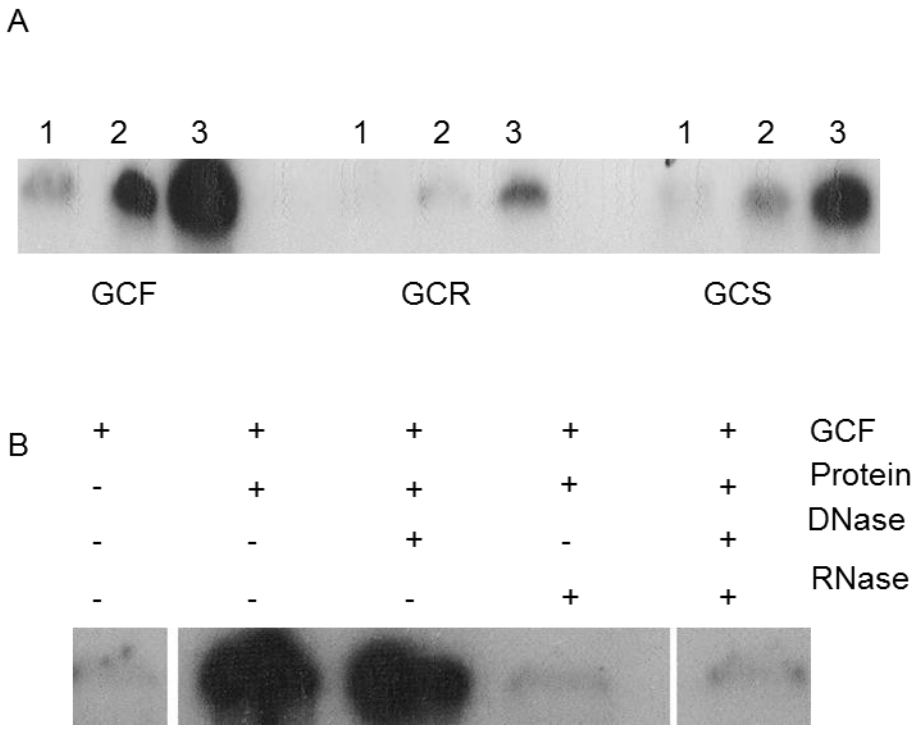


Figure 8.15 A large non-specific complex stays in the well during EMSA. A. The complex is not specific to GCF as it is formed with GCR and GCS; 1, probe alone; 2 & 3 transcript and total protein extract binding at room temperature (2) and 37^{o}C (3). B. The complex is an RNA:protein complex. Addition of DNase has no effect, but addition of RNase removes the complex.

for the GCF transcript, likewise those present in both the RNased control and GCF data. All of these overlapping, non-specific proteins and those with a non-significant score have been removed from the data summarised in table 8.3.

Many of the proteins in table 8.3 are involved in translation of mRNA. The ribosomal subunits clearly form part of the ribosomes which catalyse the translation of mRNA to protein. The eukaryotic translation initiation factor (EIF) 3 controls assembly of the 40s ribosomal subunit onto the mRNA, whilst EIF2β directly binds other initiation factors eIF1, eIF5, and eIF3

(322). Of the other proteins, aspartate tRNA ligase is indirectly involved in translation as it aminoacylates tRNAasp. Two histones, H2B and the variant histone H2A.Z were identified. Histone H2A.Z has been reported to be associated with epigenetic regulation of *var* genes (323). The UB05 antigen, GenBank: DQ235690, is predicted to be a transmembrane protein, expressed in trophozoites, schizonts and merozoites (324).

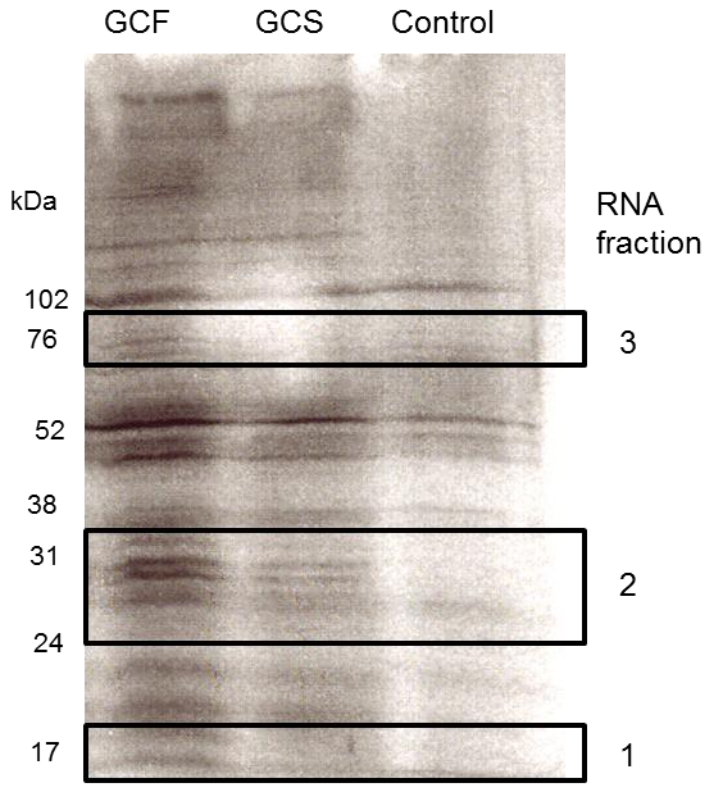


Figure 8.16 Comparative pull-down experiment 1.
RNA binding proteins eluted from Dynabeads and resolved on 10% SDS – PAGE gel. GCF and GCS are biotinylated transcripts bound to streptavidin coated beads. Control is GCF with RNase added to lysate. 1-3, gel slices processed by proteomics.

Table 8.3 Proteins identified in pull-down experiment 1.
Only those proteins with a significant score and that are not present in the scrambled or RNased controls are shown. The full data are in appendix iii.

RNA 1 < 24 kDa significant score >32

ID	score		size kDa	no. of different peptides	seq coverage %
PF3D7_0316800	59	40S ribosomal protein S15A, putative	14.8	1	6
PF3D7_1421200	55	ribosomal protein S25, putative	15.7	1	7
PF3D7_1038000.1	49	antigen UB05	13.8	1	5
PF3D7_0320900	38	Histone H2A.Z	16.5	2	10
PF3D7_1105100	38	histone H2B	13.1	1	7
PF3D7_0503800	33	60S ribosomal subunit protein L31, putative	14	1	6

RNA 2 24-35 kDa significant score >30

PF3D7_1465900	**46**	**40S ribosomal protein S3, putative**	**25**	**1**	**4**
PF3D7_0721100	45	hypothetical protein	29.2	3	13
PF3D7_1010600	39	eukaryotic translation initiation factor 2, beta, putative	25.3	2	10
PF3D7_0807100	34	RNA helicase, putative	157	2	1

RNA 3 60-70 kDa significant score >30

PF3D7_0102900	**87**	**aspartate--tRNA ligase**	**74**	**2**	**2**
PF3D7_0612100	77	eukaryotic translation initiation factor 3 subunit L, putative	78	2	3

As it had been determined that the GRE was mainly present in the cytoplasm, figure 8.2, it was decided to move to further purification of the parasite extracts. It was hoped that this would reduce the amount of background bands in the pull-downs, by using cytoplasmic and nuclear extracts. In the first attempt at a sequential pull-down, two bands were clearly enriched in the sample eluted off the Dynabeads as compared to the starting sample, figure 8.17. These were sent for analysis, and the results are presented in table 8.4 and appendix iii. Only two proteins had significant scores. The STI1-like protein fragment isolated from the 20 kDa band was considered to be too large. The STI1-like protein is 66.4 kDa, so identification of peptides that match it could be because of degradation prior to electrophoresis. The elongation factor-1 alpha (EF1-α) was isolated from the 50 kDa band, and is expected to be 49 kDa. Nine peptides were identified, covering 25% of the protein sequence. There are two copies of EF1-α adjacent to each other on chromosome 13, PF3D7_1357000 and PF3D7_1357100, and there is also a gene PF3D7_1123400 that is annotated as a putative EF1-α. EF1-α is involved in translation, delivering aminoacylated RNAs to the ribosome.

Following discussions with colleagues in the proteomics department, a further sequential pull-down was undertaken. For this, the samples bound to the GCS transcript attached to Dynabeads were also eluted from the beads, and run alongside the GCF transcript bound proteins, figure 8.18, unfortunately the GCS samples were not washed adequately and are overloaded. Even allowing for this, there are some bands visible in the

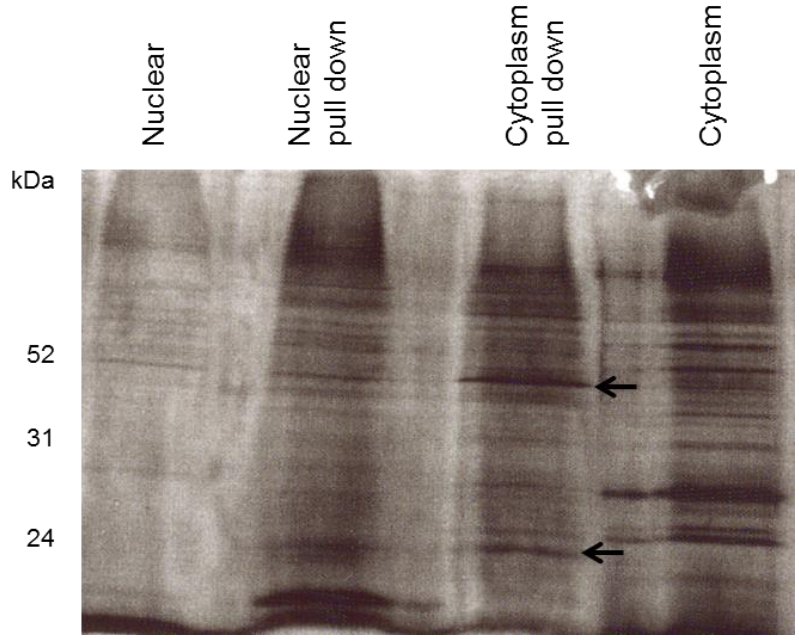


Figure 8.17 Sequential pull-down experiment 2.
The two bands marked with arrows are enriched in the proteins eluted off the Dynabeads, compared to the initial starting sample, cytoplasm.

Table 8.4 Proteins identified in pull-down experiment 2 (sequential).

RNA 1 20 kDa significant score >37

ID	score		size kDa	no. of different peptides	seq coverage %
P25407	50	STI1-like protein, flags, fragment	29.5	5	
Q8ILC1 (PF14_0324)	33	STI1-like protein PF14_0324 now annotated as PF3D7 1434300= Hsp70/Hsp90 organizing protein (HOP)	66.4	7	

RNA 2 50 kDa significant score >37

Q00080	**45**	**Elongation Factor 1-alpha**	**49**	**9**	**25**

GCF sample that are not present in the GCS lane. These bands from the GCF lane, and the corresponding GCS lane were analysed. As with the comparative experiment 1, some peptides were found in both GCF and GCS samples. The data from this experiment are in table 8.5 and appendix iii.

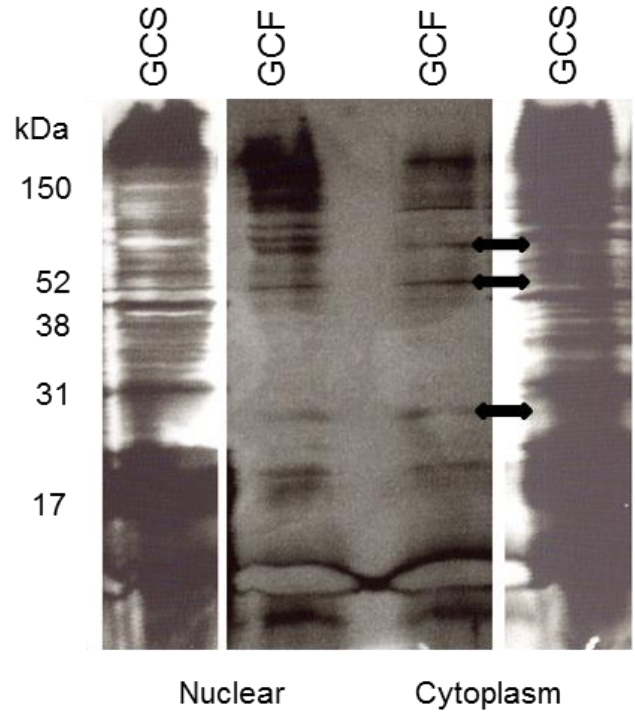


Figure 8.18 Sequential pull-down, experiment 3.
This is a combination of a sequential and comparative pull-down experiment. The proteins eluted off beads coated with GCS are run alongside those eluted off beads coated with GCF. The three bands marked with arrows were excised from both the GCF and GCS lanes and subjected to proteomic analysis.

Table 8.5 Proteins identified in pull-down experiment 3.
Only those proteins with a significant score and that are not present in the scrambled or RNased controls are shown. The full data are in appendix iii.

RNA 1 >52 kDa significant score >22

ID	score		size kDa	no. of different peptides	seq coverage %
PF3D7_0318200	29	DNA-directed RNA polymerase II, putative	280	5	<1
PF3D7_0406500	27	conserved protein unknown function	403	6	1
PF3D7_0421100 (PFD1005c)	27	VAR erythrocyte membrane protein 1 (PfEMP1)	252	5	2
PF3D7_0804600	26	U2 snRNA/tRNA pseudouridine synthase, putative	64	1	1
PF3D7_1122500	26	conserved protein unknown function	232	3	1
PF3D7_1234000	24	conserved protein unknown function	51	1	1

RNA 2 40 kDa >22 significant score

PF3D7_1410600	**75**	**eukaryotic translation initiation factor 2 gamma subunit, putative**	**51.9**	**10**	**20**
PF3D7_0715400	50	secreted ookinete protein, putative	24.6	1	3
PF3D7_0511500	35	RNA pseudouridylate synthase, putative	1195	10	2
PF3D7_0405400	30	pre-mRNA-processing-splicing factor 8, putative	368	5	2
PF3D7_1338300	30	elongation factor 1-gamma, putative	51	10	16
PF3D7_0603400	29	conserved *Plasmodium* protein, unknown function	133.5	5	3
PF3D7_1423000	27	GTPase, putative	57.4	4	3

PF3D7_0420900 (PFD1000c)	25	VAR erythrocyte membrane protein 1 (PfEMP1)	254.5	4	1
PF3D7_0701900	25	Plasmodium exported protein, unknown function	113	5	5
PFC0500w	24	conserved protein, unknown function	74	2	2
PF3D7_1411400	23	plastid replication-repair enzyme	237	6	1
PF3D7_0301300	23	alpha/beta hydrolase, putative	51.6	1	1
PF3D7_1116000 PF3D7_1116100	22	rhoptry neck protein 4 serine esterase, putative	135.5 217	2	<1
PF3D7_0821300	22	ATP-dependent RNA helicase prh1, putative	102.3	4	4
PF3D7_0707300	22	rhoptry-associated membrane antigen	95.6	4	4

RNA 3 17-31 kDa >22 significant score

PF3D7_0516200	**58**	40S ribosomal protein S14, putative	**16**	**2**	**15**
PF3D7_1317800	34	40S ribosomal protein S15/S19, putative	17	1	7
PF3D7_1304500	34	small heat shock protein, putative	25.2	2	10
PF3D7_0630600	31	conserved *Plasmodium* protein, unknown function	112	3	1
PF3D7_0930600.1	30	peptidyl-prolyl cis-trans isomerase	73	1	1
PF3D7_1326800	29	syntaxin, Qa-SNARE family	52.5	2	2
PF3D7_0514900	29	conserved *Plasmodium* protein, unknown function	52.2	3	5
PF3D7_0406500	27	conserved *Plasmodium* protein, unknown function	403	6	1
PF3D7_0617800	25	histone H2A	14.1	1	6

PF3D7_0922400	25	para-aminobenzoic acid synthetase	117.5	1	<1

The original dataset received from the proteomics laboratory contained gene identifiers from earlier editions of Plasmodb (156). Whilst updating the results to the newest version, one gene PF11_0168 returned two genes PF3D7_1116000 and PF3D7_1116100. Closer analysis of the two protein sequences revealed that one of the peptides is only in PF3D7_1116000, and the other is only in PF3D7_1116100. All of the proteins isolated from this experiment supposedly originated from a cytoplasmic fraction of *P. falciparum* cells. Two of the proteins in table 8.5, PF3D7_0318200 and PF3D7_0405400 would be expected to be in the nuclear fraction. However, their predicted sizes of 280 and 368 kDa are much larger than the size of the bands they derive from. As with the data in table 8.3, several of the proteins have a role in translation, two 40S ribosomal proteins, S14 and S15/S19, EF1-γ and the γ subunit of eukaryotic translation initiation factor 2. The data in appendix iii shows that the most prevalent protein in sample 2 was PF3D7_1357000, with peptides covering 62% of the sequence. This is EF1-α which was also found in experiment 2. This same protein was also found in the control sample, but with a lower score, and sequence coverage of 25%. As it is in the control sample, it has been excluded from table 8.4. Whilst the GTPase PF3D7_1423000 is intracellular, its conversion of GTP to GDP is required in translation. Again these data point to the GRE having a role in translation.

Two *var* genes PF3D7_0420900 (PFD1000c) and PF3D7_0421100 (PFD1005c) were also identified. Whilst these are obviously much larger than the band size they were taken from, closer analysis of the peptides revealed two identical peptides in both *var* genes. The peptide data is in table 8.6. In both cases, the peptide MLEQLEK is midway in exon 1, and the RTSTQNVAKPAR peptide is in exon 2.

Table 8.6 Peptides in *var* genes identified in proteomic analysis of GRE binding proteins.

PF3D7_0420900 (PFD1000c)		**PF3D7_0421100 (PFD1005c)**	
peptide	Position in gene	Peptide	Position in gene
		IDFNNTDNK	453
MLEQLEK	1212	MLEQLEK	1192
AENSKEDEK	1589	VNCSSGSKSR	1335
		SDDGDTVVSKVRN	1521
RTSTQNVAKPAR	2023	RTSTQNVAKPAR	2003

Of the remaining proteins, PF3D7_0804600 a putative U2 snRNA/tRNA pseudouridine synthase and PF3D7_0511500 a putative RNA pseudouridylate synthase, change U to ψ. Yeast U2 snRNA/tRNA pseudouridine synthase acts on both U2 snRNA and tRNA. The dual specificity is thought to be due to the enzyme requiring a double stranded region for binding, whilst targeting a loop (325). The RNA pseudouridylate synthase acts on tRNA and rRNA, but its predicted size of 1195 kDa when excised from a 40 kDa band makes it unlikely to be valid. Only PF3D7_0715400, PF3D7_0301300, PF3D7_1304500 and PF3D7_0617800 of the remaining proteins are of a reasonable size to be

valid predictions. The small heat shock and histone proteins are abundant within cells, so could be expected to come through the assay.

Whilst no definitive conclusions can be drawn from the data presented in tables 8.3-8.5, the results point towards the GRE having a role in protein translation.

Discussion

Structural digests

Enzymatic digests of RNA transcripts as a method to confirm their structure proved problematic. An essential tool for interpretation of data is a successful digest with RNase V1. As this enzyme cuts at base paired nucleotides, for GRE I expected to see a ladder of bands running the length of the gel, with occasional single nucleotide gaps and two longer gaps midway. The single gaps should correspond to non-base paired nucleotides along the stem, with the longer 4 then 3 nucleotide gaps corresponding to the loop and adjacent bulge. As the data point to a larger loop, I would expect a 12 nucleotide gap in place of the shorter ones. We were unable to demonstrate any digestion with RNase V1, for either GRE or tRNA. There were technical difficulties in interpreting band sizes from the gels, this would however been improved if the RNase V1 had worked, as for the GRE it may have acted as a sequencing ladder. It has been reported, that RNase A as well as cleaving 3' of single stranded C and U, will also cleave U-C, A-G and A-A mismatches (326). Likewise, under differing buffer conditions RNaseT1 will cleave a wide variety of mis-matches (326). The data obtained from GeneRacer cloning, which has no problems with data interpretation, supports the view that RNase T1 may

not be as specific as predicted. Problems with sequencing ladders are most likely due to the high GC content, of both tRNA and GRE, causing the RNA to re-conform instantaneously when removed from ice.

There have been several technological advances in RNA structure analysis, (reviewed in (327,328)). Enzymatic techniques are still used, but in combination with other methods. A new enzyme, RNase J1 has been isolated from *Bacillus subtilis* that cuts at single stranded nucleotides, with no sequence specificity (329), but this is not available commercially. Chemical mapping using kethoxal, dimethyl sulfate (DMS), and 1-cyclohexyl-3-(2-morpholinoethyl)carbodiimide metho-p-toluenesulfanate (CMCT) identifies unpaired G, U, and A and C nucleotides. Following modification, primer extension terminates 3' of the modified base which can then be analysed by electrophoresis (327). Selective 2'-Hydroxyl Acylation and Primer Extension (SHAPE), measures the relative reactivity of each nucleotide, via its ribose, to N-methylisatoic anhydride (NMIA) (330). Unpaired nucleotides e.g loops are fully reactive, those in helices have some reactivity, whereas those involved in tertiary reactions are unreactive. The data produced from chemical mapping can now be included in RNA folding programs such as RNAstructure (331), giving extra parameters other than just thermodynamics to form a prediction.

Whilst there were problems with the structural digests, the data obtained for the tRNA-Glu highlighted three single stranded regions which correlated with the known loops. The data for both GREs point to a 12 nucleotide loop, but I was unable to confirm a long stem in a hairpin structure.

EMSAs

From the results of the EMSA experiments, we were able to confirm that the GRE does bind *in vitro* to *P. falciparum* protein(s). Saponin lysed parasite pellets were used to prepare protein extracts, so should only contain parasite proteins and RBC membranes. URBC and ghost controls confirmed that there was no binding to RBC components. It is clear, both from the amount of free probe in relation to the intensity of the RNA:protein complex, and the necessity to move to radiolablelled RNA, that there is not a great deal of complex produced. Whilst more binding is achieved with DNA EMSAs (101,312,332), the only published report of RNA EMSAs in *P. falciparum* shows considerably less binding (312). On reflection, it is conceivable that by increasing the amount of protein extract from 10 µg to 40 µg with the biotinylated GRE, I may have been able to see binding. It was concluded that there was too little binding for proteins to be identified from a gel slice. An excised gel slice would also contain unbound proteins that run at the same rate as the complex. In experimental terms these could be described as “background”, in reality they could be more abundant than the complex of interest. The small amount of binding was unexpected and disappointing. As so much GRE is generated by the parasite throughout the erythrocytic stages, it was hoped that more abundant binding would be seen.

All of the EMSA binding reactions were in 1x binding buffer which contained 1 mM $MgCl_2$. The data in figure 8.13 show that more binding is achieved with addition of up to 10 mM extra $MgCl_2$. Black states that “many RNA binding proteins require magnesium ion for optimal binding” (319). It is also known that magnesium ions stabilize tertiary RNA

structures (333,334). Improved binding with 10 mM $MgCl_2$ could be due to both stabilizing and binding effects. The GRE structure may be kept in the correct conformation, therefore allowing optimal binding.

A large complex was seen to stay in the wells following electrophoresis. A similar complex is visible in DNA EMSAs that identified the PfAlba protein family (312). The complex in the data presented here, must be due to RNA binding proteins, as it is abolished with the addition of RNase. However, in the GRE EMSAs it is non-specific as it is present with both the forward and scrambled transcripts.

Pull downs

The positive results of the EMSA experiments confirmed the existence of a GRE:protein complex. The next logical step from this was to try and identify protein(s) in the complex. Biotinylated transcripts had already been used in the earlier EMSA experiments so it was decided to use these as the bait in the pull-downs. Proteins bound to them could then be purified from binding reactions by isolating the biotin tagged RNA with streptavidin coated Dynabeads. Sensitivity was always an issue and although labelling cellular proteins with ^{35}S methionine was considered, downstream proteomics analysis meant this was impossible. Similarly, dual labelling either end of the RNA with both biotin and ^{32}P, to act as a tracer during electrophoresis, was discounted.

The three experiments for which mass spectrometry data was produced differed slightly. The first compared three different pull-downs, the GRE, a scrambled GRE, and GRE with lysate that had been treated with RNase. The sequential pull-down, figure 8.16, resulted in the only

experiment with visible enrichment of some protein bands. As non-specific RNA binding proteins should have been removed by the scrambled RNA steps, no comparative lanes were run. The experiment was repeated, with the eluate from the GCS beads run as a control. Unfortunately these beads were not washed adequately and hence the lanes are overloaded. Although there were noticeable differences in the banding patterns, and these samples were processed further, it is possible that if the GCS beads had been processed correctly, that more of the background proteins could have been removed.

Considering the three sets of results, many of the proteins are involved in protein translation. As the data points to GRE being in the cytoplasm, figure 8.2, this would support translation. Six different ribosomal proteins, four of which are in the 40S subunit were identified. As ribosomes require Mg^{2+} for their structure and function (335,336), that the GRE:protein complex was stabilized by addition of such would be further evidence towards ribosome involvement. Elongation and initiation factors EIF3, EIF2-β, EIF2-γ, and EF1-α which are all involved in translation of mRNA to proteins were also found. An alternative that the GRE binding results could point to is being part of a multi-aminoacyl-tRNA synthetase complex known as MSC or MARS. In humans cells, MARS contains nine aminoacyl-tRNA synthetases, including aspartate-tRNA synthetase and 3 other protein components (337). This complex co-purifies with ribosomes and elongation factors (338). Analysis of MSC components in *Thermococcus kodakarensis* identified tRNA synthetases, ribosomal proteins, elongation factors, translation initiation factors, RNA modifying enzymes, and a GTPase (339). Broadly speaking, these are a similar list

to those identified by the GRE. Amino-acylated tRNAs are known to be channelled to ribosomes with EF1-α (340,341). In archaea it has been suggested that the MSC allows both enhanced amino-acylation of tRNAs, and channelling of these directly to ribosomes (342).

How could we fit the GRE into a role either as part of a PfEMP1 translating ribosomal complex, or a component of MSC/MARS? Several studies have reported detailed structure interactions for 80S ribosomes. Can we use this information to speculate on a role for the GRE within this well conserved complex? Amalgamation of rRNA and protein molecular models with cryo electron microscopy (cryo-EM) in *S. cerevisae* predicted 80S interactions with tRNA (343). Interactions were predicted for three ribosomal proteins, but none of these were identified by the GRE pull-downs. Cryo-EM studies with 40S and EF1 and EF1α propose a model whereby recruitment of these two translation factors opens a “latch” on the ribosome (344). This opening involves a connection with helix 16 and ribosomal protein S3. Once opened, it allows mRNA and ternary complex binding. Ribosomal protein S3 and EF1α have both been identified as potential GRE binding proteins. Could GRE be involved in this initial complex, perhaps directing the ribosome to *var* transcripts? Most studies of MSC/MARS have looked at protein:protein interactions (339,342), but methods used to analyse proteins would preclude identification of any nucleic acid component. So for the GRE to be a component of the MSC/MARS could be possible.

Whilst undertaking the pull-down experiments, two reports were published which used a similar methodology to isolate DNA binding

proteins from *P. falciparum* (99,312). Closer inspection of the data shows that in both experiments EF1-α was found with both probes and control probes, in one case this was a KAHRP sequence (312), the other a scrambled sequence (99). Similarly, ribosomal subunits were found with all of the probes except KAHRP, and histones with all probes except the scrambled. Whilst it may be expected to find histones bound to DNA, one would expect ribosomes and EF1-α to be associated with RNA. Could it be that these aforementioned proteins are not specific for GRE RNA, but are in fact a contaminant, whether by experimental design, or by their abundance within the cell? If this is the case, we must also consider the other proteins identified in tables 8.3-8.5.

Of the proteins identified that aren't directly involved in translation, two predictions were for enzymes that change U to ψ in other RNA species. As they act on U residues within loops, but require double stranded regions for binding, the GRE fulfil these requirements. The conserved region of GRE includes the loop, which if mfold predictions are correct, contain a U (202). Another pair of predictions that may be related are GTPase and syntaxin. GTPase, whilst being required in translation, is also required in other cellular processes including translocation of proteins through membranes. Syntaxins are part of the SNARE family of proteins which mediate fusion of vesicle membranes to other membranes. Could GTPase and syntaxin point to a role for GRE in trafficking of proteins? It is tempting to postulate that the identification of two PfEMP-1 proteins in the same experiment could be related to the GTPase and syntaxin. That two PfEMP-1s were identified is not an issue. Firstly the culture was not clonal, and secondly, the two genes are almost identical. As PfEMP1 is larger

than the estimated size range of the excised gel band, and has only been identified once, a replicate so that n=>1, is required before this is pursued further.

In conclusion, the EMSA results described in this chapter confirm that the GRE forms a complex *in vitro* with parasite proteins. Pull-down experiments identified many proteins as components of a potential GRE:protein complex. The data point to several possible roles for the GRE. One possibility is a role in protein translation either directly, or as component of MSC/MARS, as many of the proteins are ribosomal and translation factors. Another possibility is a role in trafficking of PfEMP1 proteins.

Chapter 9

Conclusions

One aim of this thesis was to clone novel ncRNAs from *P. falciparum*. The modified GeneRacer technique was successful, but this success was dependent on the method used to isolate small RNA. Representatives of the major classes of ncRNAs were cloned. Other ncRNAs that have been identified by other screening methods were also cloned (125,145,146,150,178). The high background level of rRNA led us to develop a method to reduce this, by removal of rRNA with biotinylated rRNA sequences and Dynabeads. Even with this rRNA reduction, extensive screening was required to identify the few novel clones that were found. This, coupled with the knowledge of similar sequencing efforts at the WTSI led us to terminate this part of the project.

The primary aim of this thesis was to investigate if a GC rich ncRNA found only in close proximity to central *var* genes has a role in their regulation. As two different sets of cultures were to be studied, one a set of clonal populations of HB3, the other progeny of the 3D7xHB3 cross, it was necessary to develop a set of qRT-PCR primers to study *var* gene expression in HB3. These were made and tested to ensure specificity for HB3 with no cross reactivity to 3D7. In testing the published primer set for 3D7, it was necessary to make alterations to those to ensure they did not cross react with HB3 *var* genes. The new and modified primer sets were

used to study *var* gene expression in the progeny of the 3D7xHB3 cross. No dominant expression could be found in two out of the five available clones. One explanation of this pattern could be due to recombination within *var* genes which generates new genes that are not detected with these primer sets. However, DNA sequencing data from the WTSI have revealed that this is not the case (C. Newbold, personal communication). The more likely explanation is that the progeny, whilst clonal originally, have been in culture for so long that background switching has resulted in a heterogeneous expression profile.

The HB3 primer set now allows us to study *var* gene expression in a different isolate to 3D7. This enables us to test discoveries made in 3D7 to see if they are isolate specific or more likely species wide. We made use of these HB3 primers to follow long term *var* gene expression in eleven clonal, and a parental, HB3 population. The results, presented in chapter 7, mimic the pattern seen in a comparable 3D7 study (210). This demonstrated that without immunological pressure, *var* genes switch expression over time from sub-telomeric genes to central *var* genes. Another study following both IT and 3D7 clones reported one case of a telomeric UpsB switching to an UpsC *var* gene (211). In HB3, all clones that switched *var* gene expression, switched to an UpsC *var* gene. HB3 may be different to other isolates as this switching pattern is not seen universally in others. With two other clones in the Recker study, one switched from an UpsB telomeric to a different UpsB telomeric *var* gene, and the other from an UpsC to an UpsB telomeric *var* gene (211). *Ex-vivo* studies in Papua New Guinea have reported expression of UpsC and UpsB class genes in asymptomatic patients (345,346), with UpsC

expression in asymptomatic older children with high parasite load (345). Generally speaking, it is difficult to reconcile data collected with and without immune pressure. One explanation suggested for data in the Papua New Guinea study is that PfEMP-1 expressed by UpsC type genes do not sequester in vital organs (345). An alternative explanation for *in vivo* asymptomatic infections and long term *in vitro* cultures both expressing UpsC *var* genes could be that the *var* switching patterns are “hard-wired”, as suggested by Recker et al (211). So that left in culture for long enough, the majority of parasites in a population would eventually express an UpsC *var* gene.

In HB3, central *var* genes are present in five chromosomes and as clusters in four of these. Somewhat surprisingly all of the isolates, if they switched expression, switched to expressing genes from the same central cluster on chromosome 4. Those that did not switch were already expressing the UpsB type *var* gene from the chromosome 4 cluster. When examining switch rates of the cultures there was no difference overall between central and sub-telomeric *var* genes, however, central UpsB genes have a much lower off rate than sub-telomeric UpsB genes

To investigate a potential role for GRE in switching, the expression of specific GRE at different time points during the *var* switching experiment was examined. There was a clear relationship between transcription of a GRE from the same cluster as a dominant *var* gene, when UpsC type *var* genes were expressed. This pattern was not repeated with expression of UpsB type central *var* genes. However if these clones switched to an

UpsC *var* gene, the GRE transcription switched to a dominant neighbouring GRE.

A quicker method for looking at GRE expression would have enabled us to follow GRE expression over time more closely. In the case of switching to UpsC, it would be interesting to know which switch occurs first, GRE or *var*, or if they happen simultaneously. HRM and ARMS PCR were tested as methods to examine GRE expression, but were unsuccessful. If further studies following GRE expression are required, it may be worth re-visiting HRM. By dividing the amplicon into two, the conserved region could be used for placing a reverse primer to pair with the 5' primer gcT335, and a forward primer to pair with gcB450. A potential problem with this would be that one primer from each pair would have a SNP within it, but this could not be avoided. It would give however smaller amplicons with fewer SNPs within each, this could result in less secondary structure influencing the melt. The transfection constructs were designed to investigate whether the GRE had any effect on the expression of *var* genes. By placing the GRE near an UpsB or an UpsC type promoter, our data suggested that only the UpsC controlled reporter would be expressed. As we were unable to fuse the UpsB sequence to the reporter gene we were unable to complete this work. Attempts will be continuing to create the UpsB construct and complete the transfection experiments.

From the data described herein, and what is already known about *var* gene regulation, it is difficult to form any definite conclusions about the role of GRE in *var* gene transcription. A role as an insulator element preventing an enhancer up-regulating *var* can be discounted, as one

would expect GRE transcription to prevent *var* transcription. Our data point to the opposite scenario, that the GRE functions as a barrier insulator because increased GRE transcription leads to transcription from a nearby locus. If the GRE does function as a barrier element, it would act to stop the spread of silencing factors to the expressed *var* locus.

Even if transfection experiments had been successful, they were designed to study the role of GRE in *var* switching, and would not have elucidated the mechanism for GRE effects. To try and identify the mechanisms for GRE involvement in *var* gene regulation, or other possible roles, protein binding experiments were undertaken. EMSA studies proved that the GRE forms a complex with protein(s) when incubated with parasite protein extracts *in vitro*. Having established that there is GRE:protein complex, attempts were made to identify the components. Proteomic analysis of bands specific for GRE identified several proteins that have a role in translation of mRNA to protein. Jackson and co-authors point out that whilst there have been major advances in our knowledge of host-parasite interactions, there has been far less interest in translation in *Plasmodium* species (347). Although a few hypothetical proteins were identified, the majority were known components of ribosome complexes the ncRNA content of which i.e. tRNA and rRNA are well established. Methods exist for purification of ribosomes (348), so we could isolate these, extract RNA and northern blot. If the GRE are a component of ribosomes that are only involved in *var* gene translation, we would be likely below the limit of detection. This would also not explain their presence throughout the asexual stages. The proteomics data has little to

suggest a role for GRE in *var* gene expression at the transcriptional level. It could however be involved at the post transcriptional level.

RNA FISH studies to look at the localization of the transcript within the cell are an obvious experiment missing from this thesis. These are being undertaken by a colleague, the results so far show a signal from the nucleolus (S. Kyes, personal communication). This is consistent with GRE being a pol III transcript. Isolation of GRE from cellular protein extracts, fig 8.2, showed that GRE was present in both the nucleus and more in the cytoplasm. Proteomics data for experiment 2 and 3 were both from cytoplasmic extracts, whereas experiment 1 was from total extract. With total or cytoplasmic extract, ribosome related proteins were the most prevalent. This is in agreement with a recently published analysis of the organellar proteome (349). Oehring and colleagues reported that the majority of ribosome sub-units and translation-related factors were present in both the nuclear and cytoplasmic proteomes.

The data obtained from proteomics experiments lead us to conclude that the GRE may have dual roles within the parasite. Switching studies show an association between GRE transcription and *var* expression from the same central cluster when an UpsC *var* gene is expressed, but not when a central UpsB type is expressed. This pattern of GRE transcription and *var* gene expression follows the same pattern when *var* switching occurs. This association suggests a role in *var* gene regulation at the transcriptional level. Protein studies suggest an alternative role for the GRE.The GRE forms a complex when incubated with *P. falciparum* protein extracts that can be resolved in EMSA

experiments. Pull-down experiments of either total parasite protein or cytoplasmic extract identify mainly ribosomal proteins and translation factors. This is consistent with the GRE being found in both the nucleus and the cytoplasm, and points to a potential role in PfEMP1 translation.

Appendix i

Oligonucleotides used in this thesis

Chapter 2

Oligo name	Sequence
M13F	gtaaaacgacggccag
M13R	caggaaacagctatgac
VarCF	aaaaaacaaaatcatcagtaggaaatttatt(c/t)c
VarCR	tatcccataaatctgc(a/t)at(a/t)gg(a/g)ta

Chapter 3

GeneRacer

Oligo name	Sequence
GeneRacer dC3'A	gctgtcaacgatacgctacgtaacggcatgacagtgggggggggga
GeneRacer dC3'C	gctgtcaacgatacgctacgtaacggcatgacagtgggggggggc
GeneRacer dC3'T	gctgtcaacgatacgctacgtaacggcatgacagtgggggggggt
GeneRacer 5' Primer	cgactggagcacgaggacactga
GeneRacer 3' Primer	gctgtcaacgatacgctacgtaacg
GeneRacer 5' Nested Primer	ggacactgacatggactgaaggagta
GeneRacer 3' Nested Primer	cgctacgtaacggcatgacagtg

Screening

Oligo name	Sequence	Sequence	
Mirna451-R	actcagtaatggtaacggttt	5.8sMid	Ttgcagtgactgtgaatcatc
5.8sF	tagcacagtcttaacgatgg	5.8sEnd	gaatagtggtactcctacg

SSFU mix		INT mix	
SSFU	ccatgttacgacttcgca	LSUR	gtttcctcacagctttat
SSUD	ttggattcaacgtccagg	RNA5	cgattgatactgtaatatac
SSUE	gcgtgacgagcggtgtgt	Mito4264-4693	gttccgctcaatactcag
LSUA	agacgctgacttcctggc	PF11_0105INT2	ttcagacaagactaatca
RNA1	aactatggatagctgata	PF11_0105INT5	gttgggcatagttatcat

rRNA Depletion

Oligo name	Sequence
5sBIO	*ctgatgggatctggtgtggccactgtagtatgaac
5.8sbioA	*ccttcatcgctgtaggaaccaagacatccatcgttaa
5.8sbioB	*gtagtttacattcagcattctgatgattcacagtca
18sBIO	*gcacgcgtgcagcctagttcatctaaggacatcacaga
28sBIO	*gtgcaaagaaaacgcaccccgcgtgtaaagcctatgt

*Denotes 5’ Biotin-TEG modification

Chapter 4

Primers tested for HB3 var gene qPCR
Those shown in bold are included in the final HB3 set.

Identifier	Broad locus name	F oligonucleotide	R oligonucleotide
HB3-1000-1	PFHG_03232.1	ccctgtccacaaccatcagc	cgtcgtcgtcatcagtgtcc
HB3-1000-2alt	PFHG_03234.1	ccaaaggagaaggcaccacc	acctatggcacccctctcac
HB3-1040	PFHG_03416.1	gatgctacaaccaccccacc	gtgttaccactcgcccactc
HB3-1074-1	PFHG_03476.1	tatcccaccaaggaggcgac	cggagtggtaccttgtgctg
HB3-1704_1zc	PFHG_03476.1	gctcccaaccaccacgttcc	gcttcctgctggtggctgtc
HB3-1704_1-2zc	PFHG_03476.1	acggagccactggtggtttg	gccggagtggtaccttgtgc
HB3-1074-2alt	PFHG_03478.1	gatggcacaaaagttggcgg	tgttctgggtcgacctcctc

HB3-1074-3	PFHG_03480.1	aagaagatggcgacgaaggc	tccggtgatccctcttctgg
HB3-1107	PFHG_03516.1	tggtaaatgcaagggtgatacagg	tgcatcgttatcactcaccagc
HB3-1108 var1	PFHG_03521.1	cgcaatatgcaactaatgac	acttggcaatattctgaacg
HB3-1210	PFHG_03840.1	gctgagagtgtcttgaaagg	aagacgacgactatcattgg
HB3-1210-1zc	PFHG_03840.1	gccctaaaaccacgtggccc	cctcctccgtggtgtcctcg
HB3-1210-2zc	PFHG_03840.1	cgaggacaccacggaggagg	ttggtgctgctggttgtggc
HB3-1210alt	PFHG_03840.1	agcttccacgaatcgcactc	gtgccttcgccttcttttgc
HB3-1235	PFHG_03671.1	aaaaattggactttctcctg	ttttgcagttcacatacgtc
HB3-1296	PFHG_04012.1	atggacaaatgatggtaagg	taggagtaggtgttgcgttc
HB3-1296-2	PFHG_04014.1	agatggcgacaaaggccaag	ttgggtttggcaccactagc

HB3-1296-3	PFHG_04015.1	aaacggaaaacctggcctcc	tcgtcttggcctttggcttc
HB3-1296-3-1zc	PFHG_04015.1	cgacgaaggcggaggaaacg	gtgcgcttttcactgtggcg
HB3-1296-3-2zc	PFHG_04015.1	tgtatcccacccaggaggcg	cactctgtgcctgcgacgac
HB3-1308-1zc	PFHG_04035.1	ggtggtggtgccgatcccgcc	tgtgacgcctccgtcttagtggccc
HB3-1308-2zc	PFHG_04035.1	cagaaagtgtcgcccgctcc	gcgtgtgtggttcgtcctcc
HB3-1334	PFHG_04081.1	ggcggtgtctgtattccacc	gctgcctcaccacctgttag
HB3-1334alt	PFHG_04081.1	aaagtcccttgcgttccgac	gtccctccccctcttcttcc
HB3-1408	PFHG_04057.1	tgctatgacgtgtaatgcccc	acttacatgagtcccatctggtg
HB3-1459	fbadf	actgacacagaaggaggaac	ttaggtttgaccccacatag
HB3-1459-1zc	Fbadf	ggaaaccgcggtggactcac	acttgtgggtgctttggggc

HB3-1459-2zc	fbadf	tacaggagtgggcgagtggc	cgatgtcgacacgccgtcac
HB3-1499	PFHG_04491.1	attggatgatgcctgtcgcc	ggcacctggtttagtggtgg
HB3-1514	PFHG_04620.1	aaactgacaatggccccgac	gttgttgaggggtcttcgg
HB3-1514alt	PFHG_04620.1	aaactgacaatggccccgac	gttgttgaggggtcttcgg
HB3-1523	PFHG_04593.1	caacgacaaaaaccacccgc	gtcgtcgtcttcaccttcgg
HB3-1523-1zc	PFHG_04593.1	gcggctcacccgacatcttc	gccgcctcgtcttcttcgtc
HB3-1523-2zc	PFHG_04593.1	gtcctggtgtcgctcgttcg	gccgcctcgtcttcttcgtc
HB3-1587	PFHG_04749.1	gcttttggtaccggtgttgc	aggcatgattcggcttcgtc
HB3-1587-1zc	PFHG_04749.1	accactcgtgccaccacctc	gagtttgtacctggcaccccc
HB3-1587-2zc	PFHG_04749.1	acctctccctgtcaaacctggc	tggttgtgattgtggcggcg

HB3-1604-1	PFHG_04769.1	agcgagtggtactcaggagg	gatggaccacgagatgtgcc
HB3-1604-2	PFHG_04770.1	acgaagaagacgatgccacc	gaagtcttcggagcgaccac
HB3-1671	PFHG_04910.1	cccgaatgtggggtcgattg	gtgtggtctacgtcggttgg
HB3-1703	PFHG_04861.1	catagaggacgaaatgaacc	gaggtcatcttggttgtcag
HB3-1703-1zc	PFHG_04861.1	cgtcgtcgtcgaggtgttcg	ttcctcctgctcttcggggc
HB3-1703-2zc	PFHG_04861.1	tggtgccaaagacccctccc	ggccactcgctgtgtctgtg
HB3-1727zc	PFHG_05046.1	gggggaaatgtggggtgccg	gggggataccccacactcattaccag
HB3-1737	PFHG_05052.1	gaaaaagggacattttatcg	tatcactgtttttggtgtgc
HB3-1737alt	PFHG_05052.1	gctggtaatggtggtcctgg	atgagtgtcaggaggggggac
HB3-1737hr	PFHG_05052.1	aaagtgcgaagcacctcccc	cgccactgcagggattagctg

HB3-1817zc	PFHG_05155.1	tggtacagctgatggtggtacttccg	tgtgcccgctttacggtttcg
HB3-2007zc	efbe	agccattacgtgcgaagctgg	agcggcacatcggcatttttg
HB3-209	PFHG_00592.1	tgctttttattatccaacagg	ccataatttttcgtttcacag
HB3-699-1	PFHG_02272.1	gaacccttgacgacgacac	ctcaacacacgtcaaaggcg
HB3-699-2	PFHG_02273.1	aagacgacaaacctggcacc	gtcgttgcttttggcttcgg
HB3-699-3	PFHG_02274.1	attcacagcactgaaagtcc	tcacaatcattaaaagcatcc
HB3-699-4	PFHG_02276.1	aagcagctgatggaacggac	tggttgttgtgggtcttggc
HB3-699-5	PFHG_02277.1	cgcgaagacgaaaacgtcac	gtttcatccggaccgtcctc
HB3-752-1	PFHG_02419.1	tggtaatgatgaagatgacg	gaattggcttcactttgttc
HB3-752-2	PFHG_02421.1	tggtagtgttgctagtggtg	atattcatcgtcctccacac

HB3-752-2 -1zc	PFHG_02421.1	gctcgctctttaccacccgc	ttccgtctcctccttcgccg
HB3-752-2 -2zc	PFHG_02421.1	cgacgacacacccgacgaac	agcacgttccgcatcttccc
HB3-752-3	PFHG_02423.1	agtggtgccaaaactgtcgg	accacaaaagtcgcttcccc
HB3-752-4	PFHG_02425.1	ccaccacaaaacccctccag	tccgcttgtggttcgtcttc
HB3-752-5	PFHG_02429.1	acagaaagttggacaggatg	atggttgttttgagaattgc
HB3-752-5alt	PFHG_02429.1	ttccgaagacgattccgcac	tcattgtcggaggcaacacc
HB3-209b*	Improved 209	agtggtgctgtagagccaaaagac	cctgcggcggtgctgtaagg
HB3-1235c*	Improved 1235	aggtctgctccttcagatgcgtg	tgttttccctaccatgacaaggatgcc

*Improved in Chapter 6 and replaced earlier 209 and 1235 pairs

Control primers for qPCR

Primer pair name	Forward primer	Reverse Primer
PFB0100	acggatccggtgactccttcg	tgcttggtgtggagcttggtg
PFI1090	tgttcttggaggaccaggagc	tcctgaaaaggcaccaccacc
seryl1a	agctacctcagaacaaccattatgtgc	atcctttccatgtgcccctgc
Aldolase	tgtaccaccagccttaccag	ttccttgccatgtgttcaat
Actin	agcagcaggaatccacaca	tgatggtgcaagggttgtaa

qPCR primers for 3D7

Only those that were redesigned from the published set (56) and a previous student's thesis (165) are shown.

Identifier	*var* gene	F oligonucleotide	R oligonucleotide
3D7-27altzc	PFD0625c	gcggtgtcgagtgaagccac	tttccacagcggccgtttc
3D7-95altzc	PFD0995c	accgtcacaacctgaccccc	gctcccgctccaacaccttg
3D7-34altzc	PFD1015c	accgacagtggcgcaaaagg	atgttgctttcgggtccggc
3D7-116altzc	PFD1005C	cccgactgtggggtcgaaaag	cgcctttatcaccgaaggaaaggac
3D7-115alt2zc	PFE1640w	tcaccgtcacgacgggcatc	tggtgctgctcgtggggttc

3D7-44altzc	PFF1580c	tgcggatataaactggctggagg	tgttccccattcagttagccatcg
3D7-93altzc	MAL6P1.1	attggggttggcggtgtgtg	cgtgtgggttcggcaccatc
3D7-45-2altzc	PF07_0048	aagaggacgaagacgccccc	cctttgccggctccaccttg
3D7-99altzc	MAL7P1.50	aaaacccgaaggtgccgctg	tcctccccctgttccgtgtg
3D7-113altzc	MAL7P1.212	acatcctcccacacctcctg	tgccctcttcgccatctttg
3D7-114zc	MAL7P1.187	gacacaagtggcgcaaaagg	gtgtgacggtgtcgtagtgg
3D7-92alt2zc	PF08_0103	acgcctacgttccgccgaccc	ccccctgtcccgtgtcttgtgcg
3D7-110altzc	MAL8P1.220	agacgtctctcctccgcgac	tggtccggctccaggaactc
3D7-112altzc	PFI0005	aggacaacacggagaacaag	actgtttcgcacaccttctc
3D7-5altzc	PF10_0406	ctccggcgtcaacttcgtcg	aagtcctcccgctccaccac
3D7-6altzc	PF11_0007	acgccggacccaaaagcaac	acgtgaggggttgggtggag
3D7-111altzc	PFL0005	cggcacagacgacaacctcg	tttcctcctccgacgacccc
3D7-91altzc	PF13_0001	cacaagcgggtggtgacgag	ccgctgcagactcaacaaatgc
3D7-short	PFA0015c/PFI1820w/MAL6P1.314	tgcagcaattggattacaagcagg	ccagcaggaaccataatgttgggg

Hybridization probes

Probe name	Forward primer	Reverse primer
HB3-1074-1	agtggtggtaatagcgaaaa	cccttctggaataccttctt
HB3-1074-2	acgtctgcccaatagtaaaa	actatttggttgtgagggtg
HB3-1074-3	cttgagagagcaatggaatc	ttttccatatttggttggac
HB3-1296-3	tttggtgataaagagctggt	tacaagaattcgcacaactg
HB3-699-2	gtggtggaaaacacagaagt	aaaatgcctctttcagatca
HB3-699-5	cctgattttgttcttgaagg	ttcttccacgacttctgttt
HB3-1499	tgtgagtgataacggtgaaa	tataaaacggttcctgctgt
HB3-1107	ctagtgtccaacaacccagt	ttccattgcacttatttcct

HB3-1737	atctgttaccgcacctctta	aaacttttgtcgtttttcca
HB3-1235	cttattcagatactagaagg	gtttctttcattacggcca
HB3-1459	gggcatcaataacacttcat	ttgtccacgacatttgtcta
Var2csa	atactataatacatggag	cattattagtgcatgcgtc

Chapter 5

HRM Primers

name	sequence
gcT-RTA	gctgccycagtagcccart
gcT-RTB	ccycagtagcccartbgtt
gcT335	agctgccycagtagccca
gcB-RTC	gccrccccctccgrg

gcB-RTD	crccccccctccgrgtaa
gcB450	tgcgccrccccccctccg

3D7 Mis-match PCR primers

Oligo name	Sequence	Pairs with	Oligo name	Sequence	Pairs with
gcTcontrol	tttccttgtgagaacgttgg	gcB450	gc4-4/4-5T	ccagtcgttaggtatgtag	gcB450
gcBcontrol	ggcaagcggagtcgaa	gcT355	gc7-4/12-2T	cgttggttcgactccgcg	gcB450
gc12-3F	agctgcctcagtagcccaa	gcB450	gc4-1/7-1/7-2B	gtctgcaacttttctcag	gcT355
gc12-2T	ctgcctcagtagcccagtt	gcB450	gc8-1B	ctccgggtaactcagcaa	gcT355
gc4-6/7-3T	ctgcctcagtagcccagtg	gcB450	gc7-3B	tgcgccacccccctccga	gcT355
gc4-2/4-3T	tagcccagtcgttaggtag	gcB450	gc8-2B	taactcagcagctcgatc	gcT355
gc12-1T	ccagtcgttaggtatgtaact	gcB450			

Primers for confirming PCR chimeras

gc12_2F	gtacacaaaatacttcatatacat	gc12_2R	ccttttttatattttattttaattatc
gc7_4F	ccagatgataatataacatatttt	gc7_4R	gtatgaaatgttatatggtcatag

gc4_6F	catagagagttacaattacatac	gc4_6R	ccccttttgtaatttattataatg
gc12_3F	gatgttaattacacacatcatctc	gc12_3R	gacacctattatggataaaatatt
gc8_2F	cgtccaatctacataaactttat	gc8_2R	attagttacgaatcagcagaac
gc7_3F	ggatatgtatatatatttgggtg	gc7_3R	cgtattcacacaagaatatacta
gc4_4F	cccaccctgacgacctcct	gc4_4R	cctatcgaaacaaatagacaaa

Chapter 6

Transfection constructs

Oligo name	Sequence	Oligo name	Sequence
IntronF	tccattgctttttattttgaag	IntronF+Sal	cgccaggtcgactccattgctttttat
IntronR	ggtcgacagggtgtttagttttt	IntronR+Sac	ttcattgagctcggtcgacagggtgtt
LucF	atgctgcatgaagacgccaaaaac	HrptermR	caggaaacagctatgaccatgatta
HB3UpsCF	gtatagacaaactagaccatattcc	HB3UpsCR	tttgattattcgggactacattatg
Ups+LucF	gtcccgaataatcaaaatgctgcatgaagacg	Luc+UpsR	cgtcttcatgcagcattttgattattcgggac
HRPtermR+Pst	agcttactgcagcaggaaacagctatga	HB3UpsCF+Spe	ccaggcactagtgtatagacaaactaga

HB3UpsBcenF	gatgcataaatgtgatgatgataa	HB3UpsBcenR	tttgattattcgggactacatgat
HB3UpsBR+Spe	tataacactagtgatgcataaatgtgatgatga	LucF2	atgcatgaagacgccaaaaacataag
HrpTermR2	cagtttaataaatatgttcttatatataa	HrpTerm+Pst-2	agcttactgcagcagtttaataaatatg
Ups+LucF-2	gtcccgaataatcaaaatgcatgaagacgcca	Luc+UpsR-2	tggcgtcttcatgcattttgattattcgggac
pGL-2LucF	aaatggaagacgccaaaaacataaa	UpsCF-2	aagtatagacaaactagaccatattc
pGC-2LucR	tttacaatttggactttccgccc	UpsCF-2+Spe	taatatactagtaagtatagacaaactag
pGL-2LucR+H3	ggtctcaagcttacaatttggac	UpsCF-3	gtgtattgttgggaatatgacaa
Ups+pGL2F	gtcccgaataatcaaaatggaagacgccaaaac	UpsCF-3+Spe	taatatactagtgtgtattgttgggaat
pGL2+UpsR	gttttggcgtcttccattttgattattcgggac		
699-1Rcloning	tgtgccgacaacatggtgtta	752-4Rcloning	tttcgttattggtgcactacat
699-1Fcloning	gttgttgacttactatatacaatg	752-4Fcloning	cagctagtggtaaaaacacac
699-1F+Spe	gatggcactagttgttgacttactat	752-4F+spe	ggtagcactagtcagctagtggtaaa
699-1+pGL2F	ccatgttgtcggcacaatggaagacgccaaaa	752-4+pGL2F	tgcaccaataacgaaaatggaagacgccaaaa
pGL2+699-1R	ttttggcgtcttccattgtgccgacaacatgg	pGL2+752-4R	ttttggcgtcttccattttcgttattggtgca

Sequencing primers

Oligo name	Sequence	Oligo name	Sequence
hrpseq1a	cttttacaatatgaacataaag	tRNAseq9	cacgaaagtattagaagcttttc
Lucseq2a	aattgttccaggaaccagggcg	gcseq10	gtccgtatatatgtttttata
1074-1seq3	catcaaaatattacaatacccc	1074-1seq11	ctgtatttgaattaggtatgc
1074-3seq4	ccttattgtttgttgtgcatac	1074-1seq12	ggtaatatttccttttttcttatc
1074-1seq5	aatacattatcatgtgttaattac	pGL2Lucseq13	gttccaggaaccagggcgtat
1074-3seq6	cttcatgtataatataccaaca	HRPseq14	ttctatatttataaggaag
tRNAseq7	caatctatgttgtaaaaatt	UpsBseq15	gtatatgatatagttataagtt
gcseq8	gtatatacaatggtatacata	UpsBseq16	ggttgggtgtatataaggctat

Screening

Oligo name	Sequence	Oligo name	Sequence
LUCscreenF	cactgataatgaattcctctgg	LUCscreenR	tcgtatccctggaagatggaa
UpsScreenF	gttggcatgtgaatgcatgca	UpsScreenR	ctcgaaatagtaatatgtctatata

Chapter 8

Oligo name	Sequence	Oligo name	Sequence
GCF	aagctgccccagtagcccagt	T7+GCF	taatacgactcactatagggaagctgccccag
GCR	ttgcgccacccccctccg	T7+GCR	taatacgactcactatagggttgcgccaccccc
SRPF	gtccggctactgttctttttaagttcag	SRPR	tggtgtacatattatacttatgag
T7SRPF	gaatttaatacgactcactatagggtccggctact	T7tRNAgluF	taatacgactcactatagggtcccacgtggtc
tRNAgluF	tcccacgtggtctagtggc	tRNAgluR	ttcccacgccgggaattgaa
T7+GCSF	taatacgactcactatagggcccggatcaggccgagaccattgccactgcagacgtgaagtatacactgctttgccccgaaggag		
GCSR	tcgaagcacatcgtcttagccgcttccaccacgtatggaaaaaatgccatcctaagaggcgtagacccacctccttcggggcaaagcagt		
GCScrambled full sequence	cccggatcaggccgagaccattgccactgcagacgtgaagtatacactgctttgccccgaaggaggtgggtctacgcctcttaggatggcatttttttccatacgtggtggaagcggctaagacgatgtgcttcga		

Appendix ii

Sequences of ncRNA identified in Chapter 3

Sequence identifier	Sequence
CDS-1	CATATCTCCCACCTTTGATTTATATATCAATAATTTATTTAATTTTATTCTATATTATTCAAAATGAAACTCTCCAAAATCTTATATTTCTTCGCAGCCCTCTTAGCTCTCAACTTCATTG
CDS-2	TAAATACCTATTTTACTAATTCTCATATTATTGTTTTCAAAGGAGAAAATCAG
CDS-3	TAAAAGAAAAAGAAAAAAAATGGAAGCTCAAAAATCTGTGTCCTT
CDS-4	TTGGAAGTCAAATACAAGACCAAGAAACCATATTAGGAGAAAGTGAACCCCNNNCNNCTACCTCTCCTCCAGAACAT
CDS-5	TATGTCAGGTTTTAAGTGTGGAATTATTATATTAGTTGTTAATACTGTTGGTGGGTTAT
CDS-6	TATGCCAAAAGATATCCAATTAGCCAGACGTATTCGTGGAGAAAGA

CDS-7	GATATAGTCCGTTTATTACTAAAACCACAAACAGTAATATTAACGGTCATACAATCCTTTATGCTAATGA
CDS-8a*	AAAAGCTTGTTTGATTTCTATTTTCA
CDS-9	AAATCACAAGATCCTACGAAATCTGTTCAAATACCAAAAGTTCCT
CDS-10	AAATACAATGCCAGTTATGCTTATTTTACTTTAATGAATATGATCGGAAAAAATCATGATATCTACTCAAAGGGTTCTCGTTTGCTTATGCATCATATATTTTAGGATTAGTCTTTTA
CDS-11	AATAATATCTCTATCGTTGGACAAGATGTGCCTAT
CDS-12	ATGGTAATATACTTTCTCCCTTCCAATTACATAATATTAATACTTGTAT
CDS-13	ATCTAGAGAACAATCCAAGACGCGCAATGAATCCTATATTCAGCAAAGAAATAGCTGAAATAAGATATGCCTTAAGCACAATTTCTTTTACCAT
CDS-14	ACCCTTCTCGGCCTCTTGGCTAAGATCAATTTGTAGTACCTGTTCTTATCAGCGTGATAGCTGATATGGTCTCAATAGAGGCCTTATCAATTTCACAAAAATTTTTGATAGGGGAAAGGATAATTTGAAGCTTGCTTCTTATAACCTTTCGCGCGTTGCCTTACCCTTGCACTAAAGGTTTGTACAGTGCACCCTTAAA
CDS-15	GAAAAAATATATTCTCCTGGTCGTACCACTGGTATTGTGTTAGAT
CDS-16	ATGGTTTAATTGTCGCTAA
CDS-17	TAAGAGTGATAATAATGTAGTGCCCACCATAAATACA
CDS-18	TCGGCTTGCTATGCCTTTAACCATAGAAGAATATAAAGTATGT
CDS-19a	AATAACATCCGTTCATTAAAAGGGGTAAAATATGAAAAAGTTAGGAAACAAGCAAGAAATGCAAT

CDS-19b	AATAACATCCGTTCATTAAAAGGGGTAAAATATGAAAAAGTTAGGAAACAAGCAAGAAATGCAAT
CDS-20a	TAGTCTTACTTTTGAGTTCAGCATCTAGAATGGGGAAAAGTAATGAAGAATATGATATAGGAGAATCTAATCT
CDS-20b	TAGTCTTACTTTTGAGTTCAGCATCTAGAATGGGGAAAAGTAATGAAGAATATGATATAGGAGAATCTAATCT
CDS-21	ATAACGCCTTTAGAACCATCAGAATGTATTGAAATAAATTTTGAAGAAACGTCCT
CDS-22	ATATTAAATCAAGATGTGGTTTACAACTAGACAAGACGGTCTTGACGATAATTGTCATACCAACAGAGGTCCATGTTTTCAAGT TAGTGGTT
INTER-1a	TTGCAAATCGAGCTGCTGAGTTACCCGGAGGGGGGTGGCGCCCA
INTER-1b	TTGCAAATCGAGCTGCTGAGTTACCCGGAGGGGGGTGGCGCCCA
INTER-1c	TTGCAAATCGAGCTGCTGAGTTACCCGGAGGGGGGTGGCGCCCA
INTER-2a	AAGCTGCCTCAGTAGCCCAGTCGTTAGGTAGGTAACTTTTCCTTGTGAGAACGTTGGTTCGACTCCGCTTGCCAACAATTTCA TAAGAAAAGTTGCAAATCGAGCTGCTGAGTTACCCGGAGGGGGGTGGCA
INTER-2b	AAGCTGCCTCAGTAGCCCAGTCGTTAGGTAGGTAACTTTTCCTTGTGAGAACGTTGGTTCGACTCCGCTTGCCAACAATTTCA TAAGAAAAGTTGCAAATCGAGCTGCTGAGTTACCCGGAGGGGGGTGGCA
INTER-3	GTGTGCTTCTTTATTTTGCATAGTGTGAGTA
CDS-8b*	AAAAGCTTGTTTGATTTCTATTTTCA
CDS-8c*	AAAAGCTTGTTTGATTTCTATTTTCA
INTER-4	TGATATGAAGAACTTGGTCTGTGTTACTGAA

INTER-5	TTGTGAATGATGAAAAAAGTTCTCAAATGTTCGCAGTTGAGAGTTTTTAGAAAAAAAGCTCAATGAATTTTTATTGTAGCCTTC ACTTAGATGTCTGACG
INTER-6	ATAAAATACATTTTANAAATAGAGAAAAAAAAAAAAAAAAAATATATATATATAAATATATATATATATATATCTACATGTAT
INTER-7	TTTTAATGATGATATGAAGAACTTGGTCTGTGTTACTGAAATAATATAGAGATGAAAAAAAAACGACTGAAAATCCT
INTER-8a	ATTTTTATTGTAGCCTTCACTTAGATGTCTA
INTER-8b	ATTTTTATTGTAGCCTTCACTTAGATGTCTA
INTRON-1	ATTACATGATGAGTAAGCTTCTACGAATCACGACGGTCTTCTGACATTATCAATGGAGATGGTAGAACGTTCTGAT
INTRON-2	TTTATATGATGACAAGTGACTATCCCAGCTCACTCTGATTTTTATTTTTAAAATGAAGAGAAAATAGCTCATATTATTTTATTAAT TTTTCTGATA
INTRON-3	TTTATATGATGATTAGTCTTGTCTGAATATTTAATTTTAAAAAATGAAGACAATAGTACTA
INTRON-4	TTTATATGATGATTAGTCTTGTCTGAATATTTAATTTTAAAAAATGAAGACAATAGTACGTG
INTRON-5	AATGAAATGATGAAACAAAACAGTTCTGCTTCTGAATTTATTTTGATGATAACTATGCCCAACTGATCTCCCA
INTRON-6	TATAAAAAAAAAGCACATTTGTTAGAACAGAATAAAAAGTGTGTTGAATTTTGTTAGTTTCTGN
INTRON-7	ATATATTACTTTATTTCATTTTTTGTTGTTTTATTAATACATAAAAAAAAATAGTATATTTCTTCAACAACA
INTRON-8	TTTTAATTTTTTCCTGGTCCAACACCTTTTTTGGT
MITO-1	ATGAGAGTTCACCGTTAGAAGCGATGCGTGAGCTGGGTTAAGAACGTCTTGAGGCAGTTTGTTCCCT

MITO-2	CCTTCATATATACTATGCTGACTTGAGTAATGATAAATTGATAGTATCAGCTATCCATAGTTAATTGATTCCGTTTTGACCGGTCATTAAAAAAAAAAA
MITO-3	TATGTCCTGTTTCAAATATATATATGAATAATTGTACGAATAGCAGATTGTGTTCATAGCTAGAGTACGTAAGGAAAAGGAAAGGTTAACCGCTATCAAAAAAAAAAAA
MITO-4	TAGCGTGTATTGTTGCCTTGTACACACCGCTCGTCACGCAATATCAATATACTGGGTATAAAAAAAAAAAA
MITO-5	TATTACCATACAAGAGATCGCGTACTTTGGACCGAATAAAGCTGTGAGGAAACTACACTAAAGGAACTCGACTGGCCTACAAAAAAAAAA
MITO-6	TGAGGCTCGGATATAAATGAGAAATATTTGATCCATACAGTCCCAGCGACAGCGGTTATACTTTGGAAGAGTCGAGTATTATCCATCCATGTCAGGCGTTAAAAGCGTTCGTTCTAAAAAAAAAAAAAAAAAAA
MITO-7	TGAGGCTCGGATATAAATGAGAAATATTTGATCCATACAGTCCCAGCGACAGCGGTTATACTTTGGAAGAGTCGAGTATTATCCATCCATGTCAGGCGTTAAAAGCGTTCGTTTTAAAAAAAAAAAAAAAAAAA
MITO-8	XXTTTGATCCATACAGTCCCAGCGACAGCGGTTATACTTTGGAAGAGTCGAGTATTATCCATCCATGTCAGGCGTTAAAAGCGTTCGTTCT
MITO-9	ATTACAAGATTGTGATAAGATGACATTTCTGAGTATTGAGCGGAACAAATCAGACCGTAAGGTTATAATTATGAAA
MITO-10	ATTTCTGAGTATTGAGCGGAACAAATCAGACCGTAAGGTTATAATTATGAAA
MITO-11	ATAGTTACCATAGCTGTAGATGGATGCTTCGATATATAGTATATTACAGTATCAATCGGATTTACATGCTCAGCCGCCAAAAAAAA
MITO-12	TGTTCGGTATTGCATGCCTGGTGTTAAA
MITO-13	ATTGGGAAGTTTAGCCAGGAAGTCAGCGTCTAAAAAAAAAAAAAAAAAA

MITO-14	GGAAAATTGGGAAGTTTAGCCAGGAAGTCAGCGTCTAAAAAAAAAAAAA
MITO-15	TGGCTGCTGGAAGTAGAAATCGTTATTAAGCGTCAGGAAGTCCTGGACGTTGAATCCAATAGCATTGAAAAAAAAAAAAA
MITO-16	AACGATTTCTACTTCAAAAAAAAAAAAAA
MITO-17	AACCAGTCGGTGCGAAGTCGTGACATGGTAGTTGACAGTGAACCTGTAGCTGAACCAAAAAATAAAAAACCT
MITO-18	AAACAGTCGGTGCGAAGTCGTAACATGGTAGTTGACAGTGAACTTGTAGCTGAACCAAAAAAAAAAAA
MITO-19	AAACAGTCGGTGCGAAGTCGTAACATGGTA
MITO-20	TAAGGATGAAACCTTCCTGATCGACTCGTGAGGTAAAAGAAAAAAAAAAAA
MITO-21	TGATAAATGGCGGCTGTATTTTAAACGGTCCTAAGGTAGCAAAATTCCTTGTCGGGTAATCTCCGTCCTGCATGAACGGTGTA ACGACTTCCCCATTGTCGCTAGTGTGAGACTCCTAATAAATAGAATTATCCATGAATATGTGGAATCATACGGCCCGACGGTA AGA
rRNA-1a	AACCTGGTTGATCTTGCCAGTAGTCATATGCTTGTCTCAAAGATTAAGCCATGCAAGTGAAAGTATATATATATTTTATATGTA GAAACTGCGAACGGCTCATTAAAACAGTTATAGTCTACTTGACATTTTTATTATAAGGATAACTACGGAAAAGCTGTAGCTAAT ACTTGCTTTATTATCCTTTGATT
rRNA-1b	AACCTGGTTGATCTTGCCAGTAGTCATATGCTTGTCTCAAAGATTAAGCCATGCAAGTGAAAGTATATATATATTTTATATGTA GAAACTGCGAACGGCTCATTAAAACAGTTATAGTCTACTTGACATTTTTATTATAAGGATAACTACGGAAAAGCTGTAGCTAAT ACTTGCTTTATTATCCTTTGATT
rRNA-2a	CACCAATTCCCTTCGGGGAATAGTGTTACTCCTACAGAAAAA
rRNA-2b	CACCAATTCCCTTCGGGGAATAGTGTTACTCCTACAGAAAAA

rRNA-3a	CTTAAGCCAGAAACCATGCTGATTAAACAATACTATTTCGTTCTTTTTTGTTTCCTCCCCA
rRNA-3b	CTTAAGCCAGAAACCATGCTGATTAAACAATACTATTTCGTTCTTTTTTGTTTCCTCCCCA
rRNA-4a	TCAGAATGCTGAATGTAAACTACACCAATTCCCTTCGGGGAATAGTGGTACTCCTACAG
rRNA-4b	TCAGAATGCTGAATGTAAACTACACCAATTCCCTTCGGGGAATAGTGGTACTCCTACAG
rRNA-5a	ATAACTATGATTCCTTTAGTAACGGCGAGTGAACAAGGACTAGCTCAAACGGTTAATCCTTCGTTTGGTGCAAG
rRNA-5b	ATAACTATGATTCCTTTAGTAACGGCGAGTGAACAAGGACTAGCTCAAACGGTTAATCCTTCGTTTGGTGCAAG
rRNA-6a	GACTCGTTCATACTACAGTGGCCACACCAGATCCCATCAGAACTCTGAAGTTAAGCACTGTAAGGCTTGGCTAGTACTGAGG TGGGAGACCGCTCGGGAACACCAGGTGATGAGTCATTCCCTCCT
rRNA-6b	GACTCGTTCATACTACAGTGGCCACACCAGATCCCATCAGAACTCTGAAGTTAAGCACTGTAAGGCTTGGCTAGTACTGAGG TGGGAGACCGCTCGGGAACACCAGGTGATGAGTCATTCCCTCCT
rRNA-6c	GACTCGTTCATACTACAGTGGCCACACCAGATCCCATCAGAACTCTGAAGTTAAGCACTGTAAGGCTTGGCTAGTACTGAGG TGGGAGACCGCTCGGGAACACCAGGTGATGAGTCATTCCCTCCT
rRNA-7a	TTAGCACAGTCTTAACGATGGATGTCTTGGTTCCTACAGCGATGAAGGCCGCAGCAAAATGCGATACGCAATGAAAATTGCA GTGACTGTGAATCATCAGAATGCTGAATGTAAACTACACCAATTCCCTTCGGGGAATAGTGGTACTCCTACAGAAAATCCCA
rRNA-7b	TTAGCACAGTCTTAACGATGGATGTCTTGGTTCCTACAGCGATGAAGGCCGCAGCAAAATGCGATACGCAATGAAAATTGCA GTGACTGTGAATCATCAGAATGCTGAATGTAAACTACACCAATTCCCTTCGGGGAATAGTGGTACTCCTACAGAAAATCCCA
rRNA-8a	TTAGCACAGTCTTAACGATGGATGTCTTGGTTCCTACTGCGATGAAGGCCGCAGCAAAATGCGATACGCAATGAAAATTGCA GTGACTGTGAATCATCAGAATGCTGAATGTAAACTACACCAATTCCCTTCGGGGAATAGTGGTACTCCTACAGAAAAT

rRNA-8b	TTAGCACAGTCTTAACGATGGATGTCTTGGTTCCTACTGCGATGAAGGCCGCAGCAAAATGCGATACGCAATGAAAATTGCA GTGACTGTGAATCATCAGAATGCTGAATGTAAACTACACCAATTCCCTTCGGGGAATAGTGGTACTCCTACAGAAAAT
rRNA-9a	AATCTGCTTTTTCTTTTTCTTCTTTCNTTTTTTCATCCCCCNCNNNTCA
rRNA-9b	AATCTGCTTTTTCTTTTTCTTCTTTCNTTTTTTCATCCCCCNCNNNTCA
rRNA-10a	TTATTAAAATATCCTTTTCCCTGTTCTACTAATAATTTGTTTTTTACTCTATTTCTCTCTTCTTTTAAGAATGTACTTGCTTGATTG AAAAGCTTCTTAGAGGAACATTGTGTGTCTAACACAAGGAAGTTTAAGGCAACAACAGGTCTGTGATGTCCTTAGATGAACTA GGCTGCACGCGTGCTACACTGATATATATAACGGGTA
rRNA-10b	TTATTAAAATATCCTTTTCCCTGTTCTACTAATAATTTGTTTTTTACTCTATTTCTCTCTTCTTTTAAGAATGTACTTGCTTGATTG AAAAGCTTCTTAGAGGAACATTGTGTGTCTAACACAAGGAAGTTTAAGGCAACAACAGGTCTGTGATGTCCTTAGATGAACTA GGCTGCACGCGTGCTACACTGATATATATAACGGGTA
rRNA-11a	TTAGCACAGTCTTAACGATGGATGTCTTGGTTCCTACAGCGATGAAGGCCGCAGCAAAATGCGATACGCAATGAAAATTGCA GTGACTGTGAATCATCAGAATGCTGAATGTAAACTACACCAATTCCCTTCGGGGAATAGTGGTACTCCTACAGAAAA
rRNA-11b	TTAGCACAGTCTTAACGATGGATGTCTTGGTTCCTACAGCGATGAAGGCCGCAGCAAAATGCGATACGCAATGAAAATTGCA GTGACTGTGAATCATCAGAATGCTGAATGTAAACTACACCAATTCCCTTCGGGGAATAGTGGTACTCCTACAGAAAA
rRNA-12a	GACTCGTTCATACTACAGTGGCCACACCAGATCCCATCAGAACCACTCTGATCTGAAGTTAAGCACTGTAAGGCTTGGCTAG TACTGAGGTGGGAGACCGCTCGGGAACACCAGGTGATGAGTCA
rRNA-12b	GACTCGTTCATACTACAGTGGCCACACCAGATCCCATCAGAACCACTCTGATCTGAAGTTAAGCACTGTAAGGCTTGGCTAG TACTGAGGTGGGAGACCGCTCGGGAACACCAGGTGATGAGTCA
rRNA-12c	GACTCGTTCATACTACAGTGGCCACACCAGATCCCATCAGAACCACTCTGATCTGAAGTTAAGCACTGTAAGGCTTGGCTAG TACTGAGGTGGGAGACCGCTCGGGAACACCAGGTGATGAGTCA

rRNA-13a	CTTTCGATTTGCTTGTAATCTGCTTTTTCTTTTTCTTCTTTCTTTTTTTCA
rRNA-13b	CTTTCGATTTGCTTGTAATCTGCTTTTTCTTTTTCTTCTTTCTTTTTTTCA
rRNA-14a	GTTAGCCGCTCCTAAGGGATAGCTGAAAAGTGTTTAAAAGGGGT
rRNA-14b	GTTAGCCGCTCCTAAGGGATAGCTGAAAAGTGTTTAAAAGGGGT
rRNA-14c	GTTAGCCGCTCCTAAGGGATAGCTGAAAAGTGTTTAAAAGGGGT
rRNA-15a	TCTACTAATAATTTGTTTTTTACTCTATTTCTCTCTTCTTTTAAGAATGTACTTGCTTGAT
rRNA-15b	TCTACTAATAATTTGTTTTTTACTCTATTTCTCTCTTCTTTTAAGAATGTACTTGCTTGAT
CDS-8d*	AAAAGCTTGTTTGATTTCTATTTTCA
CDS-8e*	AAAAGCTTGTTTGATTTCTATTTTCA
tRNA-1	TGCATTCAAGAGGTCCGGGGTTCAATTCCCCGTCTGTCCNCCACCTCNCTCNCCGC
tRNA-2	GCGTCAATTGTCTAACGGCCATGATA
tRNA-3	TCCCACGTGGTCTAGTGGCTAGGATATTCGGCTCTCACCCGAAAGGCCCGGGTTCAATTCCCGGCGTGGGAA
tRNA-4	TAAGCGGAAGGTCATCGGTTCGTCTCCGGTTGTAGGA
A-1	CCTCCCGACCCCTCCACCCGCCCATCCCATCCCCCGCCG
A-2	CGGGGCGGGGCGGTTCGTCCCCCCGCCCTACCCCCCCGGCCCCGTC
A-3	CTCTCTCTCTCCCCCGCTCCCCGTCCT

A-4	CCTCGGATAGCCGGTCCCCCGCCTA
A-5	AGTTCACCCCTGCGGTGCACGCCTCCCTGGACAAGTTCCTGGCTTCTGTGAGCACCGTGCTGACCTCCAAATACCGTTAAG CTGGAGCCTCGGTGGCCATGCTTCTTGCCCCTTGGGCCTCCCCCCAACCCT
A-6	CAAAACCCCAAAACACACAGCAAAACCACCACCACCAAACAACCCAAACAAAACCACACCACACCAAAACAAAAAACAAACA CAA
A-7	CCAACCCCCACACACACACAAACCCACCCGACAAACAACCCCAACACCACCACCCAACAAAACACACCCCCAACCAA
A-8	CCACCCACCCACACAACCCAACCCACAACCCCCCAACAACCACCAACCCCAACAACCACG
A-9	CCCAACACAACTCCAACCCCAAAAACAAAACCAAACACAAACCCCCACCAACAAACACCCACAAAACAAA
A-10	AGTGGTACTCNTACAGAAANN
A-11	TCCTCCCGACCT
A-12	GCCCACAAACCACCCACCCAAACCACA
A-13	CTTCTGTCACCA
A-14	GCCTGTCAACGATACGCTACGTAACGGCTGAATCCCACGCAACCTCCTTCAGTCCATGTCAGTGTCCTCGTGCTCCAGTCG
MULTI-1	ACACAAACAACAAA
MULTI-2	CCCACCCCCCCCG
MULTI-3	CAAAAACAAAACAACACAAAA
MULTI-4	CCAACCAAACCCACACCAACCCT

MULTI-5	AACACAACAAACACCAACCAACA
MULTI-6	CACCCCAAAAACCAAACAACCCACA
MULTI-7	CCACACACCAACCCCACA
MULTI-8	NNNGNAAAAAAGANCCCAACACCAACCACTAAACCA
MULTI-9	TCCCGCCCCCCGTTCCA
MULTI-10	CCCCCACCCCACCAACCCCAACT
MULTI-11	CCCCCACCCCCACAACCA
MULTI-12	CGCCCACCCACTCCCT
MULTI-13	TTCGACTGGAGCACGAGGACACTGACATGGACTGAAGGGGTAGAAAATAA
MULTI-14	ACAAAAAAAACCCCG
MULTI-15	CCCACCCCTCCTCCCCGCA
MULTI-16	CCTCCCGACCCTCCAC
MULTI-17	CCTTCTCAGTACATGCACTTGACTAG
MULTI-18	CGCCCCGCTCCCCA
SHORT-1	CACAAA
SHORT-2	CCCCACCCA

SHORT-3	CCTCCCGA
SHORT-4	TTAAA

Appendix iii

Proteomics data

Those rows in bold are proteins with significant scores, are within the size range and are not present in the control samples. These data are also in tables 8.3-8.5

Experiment 1

RNA 1 < 24 kDa significant score >32

ID	score	Protein	size kDa	no. of different peptides	seq coverage %
PF3D7_0316800	59	40S ribosomal protein S15A, putative	14.8	1	6
PF3D7_1421200	55	ribosomal protein S25, putative	15.7	1	7
PF3D7_1038000.1	49	antigen UB05	13.8	1	5
PF3D7_0320900	38	Histone H2A.Z	16.5	2	10
PF3D7_1105100	38	histone H2B	13.1	1	7

PF3D7_0503800	33	60S ribosomal subunit protein L31, putative	14	1	6
PF3D7_1406600	28	ATP-dependent Clp protease, putative	156	2	2
PF3D7_1462800	28	glyceraldehyde-3-phosphate dehydrogenase	36.6	1	3
PF3D7_1443900	22	heat shock protein 90, putative	107	1	1
PF3D7_0207900	20	Serine repeat antigen 2	124	1	0

RNA 1 scrambled significant score >32

PF3D7_1433400	**24**	**membrane protein, unknown function**	**658**	**2**	**0**
PF3D7_0207900	20	Serine repeat antigen 2	124	1	0

RNA 1 control significant score >32

PF3D7_1443900	**22**	**heat shock protein 90, putative**	**107**	**1**	**1**

RNA 2 24-25 kDa significant score >30

PF3D7_1465900	46	40S ribosomal protein S3, putative	25	1	4
PF3D7_0721100	45	hypothetical protein	29.2	3	13

PF3D7_1010600	39	eukaryotic translation initiation factor 2, beta, putative	25.3	2	10
PF3D7_0807100	34	RNA helicase, putative	157	2	1
PF3D7_1320600	29	small GTPase Rab11	24.7	2	8
PF3D7_0207900	27	serine repeat antigen 2	124	1	0
PF3D7_1443900	24	heat shock protein 90, putative	107	2	1

RNA 2 scrambled significant score >30

PF3D7_0207900	**27**	**serine repeat antigen 2**	**124**	**1**	**0**
PF3D7_1421300	24	hypothetical protein	110	1	2
PF3D7_0701900	23	exported protein, unknown function	113	1	2

RNA 2 control significant score >30

PF3D7_1220400	**33**	**debranching enzyme-associated ribonuclease, putative**	**63**	**1**	**2**
PF3D7_1229800	14	myosin d	266	2	0

RNA 3 60-70 kDa significant score >30

PF3D7_0818900	**165**	**heat shock 70 kDa protein**	**74**	**1**	**1**
PF3D7_0102900	87	aspartate--tRNA ligase	74	2	2
PF3D7_0612100	77	eukaryotic translation initiation factor 3 subunit L, putative	78	2	3
PF3D7_0207900	22	serine repeat antigen 2	124	1	0

RNA 3 scrambled significant score >30

PF3D7_0216700.1	**47**	**hypothetical protein**	**167**	**2**	**1**
PF3D7_0506500	33	hypothetical protein	116	1	1
PF3D7_0307400	31	ATP-dependent CLP protease, putative	43	1	5
PF3D7_0818900	27	heat shock 70 kDa protein	74	4	7

No hits in RNA 3 control

Experiment 2

RNA 1 20 kDa significant score >37

P25407	**50**	**STI1-like protein, flags, fragment**	**29.5**	**5**	
Q8ILC1(PF14_0324)	33	STI1-like protein PF14_0324 now annotated as PF3D7 1434300= Hsp70/Hsp90 organizing protein (HOP)	66.4	7	

RNA 2 50 kDa significant score >37

Q00080	**45**	**Elongation Factor 1-alpha**	**49**	**9**	**25**

Experiment 3

RNA 1 >52 kDa significant score >22

PF3D7_0424200	**37**	**reticulocyte binding protein homologue 4**	**203**	**10**	**3**
PF3D7_0818900	30	heat shock protein 70 (Hsp70-1)	74.4	6	7
PF3D7_0318200	29	DNA-directed RNA polymerase II, putative	280	5	0
PF3D7_1220500	27	ribosome biogenesis protein TSR3, putative	38	1	2
PF3D7_0406500	27	conserved protein unknown function	403	6	1
PF3D7_0421100 (PFD1005c)	27	VAR erythrocyte membrane protein 1 (PfEMP1)	252	5	2
PF3D7_0804600	26	U2 snRNA/tRNA pseudouridine synthase, putative	64	1	1
PF3D7_1122500	26	conserved protein unknown function	232	3	1
PF3D7_1234000	24	conserved protein unknown function	51	1	1
PF3D7_0521300	19	conserved protein unknown function	164.5	2	0
PF3D7_0220100	18	DnaJ protein, putative			

PF3D7_0511300	17	MORN repeat protein, putative			
PF3D7_1471600	16	conserved protein unknown function			
PF3D7_1018300	14	conserved protein unknown function			

RNA 1 control significant score >22

PF3D7_0818900	**444**	**HSP70-1**	**74.3**	**16**	**20**
PF3D7_0917900	124	Hsp70-2	72.5	8	13
PF3D7_1134000	48	Hsp70-3	73.6	11	14
PF3D7_0424200	41	reticulocyte binding protein homologue 4	20.3	13	5
PF3D7_0420900 (PFD1000c)	39	VAR erythrocyte membrane protein 1 (PfEMP1)	254.5	4	1
PF3D7_0603400	39	conserved protein unknown function	133.5	3	2
PF3D7_0511500	35	RNA pseudouridylate synthase, putative	1195	16	1
PF3D7_0308900	34	splicing factor 3B subunit 1, putative	160	2	1

PF3D7_1322300	29	translation initiation factor EIF-2B subunit related	179	4	2
PF3D7_0721100	29	conserved protein unknown function	29.6	3	10
PF3D7_1220500	27	ribosome biogenesis protein TSR3, putative	38.6	2	5
PF3D7_1241300	27	conserved protein unknown function	67	1	1
PF3D7_1444500	26	eukaryotic initiation factor 2alpha kinase 1	185.5	2	0
PF3D7_0412700 (PFD0625c)	26	VAR erythrocyte membrane protein 1 (PfEMP1)	260	3	1
PF3D7_1452600	25	conserved protein unknown function	403	7	1
PF3D7_1360700	25	SUMO ligase, putative	65.1	2	3
PF3D7_1314200	24	telomerase reverse transcriptase, putative	306	10	1
PF3D7_1116000 PF3D7_1116100	24	rhoptry neck protein 4 serine esterase, putative	437	1	0
PF3D7_1313800	23	conserved Plasmodium membrane protein, unknown function	407	4	0
PF3D7_0205300	20	conserved protein unknown function	73.5	4	6
PF3D7_1361800	18	conserved protein unknown function	294	2	0

PF3D7_0610800	17	transketolase	76	1	1

RNA2 40 kDa >22 significant score

PF3D7_1357000	**1752**	**elongation factor 1-alpha**	**49**	**37**	**62**
PF3D7_1410600	75	eukaryotic translation initiation factor 2 gamma subunit, putative	51.9	10	20
PF3D7_0424200	53	reticulocyte binding protein homologue 4			
PF3D7_0715400	50	secreted ookinete protein, putative	24.6	1	3
PF3D7_0916700	40	RNA-binding protein musashi, putative			
PF3D7_0511500	35	RNA pseudouridylate synthase, putative	1195	10	2
PF3D7_1434500	35	dynein-related AAA-type ATPase, putative			
PF3D7_0405400	30	pre-mRNA-processing-splicing factor 8, putative	368	5	2
PF3D7_1338300	30	elongation factor 1-gamma, putative	51	10	16
PF3D7_0603400	29	conserved Plasmodium protein, unknown function	133.5	5	3
PF3D7_1423000	27	GTPase, putative	57.4	4	3
PF3D7_1015900	27	enolase			
PF3D7_0420900	25	VAR erythrocyte membrane protein 1 (PfEMP1)	254.5	4	1

(PFD1000c)					
PF3D7_0701900	25	Plasmodium exported protein, unknown function	113	5	5
PFC0500w	24	conserved protein, unknown function	74	2	2
PF3D7_1411400	23	plastid replication-repair enzyme	237	6	1
PF3D7_0301300	23	alpha/beta hydrolase, putative	51.6	1	1
PF3D7_1116000 PF3D7_1116100	22	rhoptry neck protein 4 serine esterase, putative	347	2	0
PF3D7_0821300	22	ATP-dependent RNA helicase prh1, putative	102.3	4	4
PF3D7_0707300	22	rhoptry-associated membrane antigen	95.6	4	4
PF3D7_0827900	22	protein disulfide isomerase			
PF3D7_1208100	20	conserved protein, unknown function			
PF3D7_0318200	18	DNA-directed RNA polymerase II, putative			
PF3D7_1475400	18	cysteine repeat modular protein 4			
PF3D7_1222000	14	conserved protein, unknown function			

RNA 2 control >23 significant score

PF3D7_101590	**1078**	**enolase**
PF3D7_1357000	741	elongation factor 1-alpha
PF3D7_0827900	96	protein disulfide isomerase
PF3D7_0903700	42	alpha tubulin
PF3D7_0424200	33	reticulocyte binding protein homologue 4
PF3D7_1235600	32	Serine hydroxymethyltransferase
PF3D7_0916700	31	RNA-binding protein musashi, putative
PF3D7_1122500	30	conserved Plasmodium protein, unknown function
PF3D7_1241300	29	conserved Plasmodium protein, unknown function
PF3D7_0406500	29	conserved Plasmodium protein, unknown function
PF3D7_0520800	28	conserved Plasmodium protein, unknown function
PF3D7_1212100	24	peripheral plastid protein 1, putative
PF3D7_1128000	24	conserved Plasmodium protein, unknown function

PF3D7_1118600	16	histone acetyltransferase, putative (MYST)			
PF3D7_0822300	15	small nuclear ribonucleoprotein G, putative			

RNA 3 17-31 kDa >22 significant score

PF3D7_0715400	**60**	secreted ookinete protein, putative			
PF3D7_0516200	58	40S ribosomal protein S14, putative	16	2	15
PF3D7_0424200	41	reticulocyte binding protein homologue 4			
PF3D7_0934500	40	vacuolar ATP synthase subunit e, putative			
PF3D7_1317800	34	40S ribosomal protein S15/S19, putative	17	1	7
PF3D7_1304500	34	small heat shock protein, putative	25.2	2	10
PF3D7_0630600	31	conserved Plasmodium protein, unknown function	112	3	1
PF3D7_0930600.1	30	peptidyl-prolyl cis-trans isomerase	73	1	1

PF3D7_1326800	29	syntaxin, Qa-SNARE family	52.5	2	2
PF3D7_0514900	29	conserved Plasmodium protein, unknown function	52.2	3	5
PF3D7_0406500	27	conserved Plasmodium protein, unknown function	403	6	1
PF3D7_0617800	25	histone H2A	14.1	1	6
PF3D7_0922400	25	para-aminobenzoic acid synthetase	117.5	1	0
PF3D7_1411400	23	plastid replication-repair enzyme			
PF3D7_1224200	18	conserved Plasmodium protein, unknown function			

RNA 3 control >23 significant score

PF3D7_1120100	**573**	**phosphoglycerate mutase, putative**
PF3D7_1117700	367	GTP-binding nuclear protein ran/tc4
PF3D7_1439900	282	triosephosphate isomerase
PF3D7_1012400	263	hypoxanthine-guanine phosphoribosyltransferase
PF3D7_0513300	257	purine nucleoside phosphorylase

PF3D7_1465900	219	40S ribosomal protein S3, putative
PF3D7_1343000	211	phosphoethanolamine N-methyltransferase
PF3D7_0934500	115	vacuolar ATP synthase subunit e, putative
PF3D7_1474800	113	proteosome subunit alpha type 1, putative
PF3D7_1338200	109	60S ribosomal protein L6-2, putative
PF3D7_1441200	96	60S ribosomal protein L1, putative
PF3D7_1008900	96	adenylate kinase
PF3D7_1357900	77	pyrroline carboxylate reductase
PF3D7_1353900	77	proteasome subunit, putative
PF3D7_0807500	70	proteasome subunit alpha, putative
PF3D7_0715400	65	secreted ookinete protein, putative
PF3D7_1105400	64	40S ribosomal protein S4, putative
PF3D7_1302100	63	gamete antigen 27/25
PF3D7_1010600	60	eukaryotic translation initiation factor 2 beta subunit, putative

PF3D7_0317000	59	proteasome component C8, putative
PF3D7_0315100	55	translation initiation factor 4E
PF3D7_0424200	51	reticulocyte binding protein homologue 4
PF3D7_0727400	51	proteasome subunit alpha type 5, putative
PF3D7_0420900 (PFD1000c)	47	erythrocyte membrane protein 1, PfEMP1
PF3D7_1351400	45	60S ribosomal protein L17, putative
PF3D7_1320600	45	Rab GTPase 11a
PF3D7_1142300	45	conserved Plasmodium membrane protein, unknown function
PF3D7_0707200	41	conserved Plasmodium protein, unknown function
PF3D7_1126700	38	conserved Plasmodium protein, unknown function
PF3D7_1024100	38	conserved Plasmodium protein, unknown function
PF3D7_1333000	38	20 kDa chaperonin
PF3D7_1353100	33	Plasmodium exported protein, unknown function

PF3D7_0511500	31	RNA pseudouridylate synthase, putative
PF3D7_1459600	31	conserved Plasmodium protein, unknown function
PF3D7_0627600	30	conserved Plasmodium protein, unknown function
PF3D7_0618200	30	conserved Plasmodium protein, unknown func
PF3D7_1411400	30	plastid replication-repair enzyme
PF3D7_0603400	29	conserved Plasmodium protein, unknown function
PF3D7_0814000	29	60S ribosomal protein L13-2, putative
PF3D7_0504000	29	cation transporting P-ATPase
PF3D7_1238100	28	calcyclin binding protein, putative
PF3D7_1348400	28	conserved Plasmodium membrane protein, unknown function
PF3D7_0822900	27	conserved Plasmodium protein, unknown function
PF3D7_1241300	27	conserved Plasmodium protein, unknown function
PF3D7_0721100	26	conserved Plasmodium protein, unknown function
PF3D7_1359900	26	conserved Plasmodium membrane protein, unknown function

PF3D7_0406900	26	conserved Plasmodium protein, unknown function
PF3D7_1401500	25	lysophospholipase, putative
PF3D7_1245100	25	kinesin-like protein, putative
PF3D7_0211300	25	conserved Plasmodium protein, unknown function
PF3D7_0412700 (PFD0625c)	24	erythrocyte membrane protein 1, PfEMP1
PF3D7_0527000	22	minchromosome maintenance (MCM) complex subunit, putative
PF3D7_1241200	21	conserved Plasmodium protein, unknown function
PF3D7_1041300 (PF10_0406)	20	erythrocyte membrane protein 1, PfEMP1
PF3D7_1211300	17	minichromosome maintenance protein, putative
PF3D7_1433600	16	microsomal signal peptidase protein, putative
PF3D7_0727200	15	cysteine desulfurase, putative

Bibliography

1. http://www.who.int/malaria/world_malaria_report_2011/wmr2011_. *summary_keypoints.pdf*.

2. http://www.who.int/mediacentre/factsheets/fs094/en/index.html.

3. http://www.nhs.uk/conditions/malaria/pages/introduction.aspx.

4. Singh, B., Kim Sung, L., Matusop, A., Radhakrishnan, A., Shamsul, S.S., Cox-Singh, J., Thomas, A. and Conway, D.J. (2004) A large focus of naturally acquired Plasmodium knowlesi infections in human beings. *Lancet*, **363**, 1017-1024.

5. Babady, N.E., Sloan, L.M., Rosenblatt, J.E. and Pritt, B.S. (2009) Detection of Plasmodium knowlesi by real-time polymerase chain reaction. *Am J Trop Med Hyg*, **81**, 516-518.

6. Sabbatani, S., Fiorino, S. and Manfredi, R. (2012) Plasmodium knowlesi: from Malaysia, a novel health care threat. *Infez Med*, **20**, 5-11.

7. White, N.J. (2008) Plasmodium knowlesi: the fifth human malaria parasite. *Clin Infect Dis*, **46**, 172-173.

8. http://www.cdc.gov/malaria/about/biology.

9. Bruce-Chwatt, L.J., Gilles, H.M. and Warrell, D.A. (1993) *Bruce-Chwatt's essential malariology*. 3rd ed. Edward Arnold, London.

10. Amaratunga, C., Lopera-Mesa, T.M., Brittain, N.J., Cholera, R., Arie, T., Fujioka, H., Keefer, J.R. and Fairhurst, R.M. (2011) A role for fetal hemoglobin and maternal immune IgG in infant resistance to Plasmodium falciparum malaria. *PLoS One*, **6**, e14798.

11. http://www.rbm.who.int/docs/hbsm.pdf.

12. Marsh, K., Forster, D., Waruiru, C., Mwangi, I., Winstanley, M., Marsh, V., Newton, C., Winstanley, P., Warn, P., Peshu, N. *et al.* (1995) Indicators of life-threatening malaria in African children. *N Engl J Med*, **332**, 1399-1404.

13. Doolan, D.L., Dobano, C. and Baird, J.K. (2009) Acquired immunity to malaria. *Clin Microbiol Rev*, **22**, 13-36, Table of Contents.

14. http://images.wellcome.ac.uk.

15. Mackintosh, C.L., Christodoulou, Z., Mwangi, T.W., Kortok, M., Pinches, R., Williams, T.N., Marsh, K. and Newbold, C.I. (2008) Acquisition of naturally occurring antibody responses to recombinant protein domains of Plasmodium falciparum erythrocyte membrane protein 1. *Malar J*, **7**, 155.

16. Newton, C.R., Hien, T.T. and White, N. (2000) Cerebral malaria. *J Neurol Neurosurg Psychiatry*, **69**, 433-441.

17. Luse, S.A. and Miller, L.H. (1971) Plasmodium falciparum malaria. Ultrastructure of parasitized erythrocytes in cardiac vessels. *Am J Trop Med Hyg*, **20**, 655-660.

18. Howard, R.J., Barnwell, J.W., Rock, E.P., Neequaye, J., Ofori-Adjei, D., Maloy, W.L., Lyon, J.A. and Saul, A. (1988) Two approximately 300 kilodalton Plasmodium falciparum proteins at the surface membrane of infected erythrocytes. *Mol Biochem Parasitol*, **27**, 207-223.

19. Smith, J.D., Chitnis, C.E., Craig, A.G., Roberts, D.J., Hudson-Taylor, D.E., Peterson, D.S., Pinches, R., Newbold, C.I. and Miller, L.H. (1995) Switches in expression of Plasmodium falciparum var genes correlate with changes in antigenic and cytoadherent phenotypes of infected erythrocytes. *Cell*, **82**, 101-110.

20. Su, X.Z., Heatwole, V.M., Wertheimer, S.P., Guinet, F., Herrfeldt, J.A., Peterson, D.S., Ravetch, J.A. and Wellems, T.E. (1995) The large diverse gene family var encodes proteins involved in cytoadherence and antigenic variation of Plasmodium falciparum-infected erythrocytes. *Cell*, **82**, 89-100.

21. Baruch, D.I., Pasloske, B.L., Singh, H.B., Bi, X., Ma, X.C., Feldman, M., Taraschi, T.F. and Howard, R.J. (1995) Cloning the P. falciparum gene encoding PfEMP1, a malarial variant antigen and adherence receptor on the surface of parasitized human erythrocytes. *Cell*, **82**, 77-87.

22. Taylor, D.W., Parra, M., Chapman, G.B., Stearns, M.E., Rener, J., Aikawa, M., Uni, S., Aley, S.B., Panton, L.J. and Howard, R.J. (1987) Localization of Plasmodium falciparum histidine-rich protein 1 in the erythrocyte skeleton under knobs. *Mol Biochem Parasitol*, **25**, 165-174.

23. Magowan, C., Coppel, R.L., Lau, A.O., Moronne, M.M., Tchernia, G. and Mohandas, N. (1995) Role of the Plasmodium falciparum mature-parasite-infected erythrocyte surface antigen (MESA/PfEMP-2) in malarial infection of erythrocytes. *Blood*, **86**, 3196-3204.

24. Oh, S.S., Voigt, S., Fisher, D., Yi, S.J., LeRoy, P.J., Derick, L.H., Liu, S. and Chishti, A.H. (2000) Plasmodium falciparum erythrocyte membrane protein 1 is anchored to the actin-spectrin junction and knob-associated histidine-rich protein in the erythrocyte skeleton. *Mol Biochem Parasitol*, **108**, 237-247.

25. Roberts, D.D., Sherwood, J.A., Spitalnik, S.L., Panton, L.J., Howard, R.J., Dixit, V.M., Frazier, W.A., Miller, L.H. and Ginsburg, V. (1985) Thrombospondin binds falciparum malaria parasitized erythrocytes and may mediate cytoadherence. *Nature*, **318**, 64-66.

26. Rasti, N., Wahlgren, M. and Chen, Q. (2004) Molecular aspects of malaria pathogenesis. *FEMS Immunol Med Microbiol*, **41**, 9-26.

27. Ockenhouse, C.F., Tandon, N.N., Magowan, C., Jamieson, G.A. and Chulay, J.D. (1989) Identification of a platelet membrane glycoprotein as a falciparum malaria sequestration receptor. *Science*, **243**, 1469-1471.

28. Berendt, A.R., Simmons, D.L., Tansey, J., Newbold, C.I. and Marsh, K. (1989) Intercellular adhesion molecule-1 is an endothelial cell adhesion receptor for Plasmodium falciparum. *Nature*, **341**, 57-59.

29. Smith, H., Nelson, J.A., Gahmberg, C.G., Crandall, I. and Sherman, I.W. (1992) Plasmodium falciparum: cytoadherence of malaria-infected erythrocytes to human brain capillary and umbilical vein endothelial cells--a comparative study of adhesive ligands. *Exp Parasitol*, **75**, 269-280.

30. Esslinger, C.W., Picot, S. and Ambroise-Thomas, P. (1994) Intra-erythrocytic Plasmodium falciparum induces up-regulation of inter-cellular adhesion molecule-1 on human endothelial cells in vitro. *Scand J Immunol*, **39**, 229-232.

31. Turner, G.D., Morrison, H., Jones, M., Davis, T.M., Looareesuwan, S., Buley, I.D., Gatter, K.C., Newbold, C.I., Pukritayakamee, S., Nagachinta, B. *et al.* (1994) An immunohistochemical study of the pathology of fatal malaria. Evidence for widespread endothelial activation and a potential role for intercellular adhesion molecule-1 in cerebral sequestration. *Am J Pathol*, **145**, 1057-1069.

32. Ochola, L.B., Siddondo, B.R., Ocholla, H., Nkya, S., Kimani, E.N., Williams, T.N., Makale, J.O., Liljander, A., Urban, B.C., Bull, P.C. *et al.* (2011) Specific receptor usage in Plasmodium falciparum cytoadherence is associated with disease outcome. *PLoS One*, **6**, e14741.

33. Treutiger, C.J., Heddini, A., Fernandez, V., Muller, W.A. and Wahlgren, M. (1997) PECAM-1/CD31, an endothelial receptor for binding Plasmodium falciparum-infected erythrocytes. *Nat Med*, **3**, 1405-1408.

34. Ockenhouse, C.F., Tegoshi, T., Maeno, Y., Benjamin, C., Ho, M., Kan, K.E., Thway, Y., Win, K., Aikawa, M. and Lobb, R.R. (1992) Human vascular endothelial cell adhesion receptors for Plasmodium falciparum-infected erythrocytes: roles for endothelial leukocyte adhesion molecule 1 and vascular cell adhesion molecule 1. *J Exp Med*, **176**, 1183-1189.

35. Fried, M. and Duffy, P.E. (1996) Adherence of Plasmodium falciparum to chondroitin sulfate A in the human placenta. *Science*, **272**, 1502-1504.

36. Flick, K., Scholander, C., Chen, Q., Fernandez, V., Pouvelle, B., Gysin, J. and Wahlgren, M. (2001) Role of nonimmune IgG bound to PfEMP1 in placental malaria. *Science*, **293**, 2098-2100.

37. Handunnetti, S.M., David, P.H., Perera, K.L. and Mendis, K.N. (1989) Uninfected erythrocytes form "rosettes" around Plasmodium falciparum infected erythrocytes. *Am J Trop Med Hyg*, **40**, 115-118.

38. David, P.H., Handunnetti, S.M., Leech, J.H., Gamage, P. and Mendis, K.N. (1988) Rosetting: a new cytoadherence property of malaria-infected erythrocytes. *Am J Trop Med Hyg*, **38**, 289-297.

39. Carlson, J. and Wahlgren, M. (1992) Plasmodium falciparum erythrocyte rosetting is mediated by promiscuous lectin-like interactions. *J Exp Med*, **176**, 1311-1317.

40. Rowe, J.A., Moulds, J.M., Newbold, C.I. and Miller, L.H. (1997) P. falciparum rosetting mediated by a parasite-variant erythrocyte membrane protein and complement-receptor 1. *Nature*, **388**, 292-295.

41. Chen, Q., Barragan, A., Fernandez, V., Sundstrom, A., Schlichtherle, M., Sahlen, A., Carlson, J., Datta, S. and Wahlgren, M. (1998) Identification of Plasmodium falciparum erythrocyte membrane protein 1 (PfEMP1) as the rosetting ligand of the malaria parasite P. falciparum. *J Exp Med*, **187**, 15-23.

42. Claessens, A., Ghumra, A., Gupta, A.P., Mok, S., Bozdech, Z. and Rowe, J.A. (2011) Design of a variant surface antigen-supplemented microarray chip for whole transcriptome analysis of multiple Plasmodium falciparum cytoadherent strains, and identification of strain-transcendent rif and stevor genes. *Malar J*, **10**, 180.

43. Rowe, A., Obeiro, J., Newbold, C.I. and Marsh, K. (1995) Plasmodium falciparum rosetting is associated with malaria severity in Kenya. *Infect Immun*, **63**, 2323-2326.

44. Carlson, J., Helmby, H., Hill, A.V., Brewster, D., Greenwood, B.M. and Wahlgren, M. (1990) Human cerebral malaria: association with erythrocyte rosetting and lack of anti-rosetting antibodies. *Lancet*, **336**, 1457-1460.

45. Treutiger, C.J., Hedlund, I., Helmby, H., Carlson, J., Jepson, A., Twumasi, P., Kwiatkowski, D., Greenwood, B.M. and Wahlgren, M. (1992) Rosette formation in Plasmodium falciparum isolates and anti-rosette activity of sera from Gambians with cerebral or uncomplicated malaria. *Am J Trop Med Hyg*, **46**, 503-510.

46. Roberts, D.J., Craig, A.G., Berendt, A.R., Pinches, R., Nash, G., Marsh, K. and Newbold, C.I. (1992) Rapid switching to multiple antigenic and adhesive phenotypes in malaria. *Nature*, **357**, 689-692.

47. Pain, A., Ferguson, D.J., Kai, O., Urban, B.C., Lowe, B., Marsh, K. and Roberts, D.J. (2001) Platelet-mediated clumping of Plasmodium falciparum-infected erythrocytes is a common adhesive phenotype and is associated with severe malaria. *Proc Natl Acad Sci U S A*, **98**, 1805-1810.

48. Biswas, A.K., Hafiz, A., Banerjee, B., Kim, K.S., Datta, K. and Chitnis, C.E. (2007) Plasmodium falciparum uses gC1qR/HABP1/p32 as a receptor to bind to vascular endothelium and for platelet-mediated clumping. *PLoS Pathog*, **3**, 1271-1280.

49. Chotivanich, K., Sritabal, J., Udomsangpetch, R., Newton, P., Stepniewska, K.A., Ruangveerayuth, R., Looareesuwan, S., Roberts, D.J. and White, N.J. (2004) Platelet-induced autoagglutination of Plasmodium

falciparum-infected red blood cells and disease severity in Thailand. *J Infect Dis*, **189**, 1052-1055.

50. Wassmer, S.C., Taylor, T., Maclennan, C.A., Kanjala, M., Mukaka, M., Molyneux, M.E. and Grau, G.E. (2008) Platelet-induced clumping of Plasmodium falciparum-infected erythrocytes from Malawian patients with cerebral malaria-possible modulation in vivo by thrombocytopenia. *J Infect Dis*, **197**, 72-78.

51. Arman, M., Raza, A., Tempest, L.J., Lyke, K.E., Thera, M.A., Kone, A., Plowe, C.V., Doumbo, O.K. and Rowe, J.A. (2007) Platelet-mediated clumping of Plasmodium falciparum infected erythrocytes is associated with high parasitemia but not severe clinical manifestations of malaria in African children. *Am J Trop Med Hyg*, **77**, 943-946.

52. Arman, M. and Rowe, J.A. (2008) Experimental conditions affect the outcome of Plasmodium falciparum platelet-mediated clumping assays. *Malar J*, **7**, 243.

53. Marsh, K. and Howard, R.J. (1986) Antigens induced on erythrocytes by P. falciparum: expression of diverse and conserved determinants. *Science*, **231**, 150-153.

54. Bull, P.C., Lowe, B.S., Kortok, M., Molyneux, C.S., Newbold, C.I. and Marsh, K. (1998) Parasite antigens on the infected red cell surface are targets for naturally acquired immunity to malaria. *Nat Med*, **4**, 358-360.

55. Fried, M., Nosten, F., Brockman, A., Brabin, B.J. and Duffy, P.E. (1998) Maternal antibodies block malaria. *Nature*, **395**, 851-852.

56. Salanti, A., Staalsoe, T., Lavstsen, T., Jensen, A.T., Sowa, M.P., Arnot, D.E., Hviid, L. and Theander, T.G. (2003) Selective upregulation of a single distinctly structured var gene in chondroitin sulphate A-adhering Plasmodium falciparum involved in pregnancy-associated malaria. *Mol Microbiol*, **49**, 179-191.

57. Gardner, M.J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R.W., Carlton, J.M., Pain, A., Nelson, K.E., Bowman, S. *et al.* (2002) Genome sequence of the human malaria parasite Plasmodium falciparum. *Nature*, **419**, 498-511.

58. Hall, N., Pain, A., Berriman, M., Churcher, C., Harris, B., Harris, D., Mungall, K., Bowman, S., Atkin, R., Baker, S. *et al.* (2002) Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13. *Nature*, **419**, 527-531.

59. Miller, L.H., Mason, S.J., Dvorak, J.A., McGinniss, M.H. and Rothman, I.K. (1975) Erythrocyte receptors for (Plasmodium knowlesi) malaria: Duffy blood group determinants. *Science*, **189**, 561-563.

60. Miller, L.H., Mason, S.J., Clyde, D.F. and McGinniss, M.H. (1976) The resistance factor to Plasmodium vivax in blacks. The Duffy-blood-group genotype, FyFy. *N Engl J Med*, **295**, 302-304.

61. Smith, J.D., Subramanian, G., Gamain, B., Baruch, D.I. and Miller, L.H. (2000) Classification of adhesive domains in the Plasmodium falciparum erythrocyte membrane protein 1 family. *Mol Biochem Parasitol*, **110**, 293-310.

62. Rask, T.S., Hansen, D.A., Theander, T.G., Gorm Pedersen, A. and Lavstsen, T. (2010) Plasmodium falciparum erythrocyte membrane protein 1 diversity in seven genomes--divide and conquer. *PLoS Comput Biol*, **6**.

63. Calderwood, M.S., Gannoun-Zaki, L., Wellems, T.E. and Deitsch, K.W. (2003) Plasmodium falciparum var genes are regulated by two regions with separate promoters, one upstream of the coding region and a second within the intron. *J Biol Chem*, **278**, 34125-34132.

64. Epp, C., Li, F., Howitt, C.A., Chookajorn, T. and Deitsch, K.W. (2009) Chromatin associated sense and antisense noncoding RNAs are transcribed from the var gene family of virulence genes of the malaria parasite Plasmodium falciparum. *Rna*, **15**, 116-127.

65. Voss, T.S., Thompson, J.K., Waterkeyn, J., Felger, I., Weiss, N., Cowman, A.F. and Beck, H.P. (2000) Genomic distribution and functional characterisation of two distinct and conserved Plasmodium falciparum var gene 5' flanking sequences. *Mol Biochem Parasitol*, **107**, 103-115.

66. Lavstsen, T., Salanti, A., Jensen, A.T., Arnot, D.E. and Theander, T.G. (2003) Sub-grouping of Plasmodium falciparum 3D7 var genes based on sequence analysis of coding and non-coding regions. *Malar J*, **2**, 27.

67. Jensen, A.T., Magistrado, P., Sharp, S., Joergensen, L., Lavstsen, T., Chiucchiuini, A., Salanti, A., Vestergaard, L.S., Lusingu, J.P., Hermsen, R. *et al.* (2004) Plasmodium falciparum associated with severe childhood malaria preferentially expresses PfEMP1 encoded by group A var genes. *J Exp Med*, **199**, 1179-1190.

68. Kyriacou, H.M., Stone, G.N., Challis, R.J., Raza, A., Lyke, K.E., Thera, M.A., Kone, A.K., Doumbo, O.K., Plowe, C.V. and Rowe, J.A. (2006) Differential var gene transcription in Plasmodium falciparum isolates from patients with cerebral malaria compared to hyperparasitaemia. *Mol Biochem Parasitol*, **150**, 211-218.

69. Doumbo, O.K., Thera, M.A., Kone, A.K., Raza, A., Tempest, L.J., Lyke, K.E., Plowe, C.V. and Rowe, J.A. (2009) High levels of Plasmodium falciparum rosetting in all clinical forms of severe malaria in African children. *Am J Trop Med Hyg*, **81**, 987-993.

70. Claessens, A., Adams, Y., Ghumra, A., Lindergard, G., Buchan, C.C., Andisi, C., Bull, P.C., Mok, S., Gupta, A.P., Wang, C.W. *et al.* (2012) A subset of group A-like var genes encodes the malaria parasite ligands for binding to human brain endothelial cells. *Proc Natl Acad Sci U S A*.

71. Avril, M., Tripathi, A.K., Brazier, A.J., Andisi, C., Janes, J.H., Soma, V.L., Sullivan, D.J., Jr., Bull, P.C., Stins, M.F. and Smith, J.D. (2012) A restricted subset of var genes mediates adherence of Plasmodium

falciparum-infected erythrocytes to brain endothelial cells. *Proc Natl Acad Sci U S A*.

72. Lavstsen, T., Turner, L., Saguti, F., Magistrado, P., Rask, T.S., Jespersen, J.S., Wang, C.W., Berger, S.S., Baraka, V., Marquard, A.M. *et al.* (2012) Plasmodium falciparum erythrocyte membrane protein 1 domain cassettes 8 and 13 are associated with severe malaria in children. *Proc Natl Acad Sci U S A*.

73. *www.broad.mit.edu/annotation/genome/plasmodium_falciparum_spp/MultiHome.html*

74. Kraemer, S.M., Kyes, S.A., Aggarwal, G., Springer, A.L., Nelson, S.O., Christodoulou, Z., Smith, L.M., Wang, W., Levin, E., Newbold, C.I. *et al.* (2007) Patterns of gene recombination shape var gene repertoires in Plasmodium falciparum: comparisons of geographically diverse isolates. *BMC Genomics*, **8**, 45.

75. Mu, J., Duan, J., Makova, K.D., Joy, D.A., Huynh, C.Q., Branch, O.H., Li, W.H. and Su, X.Z. (2002) Chromosome-wide SNPs reveal an ancient origin for Plasmodium falciparum. *Nature*, **418**, 323-326.

76. Ponnudurai, T., Leeuwenberg, A.D. and Meuwissen, J.H. (1981) Chloroquine sensitivity of isolates of Plasmodium falciparum adapted to in vitro culture. *Trop Geogr Med*, **33**, 50-54.

77. Bhasin, V.K. and Trager, W. (1984) Gametocyte-forming and non-gametocyte-forming clones of Plasmodium falciparum. *Am J Trop Med Hyg*, **33**, 534-537.

78. Mu, J., Awadalla, P., Duan, J., McGee, K.M., Joy, D.A., McVean, G.A. and Su, X.Z. (2005) Recombination hotspots and population structure in Plasmodium falciparum. *PLoS Biol*, **3**, e335.

79. http://www.genedb.org.

80. 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. *PLoS Biol*, **1**, E5.

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

82. Balaji, S., Babu, M.M., Iyer, L.M. and Aravind, L. (2005) Discovery of the principal specific transcription factors of Apicomplexa and their implication for the evolution of the AP2-integrase DNA binding domains. *Nucleic Acids Res*, **33**, 3994-4006.

83. Bischoff, E. and Vaquero, C. (2010) In silico and biological survey of transcription-associated proteins implicated in the transcriptional machinery during the erythrocytic development of Plasmodium falciparum. *BMC Genomics*, **11**, 34.

84. Duffy, M.F., Selvarajah, S., Josling, G.A. and Petter, M. (2012) The role of chromatin in Plasmodium gene expression. *Cell Microbiol*.

85. Trelle, M.B., Salcedo-Amaya, A.M., Cohen, A.M., Stunnenberg, H.G. and Jensen, O.N. (2009) Global histone analysis by mass spectrometry reveals a high content of acetylated lysine residues in the malaria parasite Plasmodium falciparum. *J Proteome Res*, **8**, 3439-3450.

86. Lopez-Rubio, J.J., Gontijo, A.M., Nunes, M.C., Issar, N., Hernandez Rivas, R. and Scherf, A. (2007) 5' flanking region of var genes nucleate histone modification patterns linked to phenotypic inheritance of virulence traits in malaria parasites. *Mol Microbiol*, **66**, 1296-1305.

87. Perez-Toledo, K., Rojas-Meza, A.P., Mancio-Silva, L., Hernandez-Cuevas, N.A., Delgadillo, D.M., Vargas, M., Martinez-Calvillo, S., Scherf, A. and Hernandez-Rivas, R. (2009) Plasmodium falciparum heterochromatin protein 1 binds to tri-methylated histone 3 lysine 9 and is linked to mutually exclusive expression of var genes. *Nucleic Acids Res*, **37**, 2596-2606.

88. Flueck, C., Bartfai, R., Volz, J., Niederwieser, I., Salcedo-Amaya, A.M., Alako, B.T., Ehlgen, F., Ralph, S.A., Cowman, A.F., Bozdech, Z. *et al.* (2009) Plasmodium falciparum heterochromatin protein 1 marks genomic loci linked to phenotypic variation of exported virulence factors. *PLoS Pathog*, **5**, e1000569.

89. Miao, J., Fan, Q., Cui, L., Li, X., Wang, H., Ning, G. and Reese, J.C. (2010) The MYST family histone acetyltransferase regulates gene expression and cell cycle in malaria parasite Plasmodium falciparum. *Mol Microbiol*, **78**, 883-902.

90. 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. *et al.* (2005) Telomeric heterochromatin propagation and histone acetylation control mutually exclusive expression of antigenic variation genes in malaria parasites. *Cell*, **121**, 25-36.

91. Volz, J.C., Bartfai, R., Petter, M., Langer, C., Josling, G.A., Tsuboi, T., Schwach, F., Baum, J., Rayner, J.C., Stunnenberg, H.G. *et al.* (2012) PfSET10, a Plasmodium falciparum methyltransferase, maintains the active var gene in a poised state during parasite division. *Cell Host Microbe*, **11**, 7-18.

92. Tonkin, C.J., Carret, C.K., Duraisingh, M.T., Voss, T.S., Ralph, S.A., Hommel, M., Duffy, M.F., Silva, L.M., Scherf, A., Ivens, A. *et al.* (2009) Sir2 paralogues cooperate to regulate virulence genes and antigenic variation in Plasmodium falciparum. *PLoS Biol*, **7**, e84.

93. Mancio-Silva, L., Rojas-Meza, A.P., Vargas, M., Scherf, A. and Hernandez-Rivas, R. (2008) Differential association of Orc1 and Sir2 proteins to telomeric domains in Plasmodium falciparum. *J Cell Sci*, **121**, 2046-2053.

94. Freitas-Junior, L.H., Bottius, E., Pirrit, L.A., Deitsch, K.W., Scheidig, C., Guinet, F., Nehrbass, U., Wellems, T.E. and Scherf, A. (2000) Frequent ectopic recombination of virulence factor genes in telomeric chromosome clusters of P. falciparum. *Nature*, **407**, 1018-1022.

95. Ralph, S.A., Scheidig-Benatar, C. and Scherf, A. (2005) Antigenic variation in Plasmodium falciparum is associated with movement of var loci between subnuclear locations. *Proc Natl Acad Sci U S A*, **102**, 5414-5419.

96. Duraisingh, M.T., Voss, T.S., Marty, A.J., Duffy, M.F., Good, R.T., Thompson, J.K., Freitas-Junior, L.H., Scherf, A., Crabb, B.S. and Cowman, A.F. (2005) Heterochromatin silencing and locus repositioning linked to regulation of virulence genes in Plasmodium falciparum. *Cell*, **121**, 13-24.

97. Dzikowski, R., Li, F., Amulic, B., Eisberg, A., Frank, M., Patel, S., Wellems, T.E. and Deitsch, K.W. (2007) Mechanisms underlying mutually exclusive expression of virulence genes by malaria parasites. *EMBO Rep*, **8**, 959-965.

98. Lopez-Rubio, J.J., Mancio-Silva, L. and Scherf, A. (2009) Genome-wide analysis of heterochromatin associates clonally variant gene regulation with perinuclear repressive centers in malaria parasites. *Cell Host Microbe*, **5**, 179-190.

99. Zhang, Q., Huang, Y., Zhang, Y., Fang, X., Claes, A., Duchateau, M., Namane, A., Lopez-Rubio, J.J., Pan, W. and Scherf, A. (2011) A critical role of perinuclear filamentous actin in spatial repositioning and mutually exclusive expression of virulence genes in malaria parasites. *Cell Host Microbe*, **10**, 451-463.

100. Frank, M., Dzikowski, R., Costantini, D., Amulic, B., Berdougo, E. and Deitsch, K. (2006) Strict pairing of var promoters and introns is required for var gene silencing in the malaria parasite Plasmodium falciparum. *J Biol Chem*, **281**, 9942-9952.

101. Voss, T.S., Kaestli, M., Vogel, D., Bopp, S. and Beck, H.P. (2003) Identification of nuclear proteins that interact differentially with Plasmodium falciparum var gene promoters. *Mol Microbiol*, **48**, 1593-1607.

102. Flueck, C., Bartfai, R., Niederwieser, I., Witmer, K., Alako, B.T., Moes, S., Bozdech, Z., Jenoe, P., Stunnenberg, H.G. and Voss, T.S. (2010) A major role for the Plasmodium falciparum ApiAP2 protein PfSIP2 in chromosome end biology. *PLoS Pathog*, **6**, e1000784.

103. Broadbent, K.M., Park, D., Wolf, A.R., Van Tyne, D., Sims, J.S., Ribacke, U., Volkman, S., Duraisingh, M., Wirth, D., Sabeti, P.C. *et al.* (2011) A global transcriptional analysis of Plasmodium falciparum malaria reveals a novel family of telomere-associated lncRNAs. *Genome Biol*, **12**, R56.

104. Witmer, K., Schmid, C.D., Brancucci, N.M., Luah, Y.H., Preiser, P.R., Bozdech, Z. and Voss, T.S. (2012) Analysis of subtelomeric virulence gene families in Plasmodium falciparum by comparative transcriptional profiling. *Mol Microbiol*.

105. Janssen, C.S., Phillips, R.S., Turner, C.M. and Barrett, M.P. (2004) Plasmodium interspersed repeats: the major multigene superfamily of malaria parasites. *Nucleic Acids Res*, **32**, 5712-5720.

106. Cunningham, D., Lawton, J., Jarra, W., Preiser, P. and Langhorne, J. (2010) The pir multigene family of Plasmodium: antigenic variation and beyond. *Mol Biochem Parasitol*, **170**, 65-73.

107. Sam-Yellowe, T.Y., Florens, L., Johnson, J.R., Wang, T., Drazba, J.A., Le Roch, K.G., Zhou, Y., Batalov, S., Carucci, D.J., Winzeler, E.A. *et al.* (2004) A Plasmodium gene family encoding Maurer's cleft membrane proteins: structural properties and expression profiling. *Genome Res*, **14**, 1052-1059.

108. Lavazec, C., Sanyal, S. and Templeton, T.J. (2006) Hypervariability within the Rifin, Stevor and Pfmc-2TM superfamilies in Plasmodium falciparum. *Nucleic Acids Res*, **34**, 6696-6707.

109. Tsarukyanova, I., Drazba, J.A., Fujioka, H., Yadav, S.P. and Sam-Yellowe, T.Y. (2009) Proteins of the Plasmodium falciparum two transmembrane Maurer's cleft protein family, PfMC-2TM, and the 130 kDa Maurer's cleft protein define different domains of the infected erythrocyte intramembranous network. *Parasitol Res*, **104**, 875-891.

110. Gardner, M.J., Tettelin, H., Carucci, D.J., Cummings, L.M., Aravind, L., Koonin, E.V., Shallom, S., Mason, T., Yu, K., Fujii, C. *et al.* (1998) Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum. *Science*, **282**, 1126-1132.

111. Kyes, S.A., Rowe, J.A., Kriek, N. and Newbold, C.I. (1999) Rifins: a second family of clonally variant proteins expressed on the surface of red cells infected with Plasmodium falciparum. *Proc Natl Acad Sci U S A*, **96**, 9333-9338.

112. Cheng, Q., Cloonan, N., Fischer, K., Thompson, J., Waine, G., Lanzer, M. and Saul, A. (1998) stevor and rif are Plasmodium falciparum multicopy gene families which potentially encode variant antigens. *Mol Biochem Parasitol*, **97**, 161-176.

113. Wang, C.W., Magistrado, P.A., Nielsen, M.A., Theander, T.G. and Lavstsen, T. (2009) Preferential transcription of conserved rif genes in two phenotypically distinct Plasmodium falciparum parasite lines. *Int J Parasitol*, **39**, 655-664.

114. Joannin, N., Abhiman, S., Sonnhammer, E.L. and Wahlgren, M. (2008) Sub-grouping and sub-functionalization of the RIFIN multi-copy protein family. *BMC Genomics*, **9**, 19.

115. Bultrini, E., Brick, K., Mukherjee, S., Zhang, Y., Silvestrini, F., Alano, P. and Pizzi, E. (2009) Revisiting the Plasmodium falciparum RIFIN family: from comparative genomics to 3D-model prediction. *BMC Genomics*, **10**, 445.

116. Petter, M., Haeggstrom, M., Khattab, A., Fernandez, V., Klinkert, M.Q. and Wahlgren, M. (2007) Variant proteins of the Plasmodium falciparum RIFIN family show distinct subcellular localization and developmental expression patterns. *Mol Biochem Parasitol*, **156**, 51-61.

117. Petter, M., Bonow, I. and Klinkert, M.Q. (2008) Diverse expression patterns of subgroups of the rif multigene family during Plasmodium falciparum gametocytogenesis. *PLoS One*, **3**, e3779.

118. Kaviratne, M., Khan, S.M., Jarra, W. and Preiser, P.R. (2002) Small variant STEVOR antigen is uniquely located within Maurer's clefts in Plasmodium falciparum-infected red blood cells. *Eukaryot Cell*, **1**, 926-935.

119. Blythe, J.E., Yam, X.Y., Kuss, C., Bozdech, Z., Holder, A.A., Marsh, K., Langhorne, J. and Preiser, P.R. (2008) Plasmodium falciparum STEVOR proteins are highly expressed in patient isolates and located in the surface membranes of infected red blood cells and the apical tips of merozoites. *Infect Immun*, **76**, 3329-3336.

120. Khattab, A., Bonow, I., Schreiber, N., Petter, M., Schmetz, C. and Klinkert, M.Q. (2008) Plasmodium falciparum variant STEVOR antigens are expressed in merozoites and possibly associated with erythrocyte invasion. *Malar J*, **7**, 137.

121. Niang, M., Yan Yam, X. and Preiser, P.R. (2009) The Plasmodium falciparum STEVOR multigene family mediates antigenic variation of the infected erythrocyte. *PLoS Pathog*, **5**, e1000307.

122. McRobert, L., Preiser, P., Sharp, S., Jarra, W., Kaviratne, M., Taylor, M.C., Renia, L. and Sutherland, C.J. (2004) Distinct trafficking and localization of STEVOR proteins in three stages of the Plasmodium falciparum life cycle. *Infect Immun*, **72**, 6597-6602.

123. Sanyal, S., Egee, S., Bouyer, G., Perrot, S., Safeukui, I., Bischoff, E., Buffet, P., Deitsch, K.W., Mercereau-Puijalon, O., David, P.H. *et al.* (2012) Plasmodium falciparum STEVOR proteins impact erythrocyte mechanical properties. *Blood*, **119**, e1-8.

124. Winter, G., Kawai, S., Haeggstrom, M., Kaneko, O., von Euler, A., Kawazu, S., Palm, D., Fernandez, V. and Wahlgren, M. (2005) SURFIN is a polymorphic antigen expressed on Plasmodium falciparum merozoites and infected erythrocytes. *J Exp Med*, **201**, 1853-1863.

125. Chakrabarti, K., Pearson, M., Grate, L., Sterne-Weiler, T., Deans, J., Donohue, J.P. and Ares, M., Jr. (2007) Structural RNAs of known and unknown function identified in malaria parasites by comparative genomics and RNA analysis. *Rna*, **13**, 1923-1939.

126. Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res*, **31**, 3406-3415.

127. Mattick, J.S. (2009) The genetic signatures of noncoding RNAs. *PLoS Genet*, **5**, e1000459.

128. Costa, F.F. (2010) Non-coding RNAs: Meet thy masters. *Bioessays*, **32**, 599-608.

129. Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M.C., Maeda, N., Oyama, R., Ravasi, T., Lenhard, B., Wells, C. *et al.* (2005) The transcriptional landscape of the mammalian genome. *Science*, **309**, 1559-1563.

130. Collins, L.J. and Penny, D. (2009) The RNA infrastructure: dark matter of the eukaryotic cell? *Trends Genet*, **25**, 120-128.

131. Gardner, P.P., Daub, J., Tate, J., Moore, B.L., Osuch, I.H., Griffiths-Jones, S., Finn, R.D., Nawrocki, E.P., Kolbe, D.L., Eddy, S.R. *et al.* (2011) Rfam: Wikipedia, clans and the "decimal" release. *Nucleic Acids Res*, **39**, D141-145.

132. http://www.noncode.org.

133. Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E. and Mello, C.C. (1998) Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature*, **391**, 806-811.

134. Lee, R.C., Feinbaum, R.L. and Ambros, V. (1993) The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell*, **75**, 843-854.

135. Lewis, B.P., Burge, C.B. and Bartel, D.P. (2005) Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell*, **120**, 15-20.

136. Lares, M.R., Rossi, J.J. and Ouellet, D.L. (2010) RNAi and small interfering RNAs in human disease therapeutic applications. *Trends Biotechnol*, **28**, 570-579.

137. Pratt, A.J. and MacRae, I.J. (2009) The RNA-induced silencing complex: a versatile gene-silencing machine. *J Biol Chem*, **284**, 17897-17901.

138. Rathjen, T., Nicol, C., McConkey, G. and Dalmay, T. (2006) Analysis of short RNAs in the malaria parasite and its red blood cell host. *FEBS Lett*, **580**, 5185-5188.

139. Rana, T.M. (2007) Illuminating the silence: understanding the structure and function of small RNAs. *Nat Rev Mol Cell Biol*, **8**, 23-36.

140. Behm-Ansmant, I., Rehwinkel, J., Doerks, T., Stark, A., Bork, P. and Izaurralde, E. (2006) mRNA degradation by miRNAs and GW182

requires both CCR4:NOT deadenylase and DCP1:DCP2 decapping complexes. *Genes Dev*, **20**, 1885-1898.

141. Iwasaki, S., Kawamata, T. and Tomari, Y. (2009) Drosophila argonaute1 and argonaute2 employ distinct mechanisms for translational repression. *Mol Cell*, **34**, 58-67.

142. Matera, A.G., Terns, R.M. and Terns, M.P. (2007) Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs. *Nat Rev Mol Cell Biol*, **8**, 209-220.

143. Fabrizio, P., Dannenberg, J., Dube, P., Kastner, B., Stark, H., Urlaub, H. and Luhrmann, R. (2009) The evolutionarily conserved core design of the catalytic activation step of the yeast spliceosome. *Mol Cell*, **36**, 593-608.

144. Upadhyay, R., Bawankar, P., Malhotra, D. and Patankar, S. (2005) A screen for conserved sequences with biased base composition identifies noncoding RNAs in the A-T rich genome of Plasmodium falciparum. *Mol Biochem Parasitol*, **144**, 149-158.

145. Mishra, P.C., Kumar, A. and Sharma, A. (2009) Analysis of small nucleolar RNAs reveals unique genetic features in malaria parasites. *BMC Genomics*, **10**, 68.

146. Raabe, C.A., Sanchez, C.P., Randau, G., Robeck, T., Skryabin, B.V., Chinni, S.V., Kube, M., Reinhardt, R., Ng, G.H., Manickam, R. *et al.* (2010) A global view of the nonprotein-coding transcriptome in Plasmodium falciparum. *Nucleic Acids Res*, **38**, 608-617.

147. Kyes, S., Christodoulou, Z., Pinches, R. and Newbold, C. (2002) Stage-specific merozoite surface protein 2 antisense transcripts in Plasmodium falciparum. *Mol Biochem Parasitol*, **123**, 79-83.

148. Gunasekera, A.M., Patankar, S., Schug, J., Eisen, G., Kissinger, J., Roos, D. and Wirth, D.F. (2004) Widespread distribution of antisense transcripts in the Plasmodium falciparum genome. *Mol Biochem Parasitol*, **136**, 35-42.

149. Lopez-Barragan, M.J., Lemieux, J., Quinones, M., Williamson, K.C., Molina-Cruz, A., Cui, K., Barillas-Mury, C., Zhao, K. and Su, X.Z. (2011) Directional gene expression and antisense transcripts in sexual and asexual stages of Plasmodium falciparum. *BMC Genomics*, **12**, 587.

150. Otto, T.D., Wilinski, D., Assefa, S., Keane, T.M., Sarry, L.R., Bohme, U., Lemieux, J., Barrell, B., Pain, A., Berriman, M. *et al.* New insights into the blood-stage transcriptome of Plasmodium falciparum using RNA-Seq. *Mol Microbiol*, **76**, 12-24.

151. Li, Y.Y., Qin, L., Guo, Z.M., Liu, L., Xu, H., Hao, P., Su, J., Shi, Y., He, W.Z. and Li, Y.X. (2006) In silico discovery of human natural antisense transcripts. *BMC Bioinformatics*, **7**, 18.

152. Lehner, B., Williams, G., Campbell, R.D. and Sanderson, C.M. (2002) Antisense transcripts in the human genome. *Trends Genet*, **18**, 63-65.

153. Figueiredo, L.M., Rocha, E.P., Mancio-Silva, L., Prevost, C., Hernandez-Verdun, D. and Scherf, A. (2005) The unusually large Plasmodium telomerase reverse-transcriptase localizes in a discrete compartment associated with the nucleolus. *Nucleic Acids Res*, **33**, 1111-1122.

154. Deng, X. and Meller, V.H. (2006) Non-coding RNA in fly dosage compensation. *Trends Biochem Sci*, **31**, 526-532.

155. http://www.sanger.ac.uk/resources/downloads/protozoa/plasmodium-reichenowi.html.

156. http://plasmodb.org/plasmo/.

157. Escalante, A.A. and Ayala, F.J. (1994) Phylogeny of the malarial genus Plasmodium, derived from rRNA gene sequences. *Proc Natl Acad Sci U S A*, **91**, 11373-11377.

158. Petes, T.D. (2001) Meiotic recombination hot spots and cold spots. *Nat Rev Genet*, **2**, 360-369.

159. Kooij, T.W., Carlton, J.M., Bidwell, S.L., Hall, N., Ramesar, J., Janse, C.J. and Waters, A.P. (2005) A Plasmodium whole-genome synteny map: indels and synteny breakpoints as foci for species-specific genes. *PLoS Pathog*, **1**, e44.

160. Trager, W. and Jensen, J.B. (1976) Human malaria parasites in continuous culture. *Science*, **193**, 673-675.

161. Walliker, D., Quakyi, I.A., Wellems, T.E., McCutchan, T.F., Szarfman, A., London, W.T., Corcoran, L.M., Burkot, T.R. and Carter, R. (1987) Genetic analysis of the human malaria parasite Plasmodium falciparum. *Science*, **236**, 1661-1666.

162. Haynes, J.D., Diggs, C.L., Hines, F.A. and Desjardins, R.E. (1976) Culture of human malaria parasites Plasmodium falciparum. *Nature*, **263**, 767-769.

163. Lambros, C. and Vanderberg, J.P. (1979) Synchronization of Plasmodium falciparum erythrocytic stages in culture. *J Parasitol*, **65**, 418-420.

164. Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular cloning : a laboratory manual*. 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.

165. Serazin, a. (2006), D. Phil. Thesis, University of Oxford.

166. Kyes, S., Pinches, R. and Newbold, C. (2000) A simple RNA analysis method shows var and rif multigene family expression patterns in Plasmodium falciparum. *Mol Biochem Parasitol*, **105**, 311-315.

167. Feinberg, A.P. and Vogelstein, B. (1983) A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal Biochem*, **132**, 6-13.

168. http://www.gelifesciences.com.

169. Rubio, J.P., Thompson, J.K. and Cowman, A.F. (1996) The var genes of Plasmodium falciparum are located in the subtelomeric region of most chromosomes. *EMBO J*, **15**, 4069-4077.

170. Su, X.Z., Wu, Y., Sifri, C.D. and Wellems, T.E. (1996) Reduced extension temperatures required for PCR amplification of extremely A+T-rich DNA. *Nucleic Acids Res*, **24**, 1574-1575.

171. www.technelysium.com.au/chromas_lite.html.

172. www.corbettresearch.com.

173. Untergasser, A., Nijveen, H., Rao, X., Bisseling, T., Geurts, R. and Leunissen, J.A. (2007) Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Res*, **35**, W71-74.

174. http://www.sanger.ac.uk/resources/software/artemis/.

175. http://blast.ncbi.nlm.nih.gov/Blast.cgi.

176. http://www.ebi.ac.uk/Tools/msa/clustalw2/.

177. http://www.nrbsc.org/gfx/genedoc/.

178. Mourier, T., Carret, C., Kyes, S., Christodoulou, Z., Gardner, P.P., Jeffares, D.C., Pinches, R., Barrell, B., Berriman, M., Griffiths-Jones, S. *et al.* (2008) Genome-wide discovery and verification of novel structured RNAs in Plasmodium falciparum. *Genome Res*, **18**, 281-292.

179. Klein, R.J., Misulovin, Z. and Eddy, S.R. (2002) Noncoding RNA genes identified in AT-rich hyperthermophiles. *Proc Natl Acad Sci U S A*, **99**, 7542-7547.

180. Aspegren, A., Hinas, A., Larsson, P., Larsson, A. and Soderbom, F. (2004) Novel non-coding RNAs in Dictyostelium discoideum and their expression during development. *Nucleic Acids Res*, **32**, 4646-4656.

181. Eichinger, L., Pachebat, J.A., Glockner, G., Rajandream, M.A., Sucgang, R., Berriman, M., Song, J., Olsen, R., Szafranski, K., Xu, Q. *et al.* (2005) The genome of the social amoeba Dictyostelium discoideum. *Nature*, **435**, 43-57.

182. Washietl, S., Hofacker, I.L. and Stadler, P.F. (2005) Fast and reliable prediction of noncoding RNAs. *Proc Natl Acad Sci U S A*, **102**, 2454-2459.

183. Schattner, P., Brooks, A.N. and Lowe, T.M. (2005) The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res*, **33**, W686-689.

184. www.ambion.com/techlib/prot/bp_13100.pdf.

185. www.invitrogen.com.

186. Taylor, H.M., Kyes, S.A., Harris, D., Kriek, N. and Newbold, C.I. (2000) A study of var gene transcription in vitro using universal var gene primers. *Mol Biochem Parasitol*, **105**, 13-23.

187. Horrocks, P., Pinches, R., Christodoulou, Z., Kyes, S.A. and Newbold, C.I. (2004) Variable var transition rates underlie antigenic variation in malaria. *Proc Natl Acad Sci U S A*, **101**, 11129-11134.

188. Viebig, N.K., Levin, E., Dechavanne, S., Rogerson, S.J., Gysin, J., Smith, J.D., Scherf, A. and Gamain, B. (2007) Disruption of var2csa gene impairs placental malaria associated adhesion phenotype. *PLoS One*, **2**, e910.

189. Bustin, S.A., Benes, V., Garson, J.A., Hellemans, J., Huggett, J., Kubista, M., Mueller, R., Nolan, T., Pfaffl, M.W., Shipley, G.L. *et al.* (2009) The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin Chem*, **55**, 611-622.

190. Kyes, S., Christodoulou, Z., Pinches, R., Kriek, N., Horrocks, P. and Newbold, C. (2007) Plasmodium falciparum var gene expression is developmentally controlled at the level of RNA polymerase II-mediated transcription initiation. *Mol Microbiol*, **63**, 1237-1247.

191. Lanzer, M., Wertheimer, S.P., de Bruin, D. and Ravetch, J.V. (1994) Chromatin structure determines the sites of chromosome breakages in Plasmodium falciparum. *Nucleic Acids Res*, **22**, 3099-3103.

192. Brolin, K.J., Ribacke, U., Nilsson, S., Ankarklev, J., Moll, K., Wahlgren, M. and Chen, Q. (2009) Simultaneous transcription of duplicated var2csa gene copies in individual Plasmodium falciparum parasites. *Genome Biol*, **10**, R117.

193. Horrocks, P., Pinches, R.A., Chakravorty, S.J., Papakrivos, J., Christodoulou, Z., Kyes, S.A., Urban, B.C., Ferguson, D.J. and Newbold, C.I. (2005) PfEMP1 expression is reduced on the surface of knobless Plasmodium falciparum infected erythrocytes. *J Cell Sci*, **118**, 2507-2518.

194. Taylor, H.M., Kyes, S.A. and Newbold, C.I. (2000) Var gene diversity in Plasmodium falciparum is generated by frequent recombination events. *Mol Biochem Parasitol*, **110**, 391-397.

195. Frank, M., Kirkman, L., Costantini, D., Sanyal, S., Lavazec, C., Templeton, T.J. and Deitsch, K.W. (2008) Frequent recombination events generate diversity within the multi-copy variant antigen gene families of Plasmodium falciparum. *Int J Parasitol*, **38**, 1099-1109.

196. Wittwer, C.T., Reed, G.H., Gundry, C.N., Vandersteen, J.G. and Pryor, R.J. (2003) High-resolution genotyping by amplicon melting analysis using LCGreen. *Clin Chem*, **49**, 853-860.

197. Andriantsoanirina, V., Lascombes, V., Ratsimbasoa, A., Bouchier, C., Hoffman, J., Tichit, M., Rabarijaona, L.P., Durand, R. and Menard, D. (2009) Rapid detection of point mutations in Plasmodium falciparum genes associated with antimalarial drugs resistance by using High-Resolution Melting analysis. *J Microbiol Methods*, **78**, 165-170.

198. Kristensen, L.S. and Dobrovic, A. (2008) Direct genotyping of single nucleotide polymorphisms in methyl metabolism genes using probe-free high-resolution melting analysis. *Cancer Epidemiol Biomarkers Prev*, **17**, 1240-1247.

199. Newton, C.R., Graham, A., Heptinstall, L.E., Powell, S.J., Summers, C., Kalsheker, N., Smith, J.C. and Markham, A.F. (1989) Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). *Nucleic Acids Res*, **17**, 2503-2516.

200. Sommer, S.S., Cassady, J.D., Sobell, J.L. and Bottema, C.D. (1989) A novel method for detecting point mutations or polymorphisms and its application to population screening for carriers of phenylketonuria. *Mayo Clin Proc*, **64**, 1361-1372.

201. Qiang, Y.Z., Qin, T., Fu, W., Cheng, W.P., Li, Y.S. and Yi, G. (2002) Use of a rapid mismatch PCR method to detect gyrA and parC mutations in ciprofloxacin-resistant clinical isolates of Escherichia coli. *J Antimicrob Chemother*, **49**, 549-552.

202. http://rna.tbi.univie.ac.at/.

203. http://www.ebi.ac.uk/Tools/msa/clustalw2/.

204. http://stattrek.com.

205. Janukonyte, J., Vestergaard, E.M., Ladefoged, S.A. and Nissen, P.H. (2010) High-resolution melting analysis using unlabeled probe and amplicon scanning simultaneously detects several lactase persistence variants. *Scand J Clin Lab Invest*, **70**, 535-540.

206. http://www.kapabiosystems.com/products/name/kapa-hifi-pcr-kits.

207. Meyerhans, A., Vartanian, J.P. and Wain-Hobson, S. (1990) DNA recombination during PCR. *Nucleic Acids Res*, **18**, 1687-1691.

208. Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B. and Erlich, H.A. (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science*, **239**, 487-491.

209. Peters, J.M., Fowler, E.V., Krause, D.R., Cheng, Q. and Gatton, M.L. (2007) Differential changes in Plasmodium falciparum var transcription during adaptation to culture. *J Infect Dis*, **195**, 748-755.

210. Frank, M., Dzikowski, R., Amulic, B. and Deitsch, K. (2007) Variable switching rates of malaria virulence genes are associated with chromosomal position. *Mol Microbiol*, **64**, 1486-1498.

211. Recker, M., Buckee, C.O., Serazin, A., Kyes, S., Pinches, R., Christodoulou, Z., Springer, A.L., Gupta, S. and Newbold, C.I. (2011) Antigenic variation in plasmodium falciparum malaria involves a highly structured switching pattern. *PLoS Pathog*, **7**, e1001306.

212. Zhang, Q., Zhang, Y., Huang, Y., Xue, X., Yan, H., Sun, X., Wang, J., McCutchan, T.F. and Pan, W. (2011) From in vivo to in vitro: dynamic analysis of Plasmodium falciparum var gene expression patterns of patient isolates during adaptation to culture. *PLoS One*, **6**, e20591.

213. Bachmann, A., Predehl, S., May, J., Harder, S., Burchard, G.D., Gilberger, T.W., Tannich, E. and Bruchhaus, I. (2011) Highly co-ordinated var gene expression and switching in clinical Plasmodium falciparum isolates from non-immune malaria patients. *Cell Microbiol*, **13**, 1397-1409.

214. Goonewardene, R., Daily, J., Kaslow, D., Sullivan, T.J., Duffy, P., Carter, R., Mendis, K. and Wirth, D. (1993) Transfection of the malaria parasite and expression of firefly luciferase. *Proc Natl Acad Sci U S A*, **90**, 5234-5236.

215. 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. *Proc Natl Acad Sci U S A*, **92**, 973-977.

216. Wu, Y., Kirkman, L.A. and Wellems, T.E. (1996) Transformation of Plasmodium falciparum malaria parasites by homologous integration of plasmids that confer resistance to pyrimethamine. *Proc Natl Acad Sci U S A*, **93**, 1130-1134.

217. Crabb, B.S. and Cowman, A.F. (1996) Characterization of promoters and stable transfection by homologous and nonhomologous recombination in Plasmodium falciparum. *Proc Natl Acad Sci U S A*, **93**, 7289-7294.

218. Duraisingh, M.T., Triglia, T. and Cowman, A.F. (2002) Negative selection of Plasmodium falciparum reveals targeted gene deletion by double crossover recombination. *Int J Parasitol*, **32**, 81-89.

219. Fidock, D.A. and Wellems, T.E. (1997) Transformation with human dihydrofolate reductase renders malaria parasites insensitive to WR99210 but does not affect the intrinsic activity of proguanil. *Proc Natl Acad Sci U S A*, **94**, 10931-10936.

220. Mamoun, C.B., Gluzman, I.Y., Goyard, S., Beverley, S.M. and Goldberg, D.E. (1999) A set of independent selectable markers for transfection of the

human malaria parasite Plasmodium falciparum. *Proc Natl Acad Sci U S A*, **96**, 8716-8720.

221. Deitsch, K., Driskill, C. and Wellems, T. (2001) Transformation of malaria parasites by the spontaneous uptake and expression of DNA from human erythrocytes. *Nucleic Acids Res*, **29**, 850-853.

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

223. Dumay-Odelot, H., Marck, C., Durrieu-Gaillard, S., Lefebvre, O., Jourdain, S., Prochazkova, M., Pflieger, A. and Teichmann, M. (2007) Identification, molecular cloning, and characterization of the sixth subunit of human transcription factor TFIIIC. *J Biol Chem*, **282**, 17179-17189.

224. Moqtaderi, Z. and Struhl, K. (2004) Genome-wide occupancy profile of the RNA polymerase III machinery in Saccharomyces cerevisiae reveals loci with incomplete transcription complexes. *Mol Cell Biol*, **24**, 4118-4127.

225. Kendall, A., Hull, M.W., Bertrand, E., Good, P.D., Singer, R.H. and Engelke, D.R. (2000) A CBF5 mutation that disrupts nucleolar localization of early tRNA biosynthesis in yeast also suppresses tRNA gene-mediated transcriptional silencing. *Proc Natl Acad Sci U S A*, **97**, 13108-13113.

226. Kinsey, P.T. and Sandmeyer, S.B. (1991) Adjacent pol II and pol III promoters: transcription of the yeast retrotransposon Ty3 and a target tRNA gene. *Nucleic Acids Res*, **19**, 1317-1324.

227. Hull, M.W., Erickson, J., Johnston, M. and Engelke, D.R. (1994) tRNA genes as transcriptional repressor elements. *Mol Cell Biol*, **14**, 1266-1277.

228. Wang, L., Haeusler, R.A., Good, P.D., Thompson, M., Nagar, S. and Engelke, D.R. (2005) Silencing near tRNA genes requires nucleolar localization. *J Biol Chem*, **280**, 8637-8639.

229. Hiraga, S., Botsios, S., Donze, D. and Donaldson, A.D. (2012) TFIIIC localizes budding yeast ETC sites to the nuclear periphery. *Mol Biol Cell*, **23**, 2741-2754.

230. Sun, F.L. and Elgin, S.C. (1999) Putting boundaries on silence. *Cell*, **99**, 459-462.

231. Simms, T.A., Dugas, S.L., Gremillion, J.C., Ibos, M.E., Dandurand, M.N., Toliver, T.T., Edwards, D.J. and Donze, D. (2008) TFIIIC binding sites function as both heterochromatin barriers and chromatin insulators in Saccharomyces cerevisiae. *Eukaryot Cell*, **7**, 2078-2086.

232. Raab, J.R., Chiu, J., Zhu, J., Katzman, S., Kurukuti, S., Wade, P.A., Haussler, D. and Kamakaka, R.T. (2012) Human tRNA genes function as chromatin insulators. *EMBO J*, **31**, 330-350.

233. Donze, D. and Kamakaka, R.T. (2001) RNA polymerase III and RNA polymerase II promoter complexes are heterochromatin barriers in Saccharomyces cerevisiae. *EMBO J*, **20**, 520-531.

234. http://www.mr4.org/MR4ReagentsSearch/Results.aspx?BEINum=MRA-99&Template=MR4PlasmidVectors.

235. Horton, R.M., Hunt, H.D., Ho, S.N., Pullen, J.K. and Pease, L.R. (1989) Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension. *Gene*, **77**, 61-68.

236. Ho, S.N., Hunt, H.D., Horton, R.M., Pullen, J.K. and Pease, L.R. (1989) Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene*, **77**, 51-59.

237. Voss, T.S., Tonkin, C.J., Marty, A.J., Thompson, J.K., Healer, J., Crabb, B.S. and Cowman, A.F. (2007) Alterations in local chromatin environment are involved in silencing and activation of subtelomeric var genes in Plasmodium falciparum. *Mol Microbiol*, **66**, 139-150.

238. Shevchuk, N.A., Bryksin, A.V., Nusinovich, Y.A., Cabello, F.C., Sutherland, M. and Ladisch, S. (2004) Construction of long DNA molecules using long PCR-based fusion of several fragments simultaneously. *Nucleic Acids Res*, **32**, e19.

239. Hasenkamp, S., Wong, E.H. and Horrocks, P. (2012) An improved single-step lysis protocol to measure luciferase bioluminescence in Plasmodium falciparum. *Malar J*, **11**, 42.

240. Colin, M., Moritz, S., Schneider, H., Capeau, J., Coutelle, C. and Brahimi-Horn, M.C. (2000) Haemoglobin interferes with the ex vivo luciferase luminescence assay: consequence for detection of luciferase reporter gene expression in vivo. *Gene Ther*, **7**, 1333-1336.

241. Partridge, J.F., Borgstrom, B. and Allshire, R.C. (2000) Distinct protein interaction domains and protein spreading in a complex centromere. *Genes Dev*, **14**, 783-791.

242. Cam, H.P., Sugiyama, T., Chen, E.S., Chen, X., FitzGerald, P.C. and Grewal, S.I. (2005) Comprehensive analysis of heterochromatin- and RNAi-mediated epigenetic control of the fission yeast genome. *Nat Genet*, **37**, 809-819.

243. Oki, M. and Kamakaka, R.T. (2005) Barrier function at HMR. *Mol Cell*, **19**, 707-716.

244. Eichler, E.E. and Sankoff, D. (2003) Structural dynamics of eukaryotic chromosome evolution. *Science*, **301**, 793-797.

245. Dieci, G., Fiorino, G., Castelnuovo, M., Teichmann, M. and Pagano, A. (2007) The expanding RNA polymerase III transcriptome. *Trends Genet*, **23**, 614-622.

246. Alberts, B., Wilson, J.H. and Hunt, T. (2008) *Molecular biology of the cell*. 5th ed. Garland Science, New York.

247. Walliker, D., Carter, R. and Morgan, S. (1971) Genetic recombination in malaria parasites. *Nature*, **232**, 561-562.

248. Kerr, P.J., Ranford-Cartwright, L.C. and Walliker, D. (1994) Proof of intragenic recombination in Plasmodium falciparum. *Mol Biochem Parasitol*, **66**, 241-248.

249. Su, X., Ferdig, M.T., Huang, Y., Huynh, C.Q., Liu, A., You, J., Wootton, J.C. and Wellems, T.E. (1999) A genetic map and recombination parameters of the human malaria parasite Plasmodium falciparum. *Science*, **286**, 1351-1353.

250. Duffy, M.F., Byrne, T.J., Carret, C., Ivens, A. and Brown, G.V. (2009) Ectopic recombination of a malaria var gene during mitosis associated with an altered var switch rate. *J Mol Biol*, **389**, 453-469.

251. White, M.A., Dominska, M. and Petes, T.D. (1993) Transcription factors are required for the meiotic recombination hotspot at the HIS4 locus in Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A*, **90**, 6621-6625.

252. Kirkpatrick, D.T., Fan, Q. and Petes, T.D. (1999) Maximal stimulation of meiotic recombination by a yeast transcription factor requires the transcription activation domain and a DNA-binding domain. *Genetics*, **152**, 101-115.

253. Gerton, J.L., DeRisi, J., Shroff, R., Lichten, M., Brown, P.O. and Petes, T.D. (2000) Global mapping of meiotic recombination hotspots and coldspots in the yeast Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A*, **97**, 11383-11390.

254. Jensen-Seaman, M.I., Furey, T.S., Payseur, B.A., Lu, Y., Roskin, K.M., Chen, C.F., Thomas, M.A., Haussler, D. and Jacob, H.J. (2004) Comparative recombination rates in the rat, mouse, and human genomes. *Genome Res*, **14**, 528-538.

255. Nishant, K.T., Ravishankar, H. and Rao, M.R. (2004) Characterization of a mouse recombination hot spot locus encoding a novel non-protein-coding RNA. *Mol Cell Biol*, **24**, 5620-5634.

256. Wahls, W.P., Siegel, E.R. and Davidson, M.K. (2008) Meiotic recombination hotspots of fission yeast are directed to loci that express non-coding RNA. *PLoS One*, **3**, e2887.

257. Myers, S., Freeman, C., Auton, A., Donnelly, P. and McVean, G. (2008) A common sequence motif associated with recombination hot spots and genome instability in humans. *Nat Genet*, **40**, 1124-1129.

258. Pratt-Hyatt, M.J., Kapadia, K.M., Wilson, T.E. and Engelke, D.R. (2006) Increased recombination between active tRNA genes. *DNA Cell Biol*, **25**, 359-364.

259. Aguilera, A., Chavez, S. and Malagon, F. (2000) Mitotic recombination in yeast: elements controlling its incidence. *Yeast*, **16**, 731-754.

260. van de Vondervoort, P.J., Langeveld, S.M., Visser, J., van Peij, N.N., Pel, H.J., van den Hondel, C.A. and Ram, A.F. (2007) Identification of a mitotic recombination hotspot on chromosome III of the asexual fungus Aspergillus niger and its possible correlation with [corrected] elevated basal transcription. *Curr Genet*, **52**, 107-114.

261. Tang, W., Dominska, M., Greenwell, P.W., Harvanek, Z., Lobachev, K.S., Kim, H.M., Narayanan, V., Mirkin, S.M. and Petes, T.D. (2011) Friedreich's ataxia (GAA)n*(TTC)n repeats strongly stimulate mitotic crossovers in Saccharomyces cerevisae. *PLoS Genet*, **7**, e1001270.

262. Campuzano, V., Montermini, L., Molto, M.D., Pianese, L., Cossee, M., Cavalcanti, F., Monros, E., Rodius, F., Duclos, F., Monticelli, A. *et al.* (1996) Friedreich's ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion. *Science*, **271**, 1423-1427.

263. Lee, P.S., Greenwell, P.W., Dominska, M., Gawel, M., Hamilton, M. and Petes, T.D. (2009) A fine-structure map of spontaneous mitotic crossovers in the yeast Saccharomyces cerevisiae. *PLoS Genet*, **5**, e1000410.

264. Jiang, H., Li, N., Gopalan, V., Zilversmit, M.M., Varma, S., Nagarajan, V., Li, J., Mu, J., Hayton, K., Henschen, B. *et al.* (2011) High recombination rates and hotspots in a Plasmodium falciparum genetic cross. *Genome Biol*, **12**, R33.

265. Deitsch, K.W., del Pinal, A. and Wellems, T.E. (1999) Intra-cluster recombination and var transcription switches in the antigenic variation of Plasmodium falciparum. *Mol Biochem Parasitol*, **101**, 107-116.

266. Puget, N., Stoppa-Lyonnet, D., Sinilnikova, O.M., Pages, S., Lynch, H.T., Lenoir, G.M. and Mazoyer, S. (1999) Screening for germ-line rearrangements and regulatory mutations in BRCA1 led to the identification of four new deletions. *Cancer Res*, **59**, 455-461.

267. Deininger, P. (2011) Alu elements: know the SINEs. *Genome Biol*, **12**, 236.

268. Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W. *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature*, **409**, 860-921.

269. Svoboda, P. and Di Cara, A. (2006) Hairpin RNA: a secondary structure of primary importance. *Cell Mol Life Sci*, **63**, 901-908.

270. Staple, D.W. and Butcher, S.E. (2005) Pseudoknots: RNA structures with diverse functions. *PLoS Biol*, **3**, e213.

271. Lee, F. and Yanofsky, C. (1977) Transcription termination at the trp operon attenuators of Escherichia coli and Salmonella typhimurium: RNA

secondary structure and regulation of termination. *Proc Natl Acad Sci U S A*, **74**, 4365-4369.

272. Zurawski, G., Elseviers, D., Stauffer, G.V. and Yanofsky, C. (1978) Translational control of transcription termination at the attenuator of the Escherichia coli tryptophan operon. *Proc Natl Acad Sci U S A*, **75**, 5988-5992.

273. Grundy, F.J. and Henkin, T.M. (1993) tRNA as a positive regulator of transcription antitermination in B. subtilis. *Cell*, **74**, 475-482.

274. Nahvi, A., Sudarsan, N., Ebert, M.S., Zou, X., Brown, K.L. and Breaker, R.R. (2002) Genetic control by a metabolite binding mRNA. *Chem Biol*, **9**, 1043.

275. Jang, S.K., Krausslich, H.G., Nicklin, M.J., Duke, G.M., Palmenberg, A.C. and Wimmer, E. (1988) A segment of the 5' nontranslated region of encephalomyocarditis virus RNA directs internal entry of ribosomes during in vitro translation. *J Virol*, **62**, 2636-2643.

276. Jang, S.K. and Wimmer, E. (1990) Cap-independent translation of encephalomyocarditis virus RNA: structural elements of the internal ribosomal entry site and involvement of a cellular 57-kD RNA-binding protein. *Genes Dev*, **4**, 1560-1572.

277. Lunde, B.M., Moore, C. and Varani, G. (2007) RNA-binding proteins: modular design for efficient function. *Nat Rev Mol Cell Biol*, **8**, 479-490.

278. Kishore, S., Luber, S. and Zavolan, M. (2010) Deciphering the role of RNA-binding proteins in the post-transcriptional control of gene expression. *Brief Funct Genomics*, **9**, 391-404.

279. Auweter, S.D., Oberstrass, F.C. and Allain, F.H. (2006) Sequence-specific binding of single-stranded RNA: is there a code for recognition? *Nucleic Acids Res*, **34**, 4943-4959.

280. Adam, S.A., Nakagawa, T., Swanson, M.S., Woodruff, T.K. and Dreyfuss, G. (1986) mRNA polyadenylate-binding protein: gene isolation and sequencing and identification of a ribonucleoprotein consensus sequence. *Mol Cell Biol*, **6**, 2932-2943.

281. Swanson, M.S., Nakagawa, T.Y., LeVan, K. and Dreyfuss, G. (1987) Primary structure of human nuclear ribonucleoprotein particle C proteins: conservation of sequence and domain structures in heterogeneous nuclear RNA, mRNA, and pre-rRNA-binding proteins. *Mol Cell Biol*, **7**, 1731-1739.

282. Dreyfuss, G., Swanson, M.S. and Pinol-Roma, S. (1988) Heterogeneous nuclear ribonucleoprotein particles and the pathway of mRNA formation. *Trends Biochem Sci*, **13**, 86-91.

283. Nagai, K., Oubridge, C., Jessen, T.H., Li, J. and Evans, P.R. (1990) Crystal structure of the RNA-binding domain of the U1 small nuclear ribonucleoprotein A. *Nature*, **348**, 515-520.

284. Murata, Y. and Wharton, R.P. (1995) Binding of pumilio to maternal hunchback mRNA is required for posterior patterning in Drosophila embryos. *Cell*, **80**, 747-756.

285. Zhang, B., Gallegos, M., Puoti, A., Durkin, E., Fields, S., Kimble, J. and Wickens, M.P. (1997) A conserved RNA-binding protein that regulates sexual fates in the C. elegans hermaphrodite germ line. *Nature*, **390**, 477-484.

286. Cui, L., Fan, Q. and Li, J. (2002) The malaria parasite Plasmodium falciparum encodes members of the Puf RNA-binding protein family with conserved RNA binding activity. *Nucleic Acids Res*, **30**, 4607-4617.

287. Miao, J., Li, J., Fan, Q., Li, X. and Cui, L. (2010) The Puf-family RNA-binding protein PfPuf2 regulates sexual development and sex differentiation in the malaria parasite Plasmodium falciparum. *J Cell Sci*, **123**, 1039-1049.

288. Muller, K., Matuschewski, K. and Silvie, O. (2011) The Puf-family RNA-binding protein Puf2 controls sporozoite conversion to liver stages in the malaria parasite. *PLoS One*, **6**, e19860.

289. Miller, J., McLachlan, A.D. and Klug, A. (1985) Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes. *EMBO J*, **4**, 1609-1614.

290. Pelham, H.R. and Brown, D.D. (1980) A specific transcription factor that can bind either the 5S RNA gene or 5S RNA. *Proc Natl Acad Sci U S A*, **77**, 4170-4174.

291. Engelke, D.R., Ng, S.Y., Shastry, B.S. and Roeder, R.G. (1980) Specific interaction of a purified transcription factor with an internal control region of 5S RNA genes. *Cell*, **19**, 717-728.

292. Searles, M.A., Lu, D. and Klug, A. (2000) The role of the central zinc fingers of transcription factor IIIA in binding to 5 S RNA. *J Mol Biol*, **301**, 47-60.

293. Clemens, K.R., Wolf, V., McBryant, S.J., Zhang, P., Liao, X., Wright, P.E. and Gottesfeld, J.M. (1993) Molecular basis for specific recognition of both RNA and DNA by a zinc finger protein. *Science*, **260**, 530-533.

294. Burdach, J., O'Connell, M.R., Mackay, J.P. and Crossley, M. (2012) Two-timing zinc finger transcription factors liaising with RNA. *Trends Biochem Sci*, **37**, 199-205.

295. Chen, Y. and Varani, G. (2005) Protein families and RNA recognition. *FEBS J*, **272**, 2088-2097.

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

297. Lai, W.S., Carballo, E., Thorn, J.M., Kennington, E.A. and Blackshear, P.J. (2000) Interactions of CCCH zinc finger proteins with mRNA. Binding of tristetraprolin-related zinc finger proteins to Au-rich elements and destabilization of mRNA. *J Biol Chem*, **275**, 17827-17837.

298. Siomi, H., Matunis, M.J., Michael, W.M. and Dreyfuss, G. (1993) The pre-mRNA binding K protein contains a novel evolutionarily conserved motif. *Nucleic Acids Res*, **21**, 1193-1198.

299. Tian, B., Bevilacqua, P.C., Diegelman-Parente, A. and Mathews, M.B. (2004) The double-stranded-RNA-binding motif: interference and much more. *Nat Rev Mol Cell Biol*, **5**, 1013-1023.

300. Ryter, J.M. and Schultz, S.C. (1998) Molecular basis of double-stranded RNA-protein interactions: structure of a dsRNA-binding domain complexed with dsRNA. *EMBO J*, **17**, 7505-7513.

301. Lamontagne, B. and Elela, S.A. (2004) Evaluation of the RNA determinants for bacterial and yeast RNase III binding and cleavage. *J Biol Chem*, **279**, 2231-2241.

302. MacRae, I.J. and Doudna, J.A. (2007) Ribonuclease revisited: structural insights into ribonuclease III family enzymes. *Curr Opin Struct Biol*, **17**, 138-145.

303. Boni, I.V., Isaeva, D.M., Musychenko, M.L. and Tzareva, N.V. (1991) Ribosome-messenger recognition: mRNA target sites for ribosomal protein S1. *Nucleic Acids Res*, **19**, 155-162.

304. http://pfam.sanger.ac.uk.

305. Gribskov, M. (1992) Translational initiation factors IF-1 and eIF-2 alpha share an RNA-binding motif with prokaryotic ribosomal protein S1 and polynucleotide phosphorylase. *Gene*, **119**, 107-111.

306. Garner, M.M. and Revzin, A. (1981) A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the Escherichia coli lactose operon regulatory system. *Nucleic Acids Res*, **9**, 3047-3060.

307. Fried, M. and Crothers, D.M. (1981) Equilibria and kinetics of lac repressor-operator interactions by polyacrylamide gel electrophoresis. *Nucleic Acids Res*, **9**, 6505-6525.

308. Konarska, M.M. and Sharp, P.A. (1986) Electrophoretic separation of complexes involved in the splicing of precursors to mRNAs. *Cell*, **46**, 845-855.

309. Stenger, D., Gruissem, W. and Baginsky, S. (2004) Mass spectrometric identification of RNA binding proteins from dried EMSA gels. *J Proteome Res*, **3**, 662-664.

310. De Silva, E.K., Gehrke, A.R., Olszewski, K., Leon, I., Chahal, J.S., Bulyk, M.L. and Llinas, M. (2008) Specific DNA-binding by apicomplexan AP2 transcription factors. *Proc Natl Acad Sci U S A*, **105**, 8393-8398.

311. Goyal, M., Alam, A., Iqbal, M.S., Dey, S., Bindu, S., Pal, C., Banerjee, A., Chakrabarti, S. and Bandyopadhyay, U. (2012) Identification and molecular characterization of an Alba-family protein from human malaria parasite Plasmodium falciparum. *Nucleic Acids Res*, **40**, 1174-1190.

312. Chene, A., Vembar, S.S., Riviere, L., Lopez-Rubio, J.J., Claes, A., Siegel, T.N., Sakamoto, H., Scheidig-Benatar, C., Hernandez-Rivas, R. and Scherf, A. (2012) PfAlbas constitute a new eukaryotic DNA/RNA-binding protein family in malaria parasites. *Nucleic Acids Res*, **40**, 3066-3077.

313. Rosenblad, M.A., Zwieb, C. and Samuelsson, T. (2004) Identification and comparative analysis of components from the signal recognition particle in protozoa and fungi. *BMC Genomics*, **5**, 5.

314. Tuteja, R. (2007) Unraveling the components of protein translocation pathway in human malaria parasite Plasmodium falciparum. *Arch Biochem Biophys*, **467**, 249-260.

315. Knapp, G. (1989) Enzymatic approaches to probing of RNA secondary and tertiary structure. *Methods Enzymol*, **180**, 192-212.

316. http://www.invitrogen.com/site/us/en/home/References/protocols/nucleic-acid-purification-and-analysis/rna-protocol/rna-structure-function-protocols.html.

317. Voss, T.S., Mini, T., Jenoe, P. and Beck, H.P. (2002) Plasmodium falciparum possesses a cell cycle-regulated short type replication protein A large subunit encoded by an unusual transcript. *J Biol Chem*, **277**, 17493-17501.

318. Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, **227**, 680-685.

319. Smith, C.W.J. (ed.) (1998) *RNA:protein interactions : a practical approach*. Oxford University Press, Oxford; New York.

320. http://www.ccmp.ox.ac.uk/proteomics.

321. http://gtrnadb.ucsc.edu/Pfalc/Pfalc-structs.html.

322. Gai, Z., Kitagawa, Y., Tanaka, Y., Shimizu, N., Komoda, K., Tanaka, I. and Yao, M. (2012) The binding mechanism of eIF2beta with its partner proteins, eIF5 and eIF2Bepsilon. *Biochem Biophys Res Commun*, **423**, 515-519.

323. Petter, M., Lee, C.C., Byrne, T.J., Boysen, K.E., Volz, J., Ralph, S.A., Cowman, A.F., Brown, G.V. and Duffy, M.F. (2011) Expression of P. falciparum var genes involves exchange of the histone variant H2A.Z at the promoter. *PLoS Pathog*, **7**, e1001292.

324. Titanji, V.P., Amambua-Ngwa, A., Anong, D.N., Mbandi, S.K., Tangie, E., Tening, I. and Yengo, R. (2009) Isolation and expression of UB05, a Plasmodium falciparum antigen recognised by antibodies from semi-immune adults in a high transmission endemic area of the Cameroonian rainforest. *Clin Chem Lab Med*, **47**, 1147-1158.

325. Massenet, S., Motorin, Y., Lafontaine, D.L., Hurt, E.C., Grosjean, H. and Branlant, C. (1999) Pseudouridine mapping in the Saccharomyces cerevisiae spliceosomal U small nuclear RNAs (snRNAs) reveals that pseudouridine synthase pus1p exhibits a dual substrate specificity for U2 snRNA and tRNA. *Mol Cell Biol*, **19**, 2142-2154.

326. Goldrick, M.M. (2001) RNase cleavage-based methods for mutation/SNP detection, past and present. *Hum Mutat*, **18**, 190-204.

327. Mathews, D.H., Moss, W.N. and Turner, D.H. (2010) Folding and finding RNA secondary structure. *Cold Spring Harb Perspect Biol*, **2**, a003665.

328. Weeks, K.M. (2010) Advances in RNA structure analysis by chemical probing. *Curr Opin Struct Biol*, **20**, 295-304.

329. Daou-Chabo, R. and Condon, C. (2009) RNase J1 endonuclease activity as a probe of RNA secondary structure. *Rna*, **15**, 1417-1425.

330. Merino, E.J., Wilkinson, K.A., Coughlan, J.L. and Weeks, K.M. (2005) RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE). *J Am Chem Soc*, **127**, 4223-4231.

331. http://rna.urmc.rochester.edu/RNAstructure.html.

332. Lindner, S.E., De Silva, E.K., Keck, J.L. and Llinas, M. (2010) Structural determinants of DNA binding by a P. falciparum ApiAP2 transcriptional regulator. *J Mol Biol*, **395**, 558-567.

333. Draper, D.E. (2004) A guide to ions and RNA structure. *Rna*, **10**, 335-343.

334. Romer, R. and Hach, R. (1975) tRNA conformation and magnesium binding. A study of a yeast phenylalanine-specific tRNA by a fluorescent indicator and differential melting curves. *Eur J Biochem*, **55**, 271-284.

335. Chao, F.C. (1957) Dissociation of macromolecular ribonucleoprotein of yeast. *Arch Biochem Biophys*, **70**, 426-431.

336. Klein, D.J., Moore, P.B. and Steitz, T.A. (2004) The contribution of metal ions to the structural stability of the large ribosomal subunit. *Rna*, **10**, 1366-1379.

337. Kaminska, M., Havrylenko, S., Decottignies, P., Gillet, S., Le Marechal, P., Negrutskii, B. and Mirande, M. (2009) Dissection of the structural organization of the aminoacyl-tRNA synthetase complex. *J Biol Chem*, **284**, 6053-6060.

338. Kaminska, M., Havrylenko, S., Decottignies, P., Le Marechal, P., Negrutskii, B. and Mirande, M. (2009) Dynamic Organization of Aminoacyl-tRNA Synthetase Complexes in the Cytoplasm of Human Cells. *J Biol Chem*, **284**, 13746-13754.

339. Raina, M., Elgamal, S., Santangelo, T.J. and Ibba, M. (2012) Association of a multi-synthetase complex with translating ribosomes in the archaeon Thermococcus kodakarensis. *FEBS Lett*, **586**, 2232-2238.

340. Negrutskii, B.S. and Deutscher, M.P. (1991) Channeling of aminoacyl-tRNA for protein synthesis in vivo. *Proc Natl Acad Sci U S A*, **88**, 4991-4995.

341. Stapulionis, R. and Deutscher, M.P. (1995) A channeled tRNA cycle during mammalian protein synthesis. *Proc Natl Acad Sci U S A*, **92**, 7158-7161.

342. Hausmann, C.D., Praetorius-Ibba, M. and Ibba, M. (2007) An aminoacyl-tRNA synthetase:elongation factor complex for substrate channeling in archaeal translation. *Nucleic Acids Res*, **35**, 6094-6102.

343. Spahn, C.M., Beckmann, R., Eswar, N., Penczek, P.A., Sali, A., Blobel, G. and Frank, J. (2001) Structure of the 80S ribosome from Saccharomyces cerevisiae--tRNA-ribosome and subunit-subunit interactions. *Cell*, **107**, 373-386.

344. Passmore, L.A., Schmeing, T.M., Maag, D., Applefield, D.J., Acker, M.G., Algire, M.A., Lorsch, J.R. and Ramakrishnan, V. (2007) The eukaryotic translation initiation factors eIF1 and eIF1A induce an open conformation of the 40S ribosome. *Mol Cell*, **26**, 41-50.

345. Kaestli, M., Cockburn, I.A., Cortes, A., Baea, K., Rowe, J.A. and Beck, H.P. (2006) Virulence of malaria is associated with differential expression of Plasmodium falciparum var gene subgroups in a case-control study. *J Infect Dis*, **193**, 1567-1574.

346. Falk, N., Kaestli, M., Qi, W., Ott, M., Baea, K., Cortes, A. and Beck, H.P. (2009) Analysis of Plasmodium falciparum var genes expressed in children from Papua New Guinea. *J Infect Dis*, **200**, 347-356.

347. Jackson, K.E., Habib, S., Frugier, M., Hoen, R., Khan, S., Pham, J.S., Ribas de Pouplana, L., Royo, M., Santos, M.A., Sharma, A. *et al.* (2011) Protein translation in Plasmodium parasites. *Trends Parasitol*, **27**, 467-476.

348. Belin, S., Hacot, S., Daudignon, L., Therizols, G., Pourpe, S., Mertani, H.C., Rosa-Calatrava, M. and Diaz, J.J. (2010) Purification of ribosomes from human cell lines. *Curr Protoc Cell Biol*, **Chapter 3**, Unit 3 40.

349. Oehring, S.C., Woodcroft, B.J., Moes, S., Wetzel, J., Dietz, O., Pulfer, A., Dekiwadia, C., Maeser, P., Flueck, C., Witmer, K. *et al.* (2012) Organellar proteomics reveals hundreds of novel nuclear proteins in the malaria parasite Plasmodium falciparum. *Genome Biol*, **13**, R108.