Characterization of transcription factors in plasmodium falciparum

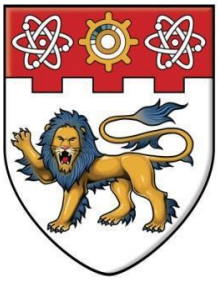
Luah, Yen Hoon

2014

Luah, Y. H. (2014). Characterization of transcription factors in plasmodium falciparum. Doctoral thesis, Nanyang Technological University, Singapore.

https://hdl.handle.net/10356/61728

https://doi.org/10.32657/10356/61728



NANYANG TECHNOLOGICAL UNIVERSITY

CHARACTERIZATION OF TRANSCRIPTION FACTORS IN *PLASMODIUM FALCIPARUM*

LUAH YEN HOON

SCHOOL OF BIOLOGICAL SCIENCES

2014

CHARACTERIZATION OF TRANSCRIPTION FACTORS IN *PLASMODIUM FALCIPARUM*

LUAH YEN HOON

SCHOOL OF BIOLOGICAL SCIENCES

A thesis submitted to the Nanyang Technological University

in partial fulfillment of the requirement for the degree of

Doctor of Philosophy

2014

Acknowledgement

I would like to express my sincere gratitude and appreciation to Associate Professor Zbynek Bozdech for his mentorship, constant support and guidance, and for providing an inspiring environment for me to pursue my research interests. I am thankful to my thesis advisory committee, Professor Mark Featherstone and Dr. Xavier Roca, who have been extremely helpful in contributing their expertise to my project. I am grateful to Dr. Balbir Chaal for being an inspiring mentor, for her patience and guidance. I thank Sabna and Ramya for being such great friends and constantly bringing laughter to the people around them. I thank all ZB lab members for their assistance and encouragement. Most importantly, I thank my family for their patience and support during the course of my Ph.D. study. Finally, I thank Eng How for his constant support, encouragement, and invaluable and constructive advices.

Abstract

Malaria is an important human infectious disease affecting millions of people every year. Despite years of research, transcriptional regulatory mechanisms in *Plasmodium* are still poorly understood. The presence of a large number of zinc finger proteins amongst the potential specific transcription factors suggests the importance of these proteins. In this thesis, we characterized PF10_0083, a putative transcription factor that belongs to the CCCH type zinc finger family. Using over-expression and knock-out cell lines, we carried out molecular characterization of this protein. PF10_0083 protein level was found to peak during the schizont stage, and the peak protein levels coincided with its nuclear localization. PF10_0083 was found to be expressed in only a sub-population of schizonts. We demonstrated that the over-expression of PF10_0083 resulted in enhanced gametocyte production, while the knock-out of PF10_0083 caused decreased gametocyte production. In addition, we showed that PF10_0083 is capable of associating with DNA. We demonstrated that PF10_0083 preferentially binds to the IGRs on the genome, and that this binding positively correlated with the transcriptional up-regulation of the gametocyte genes. As such, we conclude that PF10_0083 promotes the differentiation of asexual parasites to sexual forms under stress conditions.

Table of contents

List of publications

1. Gupta AP, Chin WH, Zhu L, Mok S, **Luah YH**, Lim EH, Bozdech Z: **Dynamic epigenetic regulation of gene expression during the life cycle of malaria parasite Plasmodium falciparum.** *PLoS Pathog* 2013, 9:e1003170.

2. Witmer K, Schmid CD, Brancucci NM, **Luah YH**, Preiser PR, Bozdech Z, Voss TS: **Analysis of subtelomeric virulence gene families in Plasmodium falciparum by comparative transcriptional profiling.** *Mol Microbiol* 2012, 84:243-259.

3. **Luah YH**, Chaal BK, Ong EZ, Bozdech Z: **A moonlighting function of Plasmodium falciparum histone 3, mono-methylated at lysine 9?** *PLoS One* 2010, 5:e10252.

4. Chaal BK, Gupta AP, Wastuwidyaningtyas BD, **Luah YH**, Bozdech Z: **Histone deacetylases play a major role in the transcriptional regulation of the Plasmodium falciparum life cycle.** *PLoS Pathog* 2010, 6:e1000737.

Conference presentation

1. Molecular Approaches to Malaria MAM 2012, Lorne, Australia, 19th – 23rd Feb '12. (Poster presentation)

2. Singapore Malaria Network Meeting 2012, Singapore, 16th – 17th Feb '12. (Poster presentation)

List of tables

List of figures

List of abbreviations

4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid	Hepes
4′,6′-diamidino-2-phenylindole	DAPI
6-cysteine protein	P230
6X histidine	his
adenine-uridine	AU
ammonium persulfate	APS
apical membrane antigen 1	AMA-1
artemisinin-based combination therapies	ACTs
Basic Local Alignment Search Tool	BLAST
bovine serum albumin	BSA
calmodulin 5' untranslated region	CAM 5’ UTR
casein kinase 1	CK1
cell lysis buffer	CLB
Chromatin immunoprecipitation coupled with microarray	ChIP-on-chip
circumsporozoite protein	CSP
cyclic adenosine monophosphate	cAMP
cysteine	C
deoxyribonucleotide	dNTPs
dichloro-diphenyl-trichloroethane	D.T.T
dimethyl sulfoxide	DMSO
dithiothreitol	DTT
DNA binding domains	DBDs
DNA replication element	DRE
downstream promoter element	DPE
early transcribed membrane protein	ETRAMP
electrophoretic mobility shift assay	EMSA
enhancer	ENH
ethylene glycol tetra-acetic acid	EGTA
ethylenediaminetetraacetic acid	EDTA
expect value	E value
finding informative regulatory elements	FIRE
gamete antigen 27/25	Pfg27
gene enrichment motif searching	GEMS
Gene Ontology	GO
general transcription factors	GTFs
glutathione S-transferases	GST
Glyceraldehyde 3-phosphate dehydrogenase	GAPDH
glycophorin binding protein 130	gbp-130
glycosylphosphatidylinositol	GPI
green fluorescent protein	gfp
Heterochromatin Protein 1	HP1
high salt buffer	HSB

histidine	H
histidine rich protein 2 3’ UTR	HRP2 3' UTR
histone 3 lysine 79 tri-methylation	H3K79Me3
histone 4 lysine 5 acetylation	H4K5Ac
histone acetyl transferases	HAT
histone deacetyl transferase	HDAC
histone H3 lysine 27 tri-methylation	H3K27Me3
histone H3 lysine 36 acetylation	H3K36Ac
histone H3 lysine 4 methylation	H3K4Me
histone H3 lysine 4 tri-methylation	H3K4Me3
histone H3 lysine 9	H3K9
histone H3 lysine 9 acetylation	H3K9Ac
histone H3 lysine 9 methylation	H3K9Me
histone H3 lysine 9 tri-methylation	H3K9Me3
histone H4 lysine 8 acetylation	H4K8Ac
hours-post invasion	hpi
human dihydrofolate reductase	hDHFR
immunofluorescence assay	IFA
Infected RBCs	iRBCs
initiator	INR
intergenic regions	IGRs
intra-erythrocytic developmental cycle	IDC
isopropyl β-D-1-thiogalactopyranoside	IPTG
knob-associated histidine-rich protein	kahrp
lithium chloride	LiCl
low salt buffer	LSB
Luria-Bertani	LB
male development gene 1	MDV1
Maturing Oocyte Expressed	MOE
merozoite surface protein	MSP
micrococcal nuclease	MNase
multiple cloning site	MCS
Myb-regulatory element	MRE
myc-streptavidin	myc
N-acetylglucosamine	GlcNAc
neuron-restrictive silencer factor	NRSF
Nonidet P-40	NP-40
nuclear localization signal	NLS
oocyst capsule protein	Cap380
ookinete surface antigen precursor	Pfs25
open reading frames	ORFs
Oryza sativa delay of the onset of senescence	OsDOS
Oryza sativa leaf and tiller angle increased controller	OsLIC
P. falciparum Chloroquine Resistance Transporter 5’ UTR promoter	PfCRT 5’ UTR
P. falciparum erythrocyte membrane protein 1	PfEMP1
P. falciparum Myb1	PfMyb1

P. falciparum proliferating cell nuclear antigen	PfPCNA
Pearson correlation coefficient	PCC
phenylmethylsulfonyl fluoride	PMSF
phosphate-buffered saline	PBS
Plasmodium dihydrofolate reductase	pDHFR
Polymerase Chain Reaction	PCR
pre-initiation complex	PIC
Protein Data Bank	PDB
Protein Information Resource	PIR
Protein Research Foundation	PRF
Real-time PCR	RT-PCR
red blood cells	RBCs
regulatory element	RE
repressor-activator protein 1	Rap1
RNA polymerase II	RNA pol II
saline-sodium citrate buffer	SSC
secreted ookinete adhesive protein	SOAP
sexual stage-specific protein precursor	Pfs16
small nuclear ribonucleoproteins	snRNPs
sodium dodecyl sulfate	SDS
Sodium dodecyl sulfate polyacrylamide gel electrophoresis	SDS-PAGE
specific TFs	STFs
switching mechanism at 5’ end of RNA template	SMART
TATA binding protein	TBP
Tetramethylethylenediamine	TEMED
time-point	tp
transcription associated proteins	TAPs
transcription factors	TFs
transcription start site	TSS
tricarboxylic acid	TCA
Tris-buffered saline with Tween-20	TBS-T
Tristetraprolin	TTP
tumor necrosis factor a	TNFa
untranslated region	UTRs
upregulated in infective sporozoites gene 3 and 4	UIS3 and UIS4
World Health Organization	WHO
Zinc finger	ZnF

Chapter 1 Introduction

1.1 Human malaria disease

Malaria is an important infectious disease of humans throughout centuries. Malaria was first described in an ancient Chinese medical book *Huang Di Nei Jing* in 2700 BC. It was also described by Hippocrates in the writings *Corpus Hippocratorum* in 400 BC. The characteristic malaria fevers were also mentioned in ancient Egyptian, Greek and Indian documents in 1570 BC, 850 BC and 500 BC, respectively [1-3]. However, the causative agent of malaria, the *Plasmodium* parasite, was only identified in 1880s [3]. The malaria parasite was thought to be present before human history because of the discovery of a fossilized mosquito carrying the malaria parasite from 30 million years ago [4]. Malaria is an Italian word which means "bad air" because it was previously believed that malaria was caused by breathing in poisonous polluted air near marshy areas [1, 3].

Malaria is a widespread and potentially lethal disease. It has been documented in at least 100 countries and remains a major health burden in many countries today. It is estimated that approximately 3.3 billion people are at risk of malaria every year. Approximately 200-300 million clinical cases are reported every year and on average, 1 million people are killed by malaria annually. In 2010, an estimated 219 million malaria cases and 660,000 malaria deaths were reported. Approximately 80% of the malaria cases and 90% of the deaths occurred in the African Region. Children under the age of 5 accounted for approximately 86% of the total number of deaths. Malaria endemic countries face poverty, under-development, economic and social problems. In recent years, World Health Organization (WHO) focuses their malaria control efforts

on malaria prevention and eradication measures. WHO introduced the Roll Back Malaria Partnership that aims to reduce the number of malaria cases by 75% by year 2015, compared to the number of malaria cases in the year 2000. Although the malaria eradication program has made progress in recent years owing to the efforts of WHO and the ministries of Health of malaria endemic countries, complete elimination of malaria does not seems that it will happen in the near future with the current intervention strategies and the increased incidence of drug resistance, hence additional tools are required [5, 6].

1.2 Malaria clinical symptoms

Malaria is a human protozoan parasitic disease that is transmitted by the female *Anopheles* species mosquito carrying the *Plasmodium* genus parasites. There are 4 *Plasmodium* species, *P. falciparum*, *P. vivax*, *P. malariae*, and *P. ovale* which are known to establish infections in human. In addition, there is a rising number of infections with the primate parasite *P. knowlesi* in Southeast Asia that is transmitted from the long-tailed and pig-tailed macaques to humans. However, no human-to-vector-to-human transmission of *P. knowlesi* has been reported to date [7, 8]. Patients suffering from malaria will develop the symptoms such as sudden onset of fever, recurring episodes of high fever, chills, headache, muscle aches, vomiting, diarrhea and anemia.

Amongst the 4 species, *P. falciparum* caused the highest rate of mortality and accounted for 90% of the malaria cases in 2010 [6]. Infection with *P. falciparum* should be promptly treated as it will rapidly progress into fatal complications. The cytoadherence of infected erythrocytes to the blood capillaries can obstruct blood flow

and deplete organs of oxygen and nutrients. The lysis of erythrocytes causes severe anemia and also deprives organs of sufficient blood flow. These events cause organ failure and other complications including cerebral malaria and coma, respiratory distress, hypoglycemia, and metabolic acidosis. These conditions if left untreated can eventually lead to death. In pregnant women infected with malaria, the adhesion of infected erythrocytes in the placenta causes placental malaria. Placental malaria leads to premature delivery, low birth weight, and high mortality rates in the newborns, and anemia in pregnant women [9].

1.3 *Plasmodium falciparum* life cycle

The life cycle of *Plasmodium*, an apicomplexan, is a complex process involving multiple organs in both the female *Anopheles* mosquito and the human hosts. The sporogonic stage takes place in the mosquito host, while the exo-erythrocytic and intra-erythrocytic stages take place in the human host.

Infection begins when a human host is bitten by an infected female *Anopheles* mosquito. The sporozoites present in the salivary glands of the mosquito will be injected into the human host skin epidermal layer. Within 30 minutes, the sporozoites enter the blood stream, travel to the liver and invade the hepatocytes. For the next 5-16 days, these sporozoites undergo differentiation and asexual reproduction within a parasitophorous vacuole inside the hepatocytes. Each sporozoite will give rise to thousands of merozoites that are enclosed within a merosome inside the hepatocyte. This is the pre-erythrocytic phase known as the exo-erythrocytic schizogony. The exo-erythrocytic stage is a silent phase and displayed no clinical symptoms [10, 11]. The breakdown of the parasitophorous vacuole membrane releases the merozoites into the

hepatocyte cytoplasm. This is followed by the death of the hepatocytes, and the release of the merosomes, which are merozoites-filled vesicles. The merosomes get transported to the liver sinusoids and then to the lungs. The merozoites are then released into the blood stream and will invade the erythrocytes within 1 minute [12].

Within the red blood cells (RBCs), the merozoites initiate a second round of asexual multiplication known as the intra-erythrocytic developmental cycle (IDC) that is approximately 48 hours long [13]. The merozoites develop inside the parasitophorous vacuole to form three morphologically distinct stages: from a ring to a trophozoite and then to a schizont. The ring stage parasites have minimal metabolic activity. As the ring matures into a trophozoite, the metabolic activity increases. The metabolic processes involve mainly host cell hemoglobin degradation, DNA replication, protein translation, and other cellular processes involved in growth [14]. The newly synthesized DNA is packaged into 8-32 new daughter merozoites during the schizont stage. The parasite parasitophorous vacuole membrane and the infected RBC membrane will lyse to release the merozoites into the blood, where the free merozoites invade other uninfected RBCs. The IDC repeats almost indefinitely and is responsible for the malaria symptoms mentioned in Section 1.2.

In the RBC, some asexual parasites exit the IDC and differentiate into the sexual male or female gametocytes (micro- and macro-gametocytes), which will be taken up by the female *Anopheles* mosquito during a blood meal. In the gut of the mosquito, the gametocytes mature into the male and female gametes (micro- and macro-gametes). The male gamete fertilizes the female gamete to produce a zygote which develops into an ookinete. The ookinete develops and matures to form an oocyst which carries

thousands of sporozoites. The rupture of the oocyst releases the sporozoites which subsequently migrate from the gut to the salivary gland, and will be stored there. These sporozoites are ready to infect humans during the next blood meal [15]. The schematic illustration of the sporogonic, exo-erythrocytic and intra-erythrocytic cycles is shown in Figure 1.1.

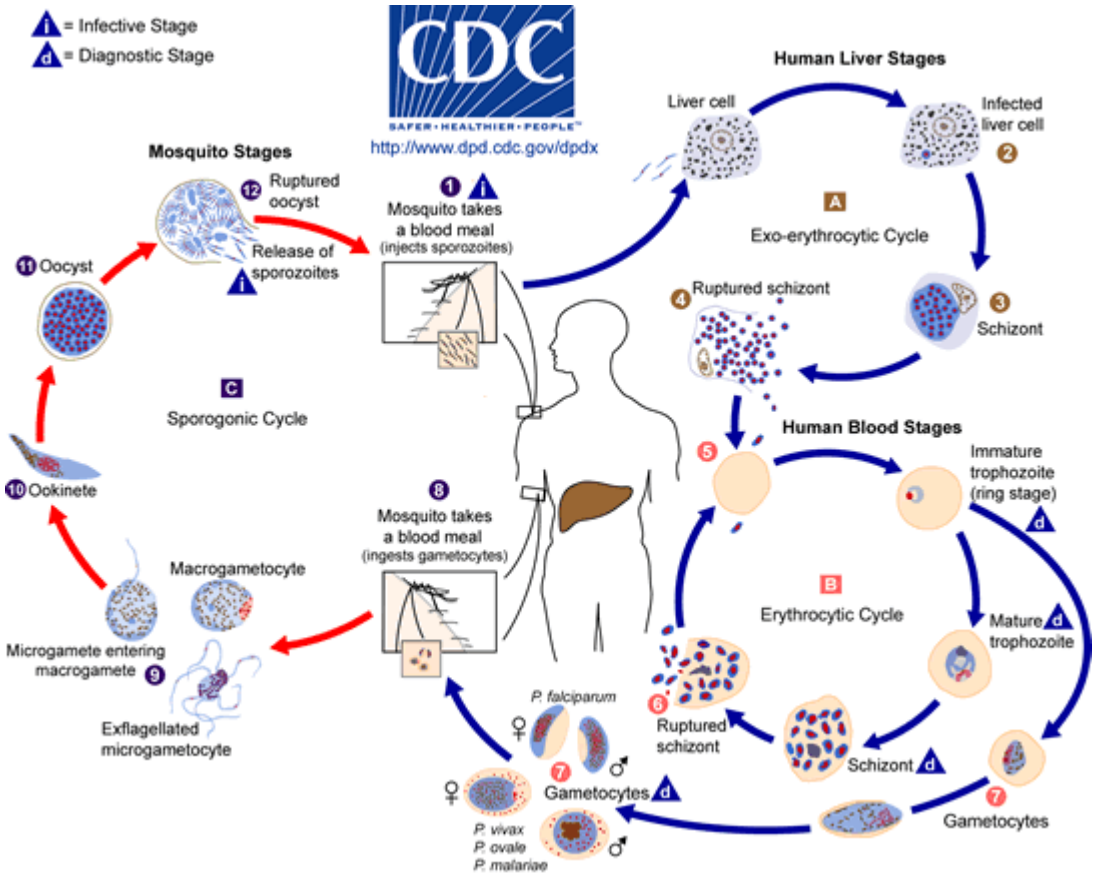


Figure 1.1 *Plasmodium* parasite life cycle.

The development of *Plasmodium* parasite occurs in both the female *Anopheles* mosquito (sporogonic cycle) and the human hosts (exo-erythrocytic and intra-erythrocytic cycle). (A) Infection begins when a human host is bitten by an infected female *Anopheles* mosquito. The sporozoites stored in the salivary glands of the mosquito will be injected into the human host skin epidermal layer. The sporozoites enter the blood stream within 30 minutes, travel to the liver and invade the hepatocytes. For the next 5-16 days, each sporozoite undergoes differentiation and asexual reproduction to give rise to thousands of merozoites enclosed within vesicles known as merosomes. This is the exo-erythrocytic

schizogony stage which is clinically silent. The death of the hepatocytes releases the merozoites-filled merosomes. The merosomes get transported to the liver sinusoids and then to the lungs. The merozoites then get released into the lung blood capillaries. (B) The intra-erythrocytic cycle (IDC) involves the asexual multiplication of the parasites in the blood. The IDC is approximately 48 hours long. The IDC involves both the intracellular (rings, trophozoites and schizonts) and extracellular (merozoites) forms. The IDC repeats almost indefinitely and is responsible for the malaria symptoms. A sub-population of parasites exits asexual multiplication and differentiates into the male and female gametocytes (micro- and macro-gametocytes). The development of the gametocytes also takes place in the blood. The gametocytes will be taken up by the mosquito during feeding. (C) The sporogonic cycle involves the maturation of the gametocytes into the gametes (micro- and macro-gametes), the fertilization of the gametes to form the ookinete and the asexual production of sporozoites. These events take place in the mosquito gut. The sporozoites travel to and get stored in the salivary glands, ready for infection during the next blood meal. Source: (Centers for Disease Control and Prevention).

1.4 *Plasmodium falciparum* gametocytogenesis

In the human host, a sub-population of parasites exits asexual multiplication and differentiates into gametocytes. Inside the mosquito host, the gametocytes develop into gametes, and undergo fertilization to form a zygote which develops into ookinete and subsequently into an oocyst carrying infectious sporozoites. The gametocytes are the only stage that can establish an infection in the mosquito host. As such, a switch from asexual to sexual reproduction is necessary in order to establish transmission. Many drugs that are effective against the asexual stages were found to be ineffective against the sexual stages. Although patients are cured from the clinical symptoms, they are still capable of transmitting the disease [16]. The mechanisms and stimuli that trigger the differentiation of asexual stages to sexual stages in *Plasmodium* are not completely understood. Both environmental and genetic factors contribute to the commitment to sexual development.

1.4.1 Gametocyte development

Gametocyte development is classified into 5 stages, termed stages I-V [17]. Stage I gametocytes are morphologically indistinguishable from the asexual trophozoite stage. Stage II gametocytes occupy about half of the area of RBCs, and have a teardrop-like or "D" shape. Stage III gametocytes are elongated and the two edges are slightly rounded. As the parasites develop into stage IV gametocytes, the two edges become sharp and pointed. Stage V parasites are fully matured and pigment granules are localized to the center of the parasite. RBCs can no longer be visualized by staining. Figure 1.2 illustrates the giemsa stained images of the 5 stages of gametocytes. These gametocytes developed from wild-type 3D7 parasites under stress conditions (refer to section 5.2).

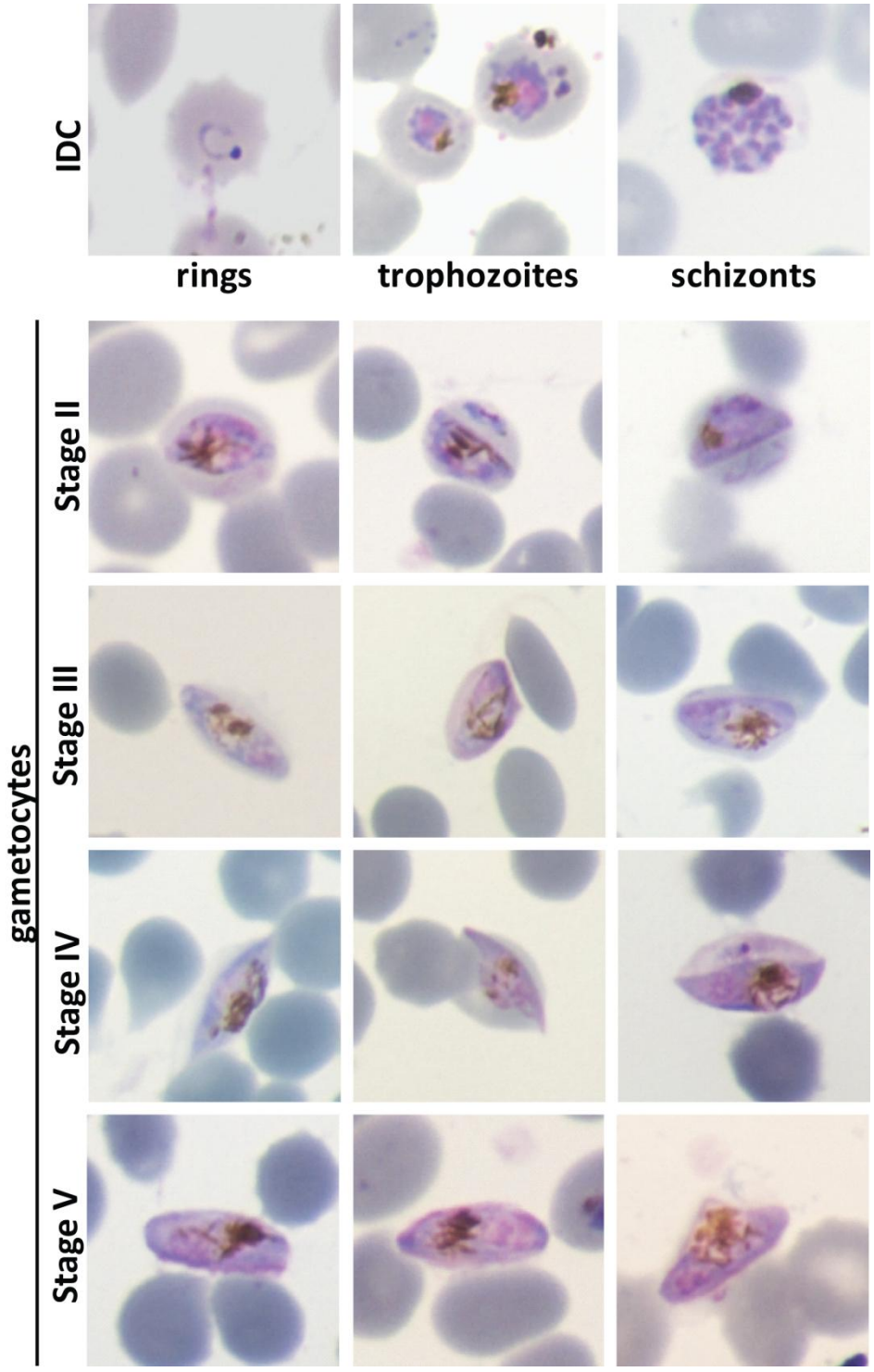


Figure 1.2 Microscope images of gametocytes.

Giemsa stained images of the 5 stages of gametocytes. The first row shows the asexual ring, trophozoite and schizont stages. Stage I gametocytes are not distinguishable from the trophozoites. Stage II gametocytes have a teardrop-like or "D" shape. Stage III gametocytes are elongated with round edges. Stage IV gametocytes have pointed edges. Stage V gametocytes are banana shaped.

1.4.2 Environmental factors affecting gametocytogenesis

Host immune factors such as tumor necrosis factor α (TNFα), interleukin 1 [18], and lymphocytes [19] have been described to affect gametocytogenesis. The addition of fresh blood to *in vitro* cultures has been reported to promote gametocytogenesis [17]. Anti-malaria drug pyrimethamine [20] and chloroquine [21] were reported to promote gametocytogenesis, while artemisinin derivatives decrease gametocytogenesis [22]. In addition, it was also suggested that higher asexual parasitemia is linked to increased gametocytogenesis [23, 24]. Signal transduction pathways involving cyclic adenosine monophosphate (cAMP)-dependent and protein kinase C-dependent pathways were also suggested to be involved in the initiation of gametocytogenesis [25-27]. However, it has been demonstrated that the number of gametocytes produced varies between different parasite lines, and also between different cultures and experiments of the same parasite lines [17, 19, 28].

1.4.3 Genetic factors affecting gametocytogenesis

The commitment of asexual parasites to male or female gametocyte development was found to be determined during the asexual stages either before schizonts maturation and merozoites release [29, 30] or during ring stage [31]. All the merozoites in a schizont either continue to develop asexually, or differentiate into gametocytes. Like in the asexual stages, gametocytes undergo development to form 5 different morphological stages, which have been described in detail [17]. There is a distinct expression pattern of gametocyte stages specific proteins [32]. Different gametocyte markers have been identified to be present in the different stages of gametocytes. For example, sexual stage-specific protein precursor (Pfs16) and gamete antigen 27/25 (Pfg27) can be detected in all stages of gametocyte, Pfg27 expression peaked during

stage II gametocytes, while Pfs230 and Pfs48/45 peaked in stage III gametocytes [33]. Alpha tubulin 2 is a male gametocyte specific marker [34], while Pfg377 is a female gametocyte specific marker [35].

The regulation of gene transcription and protein expression in the sexual stages is not completely understood. Different regulatory mechanisms have been described to be involved in the switch from asexual to sexual life cycle. For example, it was also found that abnormal transcription of Pfg27 resulted in no gametocytogenesis [36, 37]. It was suggested that 3' UTR of Pfg27 is important in transcriptional regulation as a disruption of the 3'UTR resulted in no gametocytogenesis [26]. Studies have also demonstrated that the silencing of *Pfgig* gene located at the subtelomeric region on the right arm of chromosome 9 resulted in decreased gametocytogenesis [38, 39]. It was described that the deletion of the *Pfgdv1* gene, also located on chromosome 9, resulted in defective gametocytogenesis [40]. The SET protein was found to possess 2 sets of promoter, one of which is active in the IDC stages, and the other only active in the gametocytes. A sub-population of schizonts was found to express SET under the control of the gametocyte-specific promoter. It was hence suggested that these schizonts are committed to gametocytogenesis [41]. Interestingly, SET lacks a nuclear localization signal, and has been described to be localized to both the nucleus [41] and the cytoplasm [42]. Recently, disrupting the function of PfAP2-G (PFL1085w), a potential transcriptional regulator that belongs to the ApiAP2 family, resulted in the loss of gametocyte production. PfAP2-G was hence suggested to be the master regulator of sexual development in *P. falciparum* [43]. Similarly, in *P. berghei*, mutation in PbAP2-G, the ortholog of PfAP2-G in *P. falciparum*, resulted in a loss of sexual development. PbAP2-G was shown to be essential for sexual stage commitment.

[44]. In addition, translational repression has also been suggested to be important for sexual development [45, 46]. For example, the expression of 2 zygote surface proteins, Pfs25 and Pfs28, was shown to be regulated by translational repression [47].

1.5 Malaria intervention strategies

1.5.1 Anti-malarial drugs

Malaria prophylaxis and therapy is important in the control of malaria. Most of the anti-malarial drugs target the asexual blood stage of the parasites. Quinine, extracted from the bark of the Cinchona tree, was the first effective drug against malaria and was used worldwide in the 17th century. However, the world supply of quinine was disrupted during World War II. Chloroquine, a synthetic drug, was hence formulated as a substitute for quinine in 1934. Chloroquine was widely used as a first-line treatment over the next few decades. Other quinoline drugs such as primaquine and amodiaquine were also developed in the 1940s and 1950s, and subsequently mefloquine was developed in 1975. In addition, anti-folates such as proguanil and Fansidar, a combination drug consisting of pyrimethamine and sulphadoxine, were also developed during the same period [48, 49]. In the 1990s, a combination therapy of proguanil and atovaquone was also formulated. Artemisinin is a traditional Chinese medicine extracted from the plant *Artemisia annua*. Artemisinin and its derivatives, artesunate and artemether, are used in combination with other drugs such as Fansidar, to treat multi-drug resistance malaria [50]. These drugs were all proven to be effective against malaria. However, the widespread use of these anti-malarial drugs subsequently caused the *Plasmodium* parasites to develop drug resistance [50]. As such, there is a need to overcome this problem in order to prolong the use of anti-malarial drugs. Hence, WHO and the Malaria Foundation International recommended

a global switch to the use of artemisinin-based combination therapies (ACTs) in 2004, and the use of artemisinin monotherapy was banned [51]. ACTs is the best anti-malaria therapy by far but is also one of the few remaining effective therapy against malaria [52].

1.5.2 Malaria preventive measures

The early civilizations practiced preventive and protective measures including the use of bed nets, establishing drainage system to remove swamps, and restricting human habitation near mosquito-infested areas, to reduce malaria transmission. Today, long-lasting insecticide-treated bed nets are used for all age groups at risk of malaria and have reduced the number of malaria cases by 50% [53]. Indoor residual spraying with the pesticide dichloro-diphenyl-trichloroethane (D.T.T) or other long-acting insecticide is employed to eradicate the mosquito vectors. D.D.T was very effective in controlling malaria in the 20th century. However, due to concerns over environment and human health effects, D.D.T was banned for over 30 years. In 2006, WHO implemented the use of D.D.T for indoor spraying again. Indoor spraying can reduce malaria transmission by up to 90% when D.D.T is used appropriately [54]. In addition, larvicides are also used in mosquito larval control to complement the mosquito vectors control measures [55, 56].

1.5.3 Malaria vaccine development

Although the use of anti-malarial drugs and preventive measures has had some success in controlling malaria, *Plasmodium* continues to develop drug resistance. Therefore, the call for novel drugs, more effective preventive measures and a vaccine are urgent. One of the biggest challenges in vaccine development is the huge antigenic variation

and high allelic diversity of the parasite's surface proteins. An ideal candidate would be a surface antigen but with limited antigenic diversity.

To date, a few *P. falciparum* antigens have been identified and studied as potential vaccine candidates. These antigens include apical membrane antigen 1 (AMA-1), circumsporozoite protein (CSP) and merozoite surface protein 1 (MSP-1). However, only a few antigens are in clinical trial. One of the successful vaccine candidates is RTS,S/AS02A. It is designed based on *P. falciparum* CSP to target the pre-erythrocytic stage of the parasite life cycle. The phase IIb clinical trials were carried out in young children aged 1 to 4 years from Mozambique. The vaccine efficacy of RTS,S/AS02A against severe malaria was shown to be 57.7% for at least 6 months and 48.6% for at least 18 months [57, 58]. The phase III clinical trials were carried out in 6000 children aged 5 to 17 months from 7 African countries. The vaccine efficacy of RTS,S/AS02A against severe malaria was shown to be 47.3% for up to 12 months. However, there were cases of generalized convulsive seizures and meningitis after vaccination and thus requires further investigation [59]. As such, this vaccine is still not yet ideal for large scale implementation, and further research is required.

In addition to vaccine development with surface antigen, it is also possible to develop vaccine using genetically modified malaria parasites. An example is vaccine candidates that target the parasites at the liver stage. For example, 2 genes essential for liver stage development, upregulated in infective sporozoites gene 3 and 4 (*UIS3* and *UIS4*), were deleted. The deletion of either of the 2 genes resulted in the development of sporozoites that were able to invade hepatocytes normally, but unable to mature into exo-erythrocytic schizont and hence unable to establish blood stage infection [60, 61].

A successful malaria vaccine must be safe, effective and capable of inducing long-lasting protective immune response in all age group. Hence, further research is necessary in order to produce a practical malaria vaccine. With the availability of genome sequences, expression data, protein information, immunological data, and advances in molecular genetics techniques, the identification and development of potential vaccine candidates can be expedited.

1.6 *Plasmodium falciparum* genome

The *P. falciparum* genome sequencing project was completed in 2002. The genome is 22.8 megabases long which consists of 14 linear chromosomes, a circular mitochondrial genome and a circular plastid genome. The genome is much more A+T rich than all other genomes, where the overall A+T content is 80.6% and increases to more than 90% in the introns and intergenic regions. The 14 chromosomes exhibit extensive length variation especially in the subtelomeric regions, where the genes involved in antigenic variation are concentrated. The 14 chromosomes encode ~5400 genes. The Gene Ontology (GO) database annotated ~40% of the genes. The annotation revealed that a significant proportion of the *Plasmodium* genome is involved in host cell invasion and cell adhesion (at least 1.3%), and immune evasion mechanisms (~3.9%), when compared to other eukaryotic organisms. On the other hand, there is an under-representation of gene categories involved in cell organization, cell cycle regulation, transcription factors (TFs) and transcription-associated proteins (TAPs). In addition, similarity-searching approaches were not able to identify many major eukaryotic transporters and enzymes in important biochemical pathways. Nearly 60% of the open reading frames (ORFs) translate to hypothetical proteins with no functional annotation. These proteins show no sequence or structural similarity to

those of other organisms and are unique to *Plasmodium*. It is important to elucidate the functions of these unique genes to better understand the parasite and the disease. This information will help in future drug and vaccine development [62].

1.7 *Plasmodium falciparum* transcriptome during IDC

With the availability of the *P. falciparum* genome sequence and the advances in the microarray and proteomic analysis techniques, the transcriptome during the asexual IDC was examined. The global transcriptome profiling was carried out at 1-hour intervals over 48 hours and showed that *P. falciparum* displays tight control of gene expression. At least 80% of the genes are expressed at least once during the IDC. The accumulation of mRNA transcripts occurs in a periodic and temporal fashion. The expression of every gene is linked to the developmental cycle, where each gene is only expressed prior to the time when it is required during the IDC. In addition, genes that are associated in the same cellular process or function are co-expressed (Figure 1.3). As the parasite matures, genes encoding proteins involved in protein translation machinery are maximally expressed in the ring and early trophozoite stages. As the parasite prepares for schizogony, genes encoding proteins involved in DNA replication machinery are maximally expressed during the late trophozoite and early schizont stages. In the late schizont stage where parasites prepare for invasion, genes encoding proteins involved in the merozoite invasion machinery are maximally expressed, and these mRNA transcripts will persist in the early ring stage. Every gene is regulated in a highly ordered and coordinated manner. These observations led to the "just-in-time" hypothesis which suggests that mRNA transcripts are immediately made before the translation process [14, 63].

However, the mRNA transcript accumulation does not correlate well with the protein accumulation. In general, the protein levels peak at one developmental stage later than the peak of the respective mRNA transcript levels. For example, the transcriptome during the ring stage correlates best with the proteome during the trophozoite stage. This translational delay likely reflects a regulatory system at the translational level. Nevertheless, functionally related proteins accumulate simultaneously as observed for the mRNA transcript levels. These observations have also been reported in the other eukaryotes [45, 64].

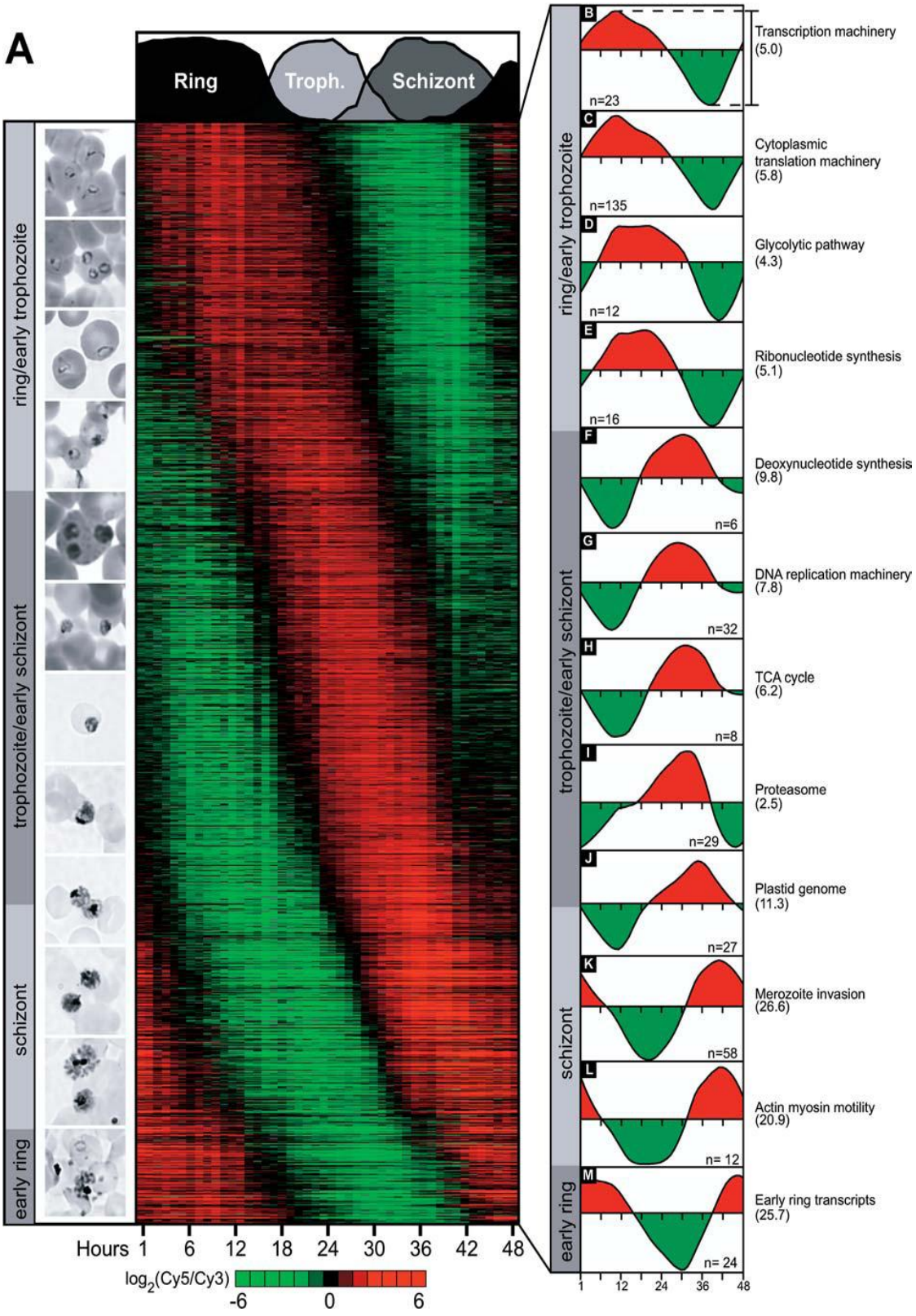


Figure 1.3 Transcriptome of *P. falciparum* intra-erythrocytic developmental cycle.
The heat map shows the genes clustered based on their timing of expression across the 48 hr life cycle. Red represents genes that had higher mRNA levels while green represents genes that had lower mRNA levels. Functionally related genes were found to be co-

expressed. The associated metabolic pathways in the ring, trophozoite and schizont stages during the intra-erythrocytic developmental cycle are shown on the right. This figure is reproduced from Bozdech et al. [14].

1.8 Gene regulation mechanisms in *Plasmodium*

The availability of the *P. falciparum* genome sequence leads to the analysis of the global transcriptome and proteome profiles of the parasite during the asexual life cycle [14, 45, 63]. The findings from these studies (refer to section 1.7) indicated that gene expression is tightly controlled throughout the life cycle. The control of gene expression is fundamental in the parasite developmental progression. However, the mechanisms underlying gene regulation in *P. falciparum* are not yet completely understood.

In addition, *P. falciparum* has evolved an intricate system of mutually exclusive expression of the virulence factors involved in erythrocyte invasion and cytoadherence. The parasites export the *P. falciparum* erythrocyte membrane protein 1 (PfEMP1) to the erythrocyte plasma membrane, and modify the infected erythrocyte surface. This event mediates the adhesion of the infected erythrocyte to blood capillaries and avoids clearance by the spleen. In addition, PfEMP1 is highly variable because there are a total of 60 polymorphic *var* genes encoding different forms of PfEMP1. The parasite is able to undergo antigenic variation to switch to other *var* gene variants periodically in order to avoid the host immune system. However, only one *var* gene is expressed at one point in time. This mutually exclusive expression of a single *var* gene has been suggested to be regulated at several levels including transcription initiation dependent on the promoter upstream of the *var* gene, histone modifications to repress or de-repress gene silencing, and changes in the chromatin structure and sub-nuclear

position. However, mutually exclusive gene expression is not yet completely understood [65-67]. Understanding the molecular mechanisms that control antigenic gene transcription will help us to understand malaria pathogenesis and support drug and vaccine development.

1.9 Transcription regulation and basic transcription machinery

The "just-in-time" hypothesis suggests that mRNA transcript accumulation occurs just before translation [14]. This hypothesis suggests the importance of controlling transcription initiation during gene expression. Hence, promoter-based control of transcription might play an important role in regulating gene expression in *Plasmodium*.

The basal transcription machinery in model eukaryotic systems is highly conserved in *Plasmodium* (Figure 1.4). As described for the other model eukaryotes, the classic model of transcription initiation entails the assembly of a pre-initiation complex (PIC) at the core promoter. The PIC comprises RNA polymerase II (RNA pol II) and the general transcription factors (GTFs) [68]. PIC formation is initiated by the binding of the TATA binding protein (TBP), to a core promoter element, the TATA box. The binding of TBP to the TATA box recruits other TFIID subunits to the PIC. The other GTFs (TFIIA-H) also get recruited sequentially. Transcription occurs in a monocistronic fashion and requires specific DNA elements located in the intergenic DNA regions such as the transcription start site (TSS), the initiator (INR) and the downstream promoter element (DPE), for initiation. Transcription is also dependent on specific interaction between the *trans*-acting factors, the specific TFs (STFs), and the *cis*-acting elements, such as the enhancer or repressor, to activate or repress gene

expression. The *cis*-acting elements are located at the proximal promoter or at a distant region. The interaction between the *trans*-acting factors and *cis*-acting elements leads to the recruitment of RNA pol II, and thereby regulating gene transcription [69, 70].

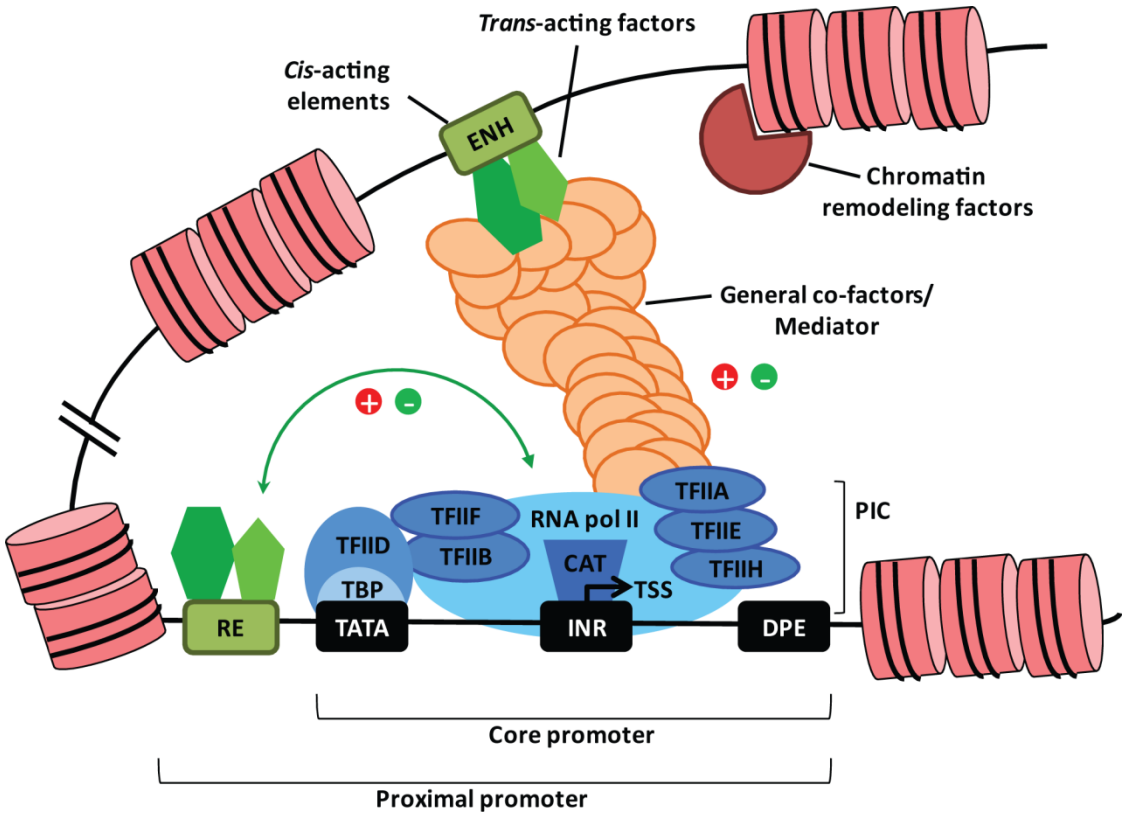


Figure 1.4 The eukaryotic transcriptional apparatus.

The classic model of transcription initiation requires the assembly of the pre-initiation complex (PIC) at the core promoter. The PIC comprises RNA polymerase II (RNA pol II) and the general transcription factors (GTFs, coloured in blue). PIC formation is initiated by the binding of the TATA binding protein (TBP), to a core promoter element, the TATA box (TATA). The binding of TBP to the TATA box recruits other TFIID subunits to the PIC, and the other GTFs (TFIIA-H) also get recruited sequentially. The GTFs also bind to other core promoter regulatory elements (coloured in black), such as the initiator (INR) and the downstream promoter element (DPE). The approximate locations of these core promoter regulatory elements relative to the transcription start site (TSS, black arrow) are shown. The *trans*-acting transcriptional regulators (coloured in green) are either activators or repressors. They bind to *cis*-acting regulatory elements (RE) at the proximal promoter and the enhancer (ENH) distal to the promoter. The transcriptional regulators can interact directly with the GTFs (green arrow), or through transcription

co-factors such as the Mediator complex (orange). In addition, epigenetic factors (coloured in pink) such as chromatin remodeling factors and histone modifying enzymes, act to affect the chromatin organization to influence the accessibility of gene for transcription, or the recruitment of *trans*-acting transcriptional regulators.

In *Plasmodium*, the homologues of all the 12 subunits of RNA pol II and the majority of the GTFs have been identified [68, 71, 72]. The binding of the GTF TFIID to the promoter triggers the recruitment of RNA pol II. The binding of the RNA pol II to the promoter is sufficient for the basal level of transcription. The TBP subunit within TFIID is also found in *Plasmodium*. The primary sequence of the *P. falciparum* TBP shares less similarity to the TBPs in other organisms. The TBP in *P. falciparum* has a 42% identity to the yeast homologue, while the TBPs in other organisms share a 76-93% sequence identity to the yeast TBP. However, the tertiary structure of *P. falciparum* TBP has the characteristic features of, and is similar to the TBPs in other organisms [73]. Interestingly, the components within the TFIID complex responsible for the selection of specific TSS are absent in *Plasmodium*. As such, the implications of the absence of these components on TSS selection in *Plasmodium* remain to be determined.

It has been also demonstrated that the native and recombinant *P. falciparum* TBP is able to bind to the core promoters of knob-associated histidine-rich protein (*kahrp*) and glycophorin binding protein 130 (*gbp-130*) genes. The same study also demonstrated that the location of the consensus TATA box sequence is more upstream of the TSS in *P. falciparum* when compared to other organisms. The consensus TATA box sequences are located at position -186 bp and -130 bp upstream of the TSS in the *kahrp* and *gbp-130* genes respectively [74]. In yeast and human, the functional TATA boxes are usually located at position -120 bp and -35 bp respectively [75]. These

observations suggest that *P. falciparum* might use a scanning mechanism to identify the correct TSS [76]. In addition, the unusual A+T composition also contributes to an increased number of TATA-like sequences on the promoter, thus leading to multiple TSS [77]. As such, how *P. falciparum* TBP recognizes and interacts with the correct TSS in a A+T rich environment remains to be determined.

1.9.1 Promoter organization

The above mentioned classic model of gene transcription is generally associated with promoters that are enriched for TATA box, exhibit a precise TSS and are usually associated with tissue-specific transcription [78, 79]. However, many mammalian promoters do not conform to this classic model of TATA-driven assembly of the PIC. Only ~10% of the human promoters contain a functional TATA box [80, 81]. Similar findings have also been reported in the *Drosophila* promoters [82] and the *Arabidopsis* promoters [83]. It was also reported that at least 50% of the human promoters contain CpG islands [84, 85]. These TATA-depleted, CpG islands associated promoters can be further classified into two classes. The first class of promoters is typically associated with one CpG island, a dispersed distribution of multiple TSS and is associated with ubiquitous expression of housekeeping genes [79]. The second class of promoters has numerous large CpG islands that extend into the gene bodies, and is generally associated with developmentally regulated gene expression [81, 86, 87]. In *Drosophila*, the promoters can be broadly classified into 5 types based on their motif content [88], or 3 types according to their functions [89]. The tissue-specific promoters are similar to the mammalian counterparts and have a TATA box followed by a initiator. The promoters associated with ubiquitous expression are enriched for Motif 1/6 or the

DNA replication element (DRE), while the promoters associated with developmental gene expression are enriched for the DPE.

In addition to the promoter sequences, the organization of the promoters in relation to the nucleosome positions and histone modifications can also be used as criteria to classify the promoters. Tissue-specific promoters are known to have histone H3 lysine 4 tri-methylation (H3K4Me3) marks present downstream of the TSS [90]. Promoters associated with ubiquitous expression have H3K4Me3 marks present on the promoters in all tissue types. The promoters associated with developmental gene expression have both H3K4Me3 and the repressive histone H3 lysine 27 tri-methylation (H3K27Me3) marks [90, 91]. In addition, the promoters associated with ubiquitous expression or developmental gene expression have a more ordered nucleosome organization compared to the tissue specific promoters [92].

Unlike the mammalian promoters, the *Drosophila* promoters lack CpG islands and use other motif sequence determinants to regulate transcription initiation in different promoter types [88]. Similarly, *Plasmodium* genome is highly AT-rich, and is not known to have CpG island. As such, it is likely that *Plasmodium* promoters also use other sequence elements to regulate transcription initiation, and further studies are required.

1.9.2 *Plasmodium* promoters and *Cis*-acting regulatory elements

In *Plasmodium*, the *cis*-regulatory sequences located in the promoter regions are important in regulating gene transcription. Both Northern blot and genomic

approaches have been utilized to characterize the 5' untranslated region (UTRs) of ORFs in *P. falciparum*. Full length cDNA library has been constructed, and the findings indicated that the average length of the 5' UTRs in *P. falciparum* is 346 nucleotides, which is 3 times longer than that in human (125 nucleotides) [77]. On the other hand, northern blot data showed that the 5' UTRs are much longer than that predicted from the genomic data. For example, the 5' UTR of the maebl transcript (codes for putative mitochondrial ATP synthase, PF11_0485) was found to be more than 2 kb long [93]. Similarly, the 5' UTR of casein kinase 1 (CK1) was predicted to be at least 1.2 kb long [94]. Recently, a modeling approach that utilized the Northern blot data of 105 genes was undertaken to examine the lengths of the 5' UTRs in *P. falciparum*. It was predicted that the length of 5' UTR in *P. falciparum* is typically 800 to 1800 bp long [95].

Furthermore, it has been shown that the deletion of specific sequences upstream of the core promoter region [96], the inversion of a promoter [97], and the deletion of certain region within the promoter [97, 98], can result in the complete abrogation, reduction or increase in the reporter gene expression. The deletion or mutation of the 5' UTRs resulted in either an increase or decrease in reporter gene activity, suggesting that the 5' UTRs may contain enhancer or repressor sequence elements that can alter gene transcription [99]. For example, a sequence motif located between positions -1640 bp and -1596 bp upstream of the ATG codon of the *P. falciparum* CDP-diacylglycerol synthase gene was able to drive luciferase reporter gene activity [100]. On the other hand, deletion of 292 bp from the 5' end of the *gbp-130* gene resulted in a 3 fold increase in the reporter gene activity, indicating the presence of a repressor sequence [101]. In addition to these experimentally identified *cis*-acting elements, *in silico*

identification of putative *cis*-acting elements using bioinformatics tools and programs, FIRE (finding informative regulatory elements) and GEMS (gene enrichment motif searching), were also developed [102, 103].

In addition, several studies have demonstrated the role of promoters in the temporal regulation, trafficking and silencing of *trans*-gene expression. Pace et al. demonstrated that the position of promoter on the genome does not affect its activity [104]. Horrocks et al. also showed that gene transcription can be regulated independently of its chromosomal position. Instead, the promoter and terminator sequences immediately upstream and downstream of gene contain regulatory elements that control the timing of expression [105]. Similarly, in the *P. berghei* model, it was demonstrated that the *P. falciparum* AMA-1 promoter is important in the temporal regulation of *trans*-gene expression. Furthermore, the precise timing of AMA-1 expression is important for its subsequent localization [106]. Likewise, the correct localization of *P. chabaudi* AMA-1 is also dependent on the correct timing of expression [107]. It was also found that the *P. falciparum* MSP-2 promoter contains sequence motifs that regulate both the timing and level of MSP-2 expression [108]. The promoters of the *var* genes were shown to contain sequence elements that are important for gene silencing, and the mutually exclusive expression of 1 *var* gene [109]. The temporal control of gene expression is important in both IDC and sexual development. Deletion mapping of the Pfs16 and Pfs25 promoters showed that specific sequence elements in the promoters act as developmental switches at the onset of gametocytogenesis and gametogenesis, respectively [110]. Conversely, one recent study has shown that the temporal regulation of *P. falciparum* proliferating cell nuclear antigen (PfPCNA) expression does not rely solely on the promoter and terminator sequences [111]. As such, it is

likely that multiple players including epigenetic factors and post-transcriptional regulatory machineries are also important for the temporal control of gene expression. More work need to be done to elucidate the interplay between promoter sequence elements and other transcriptional regulatory machineries on stage specific gene expression.

1.9.3 Transcription co-factors

Transcription co-factors play important roles in the interaction of *trans*-acting factors and RNA pol II. In general, these co-factors do not bind DNA directly, but act as a protein bridge to convey signals from *trans*-acting factors bound at distal regulatory elements to RNA pol II in order to initiate transcription (Figure 1.4). One important class of transcription cofactors is the Mediator complex. Mediator is an evolutionally conserved multiprotein complex. It is a large complex which comprises 25 protein subunits in yeast and at least 30 protein subunits in higher organisms [112]. The Mediator subunits are arranged into the head, middle, tail and CDK8/kinase modules [113]. It has been demonstrated that many Mediator subunits exhibit low sequence homology, but their structure and functions are well conserved [112, 114].

The Mediator was first isolated as part of the RNA pol II holoenzyme [115], and is an important co-activator that functions to facilitate the assembly of PIC [116]. The different protein subunits of the Mediator was shown to interact with *trans*-acting activators [117, 118], GTFs [119] and RNA pol II [120]. As such, the Mediator is capable of long range interaction between enhancers and promoters [121]. Once the Mediator is recruited to the promoter by the *trans*-acting factors, it functions to recruit RNA pol II and GTFs to form the PIC [116]. In addition, the Mediator also functions

in transcriptional activation and repression [122], transcription elongation [123] and termination [124]. The Mediator also affects chromatin remodeling through interaction with histone tails and chromatin remodeling factors [125, 126]. The Mediator was reported to interact with the Swi/Snf complexes in yeast, and is required for its recruitment at promoters [127]. It was also demonstrated that the Mediator regulates mRNA processing through interaction with mRNA processing factors [128].

In mammals, there have been reports of transcription activation independent of Mediator recruitment [129]. However, in yeast, Mediator is required for the expression of nearly all genes [130]. In yeast, a temperature-sensitive mutation of an essential Mediator subunit resulted in the loss of almost all mRNA transcription as observed in a RNA pol II temperature-sensitive mutant [131, 132]. In addition, the loss-of-function mutants of the Mediator subunits compromised the recruitment of RNA pol II to promoters [133, 134]. These results indicate the importance of the Mediator in basal transcription. To date, *Plasmodium* lacks many Mediator subunits [135]. As such, the implications of the lack of Mediator on basal transcription in *Plasmodium* remain to be determined.

1.9.4 *Trans*-acting factors

Following the publication of the *P. falciparum* genome, bioinformatics analysis was carried out to identify potential transcription associated proteins (TAPs). The transcription factors involved in PIC formation were almost complete (refer to section 1.9). However, there appears to be an apparent dearth of *trans*-acting factors, the STFs [68, 71, 135, 136].

In one of the studies, a comprehensive directory of TAPs was put together based on all the available databases, including experimentally determined TFs, transcriptome data and proteome data [135]. A total of 202 potential TAPs were identified and are grouped into GTFs, STFs and chromatin-related proteins (Figure 1.5). Amongst the 202 TAPs, 56 proteins are classified as the GTFs involved in PIC formation. They include the RNA polymerase subunits, its accessory proteins and the GTFs (TFII subunits) involved in transcription initiation and elongation. The number of proteins identified is similar to those found in the other eukaryotes [137]. Of the 202 proteins, 63 are grouped in the chromatin-related proteins category. These chromatin-related proteins include 8 histone proteins involved in nucleosome structure and organization. However, the linker histone H1, which functions to package nucleosomes into higher order structure, was not found. There are also 24 histone modifying enzymes such as the histone acetyl transferases (HAT) and histone deacetyl transferase (HDAC). In addition, chromatin structuring and organization factors such as the bromodomain and chromodomain proteins, and chromatin remodeling factors were also identified. The third class of TAPs is the specific transcription factors. This category consists of 73 proteins. These STFs have homologues in eukaryotes known to interact specifically with DNA motifs in gene promoters and regulate gene expression. There are two major families, the ApiAP2 TFs, and the zinc finger proteins, and also includes the helix-turn helix proteins such as *P. falciparum* Myb1 (PfMyb1). The number of STFs is much lower compared to other unicellular eukaryotes of comparable genome size [68, 135].

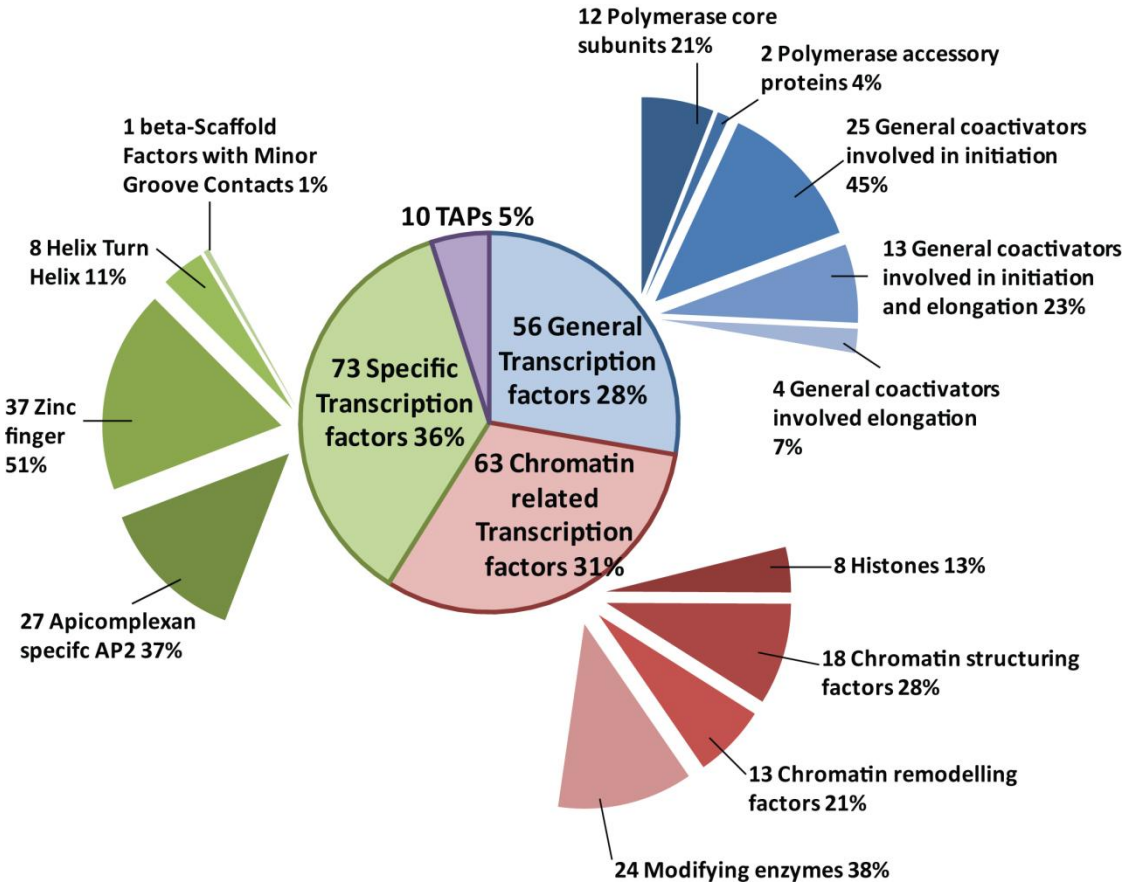


Figure 1.5 Summary of the transcription-associated proteins (TAPs) in *Plasmodium*. **The TAPs were determined from available experimental data and bioinformatics computational tools. A total of 202 TAPs have been identified. These proteins are categorized into 56 general transcription factors (GTFs), 63 chromatin-related TAPs, 73 specific transcription factors (STFs), and 10 other TAPs that do not fall into the other 3 categories [135]. The GTFs include the RNA polymerase subunits, its accessory proteins and the TFII subunits. The chromatin-related TAPs include the histone proteins, histone modifying enzymes such as histone acetyl transferases (HAT) and histone deacetyl transferase (HDAC), and chromatin remodeling complexes. The STFs include the ApiAP2 proteins, the zinc finger proteins, and the helix-turn helix protein *P. falciparum* Myb1 (PfMyb1).**

Taken together, there seems to be a lack of STFs in *Plasmodium*. These findings can be attributed to several reasons. The highly A+T rich genome makes the identification of conserved motifs in the STFs and TAPs difficult and *Plasmodium* severely lacks the DNA binding domains (DBDs) that are characteristic of STFs [62, 71, 136, 138]. With

60% ORFs encoded as hypothetical proteins, the poor annotation of the ORFs makes it difficult to identify putative STFs. In addition, the currently available bio-informatics tools might be inadequate for handling such an A+T rich genome. The dearth of STFs could also be due to a unique parasitic lifestyle, or a genuine evolutionary distance between *Plasmodium* and other eukaryotes. It is possible that *Plasmodium* may possesses a complete but diverse set of STFs that are distantly related to, and has no counterparts in the other eukaryotes [139]. Comparative genomics analysis of various eukaryotic species indicates that STFs and proteins involved in specialized processes such as reproduction and host defence are highly divergent and undergo a more lineage-specific expansion compared to proteins implicated in core cellular processes such as metabolism and protein trafficking [140-143]. It is also possible that only a few STFs are required to regulate gene transcription, or other post-transcriptional mechanisms help to regulate the process.

1.9.4.1 ApiAP2

ApiAP2 is a lineage-specific family of STFs that contains the AP2 DBD. The ApiAP2 STFs identified in *Plasmodium* has been partially characterized [144, 145]. The ApiAP2 STFs family contains 27 members. They have either one or two tandem AP2 DBD. Gene expression data showed that the ApiAP2 genes are differentially expressed in the different stages of IDC and may be important in regulating stage-specific gene expression [145]. Protein-binding microarray assays and computational predictions have been carried out to study their DNA binding specificities [146]. It was found that the recombinant proteins were able to bind to palindromic DNA sequences. For example, PFF0200c was predicted to bind to 66 genes that contain the *cis*-acting DNA sequence, GTGCAC. It was found that the expression profile of

PFF0200c and the average expression profiles of the 66 target genes were highly correlated; suggesting the transcription of PFF0200c might be involved in the transcriptional activation of the 66 target genes [145].

1.9.4.2 *P. falciparum* Myb1

The other STF that was identified and characterized was *P. falciparum* Myb1 (PfMyb1, PF13_0088) [147, 148]. The recombinant PfMyb1 protein was found to be enriched in the nuclear extracts. The recombinant PfMyb1 from the nuclear extract has DNA binding activity, and was able to bind specifically to the Myb-regulatory element (MRE) sequence [147]. In addition, the knock-down of PfMyb1 expression using double-stranded RNA technology inhibited normal IDC progression. There was a 40% growth inhibition in the rings, and the parasites did not transit from trophozoites to schizonts. The knock-down of PfMyb1 also altered the expression of 8 genes, out of which, 7 genes contain the MRE [148].

1.9.4.3 Zinc finger proteins

The zinc finger (ZnF) protein family is one of the most abundant in eukaryotes. They are known to bind DNA, RNA, other proteins and lipids. They have diverse functions, which include DNA and RNA recognition, transcriptional regulation, mRNA stability, mRNA trafficking, and protein folding [149-152]. They are known to be involved in the regulation of important cellular processes such as development, differentiation and immunological responses. Mutation of ZnF leads to interference with many important biological processes [153].

ZnF was first described in 1985 as a DNA-binding motif containing repeated zinc-binding domains in the TFIIIA TF in *Xenopus laevis* oocytes [154]. Since then, many types of ZnF structures have been described. ZnF proteins have diverse structures but always contain at least one ZnF domain. The ZnF is a structural motif that contains conserved histidine (H) and cysteine (C) residues that coordinate with one or more zinc ions in order to stabilize its structure [155]. ZnF proteins are classified mainly based on their secondary and tertiary structures, the number of H and C residues that bind zinc ions and their function. Some of the common classes of ZnF include C2H2, CCHC, Gag knuckle, treble clef finger, C3HC4, zinc ribbon, Zn2Cys6, TAZ2 domain-like, zinc binding loops, and metallothionein [156].

The C2H2 ZnF proteins are commonly found in transcription factors and function in the recognition of specific DNA sequences. They have been implicated in cell differentiation and development via mediating RNA transcriptional repression [157] or transcriptional activation [158]. The C2H2 ZnF proteins were also reported to function in protein-protein interactions and RNA binding [159]. The Gag knuckle ZnF is found in HIV-1 nucleocapsid protein and has been described to be involved in binding specific RNA sequences that are important in virus packaging [160]. The treble clef fingers are found in a variety of proteins such as ribosomal proteins, nuclear receptors, and protein kinases [161]. The zinc ribbon is found in transcription initiation factor TFIIB [162], and elongation factor TFIIS [163]. Zn2/Cys6 is found in transcriptional regulators such as *Saccharomyces cerevisiae* protein Gal4 [164]. The above examples exemplify the diversities of the ZnF proteins, in terms of their structures and functions. These structural motifs are involved in a wide range of cellular processes ranging from transcription to cell signalling. As such, ZnF proteins have also been implicated in

numerous human developmental and neurological diseases [165]. In addition, ZnF proteins have the ability to recognize and bind to specific DNA sequences. Therefore, many studies have been done to explore and develop technologies to use ZnF proteins in gene therapy [166, 167]. In *Plasmodium*, the ZnF family makes up half of the STFs identified (refer to section 1.9.4, Figure 1.5). Their roles in life cycle progression are unclear, thus much work is required to characterize and better understand them.

1.9.4.4 CCCH zinc finger

The ZnF family contains 37 members, which comprises 4 Bbox protein, 3 MYND finger domain protein, 12 CCHH protein, 2 C3HC4 protein, 15 CCCH protein, and 1 U2 snRNP auxiliary factor, small subunit [135]. In particular, the CCCH-type ZnF proteins are highly over-represented in the *P. falciparum* genome [68]. They are found in many organisms including human and yeast. These proteins are C-X_8-C-X_5-C-X_3-H type fingers where X represent variable amino acids. The CCCH-type ZnF are most commonly associated with modulating mRNA stability, mRNA decay and translation rates [168]. This finding, and their abundance in the *Plasmodium* genome, supports the view that post-transcriptional machineries play important roles in controlling gene expression in *Plasmodium*. It was also reported that the CCCH-type ZnF are highly diverse in *Plasmodium* [139].

The majority of CCCH type ZnF proteins are RNA-binding proteins that function to regulate mRNA stability and localization. They function by binding to the adenine-uridine (AU) rich elements in the mRNA and targeting the mRNA for degradation [169, 170]. Another mechanism that CCCH ZnF uses to regulate mRNA stability is by regulating the mRNA polyadenylation process. In *Drosophila melanogaster*, dZC3H3,

a CCCH ZnF, is one such example. The knock-down of dZC3H3 resulted in hyper-polyadenylated mRNAs that sequestered into cytoplasmic foci, and are therefore unavailable for translation. These findings also suggested that the length of the polyA tail is an important mechanism to regulate the quantity and availability of mRNA for translation [171, 172]. In mammals, CCCH ZnF functions to regulate many cell processes, such as inflammatory responses. Tristetraprolin (TTP) is one of the best characterized CCCH ZnF proteins. TTP functions to inhibit TNFα secretion from the macrophages by binding to and destabilizing TNFα mRNA [173]. It was found that the knock-out of TTP resulted in the stabilization of the mRNAs of TNFα and granulocyte-macrophage colony-stimulating factor [174]. In addition, CCCH ZnF proteins are also involved in embryogenesis and development. In *Xenopus*, XC3H-1, XC3H-2, XC3H-3 and XC3H-4 were found to affect mRNA stability and hence regulate oocyte maturation and early embryogenesis. Furthermore, XC3H-4 was found to be absent in adult tissues, and showed a restricted localization to the ovary and oocyte [175]. In *Caenorhabditis elegans*, POS-1, PIE-1, MEX-3 and MEX-6, all containing two CCCH motifs, function to regulate maternal mRNA expression in germline blastomeres to specify cell fate [176-178]. Another family of CCCH ZnF proteins, M*aturing Oocyte Expressed* (MOE)-1, -2, and -3, is involved in oocyte maturation in *C. elegans* [179]. *Trypanosoma* is a genus of intracellular blood parasites like *Plasmodium*. In *Trypanosoma*, most genes are constitutively active and post-translational mechanisms function to regulate gene expression. CCCH ZnF proteins have been implicated in regulating translation by affecting mRNA splicing and export during the differentiation of the parasites from the blood stage to procyclic stage [180-183].

Crop plants are important for food security. As such, many studies have been carried out to improve crop yield, including attempts to identify transcriptional regulators that affect plant architecture and crop yield. In plants, CCCH ZnF proteins are known to play important regulatory roles in plant embryogenesis and development. Rice is an important crop and rice yield is mainly determined by rice architecture. *Oryza sativa* leaf and tiller angle increased controller (OsLIC) was shown to be involved in the regulation of rice plant architecture. It was demonstrated that the knock-down of OsLIC resulted in the activation of brassinosteroids, a class of plant hormones which regulates cell division and stem growth. The activation of brassinosteroids resulted in abnormal plant phenotype and reduced stem elongation. OsLIC was also shown to bind to both double stranded DNA and RNA, and is likely to act as a transcriptional regulator [184]. In another study, *Oryza sativa* delay of the onset of senescence (OsDOS) was shown to act as a negative regulator in leaf senescence. The knock-down of OsDOS resulted in delayed leaf senescence [185]. Many of such studies have also been carried out in *Arabidopsis*, a model organism for studying plant biology. A ZnF protein, PEI1, was shown to regulate embryogenesis. The knock-down of PEI1 resulted in abnormal seed development [186]. Somnus, a CCCH ZnF which localized to the nucleus, was found to be a negative regulator of phytochrome, a photoreceptor that promotes seed germination in the presence of light. The knock-down of Somnus allowed seeds to germinate in the absence of light [187]. In addition, the knock-down of CsSEF1, containing 2 CCCH and 1 CHCH ZnF motifs, affected embryo, seed, and root development [188]. Identification of factors that regulate tolerance to salt stress has important relevance to increasing crop yield because salt stress in a major limiting factor in agriculture. In addition, bacterial and fungal diseases are also known to severely limit crop yield [189]. CCCH ZnF is also known to play important regulatory

roles in plant diseases and stress responses. The activation of C3H12, a CCCH ZnF, resulted in increased resistance to *Xanthomonas oryzae* infection, the causative agent of rice bacterial blight disease [190]. The over-expression of GhZFP1, found in cotton, resulted in increased resistance to *Rhizoctonia solani*, a plant pathogenic fungus. It was also found to enhance tolerance to salt stress [191]. In *Arabidopsis*, disrupted expression of AtSZF1 and AtSZF2 led to the increase expression of salt-responsive genes, and increased the plant sensitivity to salt stress. These findings implied the importance of negative regulation on the expression of salt-responsive genes [192]. In maize, transcriptome analysis identified 12 CCCH ZnF transcripts with differential expression in response to drought treatment, suggesting the importance of these proteins in stress response [193].

These findings demonstrated the diversity and importance of the roles CCCH ZnF proteins play in different organisms. However, in *Plasmodium*, the roles of these proteins remain elusive. In *Plasmodium*, TAPs are not only involved in regulating transcription during proliferating stages in the IDC. Some TAPs have been suggested to be involved in differentiation and arrest of gametocyte and sporozoite stages. Two CCCH-type ZnF, PF13_0314 and PFF0095c, are preferentially expressed in gametocytes, while a CCHH-type ZnF, PF14_0707, is expressed more in sporozoites [68]. Studying these TAPs is essential to understand the complex development of the parasite in the human and mosquito hosts.

1.10 Post-transcriptional regulation of gene expression

There are many studies that have contributed evidence to support a prominent role for post-transcriptional mechanisms in regulating gene expression in *Plasmodium*. Post-

transcriptional mechanisms function to control the rate of translation by mediating mRNA processing, maintaining mRNA stability, directing mRNA distribution and degradation.

Steady state mRNA level is dependent on *de novo* synthesis and degradation processes. In *Plasmodium*, studies have shown that mRNA transcripts levels remained high in the absence of *de novo* mRNA synthesis. The transcripts made during the trophozoite stage remain high in the schizont stage [194]. Another study found that mRNA transcript half-life increases during IDC. The average mRNA half-life is 9.5 min in the rings, increases to 20.5 min in the trophozoite stage, and subsequently increases to 65.4 min during the late schizont stage. These observations suggest that the mRNA half-lives were increased during the progression of the life-cycle [195]. The continual high mRNA level in the absence of *de novo* synthesis suggested that post-transcriptional regulatory events might play roles in promoting mRNA stability [196].

In addition, mature mRNA transcripts without the 7-methyl guanosine cap are present in *Plasmodium*. As in other eukaryotes, uncapped mRNA transcripts are suggested to be a form of quiescent storage that can be rapidly mobilized for translation under conditions of stress and starvation [197, 198]. In eukaryotic cells, a post-transcriptional mechanism known as translation repression is known to be implicated in early embryogenesis and neurological diseases [199]. Translation repression refers to a state where quiescent mRNAs are assembled with ribonucleoproteins into cytoplasmic foci known as P bodies [200, 201]. These mRNA transcripts can either be degraded, or can return to translation. In *Plasmodium*, translation repression has also been described in the gametogenesis and sexual developmental processes. A

comparison of the gametocyte transcriptome to the gametocyte and ookinete proteomes revealed that the mRNA transcripts of 9 genes were stored in cytoplasmic bodies during the gametocyte stage. These transcripts persist after fertilization, and only get translated during the ookinete stage in the mosquito vector [46, 202]. Together, these data put forward the association of post-transcriptional regulation with the parasite's physiological development and potential response to stress or drug challenge.

1.11 Epigenetic regulation of gene expression

In eukaryotes, DNA is not naked but is wrapped around histone octamer to form nucleosomes which are further arranged into higher-order chromatin structure. Chromatin not only helps to compact DNA in the nucleus but also act as a physical barrier to restrict the access of transcription machineries [203]. In addition to DNA *cis*-acting elements and *trans*-acting factors (section 1.9), epigenetic mechanisms also play important roles in transcriptional regulation. Epigenetic regulation refers to the control of gene expression by altering chromatin composition, structure and organization to affect the accessibility of the DNA to the transcription machineries, rather than altering the underlying DNA sequences. Epigenetic mechanisms include chromatin remodeling, histone modifications and DNA methylation.

1.11.1 Chromatin remodeling

The organization of chromatin differs between the euchromatin and heterochromatin regions. The chromatin structures are in turn regulated by mechanisms including histone modifications and chromatin remodeling [204]. The regulation of nucleosome positioning and chromatin structure is vital to regulating transcription. Chromatin

remodeling complexes such as Swi/Snf complex in yeast have been extensively studied [205, 206]. It was found that promoters with CpG islands have a more ordered nucleosome organization [92] and they often are not dependent of Swi/Snf remodeling complexes [207]. On the other hand, Swi/Snf-dependent activation of gene transcription occurs on promoters that lack CpG islands and have a less ordered nucleosome organization. However, they are exceptions where promoters lacking CpG island are Swi/Snf-independent [207]. In addition, it was found that mutations in the Swi/Snf complex blocked transcription and this was rescued by mutations that decreased the number of histones [208]. Furthermore, the Swi/Snf complex has been implicated in development and differentiation in mice [209, 210], and cancer in human [211].

1.11.2 Histone modifications

Many residues on the N-terminal tails of histones are post-translationally modified. These modifications include acetylation, methylation, phosphorylation, ubiquitination and SUMOylation. These modifications alter the charges on the histone tails and modulate the interaction between neighboring nucleosomes and hence the chromatin structure. An open chromatin conformation allows easy access to the transcription machineries, while a closed chromatin conformation results in transcription repression. The modifications also affect the recruitment of different *trans*-acting factors [204]. The different histone modifications do not work alone. They occur in distinct patterns and interact with each other. The combination of the different histone modifications forms the histone code that affects transcription [212]. The different histone modifications lead to transcription activation or repression. For example, acetylation is an activating mark [213], SUMOylation is a repressing mark [214], while methylation

can be an activating or repressing mark [215]. In addition, the different types of modifications can be co-occurring or mutually exclusive of each other. Some modifications can chemically or sterically block another modification [216]. For example, histone H3 lysine 9 (H3K9) can only have either an acetylation mark or methylation mark, and not both modifications at the same time. H3K9 acetylation (H3K9Ac) is a euchromatin and activating mark [213], while H3K9 methylation (H3K9Me) is a heterochromatin and repressing mark [217]. Different modifications on the same histone residue occur on different regions within the gene. For example, histone H3 lysine 36 acetylation (H3K6Ac) occurs on promoter regions, while H3K36 methlyation occurs on coding regions [218].

1.11.3 DNA methylation

DNA methylation is a process where a methyl group is covalently added to the 5-carbon position of cytosine, and typically occurs within CpG dinucleotides [219]. DNA methylation influences many important processes including genome stability [220], X chromosome inactivation [221], imprinting [222], and embryonic development [223]. As such, the disruption of DNA methylation can cause developmental abnormalities [224], cancer [225] and genetic diseases such as Fragile X syndrome [226]. DNA methylation on promoters typically leads to transcription repression. This is because DNA methylation results in the inhibition of binding of STFs [227], or the recruitment of methylated DNA binding proteins which compete with STFs for binding sites [228]. In addition, methylated DNA is enriched for hypoacetylated histones which compacts chromatin [229].

1.11.4 Epigenetic mechanisms in *Plasmodium*

In *Plasmodium*, epigenetic mechanisms have also been suggested to be important in gene regulation. The organization of chromatin in *Plasmodium* is similar to other eukaryotes, where ~150 bp DNA is wrapped around a histone octamer consisting of two of each of the histone proteins, H2A, H2B, H3, and H4 [230]. Mass spectrometry and chromatin immunoprecipitation techniques identified a conserved set of histones and their modifications on the N-terminal tails, with the exception of linker histone H1 [230, 231]. In addition, a set of chromatin remodeling enzymes and histone modifying enzymes was also identified [136, 232].

P. falciparum maintains its genome in a euchromatic state as characterized by the enrichment of H3K9Ac and histone H3 lysine 4 methylation (H3K4Me) marks [231, 233, 234]. Acetylation leads to an open chromatin conformation and is an activating mark, while deacetylation leads to a closed chromatin conformation and is a repressing mark [213]. The importance of these histone marks became evident when the perturbation of histone modifying enzymes resulted in a global disruption of histone modifications and an overall effect on transcriptional activity, and the progression through IDC was also affected. The inhibition of HAT with curcumin or anacardic acid resulted in a decrease in H3K9Ac, which caused a decrease in transcriptional activity, and eventually affected progress throughout IDC [235, 236]. Interestingly, the inhibition of HDAC with apicidin also resulted in a global disruption of histone modifications and an overall effect on transcriptional activity and progression through the IDC [237]. It was found that the inhibition of HDAC resulted in the activation of genes which are otherwise repressed during a particular stage of IDC. For example, tricarboxylic acid (TCA) cycle-related genes which are normally repressed during the

ring stage were expressed. In addition, sporozoite and gametocyte stage specific genes, which are otherwise repressed during IDC, were also activated by apicidin treatment. At the same time, genes that are expressed under normal growth conditions were repressed. For example, invasion-related genes that are normally expressed during the schizont stage were repressed. Taken together, these findings exemplify the importance of chromatin structure on gene regulation in *Plasmodium*.

In eukaryotes, the combination of the different histone modifications forms the histone code that affects transcription [212]. In *Plasmodium*, the histone code is also likely to play important roles in regulating gene transcription [230, 231, 238]. Chromatin immunoprecipitation experiments demonstrated that the *Plasmodium* genome is associated with different histone marks, and these marks correlated well with transcriptional activities. It was found that H3K9Ac and H3K9Me were associated with transcriptionally active and silent promoters, respectively [231]. In addition, different levels of histone modifications were found during the different stages of the IDC. For example, histone H4 lysine 8 acetylation (H4K8Ac) was found to be associated with the intergenic regions during the trophozoite and schizont stages [238]. Epigenetic mechanisms were also suggested to regulate the mutually exclusive expression and switching of the *var* gene family [239, 240], and gene families involved in host cell invasion [241]. The promoter of an active *var* gene is associated with H3K9Ac and H3K4Me3 histone marks. On the other hand, the promoters of the silent *var* genes are associated with histone H3 lysine 9 tri-methylation (H3K9Me3) [239, 240]. These findings suggest that the histone code is likely to be involved in regulating gene transcription during the progression of the parasite life cycle. Recently, DNA methylation has been described in *Plasmodium* and was suggested to play roles

in gene expression and transcription elongation [242]. Taken together, these findings imply the importance of the epigenetic machineries on transcriptional activation and repression in *Plasmodium*.

In summary, transcriptional, epigenetic, and post-transcriptional mechanisms likely all play important roles in regulating gene expression in *Plasmodium*. However, how the different mechanism work together to achieve such sophisticated gene regulation remains to be determined. Much work is still required in order to identify and characterize the *cis*-acting elements and *trans*-acting factors.

1.12 Nuclear compartmentalization

In a cell, biomolecules and biological processes are confined to specific compartments. In particular, nucleus is a specific compartment where important activities controlling gene expression are carried out. Within the nucleus, specific sub-compartments that orchestrate replication, transcription and repair have also been described [243].

In eukaryotic system, DNA has been shown to localize to specific compartments within the nucleus and that these nuclear compartments play important roles in regulating gene expression [244, 245]. DNA is organized into the euchromatin and heterochromatin in different compartments [246]. Gene activation and silencing was shown to be associated with the repositioning of genes within the nucleus. Silenced genes are typically positioned at the nuclear periphery and will be repositioned to transcription factories upon activation [247-250]. For example, during the development of B and T lymphocytes, there is a dynamic repositioning of genes away from the heterochromatin compartment [251]. Chromosomes are found to occupy

distinct territories within the nucleus [246]. Within each chromosome, there is also compartmentalization of early-, mid- and late-replicating genes, which can be visualized as chromosome banding patterns [252]. In the event of DNA damage, DNA repair machineries were demonstrated to relocate to the site of DNA damage [253]. Other nuclear compartment such as the nucleolus is a compartment where ribosomes are being assembled [254]. The small nuclear ribonucleoproteins (snRNPs) are assembled into the Cajal bodies [255]. Pre-mRNA splicing factors were reported to accumulate at nuclear speckles [256]. In addition, in diseases such as Huntington's disease and myotonic dystrophy, it has been demonstrated that mutant proteins and mRNAs sequester transcription regulatory proteins in nuclear foci, disrupt the normal transcriptional and translation programs, and contribute to disease phenotype [257, 258]. In all, these findings demonstrated the importance of nuclear compartmentalization in order to carry out the various DNA-related functions effectively.

In *Plasmodium*, DNA has also been shown to localize to specific compartments within the nucleus. Within the nucleus, compartments such as telomere clusters and perinuclear heterochromatin have been described [259]. In particular, the roles of nuclear compartmentalization and positioning in the transcriptional regulation of *var* genes have been extensively studied [260-264]. The *var* gene family is located at both inter-chromosome and subtelomeric loci. The silenced *var* genes are positioned at the nuclear periphery. The perinuclear positioning of silenced *var* genes is associated with the PfSir2-dependent H3K9Me3 modification. The activation of a *var* gene is associated with the repositioning at the nuclear periphery, where the *var* loci relocate

to a transcriptionally active site, also located at the nuclear periphery. The active and silence *var* genes localized to distinct foci at the nuclear periphery.

The distribution and dynamics of nuclear pores in *Plasmodium* during IDC has also been studied [265]. It was found that ring stage parasites have 3-7 nuclear pores clustered close together. Trophozoite stage parasites have 12-58 nuclear pores that are uniformly distributed around the nuclear envelope. In schizont stage parasites, the number of nuclear pores decreases as schizogony progresses. Also, the nuclear pores were always found to be located close to euchromatin and not heterochromatin. These findings suggest the correlation of nuclear architecture and parasite life cycle development. In addition, the organization of histones modifications within the nucleus has been relatively well characterized [266, 267]. It was shown that repressive and activating histone modifications are localized to distinct compartments within the nucleus. Also, the *in vivo* localization of twelve proteins predicted to contain known domains involved in gene regulation were studied [268]. It was found that these potential regulatory proteins localized to distinct nuclear compartments. Together, these findings demonstrated the importance of an ordered compartmentalization within the nucleus in gene activation and silencing in *Plasmodium*.

1.13 Aim of thesis

The *P. falciparum* transcriptome suggests that the parasite has a highly regulated transcriptional machinery. The presence of the near complete set of RNA pol II and GTFs subunits [135] suggests that *Plasmodium* likely shares similar transcription mechanisms with other eukaryotes. However, the lack of STFs and DNA regulatory motifs [62, 71, 136, 138], the delay in the timing of mRNA transcript and protein

accumulation [45], and the variation in RNA decay rates [195], suggest that gene expression in *Plasmodium* might be a more general and promiscuous process. These findings also suggest that precise and specific transcription initiation might play a less significant role in gene regulation than in other eukaryotes. Hence, post-transcriptional regulation might have important roles in gene regulation in *Plasmodium*. Furthermore, perturbation of epigenetic machineries disrupted IDC progression [235-237]. However, we cannot rule out the possibility that *Plasmodium* possesses a set of diverse STFs. It is likely that *P. falciparum* uses a combination of the different mechanisms to regulate gene expression. It is of interest and importance to study how the parasites achieve such a sophisticated level of gene regulation. Knowledge about transcriptional regulation will hopefully shed new light on the pathogenesis of malaria and assist in the development of novel targets for effective treatment. Hence it is of upmost importance to uncover other unknown STFs and to characterize *Plasmodium*-specific TFs. In *Plasmodium*, the largest class of STFs is the ZnF proteins. 51% is the ZnF proteins are the CCCH type. However, their functions are poorly understood. We previously demonstrated that the inhibition of HDAC resulted in a global disruption of histone modifications [237]. In addition, the expression of stage specific TFs and TAPs were also affected. PF10_0083 was one such protein that was up-regulated in ring, trophozoite and schizont stages. As such, this thesis aims to characterize this putative transcription factor, PF10_0083, a CCCH type ZnF. In particular, the relationship of PF10_0083 and gene expression during the parasite life cycle was studied.

The experimental studies were divided into three main objectives:

1) Molecular characterization of PF10_0083

Transgenic *P. falciparum* cell lines with PF10_0083 over-expression and knock-out expression were generated. Transgenic *P. falciparum* cell lines with PF10_0083 tagged at its endogenous locus were also generated. Phenotypic characterization of the generated cell lines investigating the significance of PF10_0083 on *Plasmodium* growth was carried out. In addition, characterization of PF10_0083 mRNA and protein expression, and its intracellular localization were also carried out.

2) The role of PF10_0083 in transcriptional regulation

Transcriptional profiling was carried out to investigate the global transcriptional changes caused by PF10_0083 over-expression and knock-out expression. Chromatin immunoprecipitation coupled with microarray (ChIP-on-chip) analysis was also carried out in multiple developmental stages in the over-expression and endogenous-tagged cell lines. Results from the transcriptional profiling and ChIP-on-chip studies were correlated, in order to identify the potential genomic targets that directly interact with PF10_0083. The genome occupancy and temporal binding pattern of PF10_0083 to the genome was analyzed. The DNA regions that directly interact with PF10_0083 were identified.

3) The role of PF10_0083 in gametocyte development

With the data obtained from objectives 1 and 2, functional experiments were carried out to study the effects of PF10_0083 over-expression and knock-out on gametocyte production and development.

Chapter 2 Materials and methods

2.1 Parasite culture

P. falciparum 3D7 and T996 strains were cultured in washed human RBCs purified from whole blood obtained from anonymous donors. The parasites were grown in cell culture media that was sterilized by filtration. The cell culture media is a Hepes (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid)-buffered RPMI 1640 media (GIBCO®), supplemented with 0.25% Albumax II (GIBCO®), 2 g/l sodium bicarbonate ($NaHCO_3$) (SIGMA), 10 mg/l gentamycin (GIBCO®), and 0.1 mM hypoxanthine (SIGMA). The parasite culture was maintained at 2% hematocrit, kept at pH 7.4, in sterile cell culture flasks. The culture was grown in a 5% CO_2, 3% O_2, and 92% N_2 gas mixture in a 37°C incubator. Culture media was changed every 24 hr, and fresh blood was added to the culture every cycle during the schizont stage. Parasites were synchronized by 2 consecutive 5% D-Sorbitol (SIGMA) treatments 10-12 hr apart for at least 3 generations before they were used for further experiments.

2.1.1 Gametocyte culture

To induce asexual stage parasites to develop into sexual stage gametocytes, the parasites were starved by not supplying fresh blood, and supplying limited amount of fresh culture media. Highly synchronized ring-stage parasites at 1.5% to 2% parasitemia and 2% hematocrit were prepared. Limited volume of fresh culture media was provided every 24 hr, but fresh blood was not given. Stressed schizont (small schizonts with poorly defined nuclei) stage parasites were observed after 5-7 days (Chapter 1, Figure 5.1). The following day was considered as day 0 of gametocyte

culture. On day 3, early stage gametocytes (stage I-III) were observed. On day 5, late stage gametocytes (Stage IV-V) were observed. Giemsa-stained smears were prepared every day, and parasitemia was recorded.

2.2 Genomic DNA isolation

Infected RBCs (iRBCs) were washed twice in 1X phosphate-buffered saline (PBS), pH 7 and pelleted at 2400 rpm for 5 min. The packed cells were used for genomic DNA isolation using the Easy-DNA kit (Invitrogen Protocol #2-30 Minute DNA Extraction from Blood Samples) according to the manufacturer's protocol. The genomic DNA was quantified by UV (ultraviolet) spectrometry and stored at -20°C.

2.3 Over-expression, endogenous tagging, targeted gene disruption

2.3.1 PCR amplification

Polymerase Chain Reaction (PCR) was carried out using the Expand High Fidelity PCR System from Roche Diagnostics, using the Eppendorf Thermocycler. A 25 μl PCR reaction volume containing 50 ng *P. falciparum* genomic DNA, 0.2 μM of each primer (Table 2.1), 0.2 μM of each deoxyribonucleotide (dNTPs: dATP, dTTP, dCTP, dGTP) from Fermentas, 10 μl of 10X Expand High Fidelity buffer, and 1 unit of Expand High Fidelity Polymerase (Roche) was prepared. The PCR cycling parameters used were as follows: 1 cycle at 94°C for 5 min, 30 cycles of 94°C for 1 min, 55°C for 1 min and 60°C for 2 min, 1 cycle of 60°C for 7 min and followed by pausing at 4°C. All PCR reactions were carried out in 0.2 ml tubes. PCR products were resolved on 1% agarose gels. All PCR products were purified in 15 μl elution buffer using MinElute PCR purification kit (QIAGEN) according to the manufacturer's protocol.

Construct	Restriction enzyme	Primer name	Primer sequence (5' → 3')
pARL-myc	XhoI	pARL-myc forward primer	GCGA**CTCGAG**CCAGAGCAAAA GCTCAT
	KpnI	pARL-myc reverse primer	GCAT**GGTACC**CTAGACTTTCTC AAACTGAG
pARL-gfp	XhoI	pARL-gfp forward primer	GCGA**CTCGAG**ATGAGTAAAGG AGAAGAACTTTTCACTGG
	KpnI	pARL-gfp reverse primer	GCAT**GGTACC**TTATTTGTATAG TTCATCCATGCCATGTG
pARL-PF10_0083-myc, pARL-PF10_0083-gfp	XhoI	PF10_0083 forward primer	GCGA**CTCGAG**ATGTTATACAA AATGAATGTATCGTCCG
	AvrII	PF10_0083 reverse primer	GCAT**CCTAGG**ATAGATATATAT ATTTTCATATTTTTTTGTGATTT CTT
PF10_0083-CHA (endogenous tagging)	XmaI	CherryHA XmaI forward primer	GCGA**CCCGGG**ATTTTTGGAAG AGAAAAAAATGTATG
	PstI	CherryHA PstI reverse primer	GCGA**CTGCAG**ATAGATATATAT ATTTTCATATTTTTTTGTGA
PF10_0083-KO (targeted gene disruption)	XmaI	KO forward primer	GCGA**CCCGGG**AACTATGCAAA AGCTACAAAAAA
	SpeI	KO reverse primer	GCGA**ACTAGT**GATGATTTGAAA TATATGTATCATTATAC

Table 2.1 Primers design for preparing PF10_0083 constructs for transfection in *Plasmodium*.

PF10_0083 constructs used in over-expression studies (pARL-PF10_0083-myc, pARL-PF10_0083-gfp), endogenous tagging (PF10_0083-CHA), and targeted gene disruption (PF10_0083-KO) studies. Restriction enzyme sites are highlighted in bold.

2.3.2 Construction of transfection clones

2.3.2.1 Restriction digestion

The plasmids used for cloning are pARL for over-expression (Chapter 3, Figure 3.2A), and pTCherryHA (Chapter 3, Figure 3.8A) to prepare PF10_0083-CHA (endogenous

tagging) and PF10_0083-KO (targeted gene disruption). 3 μg of plasmids and 1 μg of PCR products, generated according to section 2.3.1, were digested with the respective enzymes from New England Biolabs (Table 2.1). The plasmids and PCR products were digested in 30 μl reaction volume that contains 1 unit of enzyme at 37°C overnight. The digested PCR products were purified in 15 μl elution buffer using MinElute PCR purification kit (QIAGEN) according to the manufacturer's protocol. The digested plasmids were gel-extracted and purified in 30 μl elution buffer using the QIAquick gel extraction Kit (QIAGEN) according to the manufacturer's protocol, and quantified by UV spectrometry.

2.3.2.2 Ligation of digested products

A 1:3 molar ratio of vector to PCR products was used for each ligation reaction. The purified digested PCR products were ligated to the purified digested plasmid using the LigaFast^{TM} Rapid DNA ligation system (Promega) at 25°C for 5 min in a 10 μl reaction volume that contains 3 units of ligase enzyme.

2.3.2.3 Transformation of ligated products

10 μl of ligation reaction volume was used for transformation into 100 μl of *E. coli* DH5α competent cells (Agilent Technologies). The ligated products were incubated with the competent cells on ice for 30 min, heat-shock at 42°C for 1 min, and incubated on ice for 5 min. 1 ml of Luria-Bertani (LB) broth was added to the transformed cells and incubated at 37°C with agitation at 220 rpm for 45 min. The cells were plated on LB plates, supplemented with 100 μg/ml of ampicillin, and incubated at 37°C overnight.

2.3.2.4 Verification of recombinant plasmid-DNA

Colonies were selected and picked from the LB plates (section 2.3.2.3) and cultured in 2 ml of LB broth, supplemented with 100 μg/ml of ampicillin, and incubated at 37°C with agitation at 220 rpm overnight. Plasmid DNA was isolated and eluted in 30 μl elution buffer using the QIAprep Miniprep Plasmid Purification Kit (QIAGEN) according to manufacturer's protocol. The colonies were screened for correct inserts size by restriction digestion using the respective restriction enzymes (Table 2.1). Positive clones were sent for sequencing to confirm the sequences. The primers used for making the constructs are used for sequencing (Table 2.1). Glycerol stocks were prepared and stored in -80°C.

2.3.2.5 Preparation of recombinant plasmid-DNA

To prepare large quantities of recombinant plasmid-DNA, the positive clones were cultured in 100 ml of LB broth, supplemented with 100 μg/ml of ampicillin, and incubated at 37°C with agitation at 220 rpm overnight. Plasmid DNA was purified and eluted in 300 μl sterile water using QIAGEN Plasmid Maxiprep Purification Kit (QIAGEN) according to manufacturer's protocol. The plasmid DNA was quantified by UV spectrometry, and stored at -20°C in aliquots of 1 μg/μl.

2.4 Transfection

The plasmid DNA was introduced into *P. falciparum* T996 and/or 3D7 strains by electroporation. Parasite cultures of 20% to 25% highly synchronized rings were pelleted and resuspended in Cytomix (120 mM potassium chloride (KCl), 0.15 mM calcium chloride ($CaCl_2$), 10 mM potassium phosphate buffer (K_2HPO_4/KH_2PO_4), 25

mM Hepes/2 mM ethylene glycol tetra-acetic acid (EGTA), pH 7.6) and 100 μg plasmid DNA. Electroporations were performed using a Bio-Rad Gene Pulser Xcell Total Electroporation System (Bio-Rad Laboratories) at settings of 320 V, 950 μF, and 0 Ω in 2 mm cuvettes. Time constants were 7-12 msec. Following transfection, parasites were cultured in normal cell culture media for 10-14 hr before selection with 2.5 nM WR99210 (Sigma). WR99210 resistant parasites were detected after 2-3 weeks of continuous culture, and WR99210 concentration was increased to 20 nM. For the over-expression parasite lines, WR99210 was maintained at 20 nM concentration. For the endogenous-tagged and knock-out parasite lines, drug pressure was removed after the integration of the plasmid into the endogenous locus has been validated (section 2.5).

2.5 Verification of integration of plasmid into genome

Integration of the plasmid at the endogenous PF10_0083 locus in the transfected parasites was verified. DNA was isolated from the parasites according to section 2.2. PCR was carried out according to section 2.3.1. The primers used are listed in table 2.2.

Construct	Primer name	Primer sequence (5' → 3')
PF10_0083-CHA (endogenous tagging)	endogenous PF10_0083 5' end forward primer	ACCAAAATATGCAAACATTATCTT
	plasmid cherry reverse primer	ATGATGGCCATGTTATCCTC
PF10_0083-KO (targeted gene disruption)	endogenous PF10_0083 5' end forward primer	ACCAAAATATGCAAACATTATCTT
	plasmid 3'CAM reverse primer	GTTACATGCACGTATATTTATACGAA
	plasmid 5'CAM forward primer	AACCAATAGATAAAATTTGTAGAGAAA

	endogenous PF10_0083 3' end reverse primer	CATTTATATAATTCGTTAAAACACTCTTC

Table 2.2 Primers design for the verification of the integration of plasmid DNA into genome.

Integration of plasmid into PF10_0083 locus in the PF10_0083-CHA (endogenous tagging) and PF10_0083-KO (targeted gene disruption) transfectants.

2.6 Cloning by limiting dilution

Highly synchronized PF10_0083-CHA and PF10_0083-KO parasite culture was used for cloning by limiting dilution [269]. 12 hr prior to the invasion of RBCs to form new ring form parasites, parasite cultures were placed on a shaker at 37°C. Shaking was done to ensure that 1 RBC gets infected by only 1 merozoite. 4-6 hr after the invasion of RBCs to form new ring form parasites, giemsa-stained smear was prepared and parasitemia was checked by light microscopy. At least 95% of the iRBCs should have a single ring form parasite, and not have multiple infections.

The parasites were then serially diluted with fresh RBCs. There is 10^7 RBCs per 0.001 ml volume. The serial dilutions done were based on the parasitemia count, and calculating the expected number of parasites, such that a final diluent of 0.5 parasites per 0.1 ml volume at 5% hematocrit was obtained. Parasites were first diluted with fresh RBCs to obtain 10,000 iRBCs per 0.1 ml volume (diluent A). Diluent A was then diluted with fresh RBCs to obtain 100 iRBCs per 0.1 ml volume (diluent B). Diluent B was then diluted with fresh RBCs to obtain 0.5 iRBCs per 0.1 ml volume (diluent C). 100 μl of diluent C was transferred to 48 wells in two 96-well plates (Figure 2.1). Positive controls were prepared with 100 μl of diluent A and diluent B in the same plates.

The plates were gassed and kept in 37°C incubator. Culture medium was replaced every 48 hr and fresh RBCs was added every 6 days. On day 4 and day 6, the presence of parasites in the positive controls (diluents A and B) were checked. On day 6, the parasite cultures in all the wells were split into two wells, by transferring half the parasite culture to the empty well beside and adding fresh RBCs. From day 10 onwards, for every alternate day, giemsa-stained smears were prepared for all 48 wells to check for presence of parasites. 21 out of 48 wells were positive in PF10_0083-CHA, and 15 out of 48 wells were positive in PF10_0083-KO (Appendix, Supplementary Figure 1). All positive wells were subsequently transferred to 24-well plates, and were later cultured in cell culture flasks, and maintained at 1-2% hematocrit. All positive clones were cryopreserved in liquid N_2. PF10_0083-CHA18, PF10_0083-CHA30, PF10_0083-KO24, and PF10_0083-KO33 were randomly selected for further studies.

plate 1	1	2	3	4	5	6	7	8	9	10	11	12
A												
B		1		7		13		19		diluent A		
C		2		8		14		20		diluent A		
D		3		9		15		21		diluent A		
E		4		10		16		22				
F		5		11		17		23				
G		6		12		18		24				
H												

plate 2	1	2	3	4	5	6	7	8	9	10	11	12
A												
B		25		31		37		43		diluent B		
C		26		32		38		44		diluent B		
D		27		33		39		45		diluent B		
E		28		34		40		46				
F		29		35		41		47				
G		30		36		42		48				
H												

Figure 2.1 Cloning by limiting dilution.

48 wells in two 96-well plates were prepared with serially diluted parasite culture such that an average of 0.5 parasites per 0.1 ml volume was obtained. Diluent A and B are positive controls. Orange-coloured wells are culture media to maintain humidity in the plates.

2.7 Western analysis

2.7.1 Harvesting *Plasmodium* parasites samples

Parasites time-point samples were harvested from highly synchronized cultures. 0.1% saponin (Sigma) were added to the iRBCs in a 10:1 ratio, and incubated for 5 min at room temperature. The parasites were pelleted by centrifugation at 4500 rpm for 5 min and washed 3 times in 1X PBS. Parasite pellets were stored at -80°C.

2.7.2 Total protein lysate

Parasite pellets were lysed by boiling in 2X Laemmli sample buffer (100 mM Tris-HCl pH 6.8, 4% (w/v) sodium dodecyl sulfate (SDS), 20% (v/v) glycerol, 10% β-mercaptoethnol, and 0.2% (w/v) bromophenol blue) for 10 min. Insoluble material were pelleted by centrifugation at 13,000 rpm for 10 min. Supernantant was collected as total protein lysate. Protein concentration was quantified by Bio-Rad Protein assay (Bio-Rad Laboratories) according to manufacturer's protocol. Total protein lysate were stored at -20°C.

2.7.3 Nuclear fractionation

Parasite pellet were incubated in cell lysis buffer (CLB), containing 20 mM Hepes pH 7.9, 10 mM KCL, 1 mM ethylenediaminetetraacetic acid (EDTA), 1 mM EGTA, 0.65% Nonidet P-40 (NP-40), 1 mM dithiothreitol (DTT), and 1X protease inhibitor (Roche),

for 10 min on ice, followed by centrifugation at 5000 rpm for 5 min. Supernatant was collected as cytoplasmic fraction. Nuclear pellet was washed 3 times in CLB.

To fractionate the nuclear pellet, the nuclear pellet was resuspended in low salt buffer (LSB), containing 20 mM Tris HCl pH 7.5, 5 mM magnesium chloride ($MgCl_2$), 1 mM $CaCl_2$, 15 mM sodium chloride (NaCl), 60 mM KCl, 5 mM manganese chloride ($MnCl_2$), 300 mM sucrose, 0.4% NP-40, 1 mM DTT, and 1X protease inhibitor. 300 units of micrococcal nuclease (MNase) from Fermentas was added. The suspension was incubated at 37°C for 20 min followed by centrifugation at 13,000 rpm for 10 min. The supernatant was collected as low salt fraction. The insoluble nuclear fraction was washed 2 times in LSB. The nuclear fraction was resuspended in high salt buffer (HSB), containing 20 mM Hepes pH 7.9, 1 M KCl, 1 mM EDTA, 1 mM EGTA, 1 mM DTT, and 1X protease inhibitor, and vortexed for 20 min at 4°C. The suspension was centrifuged at 13,000 rpm for 10 min, and the supernatant was collected as high salt fraction. The insoluble nuclear pellet was washed 2 times in HSB, and solubilized in SDS extraction buffer (2% SDS and 10 mM Tris HCl pH 7.5) by vortexing for 20 min at room temperature.

All cytoplasmic and nuclear fractions were diluted with an equal volume of 2X Laemmli sample buffer, and prepared according to section 2.7.2, and stored at -20°C.

2.7.4 Gel electrophoresis

Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) gels were casted using the Mini-PROTEAN 3 electrophoresis system (Bio-Rad Laboratories). Protein lysate was separated on a 12% SDS-PAGE gel (Table 2.3) in SDS-running

buffer (0.3% (w/v) Tris, 1.44% (w/v) glycine, 0.1% (w/v) SDS) and transferred onto nitrocellulose membrane (GE Healthcare) in transfer buffer (0.3% (w/v) Tris, 1.44% (w/v) glycine, 20% (v/v) ethanol) using a wet-transfer system (Bio-Rad Laboratories).

Reagents	12% Resolving gel (ml)	4% Stacking gel (ml)
1.5 M Tris-HCl pH 8.8	2.5	-
0.6 M Tris-HCl pH 6.8	-	0.5
40% Acrylamide/Bis-acrylamide	3	0.5
10% SDS	0.1	0.05
10% ammonium persulfate (APS)	0.1	0.05
Tetramethylethylenediamine (TEMED)	0.01	0.01
Water	4.3	3.9

Table 2.3 Recipe of SDS-PAGE gels.

2.7.5 Western blot

The membrane was blocked with 5% skimmed milk (Bio-Rad Laboratories) for 1 hr at room temperature. One 15 min and two 5 min washes with Tris-buffered saline with Tween-20 (TBS-T) were carried out. Primary antibodies were incubated with the membrane at 4°C overnight (Table 2.4). One 15 min and two 5 min washes with TBS-T were carried out. The horseradish peroxidase conjugated secondary antibodies were incubated with the membrane at room temperature for 1 hr. One 15 min and four 5 min washes with TBS-T were carried out. Detection was performed with the Western Blotting Luminol Reagent (Santa Cruz Biotechnology Inc.). The X-ray film (Kodak) was developed in a X-ray developing machine.

Antibodies	Dilution	Company
Anti-myc	1:5000	Cell Signaling Technology
Anti-gfp	1:6000	Millipore
Anti-actin	1:8000	SIGMA
Anti-HA	1:1000	Roche
Anti-mCherry	1:1000	Abcam
Horseradish peroxidase conjugated anti-rabbit/anti-mouse/anti-rat	1:5000	GE Healthcare

Table 2.4 List of primary and secondary antibodies used in western analysis.

2.8 Immunoflurescence assay (IFA)

Parasite cultures were washed twice in 1X PBS and smears of the cultures were prepared on microscopic slides, air-dried, and stored in -80°C. Prior to use, the smears were thawed at room temperature for 5 min. The smears were fixed in 4% paraformaldehyde/0.0075% glutaraldehyde for 30 min. The smears were then subjected to permeabilization with 0.1% Triton X-100 for 10 min, followed by quenching with 1 mg/ml sodium borohydride ($NaBH_4$)/PBS 3 times for 15 min each time. The smears were then blocked with 5% bovine serum albumin (BSA) for 1-2 hr at room temperature. 3 washes with 1X PBS for 5 min each were carried out between each step. The smears were then air-dried at room temperature. Primary antibodies incubation was carried out overnight at 4°C (Table 2.5). Following primary antibodies incubation, 3 washes with 1X PBS for 10 min each were carried out, and the smears were air-dried at room temperature. Secondary antibodies incubation (Table 2.5) was carried out for 1-2 hr at room temperature in the dark. Parasite nuclei were stained with 4′,6′-diamidino-2-phenylindole (DAPI) for 1 min. 3 washes with 1X PBS for 5 min each were carried out between each step. After the final wash, the slides were air-

dried and mounted in VECTASHIELD® Mounting Media (Vector Laboratories, Inc.). The slides were analyzed by Carl Zeiss LSM 510 Meta Confocal Laser Scanning Microscope (Carl Zeiss). Image processing was done using the Carl Zeiss LSM 510 software.

Antibodies	Dilution	Company
Anti-myc	1:200	Cell Signaling Technology
Anti-gfp	1:200	Millipore
Anti-mCherry	1:50	Abcam
Anti-HA	1:100	Roche
Anti-H3K9Me3	1:500	Millipore
Anti-H3K79Me3	1:500	Abcam
Anti-H3K9Ac	1:1000	Millipore
Anti-H4K5Ac	1:1000	Millipore
Anti-Glyceraldehyde 3-phosphate dehydrogenase (GAPDH)	1:250	-
Anti-Heterochromatin Protein 1 (HP1)	1:200	-
Anti-Nup100	1:200	-
Anti-Pfg27	1:200	-
Anti-Pfs25	1:200	Malaria Research and Reference Reagent Resource Center
Alexa Flour® 488 conjugated anti-rabbit/anti-mouse/anti-rat Alexa Flour® 594 conjugated anti-rabbit/anti-mouse	1:1000	Invitrogen

Table 2.5 List of primary and secondary antibodies used in immunofluorescence analysis. Anti-GAPDH and anti-HP1 are kind gifts from Till Voss (Swiss Tropical & Public Health Institute). Anti-Nup100 is a kind gift from Alan Cowman and Jennifer Volz (The

Walter and Eliza Hall Institute of Medical Research). Anti-Pfg27 is in-house produced antibodies.

2.9 Transcriptional profiling

2.9.1 Harvest samples for RNA extraction

Highly synchronized parasite cultures were harvested by centrifugation at 2500 rpm for 5 min, washed in 1X PBS and pelleted at 2500 rpm for 5 min. Blood pellets were rapidly frozen in liquid nitrogen and stored at -80°C.

2.9.2 RNA isolation

Frozen blood pellet samples were thawed at 65°C for 5 min. TRIzol reagent (Invitrogen) was added to the packed blood sample in a 10:1 ratio, and mixed by inversion to resuspend the RBCs. Chloroform (Fisher Scientific) was then added in a 5:1 (TRIzol:chloroform) ratio, mixed by inversion, and incubated on ice for 5 min. This was followed by centrifugation at 4500 rpm for 10 min at 4°C. The top aqueous phase that contains the RNA was recovered. An equal volume of isopropanol (Fisher Scientific) was added to the aqueous supernatant, mixed by inversion, and kept at -20°C overnight for RNA precipitation. After centrifugation at 9000 rpm for 1 hr at 4°C, the resulting pellet was washed twice with ice-cold 70% ethanol (Merck), air-dried at room temperature and resuspended in RNAse-free water. The RNA was quantified by UV spectrometry, and stored in aliquots at -80°C.

2.9.3 cDNA synthesis by reverse transcription

12 μg total RNA was mixed with 4 μg of 1:1 mixture of 9-mer random primers (N9) and 25-mer oligo-dT (both from PROLIGO), in a total volume of 28 μl, and incubated

at 70°C for 10 min, and then on ice for 10 min. 1 mM dATP, 0.5 mM dTTP, dGTP, dCTP (Fermentas) and 0.5 mM amino-allyl-dUTP (Biotium, USA), and 300 units of RevertAid H Minus MMuLV (Moloney murine leukemia virus) reverse transcriptase (Fermentas) were added in a final volume of 45 μl. The mixture was incubated at 42°C for 2 hr for reverse transcription. 10 μl of 0.5 M EDTA and 10 μl of 1 M sodium hydroxide (NaOH) were added to the mixture and incubated at 65°C for 15 min to hydrolyze the RNA. The mixture was chilled on ice for 5 min. The cDNA products were purified in 15 μl elution buffer using MinElute PCR purification kit (QIAGEN) according to the manufacturer's protocol. The cDNA was quantified by UV spectrometry, and stored at -20°C.

2.9.4 cDNA synthesis by SMART and PCR amplification

For samples with less than 12 μg RNA, amplification of cDNA was carried out. The SMART (switching mechanism at 5' end of RNA template)-PCR method utilizes the terminal transferase activity and the template-switching property of the reverse transcriptase enzyme at the 5' end of the mRNA template. This method ensures the synthesis of full length cDNA products from the mRNA template. During reverse transcription, a template switching sequence (Table 2.6) will be added to both ends of the cDNA strand. The generated cDNA will be used as the template for PCR amplification using primers that contain the template switching sequence.

The first strand synthesis by reverse transcriptase was carried out. 500 ng total RNA was mixed with SMART primer mix (6.25 μM SMART-oligo-dT, 6.25 μM SMART-N9, and 12.5 μM TS-oligo) in a total volume of 8 μl. The sequences of the primers are listed in Table 2.6. The mixture was incubated at 65ºC for 5 min and chilled on ice for

10 min. 4 μl of 5X First Strand reverse transcription buffer, 2 μl of 0.1 M DTT, 2 μl of 3 mM dNTPs mix, 40 units of RNaseOUT™ Recombinant Ribonuclease Inhibitor (Invitrogen) were added to the mixture in a total volume of 19 μl, and incubated at 25°C for 2 min. 200 units of SuperScript® II Reverse Transcriptase (Invitrogen), was added and mixed. Reverse transcription was carried out at 42°C for 50 min (Figure 2.2A).

2.5 μl of the reaction mixture, containing the reverse transcription cDNA product, was used for PCR amplification, and the rest were stored at -20°C. 2.5 μl of cDNA product was mixed with 0.45 mM dATP, 0.225 mM dTTP, dGTP, dCTP (Fermentas) and 0.225 mM amino-allyl-dUTP (Biotium, USA), 6 μM Primer IIa (Table 2.6), 10 μl of 10X ThermoPol® Buffer, and 10 units of *Taq* DNA Polymerase (New England Biolabs), in a total volume of 100 μl. The PCR cycling parameters used were as follows: 1 cycle at 95°C for 5 min, 60°C for 1 min, 68°C for 10 min, 19 cycles of 95°C for 30 sec, 60°C for 30 sec and 68°C for 5 min, 1 cycle at 72°C for 7 min and followed by pausing at 4°C. All PCR reactions were carried out in 0.2 ml tubes, using the Eppendorf Thermocycler (Figure 2.2B).

Primer name	Primer sequence (5' → 3')
SMART-oligo-dT	**AAGCAGTGGTATCAACGCAGAGT**AC(T)30VN (N = A/C/G/T; V = A/G/C)
SMART-N9	**AAGCAGTGGTATCAACGCAGAGT**(N)9
TS-oligo	**AAGCAGTGGTATCAACGCAGAGT**ACGCrGrGrG
Primer IIa	**AAGCAGTGGTATCAACGCAGAGT**

Table 2.6 Primers used in SMART and PCR amplification.
The template switching sequence is highlighted in bold, and is found in all the 4 primers.

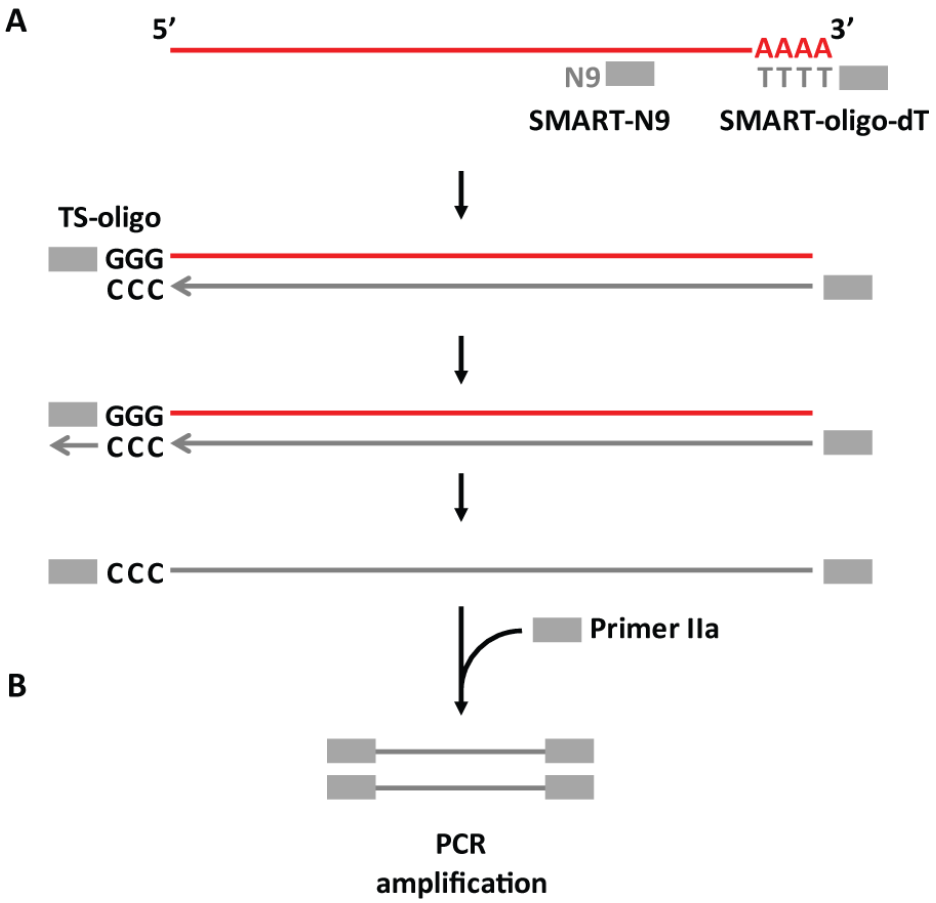


Figure 2.2 Illustration of cDNA synthesis by SMART and PCR amplification.
(A) The SuperScript® II reverse transcriptase enzyme synthesizes the first strand cDNA using the SMART-oligo-dT and SMART-N9 primers. The reverse transcriptase has terminal transferase activity and adds extra bases at the 3' end of the newly synthesized cDNA (grey line). Template switching occurs at the 3' end of the cDNA and TS-oligo binds to the RNA 5'end (red line), and serves as a template for reverse transcriptase. Reverse transcription introduces the template switching sequence (grey box) at the 3'end. (B) PCR amplification using the cDNA generated by reverse transcription.

The amplified PCR products were resolved on 1% agarose gels, and the products ranged from 100-1500 bp. All PCR products were purified and eluted in 15 μl elution buffer using MinElute PCR purification kit (QIAGEN) according to the manufacturer's protocol. The amplified DNA was quantified by UV spectrometry, and stored at -20°C.

2.9.5 Reference pool

Highly synchronized 3D7 and T996 parasite cultures were used for preparing the reference pool. Parasite samples were harvested every 8 hr for 48 hr according to section 2.9.1. Total RNA was harvested according to section 2.9.2. Equal amounts of total RNA from the time-point samples were mixed together and cDNA or amplified DNA was generated according to section 2.9.3 or 2.9.4, to generate a standard reference pool which was used for all microarray hybridization studies.

2.9.6 Microarray

2.9.6.1 Preparation of microarray chip

The microarray chips used in all the microarray studies were in-house prepared. 10,416 70-mer oligonucleotides representing the ORFs, and 5402 50-mer oligonucleotides representing the upstream intergenic regions (IGRs) of *P. falciparum* 5343 coding genes, were designed using the OligoRankPick program [270]. The oligonucleotides were synthesized by Invitrogen.

Poly-L-lysine coated gold-seal slides were also in-house prepared. The gold-seal slides (Gerhard Menzel) were cleaned by incubation with agitation, in 2.5 M NaOH/60% ethanol for 2 hr. The slides were then rinsed 6 times with double distilled water. The cleaned slides were then incubated, with agitation, in poly-L-lysine solution (300 ml poly-L-lysine solution (Sigma), 350 ml 1X PBS, 2850 ml water) for 1 hr. The slides were then rinsed 6 times with double distilled water, spun dry at 600 rpm for 5 min, and heated at 50°C for 20 min. The poly-L-lysine coated slides were aged for 2 weeks at room temperature. The slides were then spotted with the oligonucleotides mentioned above, using the VersArray ChipWriter Pro Systems (Bio-Rad Laboratories).

The microarray chips were post-processed prior to use. The oligonucleotides spots were rehydrated in a hydration chamber containing 3X saline-sodium citrate buffer (SSC) at 42°C for 1 min. The oligonucleotides were cross-linked to the slides at 80 mJ UV energy. The slides were then blocked in BSA solution (5X SSC, 0.1% SDS, 5% BSA) for 1 hr at 42°C. The slides were then rinsed 5 times with double distilled water, and spun dry at 600 rpm for 5 min.

2.9.6.2 Cy dye labeling

2 μg of cDNA and 3 μg of SMART-amplified DNA samples were dye-labeled with Cy3 or Cy5 (GE Healthcare) in the presence of 0.1 M $NaHCO_3$ pH 9 for 2-3 hr at room temperature in the dark. All time course samples were labeled with Cy5 and reference pools were labeled with Cy3. Labeled samples were purified and eluted in 15 μl elution buffer using the MinElute PCR purification kit (QIAGEN) according to the manufacturer's protocol.

2.9.6.3 MAUI® Hybridization System

Cy3-labeled reference pool and Cy5-labeled test samples were mixed together. 3X SSC, 24.5 mM Hepes pH 7, and 220 mM SDS were mixed with the labeled samples in a total volume of 46 μl. The hybridization mixture was boiled at 100°C for 5 min and cooled at room temperature for 5 min. Hybridizations were carried out on post-processed microarray chips for 16-18 hr at 65°C using the MAUI® hybridization system (BioMicro® Systems). The arrays were washed sequentially in 0.6X SSC/0.03% SDS and then in 0.06X SSC for 5 min each. The arrays were spun dry at 600 rpm for 5 min. The arrays were scanned using the GenePix® 4000B Dual Channel Microarray scanner (Axon Laboratory).

2.9.6.4 Agilent hybridization system

Cy3-labeled reference pool and Cy5-labeled samples were mixed together, boiled at 100°C for 5 min and cooled at room temperature for 5 min. The mixture was topped up to 125 µl with elution buffer and mixed with 125 µl of 2X Agilent hybridization buffer (Agilent Technologies). Hybridizations were carried out on post-processed microarray chips for 20 hr at 65°C using the Agilent hybridization system (Agilent Technologies). The arrays were washed sequentially in 0.6X SSC/0.03% SDS and then in 0.06X SSC for 5 min each. The arrays spun dry at 600 rpm for 5 min. The arrays were scanned using the GenePix® 4000B Dual Channel Microarray scanner (Axon Laboratory).

2.9.7 Statistical analysis of microarray data

The microarray images were analyzed using the GenePix Pro 6.0 program (Axon Instruments). The data was lowess normalized and filtered for signal intensity more than 2 times the background intensity for each Cy channel. The expression profile for each gene was represented by an average expression of all the probes representing the ORFs of a particular gene.

2.10 Chromatin immunoprecipitation with microarray (ChIP-on-chip)

2.10.1 Cross-linking of samples

Synchronized parasite cultures were lysed in 0.1% saponin and washed 3 times with 1X PBS. Chromatin was cross-linked by incubating the isolated parasites with 0.5% formaldehyde (Sigma) for 10 min at room temperature. The cross-linking was stopped

by incubating in 0.125 M glycine for 1 min at room temperature. The cross-linked pellets were washed twice with ice-cold 1X PBS. The pellets were stored at -80°C until use.

2.10.2 Chromatin immunoprecipitation

Parasite pellets were resuspended in lysis buffer (25 mM Tris-HCl pH 8, 1 mM EDTA, 0.25% NP-40, 2 mM phenylmethylsulfonyl fluoride (PMSF), and 1X protease inhibitor), incubated on ice for 1 hr and then lyzed by homogenization using a dounce homogenizer. The nuclei were pelleted by centrifugation at 5000 rpm for 5 min at 4°C.

The nuclei were then resuspended in sonication buffer (50 mM Tris-HCl pH 8, 10 mM EDTA, 1% SDS, 2 mM PMSF, and 1X protease inhibitor), incubated on ice for 15 min and sonicated 8 times for 10 sec at 25% amplitude with 50 sec intervals between each pulse (Sonics Vibra Cell) to obtain DNA fragments in the range of 200 to 1000 bp. After centrifugation at 13,000 rpm for 10 min at 4°C, the supernatant was collected and diluted 10 times in ChIP dilution buffer (16.7 mM Tris-HCl pH 8, 1.2 mM EDTA, 0.01% SDS, 1.1% triton X-100, 167 mM NaCl, 2 mM PMSF, and 1X protease inhibitor). An aliquot of the nuclei was kept as input DNA. The remainder was subjected to chromatin immunoprecipitation.

Protein A Agarose/Salmon Sperm DNA slurry (Millipore) was added to the nuclei, rotated at 10 rpm for 2 hr at 4°C, to remove non-specific binding. After centrifugation at 1000 rpm for 5 min at 4°C, the supernatant was collected. Antibodies (anti-myc, anti-gfp, anti-mCherry, and anti-HA) were added to the supernatant in a 1:100 dilution, and incubated with rotation at 10 rpm overnight at 4°C. Protein A slurry was blocked

in 3% BSA and added to the supernatant, and incubated with rotation at 10 rpm for 1 hr at 4°C. After centrifugation at 1000 rpm for 5 min at 4°C, the Protein A slurry was collected. The Protein A slurry was subjected to 2 washes with low salt wash buffer (20 mM Tris-HCl pH 8, 2 mM EDTA, 0.1% SDS, 1% triton X-100, and 150 mM NaCl), 3 washes with high salt wash buffer (20 mM Tris-HCl pH 8, 2 mM EDTA, 0.1% SDS, 1% triton X-100, and 500 mM NaCl), 1 wash with lithium chloride (LiCl) immune complex wash buffer (0.25 M LiCl, 1% NP-40, 1% deoxycholic acid, 10 mM Tris-HCl pH 8, and 1 mM EDTA), and 2 washes in Tris-EDTA buffer (10 mM Tris-HCl pH 8, and 1 mM EDTA). All washes were carried out at 4°C with rotation at 10 rpm for 10 min, followed by centrifugation at 1000 rpm for 5 min at 4°C. The DNA-protein complex was eluted from the Protein A Agarose slurry in 500 µl ChIP elution buffer (50 mM Tris-HCl pH 8, 10 mM EDTA, 1% SDS). The DNA in the DNA-protein complex in the eluate and input DNA was reverse cross-linked by incubation with 0.2 M NaCl overnight at 65°C. Proteinase K (Invitrogen) was used to digest the protein for 2 hr at 37°C.

The DNA was purified and eluted in 25 µl elution buffer using the QIAEX II gel extraction kit (QIAGEN) according to the manufacturer's protocol: Desalting and Concentrating DNA Solutions. The eluted DNA was quantified by UV spectrometry, and stored at -20°C.

2.10.3 Amplification of ChIP DNA

Equal amount of input DNA from different samples were combined together as the input pool DNA. 5 µl of ChIP DNA or 1 µl of input pool DNA was mixed with 5 µM Primer A (Table 2.7), 0.6 µl dimethyl sulfoxide (DMSO), and 1 µl NEB buffer 2 (New

England Biolabs) in a total volume of 10 µl. The mixture was incubated at 70°C for 5 min, 94°C for 5 min, and cooled at 5°C for 10 min. To the mixture, 2.5 µl water, 0.5 µl NEB buffer 2, 1 µl 10 mM dNTPs mixture (Fermentas), and 1 µl Klenow Fragment (3' → 5' exo$^-$) (New England Biolabs) was added. The mixture was incubated at 37°C for 30 min, 94°C for 5 min, and cooled at 5°C for 10 min. 1 µl Klenow Fragment (3' → 5' exo$^-$) was added to the mixture and incubated at 37°C for 30 min.

The DNA is now ready for amplification. All the ChIP DNA, and 5 µl of the input DNA was used. The DNA was mixed with 0.75 mM dATP, 0.375 mM dTTP, dGTP, dCTP (Fermentas) and 0.375 mM amino-allyl-dUTP (Biotium, USA), 1 µM Primer B (Table 2.7), 10 µl of 10X ThermoPol® Buffer, and 5 units *Taq* DNA Polymerase (New England Biolabs), in a total volume of 100 µl. The PCR cycling parameters used were as follows: 1 cycle at 94°C for 5 min, 30 cycles of 94°C for 30 sec, 40°C for 30 sec, 50°C for 30 sec, and 60°C for 2 min, followed by pausing at 4°C. All PCR reactions were carried out in 0.2 ml tubes, using the Eppendorf Thermocycler.

Primer name	**Primer sequence (5' → 3')**
Primer A	GTT TCC CAG TCA CGA TC N NNN NNN NN
Primer B	GTT TCC CAG TCA CGA TC

Table 2.7 Primers used in ChIP-on-chip DNA PCR amplification.

All amplification reactions were done in triplicates. The amplified DNA was purified and eluted in 15 µl elution buffer using the MinElute PCR purification kit (QIAGEN) according to the manufacturer's protocol. The eluted DNA was quantified by UV spectrometry, and stored at -20°C.

2.10.4 Microarray

2 μg of amplified ChIP and input DNA samples were dye-labeled with Cy5 or Cy3 respectively, for 2 hr in the dark. Labeled samples were purified and eluted in 15 μl elution buffer using MinElute PCR purification kit (QIAGEN) according to the manufacturer's protocol. Microarray hybridization using the MAUI® hybridization system or the Agilent hybridization system was carried out as described in section 2.9.6.3 and 2.9.6.4, but hybridization was carried out at 63.5°C.

2.10.5 Statistical analysis

The microarray images were analyzed using the GenePix Pro 6.0 program (Axon Instruments). The data was lowess normalized and filtered for signal intensity more than 2 times the background intensity for each Cy channel. For every gene, only oligonucleotide probes with data present in at least 2 out of 3 triplicates were included for analysis.

2.11 Real-time PCR (RT-PCR)

DNA was first diluted to 2 ng/μl. 10 ng (5 μl) of DNA sample was mixed with 0.5 μM of each primer (Table 2.8) and 10 μl of 2X SYBR Green PCR Master Mix (Roche), in a total volume of 20 μl, in a 96-well plate (Roche). The thermal cycling parameters used were as follows: 1 cycle of pre-incubation at 95°C for 5 min, 45 cycles of amplification at 95°C for 10 sec, 55°C for 10 sec and 70°C for 10 sec. This was followed by 1 cycle of 95°C for 5 sec and 65°C for 1 min. Acquisition of the melting curve occurred at 97°C. A no-template water control was run alongside with all reactions. All reactions were performed in triplicates. PFL0900c (arginyl-tRNA

synthetase) was used as the reference gene. The ΔΔCt method was used to analyze the relative changes in gene expression level. ΔΔCt = (Ct of test sample target gene - Ct of test sample reference gene) - (Ct of control sample target gene - Ct of control sample reference gene), where Ct is the threshold cycle. $2^{-\Delta\Delta Ct}$ = the ratio of target gene compared to reference gene.

Gene	Primer name	Primer sequence (5' → 3')
PFL0900c	PFL0900c forward primer	AAGAGATGCATGTTGGTCATTT
	PFL0900c reverse primer	ATGATGGCCATGTTATCCTC
PF10_0083	PF10_0083 RT forward primer	TGCTAGGAGAATTGAGACCTCT
	PF10_0083 RT reverse primer	GACATATCTGTCGTAGGTTGTAGG
PFD0310w	Pfs16 RT forward primer	CAAAAAAGCCCGCTGGAA
	Pfs16 RT reverse primer	CCTTGAGATAGTCCACCTTGATTAGG

Table 2.8 Primers design for real-time PCR analysis.
PFL0900C is the reference gene, arginyl-tRNA synthetase. PFD0310w is a gametocyte marker, sexual stage-specific protein precursor Pfs16.

2.12 Protein expression

2.12.1 PCR

PCR was carried out as described in 2.3.1. A total of 6 constructs were made (Chapter 3, Figure 3.18). Table 2.9 lists the different reverse primers used with the forward primer to generate 5 truncated protein and full length protein products.

Construct	Construct length (bp)	Primer name	Primer sequence (5' → 3')
		PF10_0083 forward primer	GCGACTCGAG**GGATCC**ATGTTATA CAAAATGAATGTATCGTCCG
1	1-225	225 reverse primer	GCGA**AAGCTT**TTTGTGGGCATATTT ACAATTA
2	1-273	273 reverse primer	GCGA**AAGCTT**ACTTGTTTTATAAGT AGACATATCTGTCGT
3	1-396	396 reverse primer	GCGA**AAGCTT**TTTTTGCTTTGTGTG TATATCC
4	1-420	420 reverse primer	GCGA**AAGCTT**ATTATTATTATTATT ATTATCATCTTTTTGC
5	1-630	630 reverse primer	GCGA**AAGCTT**TTCATTTTTGTCGTT CATTTTTA
6	Full length (1-1119)	PF10_0083 reverse primer	GCATCCTAGG**AAGCTT**ATAGATAT ATATATTTTCATATTTTTTTGTGATT TCTT

Table 2.9 Primers design for PF10_0083 protein expression.
PF10_0083 forward primer was paired with the 6 reverse primers for preparing the full length and 5 different truncated protein constructs. The restriction enzyme sites BamHI and HindIII are highlighted in bold.

2.12.2 Construction of expression clones

The expression vectors used in this study are pET-24a and pET-42a (Chapter 3, Figure 3.17). Restriction digestion of the PCR products and plasmids with BamHI and HindIII (New England Biolabs) was carried out as described in section 2.3.2.1. The digested PCR products were ligated to the digested plasmids according to section 2.3.2.2. Ligated products were transformed into *E. coli* DH5α competent cells (Agilent Technologies) and plated on LB plates, supplemented with 50 μg/ml of kanamycin according to section 2.3.2.3. Colonies were screened for the correct constructs and

glycerol stocks were prepared and stored in -80°C, according to section 2.3.2.4. Large quantities of constructs were prepared according to section 2.3.2.5.

2.12.3 Expression of PF10_0083

The constructs prepared above were transformed into BL21-CodonPlus-RIL expression competent cells (Agilent Technologies), and grown on LB plates supplemented with 50 μg/ml of kanamycin and 34 μg/ml of chloramphenicol, at 37°C overnight.

2.12.3.1 Induction assay

For each construct, 5-10 colonies were selected and picked from the LB plates and cultured in 5 ml of LB, supplemented with 50 μg/ml of kanamycin and 34 μg/ml of chloramphenicol, and incubated overnight at 37°C with agitation at 220 rpm. An inoculum from this overnight culture was inoculated in 3 ml of LB, supplemented with 50 μg/ml of kanamycin and 34 μg/ml of chloramphenicol, and incubated at 37°C with agitation at 220 rpm, until the optical density, OD_{600} reached 0.5-0.7. The induction of protein expression with isopropyl β-D-1-thiogalactopyranoside (IPTG) was carried out at the following conditions:

1. 37°C, 1 hr, 1 mM IPTG
2. 37°C, 3 hr, 1 mM IPTG
3. 37°C, 1 hr, 0.5 mM IPTG
4. 37°C, 3 hr, 0.5 mM IPTG
5. 16°C, 6 hr, 10 μM IPTG
6. 16°C, 6 hr, 50 μM IPTG

7. 16°C, 6 hr, 100 μM IPTG

A control culture at the same induction conditions, but without IPTG was carried out in all induction experiments.

Bacteria cells were harvested by centrifugation at 13,000 rpm for 2 min. The bacteria pellet was resuspended in 1X PBS. An equal volume of 2X Laemmli sample buffer was added. Total protein lysate was prepared according to section 2.7.2. Total protein lysate was separated in a SDS-PAGE gel (section 2.7.4). The gel was stained with Coomassie brilliant blue (0.03% (w/v) Coomassie blue, 40% (v/v) methanol, 10% (v/v) glacial acetic acid) for 20 min. The gel was then destained in destaining solution (40% (v/v) methanol and 10% (v/v) glacial acetic acid). The expression of the protein was checked, and positive clones were tested for protein solubility.

2.12.3.2 Protein solubility assay

Protein induction was carried out as described in section 2.12.3.1, in 5 ml culture volume. After induction, 2.5 ml of culture was used to check for protein expression, as described in section 2.12.3.1. 2.5 ml of culture was used to check for protein solubility. The bacteria cells were pelleted at 13,000 rpm for 2 min. The pellet was resuspened in 1X PBS, and 1X protease inhibitor was added. The bacteria cells were sonicated for 10 times for 30 sec at 25% amplitude with 30 sec intervals between each pulse (Sonics Vibra Cell). The suspension was centrifuged at 13,000 rpm for 10 min at 4°C. The supernantant were collected as soluble protein fraction. The pellet was collected as the insoluble protein fraction. Both the soluble and insoluble protein fractions were lysed in 2X Laemmli sample buffer, separated by SDS-PAGE, and stained with Coomassie blue.

Chapter 3 Molecular characterization of PF10_0083

3.1 Introduction

PF10_0083 is a hypothetical protein with no ascribed function, and has not been functionally characterized. The initial characterization of the protein involved bioinformatics studies to identify conserved domains or motifs, and its conservation to other proteins. In order to elucidate the function of the protein in the *Plasmodium* parasites, over-expression and knock-out studies were employed to determine gain and loss of function phenotypes. PF10_0083 was also tagged at its endogenous locus. With these transgenic cell lines, PF10_0083 mRNA and protein expression levels and the intracellular localization pattern have been established. In addition, phenotypic characterization of the generated cell lines was carried out to investigate the significance of PF10_0083 on *Plasmodium* parasites growth.

3.2 Bioinformatics information of PF10_0083

The *P. falciparum* genome encodes 37 ZnF proteins, which includes 4 B-box protein, 3 MYND finger domain protein, 12 CCHH ZnF, 2 C3HC4 ZnF, 15 CCCH ZnF, and 1 U2 snRNP auxiliary factor, small subunit (refer to section 1.9.4, Figure 1.5, section 1.9.4.3 and section 1.9.4.4) [135]. The CCCH-type ZnF proteins are C-X8-C-X5-C-X3-H type fingers where C represent cysteine residues, H represents histidine residues, and X represent variable amino acids. PF10_0083 is one of the 15 CCCH ZnF proteins, and contains 3 CCCH ZnF domains between positions 30 bp and 120 bp. The three CCCH ZnF domains are separated by 16 and 18 amino acids. A multiple sequence alignment using PF10_0083 protein sequence was performed using the protein-protein

Basic Local Alignment Search Tool (BLAST). The database used was non-redundant protein sequences, which includes all non-redundant GenBank coding sequence translations, Protein Data Bank (PDB), SWISS-PROT protein sequence database, Protein Information Resource (PIR) database and Protein Research Foundation (PRF) database. There are 58 proteins, all from the *Plasmodium* genus, showing similarity to PF10_0083. The expect values (E value) range from 7×10^{-4} to 1×10^{-64}. PF10_0083 protein sequence is highly conserved in the other *Plasmodium* species, including *P. vivax*, *P. knowlesi*, *P. berghei*, and *P. yoelii*. Figure 3.1 illustrates 9 examples of proteins in other *Plasmodium* species that have the lowest E values. The E values range from 5×10^{-33} to 1×10^{-64}. There is a conserved CCCH zinc finger domain containing 3 zinc finger motifs in all the 9 proteins. Other than the ZnF domain, PF10_0083 does not contain other recognizable motifs. In summary, the computational analysis provides evidence that PF10_0083 is highly conserved in the other *Plasmodium* species.

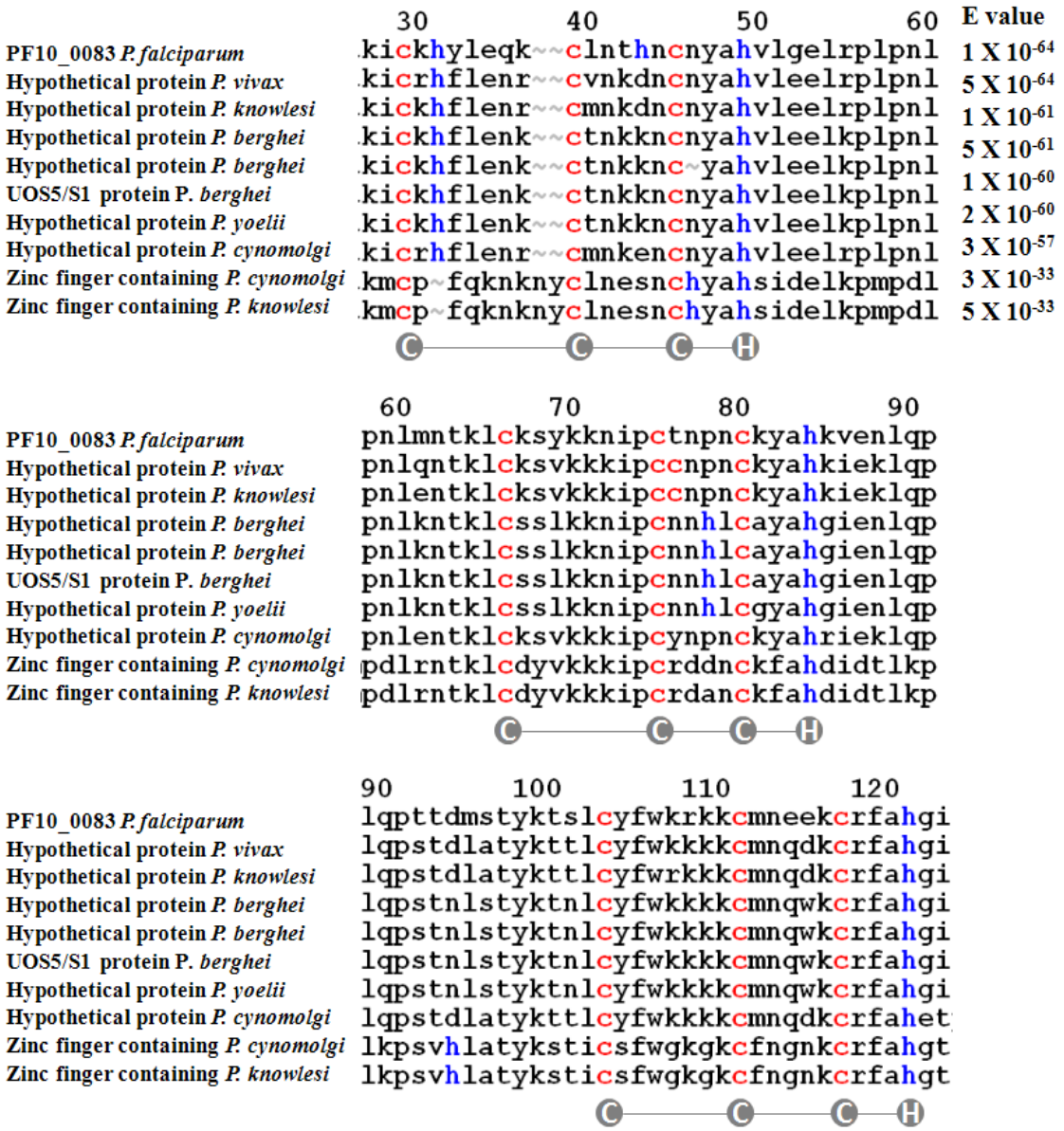


Figure 3.1 Multiple sequence alignment of PF10_0083 protein sequence.

Multiple sequence alignment was performed using the protein-protein Basic Local Alignment Search Tool (BLAST). The database used was non-redundant protein sequences, which includes all non-redundant GenBank coding sequence translations, Protein Data Bank (PDB), SWISS-PROT protein sequence database, Protein Information Resource (PIR) database and Protein Research Foundation (PRF) database. The 9 proteins, all from the *Plasmodium* genus, are the top 9 hits that showed similarity to PF10_0083. The expect values (E value) are shown on the right. The E values range from 5 X 10^{-33} to 1 X 10^{-64}. There is a conserved CCCH zinc finger domain (containing 3 zinc finger motifs) in the homologues of PF10_0083 in other *Plasmodium* species, including *P. vivax*, *P. knowlesi*, *P. berghei*, and *P. yoelii*.

3.3 Over-expression of PF10_0083

In order to study the gain of function phenotype, PF10_0083 over-expression cell lines were generated. The pARL vector was used to prepare the construct for PF10_0083 over-expression. pARL contains a human dihydrofolate reductase (hDHFR) drug selection cassette. The cassette is comprised of a calmodulin 5' untranslated region (CAM 5' UTR), hDHFR gene, and histidine rich protein 2 3' UTR (HRP2 3' UTR). WR99210 is a potent and selective inhibitor against *Plasmodium* dihydrofolate reductase (pDHFR), but does not have any effect against hDHFR. Parasites with only pDHFR will die on exposure to WR99210, while parasites carrying the pARL plasmid will be able to produce hDHFR to replace pDHFR, and will survive in the presence of WR99210 [271]. Transgene expression is driven by the *P. falciparum* Chloroquine Resistance Transporter 5' UTR promoter (PfCRT 5' UTR). A multiple cloning site (MCS) is located downstream of the promoter, and either a myc-streptavidin (myc) or green fluorescent protein (gfp) is tagged to the end of the MCS (Figure 3.2A).

PF10_0083 is a 1.2 kb gene with no introns. The entire gene was cloned into the pARL plasmid. The ATG was provided within the PF10_0083 sequence and was cloned in frame with the tag. The construct was transfected into *P. falciparum* T996 and 3D7 strains. PF10_0083 was tagged with either myc or gfp (Figure 3.2B). A total of 4 sets of PF10_0083 over-expression transfectants were prepared independently. For 3 transfectants, PF10_0083 was tagged to a myc, and for 1 transfectant, PF10_0083 was tagged to a gfp. Parasites transfected with pARL without the transgene were also prepared as control parasites (pARL-myc, and pARL-gfp). There are a total of 4 sets of control parasites: untransfected T996 and 3D7 parasites, pARL-myc, and pARL-gfp, summarized in Table 3.1. All transfected cells were cultured

under continuous WR99210 drug pressure. Parasites were first observed between 14 to 16 days after transfection. The parasites were allowed to multiply for an additional 12 to 20 days to obtain enough parasitemia for synchronization and further experiments.

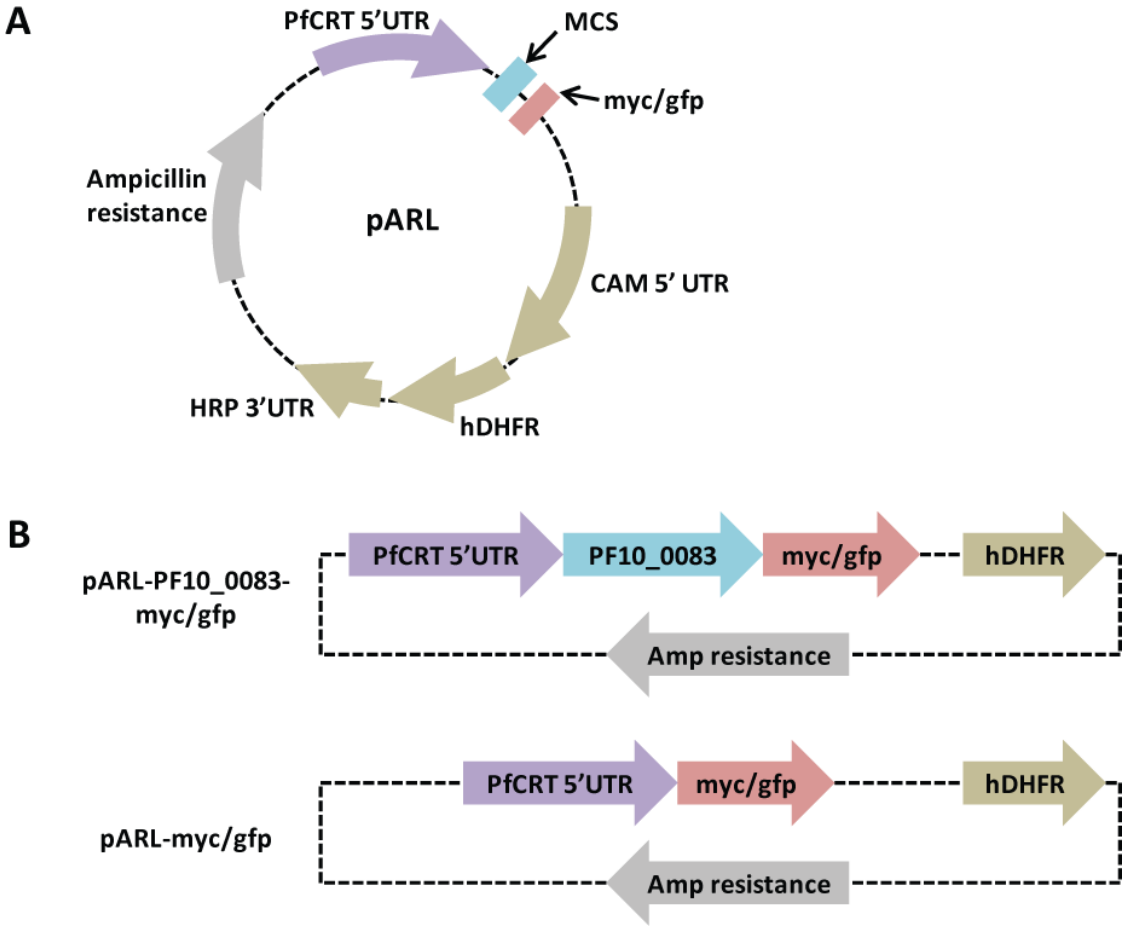


Figure 3.2 pARL vector map used in over-expression studies.

(A) Transgene expression is driven by the *P. falciparum* Chloroquine Resistant Transporter 5' untranslated region (PfCRT 5' URT). A multiple cloning site (MCS) is located downstream of the promoter, and either a myc-streptavidin (myc) or green fluorescent protein (gfp) is tagged to the end of the MCS. The human dihydrofolate reductase (hDHFR) is used as a selection marker of the parasites carrying the plasmid. The hDHFR drug cassette is comprised of a calmodulin 5' UTR (CAM 5' UTR), hDHFR gene, and histidine rich protein 2 3' UTR (HRP2 3' UTR). (B) In the over-expression cell lines, PF10_0083 was tagged with either a myc or a gfp. Parasites transfected with pARL without the transgene were also prepared as control parasites.

Transfection	Control	Over-expression
1	Untransfected wild-type T996 parasites	T996-pARL-PF10_0083-myc1
2	T996-pARL-myc	T996-pARL-PF10_0083-myc2
3	T996-pARL-gfp	T996-pARL-PF10_0083-gfp
4	Untransfected wild-type 3D7 parasites	3D7-pARL-PF10_0083-myc3

Table 3.1 Summary of the PF10_0083 over-expression and control cell lines.
4 transfections were carried out independently. 4 over-expression cell lines were prepared. They are T996-pARL-PF10_0083-myc1, T996-pARL-PF10_0083-myc2, T996-pARL-PF10_0083-gfp and 3D7-pARL-PF10_0083-myc3. The 4 control cell lines used are untransfected T996 parasites, parasites transfected with pARL without the transgene: T996-pARL-myc and T996-pARL-gfp, and untransfected 3D7 parasites.

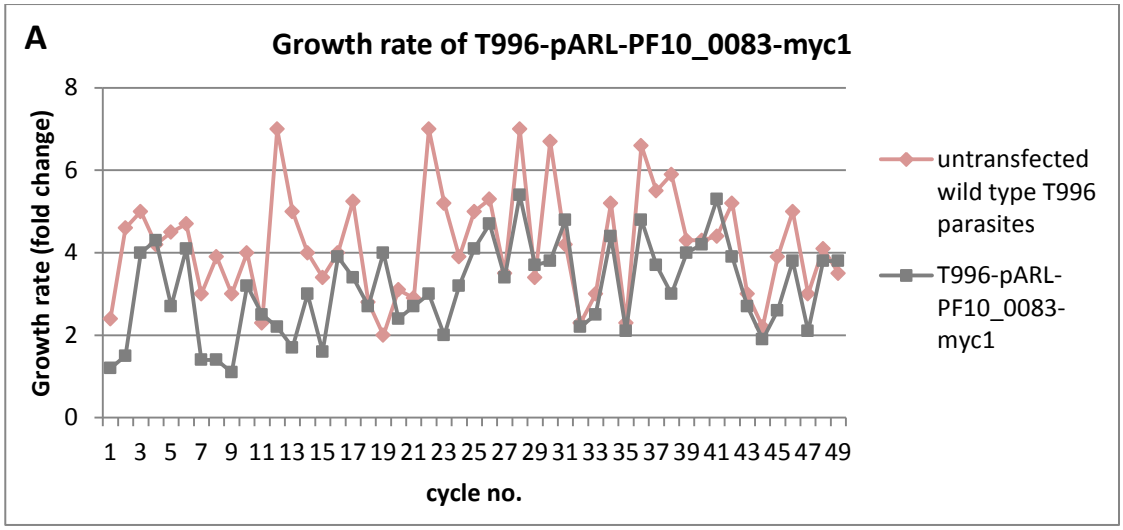
3.3.1 Growth rates of PF10_0083 over-expression transfectants, control transfectants and untransfected control parasites

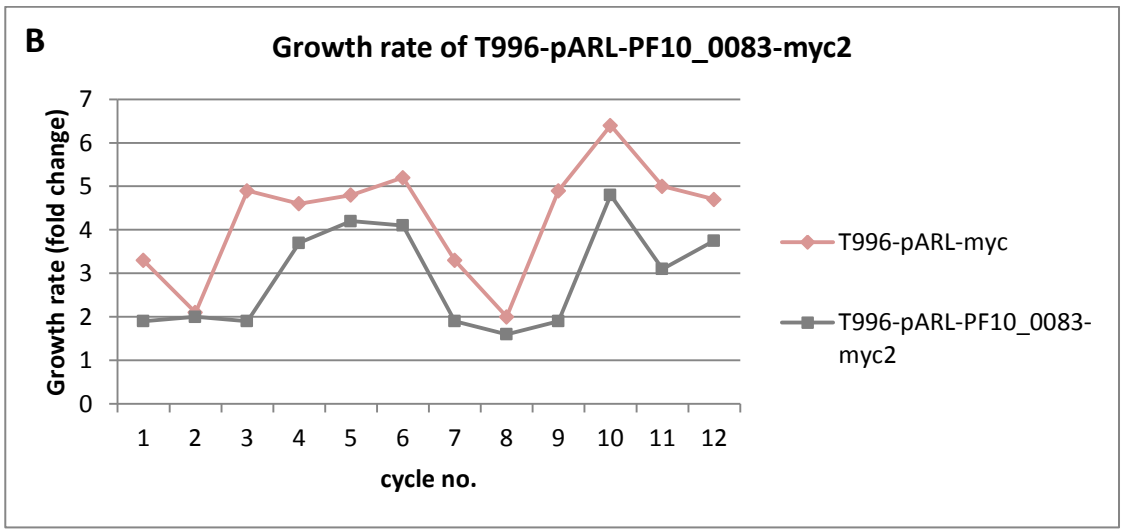
The over-expression parasite cell lines displayed no observable morphological difference from the untransfected control parasites and the transfected control parasites by means of microscopic examination. All T996 transfected and untransfected cell lines has a life cycle of 48 hours, and all 3D7 transfected and untransfected cell lines has a life cycle of 44 hours.

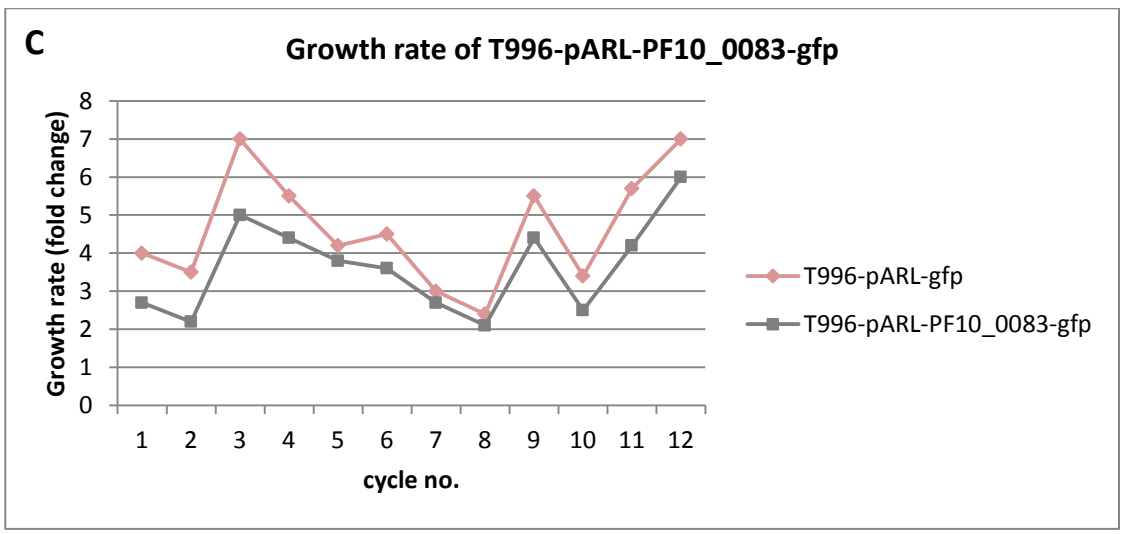
The over-expression cell lines were observed to have a lower growth rate than both the untransfected and the transfected control parasites. The growth rate of the parasites was determined by the fold change in the parasitemia from schizont stage to the next ring stage by microscopy examination. The growth rates for a total of 49 parasitic cycles were recorded for T996-pARL-PF10_0083-myc1 transfectants and untransfected wild-type T996 parasites (Figure 3.3A). The average growth rates of T996-pARL-PF10_0083-myc1 and the untransfected wild-type T996 parasites were

3.1±1.1 and 4.2±1.3 fold respectively. This difference is statistically significant with a p-value of 1.1 X 10^{-5} (Figure 3.3E). Since the transfectants were under continuous drug pressure to maintain the plasmid inside the parasites, the observed lower growth rate could be due to the drug pressure. Thus, the growth rates of T996-pARL-PF10_0083-myc2 and T996-pARL-PF10_0083-gfp transfectants and T996-pARL-myc and T996-pARL-gfp control transfectants from an independent set of transfection experiment were also recorded. The growth rates for a total of 12 parasitic cycles were recorded. It was observed that the PF10_0083-transfectants also have a lower growth rate than the control transfectants (Figure 3.3B and Figure 3.3C). The average growth rates of T996-pARL-PF10_0083-myc2 and T996-pARL-myc were 2.9±1.2 and 4.3±1.3 fold respectively. The difference is statistically significant with a p-value of 3.2 X 10^{-4}. The average growth rates of T996-pARL-PF10_0083-gfp and T996-pARL-gfp were 3.6±1.2 and 4.6±1.5 fold respectively. The difference is statistically significant with a p-value of 2.5 X 10^{-5} (Figure 3.3E). The growth rates for 20 parasitic cycles were recorded for 3D7-pARL-PF10_0083-myc3 transfectants and untransfected wild-type 3D7 parasites (Figure 3.3D). The average growth rates of 3D7-pARL-PF10_0083-myc3 transfectants and untransfected wild-type 3D7 parasites were 2.2±1.0 and 3.1±1.0 fold respectively. This difference is statistically significant with a p-value of 0.003 (Figure 3.3E).

On average, the growth rates of the PF10_0083 over-expression transfectants and the control parasites were 3.0±1.2 and 4.0±1.4 fold respectively, and the difference in the growth rates is statistically significant with a p-value of 3.5 X 10^{-12}. Based on these observations, it can be concluded that the over-expression of PF10_0083 resulted in lowered growth rates in the transfected parasites.


A
Growth rate of T996-pARL-PF10_0083-myc1
Growth rate (fold change)
cycle no.
untransfected wild type T996 parasites
T996-pARL-PF10_0083-myc1


B
Growth rate of T996-pARL-PF10_0083-myc2
Growth rate (fold change)
cycle no.
T996-pARL-myc
T996-pARL-PF10_0083-myc2


C
Growth rate of T996-pARL-PF10_0083-gfp
Growth rate (fold change)
cycle no.
T996-pARL-gfp
T996-pARL-PF10_0083-gfp

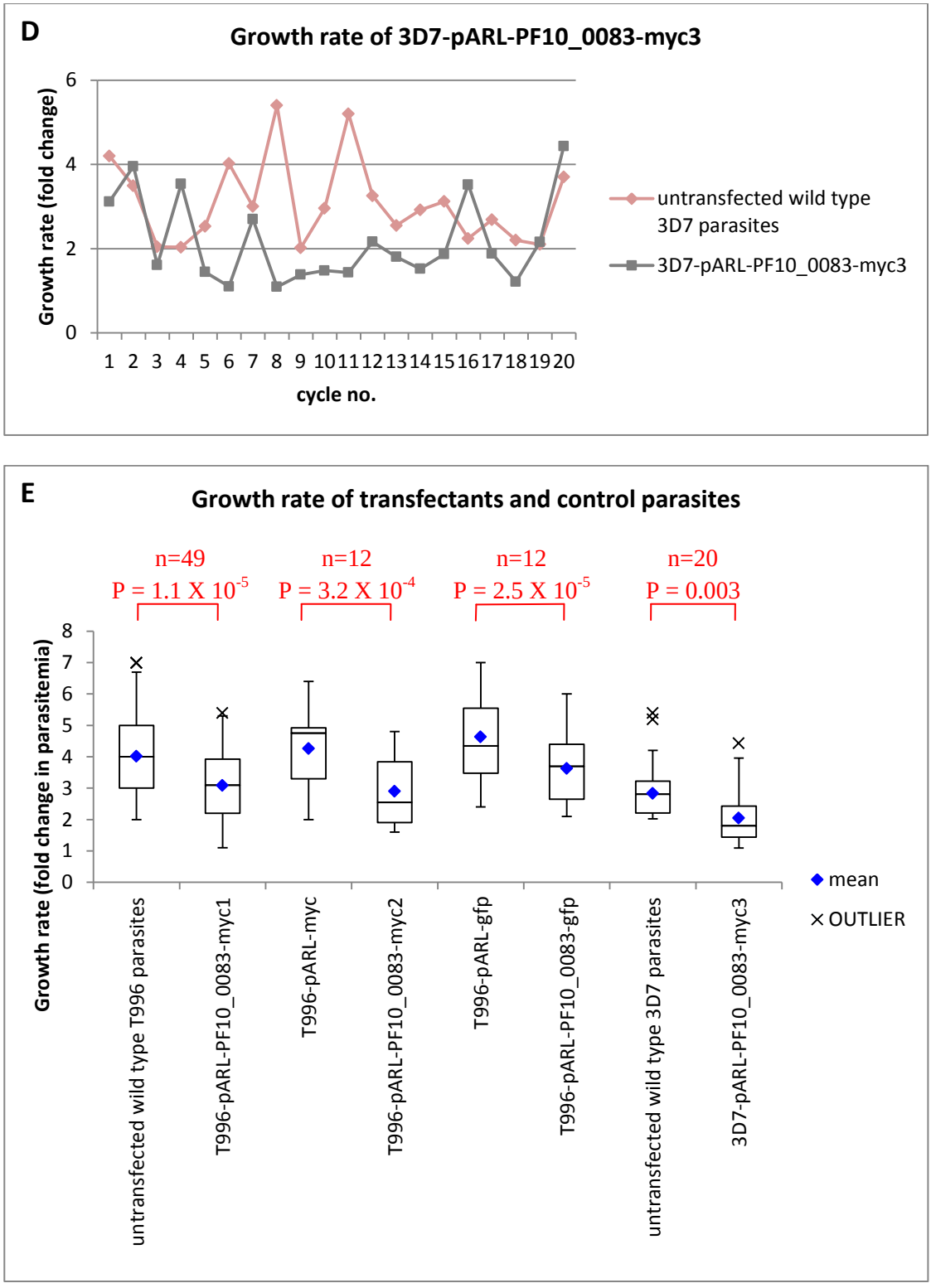


Figure 3.3 Effects of PF10_0083 over-expression on the growth rate of parasites. **Comparison of the growth rates of (A) over-expression cell line T996-pARL-PF10_0083-myc1 and the untransfected wild type T996 parasites over 49 parasitic cycles, (B) over-expression cell line T996-pARL-PF10_0083-myc2 and T996-pARL-myc control transfectants over 12 parasitic cycles, (C) over-expression cell line T996-pARL-**

PF10_0083-gfp and T996-pARL- gfp control transfectants over 12 parasitic cycles, and (D) over-expression cell line 3D7-pARL-PF10_0083-myc3 and untransfected wild type 3D7 parasites over 20 parasitic cycles. (E) The average growth rates of the over-expression and the control parasites. The difference in the growth rates between the over-expression and control parasites is statistically significant. The p-values are shown in red. The growth rate of the parasites was determined by the fold change in the parasitemia from schizont stage to the next ring stage by microscopy examination.

3.3.2 PF10_0083 over-expression levels during IDC

3.3.2.1 mRNA levels

The mRNA expression level of PF10_0083 in the over-expression cell line was then determined by real-time PCR (Figure 3.4). Real-time PCR was performed on the ring, trophozoite and schizont stages of the T996-pARL-PF10_0083-myc1 over-expression cell line. The reference control gene used was PFL0900c (arginyl-tRNA synthetase). The control sample used was untransfected T996 parasites. The fold changes were calculated by the ΔΔCt method. ΔΔCt equals to (Ct of test sample target gene - Ct of test sample reference gene) minus (Ct of control sample target gene - Ct of control sample reference gene), where Ct is the threshold cycle. The over-expression parasite lines expressed PF10_0083 mRNA at least 30 fold higher than that of the control parasites. The mRNA level is lowest during rings, at 37.75 fold higher than the control cells. PF10_0083 mRNA level during trophozoites and schizonts are about the same, at 58.44 fold and 58.01 fold higher than the control cells, respectively.

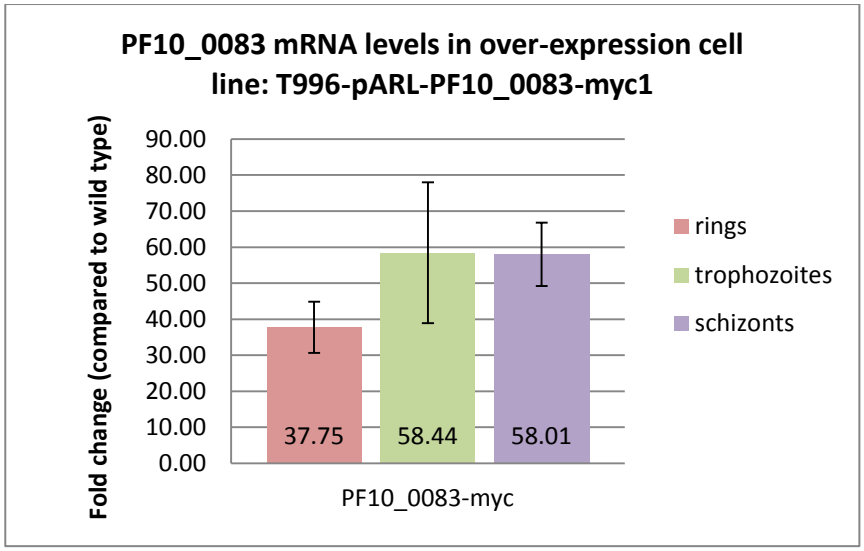


Figure 3.4 PF10_0083 mRNA levels in over-expression cell line T996-pARL-PF10_0083-myc1.

Real-time PCR was performed to determine the level of expression of PF10_0083 in ring, trophozoite, and schizont stages during the IDC. Real-time PCR experiments were performed in triplicates. The reference control gene used was PFL0900c (arginyl-tRNA synthetase). The control sample used was untransfected T996 parasites. The ΔΔCt method was used to calculate the fold change. ΔΔCt equals to (Ct of test sample target gene - Ct of test sample reference gene) minus (Ct of control sample target gene - Ct of control sample reference gene), where Ct is the threshold cycle. The calculated fold change was at least 37.75, 58.44 and 58.01 fold higher than the endogenous level in rings, trophozoites and schizonts respectively. Error bars represent standard deviation of the mean.

3.3.2.2 Protein expression

In the next step, we determined the protein expression of PF10_0083 in the over-expression cell lines using western blot analysis. The main aim was to study whether PF10_0083 expression was controlled at the protein level. Figure 3.5 illustrates the protein expression in the T996-pARL-PF10_0083-myc1, T996-pARL-PF10_0083-gfp, T996-pARL-myc, T996-pARL-gfp and untransfected T996 cell lines. Six protein samples were harvested at 8-hour intervals during the 48-hour IDC. The protein samples were separated on a 12% SDS-PAGE system. The expected size of

PF10_0083 is 44 kDa. The size of a myc-streptavidin tag is 2.5 kDa. A distinct band at ~47 kDa was observed in all the 6 time-points in T996-pARL-PF10_0083-myc1 over-expression cell line. On the other hand, no band was observed in the pARL-myc parasites as a myc-streptavidin tag (2.5 kDa) would be too small to be detected on a 12% SDS-PAGE (Figure 3.5A). The size of a gfp tag is 27 kDa. A band at the expected size of ~70 kDa was detected in the T996-pARL-PF10_0083-gfp over-expression cell line. A band at ~27 kDa (gfp protein) was observed in the pARL-gfp control transfectants (Figure 3.5B). Both the anti-myc and anti-gfp antibodies did not react with the untransfected control parasites.

Both PF10-0083-myc and PF10_0083-gfp were expressed throughout the entire IDC, but at different levels. The peak expression was between 32 and 40 hours-post invasion (hpi). The PF10_0083 protein levels decreased during the late schizont stage and during invasion (48 hpi). The over-expression of PF10_0083 was controlled by a non-endogenous promoter, PfCRT 5' UTR. The expression of PfCRT is higher during the ring and trophozoite stages than during the schizont stage under the control of PfCRT 5' UTR promoter. This suggests that the higher protein level observed in the PF10_0083 over-expression cell lines during the schizont stages was not due to the control of the PfCRT 5' UTR promoter. In addition, in the T996-pARL-gfp control transfectants, the expression of gfp was similar in all 6 time-points, and the gfp protein level in the control transfectants was much stronger than the PF10_0083 over-expression transfectants. Furthermore, there was no variable protein expression in the different time-points that was observed in the over-expression cell lines. In all, these results suggest that PF10_0083 has a specific protein abundance profile, and PF10_0083 levels might be regulated at the protein level.

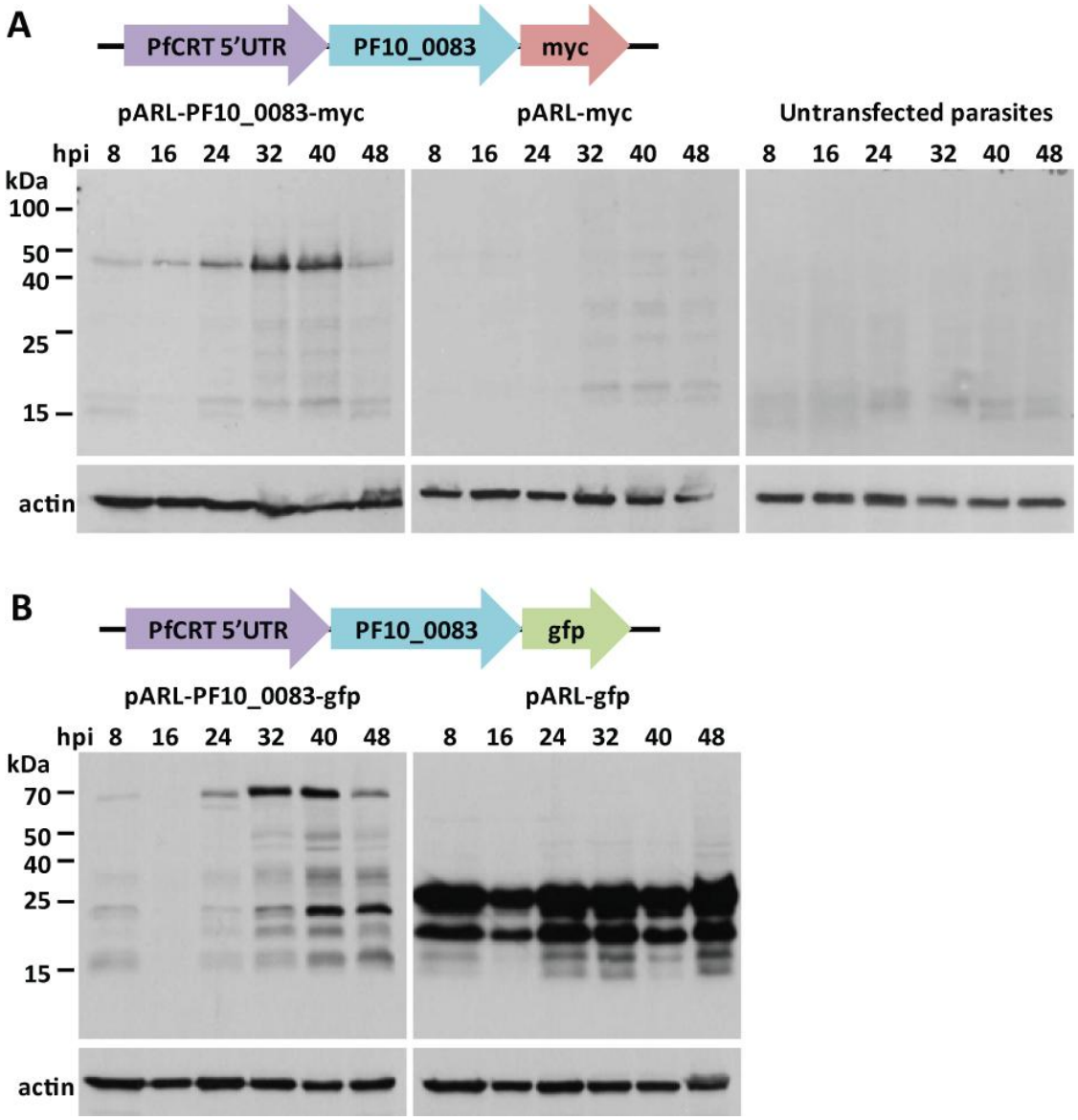


Figure 3.5 PF10_0083 protein expression during the IDC.

Western blot analysis was carried out on the T996-pARL-PF10_0083-myc1, T996-pARL-PF10_0083-gfp, T996-pARL-myc, T996-pARL-gfp and untransfected T996 cell lines. Protein samples were harvested at 8-hour intervals during the 48-hour IDC. Protein samples were separated on a 12% SDS-PAGE system. Western blot analysis was performed using anti-myc and anti-gfp antibodies. (A) A band at ~47 kDa was observed in all the 6 time points of the T996-pARL-PF10_0083-myc1 transfectants, but not in the pARL-myc and untransfected parasites. (B) A band at ~70 kDa was observed in all the 6 samples of the T996-pARL-PF10_0083-gfp transfectants. A band at ~27 kDa was observed in the pARL-gfp control parasites. In both the over-expression cell lines, the peak protein expression was between 32 and 40 hours-post invasion (hpi). Bottom lanes are the actin loading control.

3.3.3 Intracellular localization of PF10_0083

We then analyzed the intracellular localization of PF10_0083 during the IDC. An immunofluorescence assay (IFA) was carried out using the over-expression cell lines to determine the intracellular localization of PF10_0083. Figure 3.6 illustrates the intracellular localization of PF10_0083 in the T996-pARL-PF10_0083-myc1, T996-pARL-PF10_0083-gfp and T996-pARL-gfp cell lines. The parasite smears were prepared using the same parasite culture used to harvest the protein samples for western blot. The smears were also prepared at 8-hour intervals for 48 hours.

In the ring stage parasites at 8 and 16 hpi, there was only a weak staining by the anti-myc and anti-gfp antibodies in the cytoplasm. There was no signal in the nuclei. This observation is in agreement with the western blot results, where the lowest protein levels were observed in the rings. In the western blot data, PF10_0083 level increased during the trophozoite stages at 24 hpi. In the IFA experiment, although fluorescence signal was observed in the nucleus, a large amount of signal was still observed in the cytoplasm. In the western blot data, PF10_0083 protein level peaks between 32 and 40 hpi. In the IFA experiment, in the schizont stages at 32 and 40 hpi, the fluorescence signal was detected mainly in the parasites' nuclei. The fluorescence signal was not distributed evenly in the nuclei, but was focused to a small spot at the periphery. At 48 hpi, the fluorescence signal in the merozoites was located at the periphery of the nuclei. In contrast, the fluorescence signal in the parasites carrying the T996-pARL-gfp control plasmid was detected only in the cytoplasm with no distinct localization pattern. In the parasites carrying the T996-pARL-myc control plasmid, no fluorescence signal was detected (results not shown).

Interestingly, the temporal abundance pattern of PF10_0083 coincided with its nuclear translocation. In the early ring stages (8 and 16 hpi), the majority of PF10_0083 was localized to the cytoplasm. In the trophozoite stage at 24 hpi, PF10_0083 was observed to start to translocate to the nuclei. In the schizont stages (32 and 40 hpi), the peak protein expression was observed. The majority of PF10_0083 also localized to the nuclei during 32 and 40 hpi.

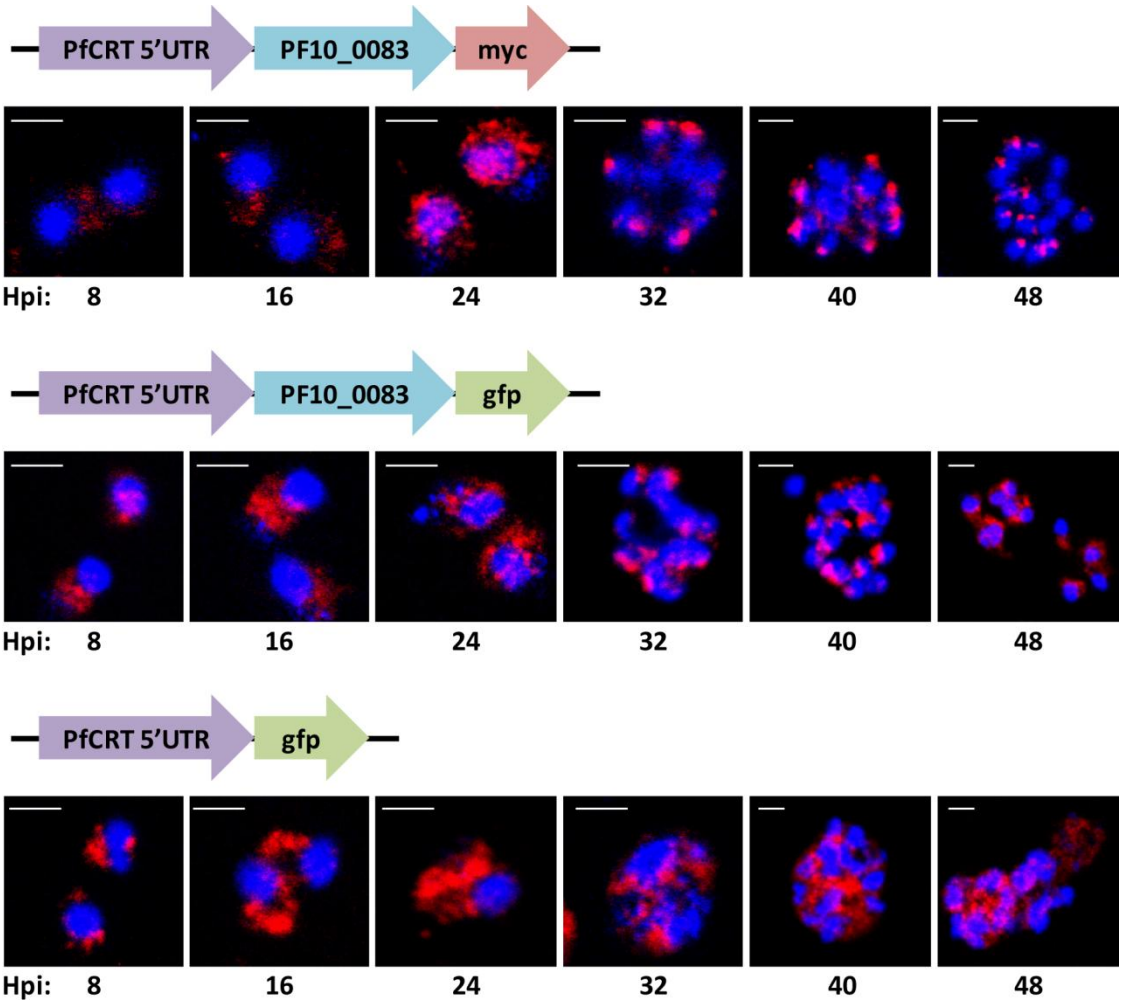


Figure 3.6 Intracellular localization of PF10_0083.

Immunofluorescence assay (IFA) was carried out to determine the intracellular localization of PF10_0083. IFA was carried out in the over-expression cell lines, T996-pARL-PF10_0083-myc1 and T996-pARL-PF10_0083-gfp, and the control cell line, T996-pARL-gfp. IFA was performed with anti-myc and anti-gfp antibodies (red), and DAPI (4′,6′-diamidino-2-phenylindole) stained the nuclei blue. In the over-expression cell lines,

fluorescence signal was detected in the ring stage parasites (8 hpi and 16 hpi) in the cytoplasm. In the trophozoites, fluorescence signal was detected in the cytoplasm, but weak fluorescence signal was also detected in the nuclei (24 hpi). In the early and mid schizonts (32 hpi and 40 hpi), PF10_0083 was mainly detected in the nuclei, and the fluorescence signal was localized to a small spot at the nuclear periphery. In the late schizonts (48 hpi), the fluorescence signal was also localized at the nuclear periphery. In T996-pARL-gfp, fluorescence signal in all the time-points was detected in the cytoplasm with no localization pattern. Scale bar = 2μm.

We also analyzed the presence of PF10_0083 in the nuclear fractions of the parasite by nuclear fractionation experiment. Figure 3.7 illustrates the nuclear fractionation experiment in the T996-pARL-PF10_0083-myc1 cell line. The protein samples were separated on a 12% SDS-PAGE system. Western blot was carried out with anti-myc antibodies. The protein samples were fractionated into cytoplasmic and nuclear fractions. The nuclear fraction was further fractionated with low and high salt buffers (refer to section 2.7.3). After digestion with micrococcal nuclease (MNase), the proteins that are loosely bound to the chromatin will be extracted in the low salt buffer. The proteins that are tightly bound to the chromatin will only be extracted in the high salt buffer (containing 1M KCl). The most tightly bound proteins will be associated with the insoluble nuclear fraction (SDS fraction) [272]. It was observed that PF10_0083 was enriched mainly in the nuclear fraction. PF10_0083 was extracted and enriched in the low salt fraction, suggesting that this protein is loosely bound to the chromatin (Figure 3.7). It is interesting to note that IFA was unable to detect PF10_0083 signal in the nuclei during the ring stage, and only weak signal in the nuclei during the trophozoite stage. However, this experiment demonstrated that PF10_0083 also localized to the nuclear fraction in the ring and trophozoite stages.

In summary, there appear to be a developmental trigger that translocates large amounts of PF10_0083 into the nucleus during the early schizont stage. It is possible that the nuclear signal observed in the rings (Figure 3.7) was carried over from the schizonts from the previous IDC. It is also possible that translocation of PF10_0083 from the cytoplasm to the nucleus is just a normal transfer that must occur after the nuclear protein has translated in the cytoplasm.

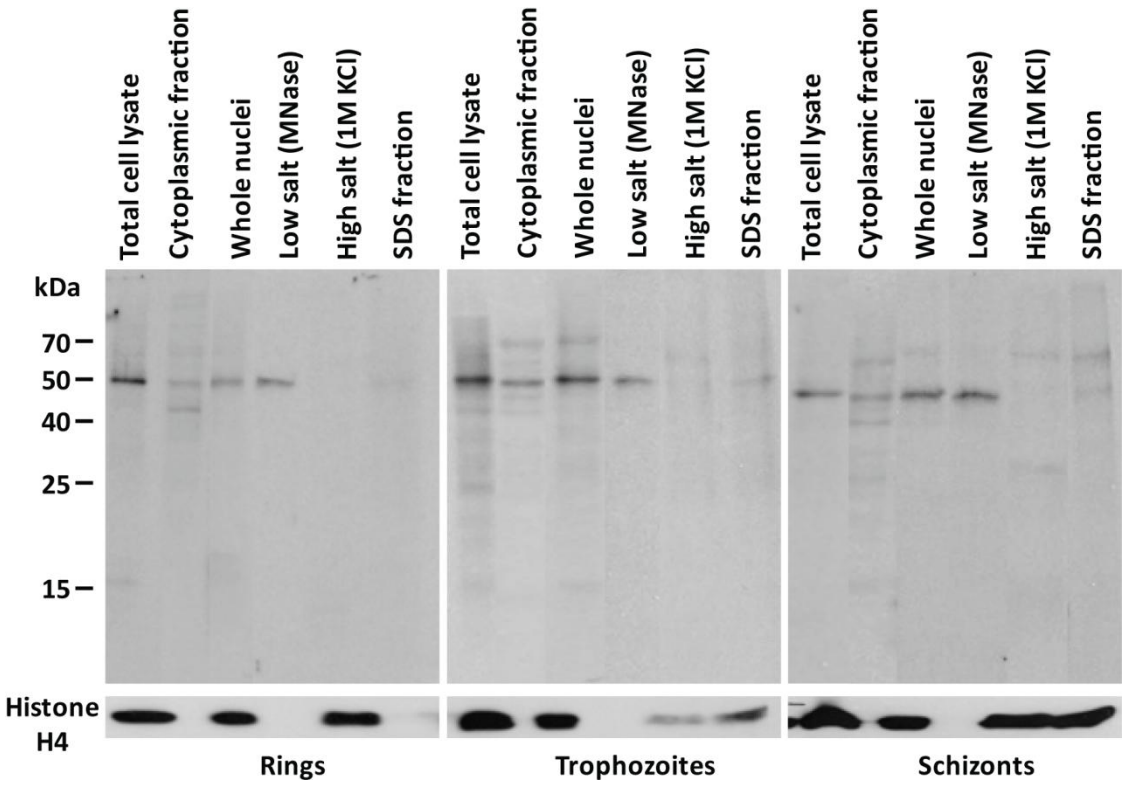


Figure 3.7 Nuclear fractionation of the protein samples.

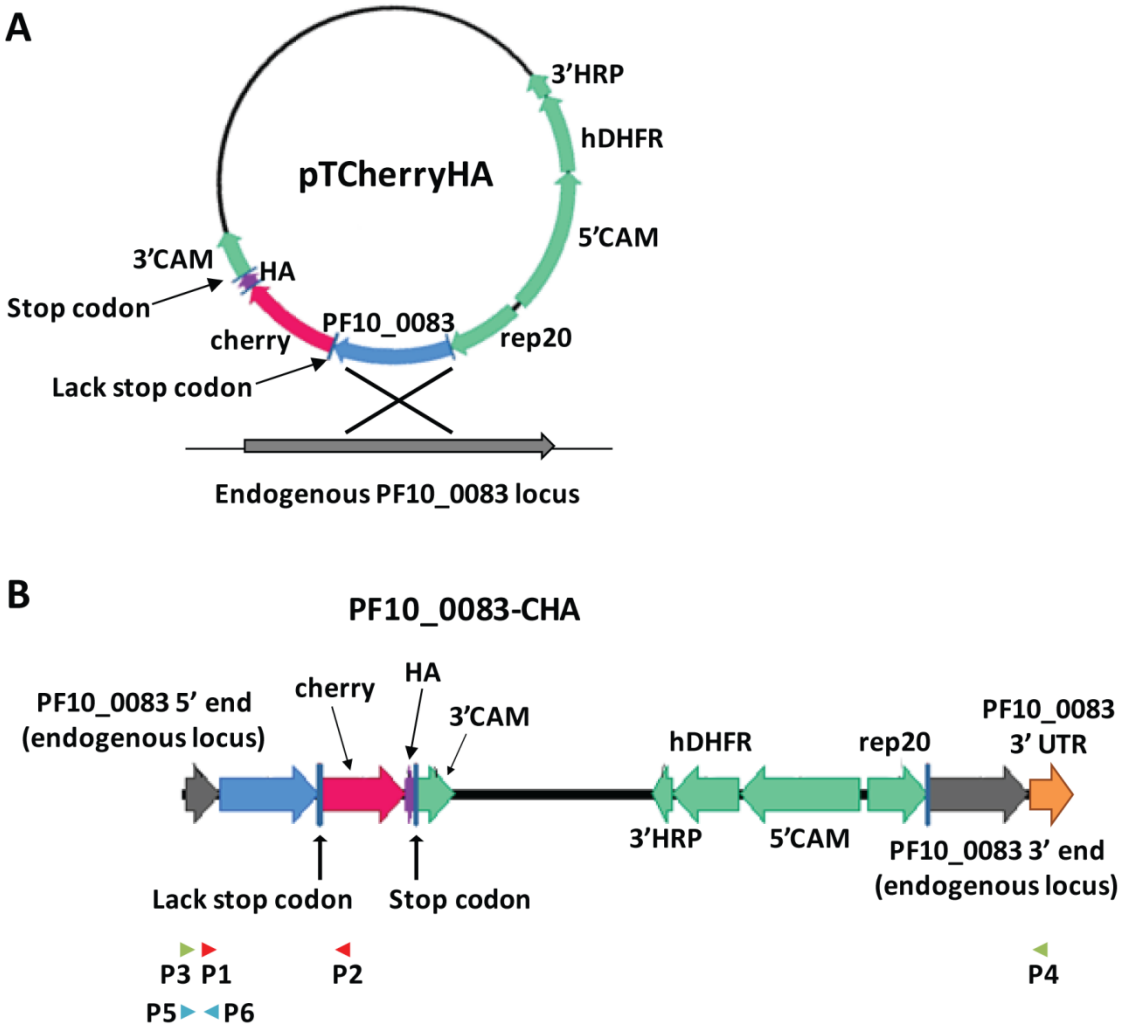
Nuclear fractionation experiment was carried out on the T996-pARL-PF10_0083-myc1 cell line to determine the intracellular localization of PF10_0083 in the ring, trophozoite and schizont stages. Protein samples were fractionated into cytoplasmic and nuclear fractions. Nuclear fraction was further fractionated with low salt and high salt buffers. The proteins that are loosely bound to the chromatin will be extracted in the low salt buffer after treatment with micrococcal nuclease (MNase). The proteins that are tightly bound to the chromatin will only be extracted in the high salt buffer (containing 1M KCl). The most tightly bound proteins will be associated with the insoluble nuclear fraction (SDS fraction). Protein samples were separated on a 12% SDS-PAGE system.

Western blot analysis was carried out using an anti-myc antibody. A band at ~47 kDa was observed in rings, trophozoites and schizonts. PF10_0083 is localized mainly to the nuclear fraction and is associated with the low salt fraction. Bottom lanes are the histone H4 loading control.

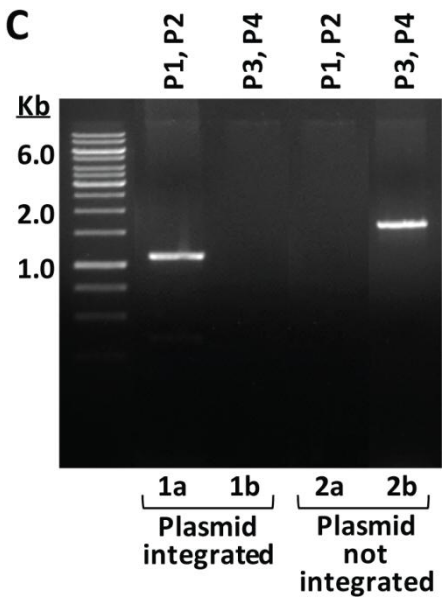
3.4 Endogenous tagging of PF10_0083

In the over-expression cell lines, the expression of PF10_0083 was driven by the PfCRT 5' URT and not the endogenous promoter. As such, the over-expression system may not reflect the PF10_0083 protein expression and localization pattern driven by its endogenous promoter. To better characterize PF10_0083, the endogenous protein was tagged with a mCherry-HA fusion protein (PF10_0083-CHA). The pTCherryHA plasmid was used to prepare the construct for PF10_0083 tagging by homologous recombination. The plasmid contains a hDHFR drug selection cassette that is comprised of a CAM 5' UTR, hDHFR gene, and HRP2 3' UTR, and WR99210 was used as a selection drug. A 800 bp region, lacking the stop codon, from the 3'end of PF10_0083 was cloned in frame with and before the mCherry-HA tag (Figure 3.8A, blue arrow). The construct was transfected into *P. falciparum* 3D7 strain. Homologous recombination by a single crossover event occurred at the endogenous PF10_0083 locus. The stop codon of PF10_0083 was replaced by the mCherry-HA tag, and the stop codon was provided by the mCherry-HA tag. As a result, the mCherry-HA tag was introduced after and in frame with PF10_0083, along with CAM 3' UTR and the hDHFR drug cassette (Figure 3.8B). As such, a functional tagged PF10_0083 protein was generated. Parasites were cultured under WR99210 drug pressure and infected RBCs were first observed about 14 days after transfection. PCR was performed to confirm the integration of the plasmid into the genome (Figure 3.8B, red arrowheads and Figure 3.8C).

WR99210 drug pressure was removed to allow the parasites still containing the episomal plasmid to lose the plasmid. However, the parasite culture may still contain a mixed population of parasites with integration, and parasites without the integration but carrying the plasmid episomally. In order to select only for parasites that have the integrated plasmid, the parasite population was cloned out to single cell by a technique known as cloning by limiting dilution (refer to section 2.6). A total of 48 wells were prepared and out of which 21 wells contained parasites (Appendix, Supplementary Figure 1). PCR was carried out to check for the presence of integrated plasmid (Figure 3.8B, red arrowheads and Figure 3.8C, lanes 1a and 1b). PCR was also carried out to identify the parasites carrying the plasmid episomally without integration (Figure 3.8B, green arrowheads and Figure 3.8C, lanes 2a and 2b). 18 out of 21 clones were positive for parasites with the integrated plasmid. Two clones, PF10_0083-CHA18 and PF10_0083-CHA30, were randomly selected for further studies. Real-time PCR was carried out to determine the expression level of the endogenously tagged protein compared to wild type parasites. Real-time PCR was performed on the ring, trophozoite and schizont stages. The reference control gene used was PFL0900c (arginyl-tRNA synthetase). The control sample used was untransfected 3D7 parasites. The fold changes were calculated by the ΔΔCt method. It was found that the expression of PF10_0083-CHA18 and PF10_0083-CHA30 was similar to that of the wild type parasites (Figure 3.8B, blue arrowheads and Figure 3.8D).



P1, P2: PCR to check for integration of plasmid
P3, P4: PCR to identify parasites without integration, but carrying plasmid episomally
P5, P6: RT-PCR



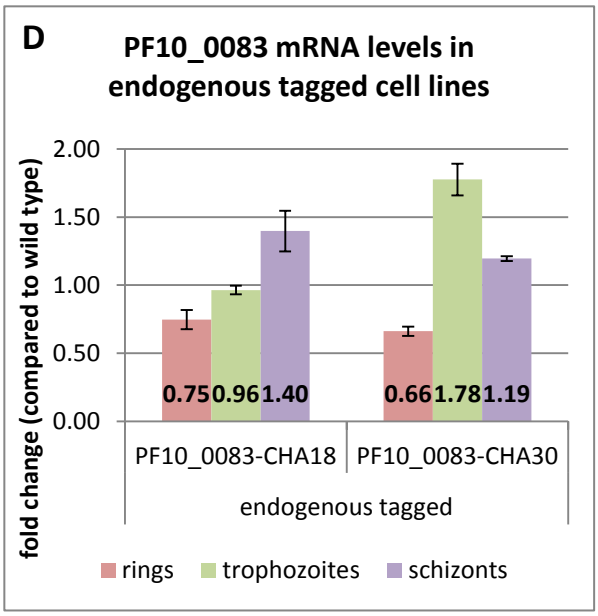


Figure 3.8 Endogenous tagging of PF10_0083.

Targeted gene tagging by homologous integration of plasmid DNA by a single crossover event resulted in the insertion of the resistance gene cassette (green arrows) and the tagging of the endogenous PF10_0083 gene. (A) The pTCherryHA plasmid was used to prepare the construct for PF10_0083 tagging. A 800 bp region, lacking the stop codon, from the 3'end of PF10_0083 (blue arrow) was cloned in frame with and before the mCherry-HA tag (pink and purple arrows). Homologous recombination occurred at the endogenous PF10_0083 locus (grey arrow). (B). The PF10_0083 region lacking the stop codon (blue arrow) was introduced together with the mCherry-HA tag (containing the stop codon, pink and purple arrows) in frame into the endogenous PF10_0083 locus (grey arrow). The red arrowheads represent the positions of primers for PCR to check for plasmid integration. The green arrowheads represent the positions of primers for PCR to check for parasites without integration. The blue arrowheads represent the positions of primers for RT-PCR. (C) PCR confirmed the integration of the plasmid DNA at the correct locus (lane 1a and 1b). PCR was carried out to identify the parasites carrying the plasmid episomally without integration (lanes 2a and 2b). (D) Real-time PCR (RT-PCR) determined the level of expression of tagged PF10_0083 in ring, trophozoite, and schizont stages during IDC. Real-time PCR experiments were performed in triplicates. The reference control gene used was PFL0900c (arginyl-tRNA synthetase). The control sample used was untransfected 3D7 parasites. The ΔΔCt method was used to calculate the fold change. ΔΔCt equals to (Ct of test sample target gene - Ct of test sample reference gene) minus (Ct of control sample target gene - Ct of control sample reference gene), where Ct is the threshold cycle. The calculated fold change of tagged PF10_0083 was close to one which is the endogenous level. Error bars represent standard deviation of the mean.

3.4.1 PF10_0083 protein endogenous levels

In the next step, we determined the endogenous protein levels of PF10_0083 using western blot analysis. The main aim was to validate the findings from the over-expression cell lines that PF10_0083 protein levels peaked during the schizont stages (section 3.3.2.2, Figure 3.5). Rings, trophozoites and schizonts protein samples were harvested from PF10_0083-CHA18 and PF10_0083-CHA30 cell lines. The protein samples were separated on a 12% SDS-PAGE system. Western blot was carried out

with anti-mCherry and anti-HA antibodies. The size of a mCherry-HA tag is 30 kDa. Both the anti-mCherry and anti-HA antibodies detected a distinct band at the expected size of ~70 kDa. PF10_0083 was only detected in the schizont stage parasites, and not in the ring and trophozoite stage parasites, in both PF10_0083-CHA18 and PF10_0083-CHA30 cell lines (Figure 3.9).

In contrast to the over-expression cell lines, PF10_0083 protein in the endogenous-tagged cell lines was undetectable in the ring and trophozoite stages. In the over-expression cell lines, it was observed that the peak expression of PF10_0083 was between 32 and 40 hpi during the schizont stage (section 3.3.2.2, Figure 3.5). In agreement with the over-expression cell lines, PF10_0083 protein level in the endogenous-tagged cell lines was the highest in the schizont stage. In summary, the protein levels of PF10_0083 is the lowest in the ring and trophozoite stages, and peaks during the schizont stage in both the over-expression and the endogenous-tagged cell lines.

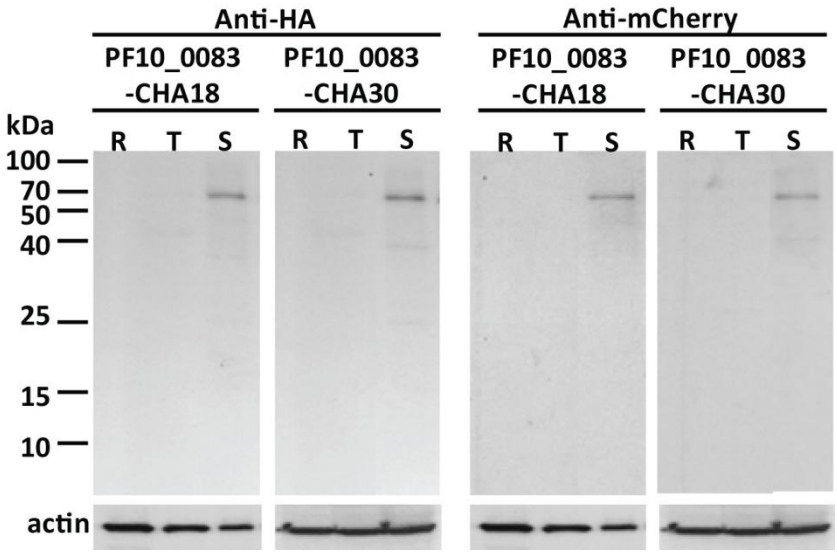


Figure 3.9 Endogenous protein level of PF10_0083 during the IDC.

Protein samples were harvested from rings, trophozoites and schizonts from the PF10_0083-CHA18 and PF10_0083-CHA30 endogenous-tagged cell lines. Protein

samples were separated on a 12% SDS-PAGE system. Western blot analysis was performed using anti-HA and anti-mCherry antibodies. A band at ~70 kDa was observed in the schizont stage parasites, and not in the ring and trophozoite stages parasites. Bottom lanes are the actin loading control. R: rings, T: trophozoites, S: schizonts.

3.4.2 PF10_0083 intracellular localization in the endogenous-tagged cell lines

IFA was performed to determine the localization of PF10_0083 in the PF10_0083-CHA parasites. The main aim was to validate the findings from the over-expression cell lines that PF10_0083 translocates into the nucleus during the schizont stages (section 3.3.3, Figure 3.6). Figure 3.10 and Figure 3.11 illustrates the intracellular localization of PF10_0083 in the PF10_0083-CHA18 cell line. Parasite smears were prepared from ring, trophozoite and schizont stages. In the schizont stage parasites, the fluorescence signal detected by both the anti-mCherry and anti-HA antibodies was localized predominantly in the nuclear compartment (Figure 3.10A). These results are in agreement with the over-expression cell lines that the majority of PF10_0083 localizes to the nucleus during the schizont stage. In contrast to the over-expression parasites (pARL-PF10_0083-myc and pARL-PF10_0083-gfp, Figure 3.6), in PF10_0083-CHA18, the fluorescence staining pattern of PF10_0083 was not localized to a focused small spot at the periphery of the nuclei. Instead, several distinct small spots were observed to be surrounding each blue DAPI nuclear signal. In addition, the fluorescence signal of both anti-mCherry and anti-HA antibodies co-localized to the same regions (Figure 3.10B). It was also found that the localization of PF10_0083 during rings and trophozoites was not detectable by either the anti-mCherry or anti-HA antibodies (Figure 3.11).

In summary, similar to the over-expression cell lines, we observed nuclear localization of PF10_0083 during the schizont stage in both the over-expression and endogenous-tagged cell lines.

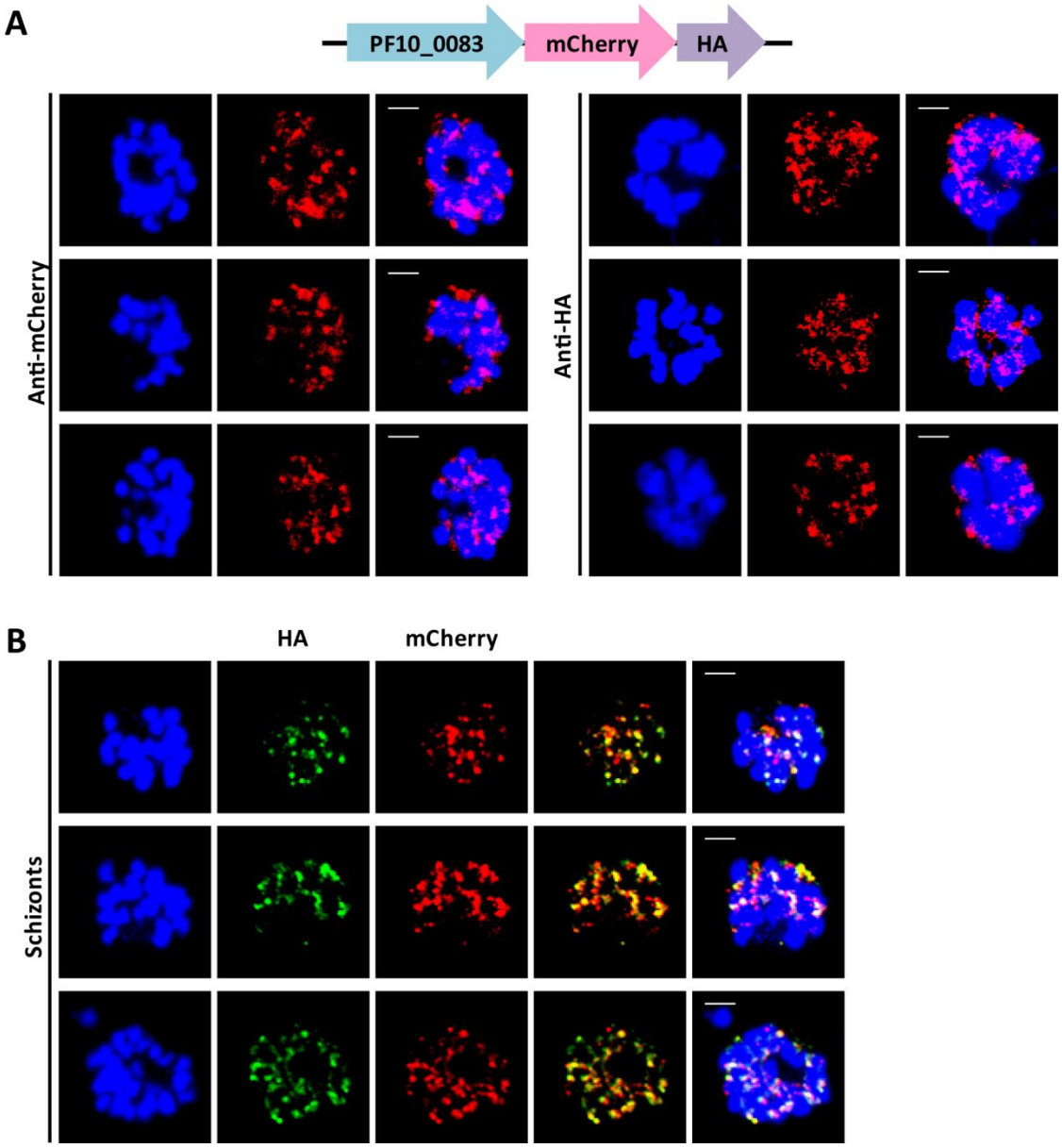


Figure 3.10 Localization of PF10_0083 during schizont stage.

IFA was carried out to determine the intracellular localization of PF10_0083 during the schizont stage in the PF10_0083-CHA18 cell line. (A) IFA was performed with the anti-mCherry and anti-HA antibody (red), and DAPI stained the nuclei blue. The fluorescence signal by both the anti-mCherry and anti-HA antibodies was localized predominantly in the nuclear compartment. Several distinct fluorescence spots were

observed to be surrounding each blue DAPI nuclear signal. (B) Anti-mCherry (red) and anti-HA (green) fluorescence signals co-localized to the same region (yellow). Scale bar = 2μm.

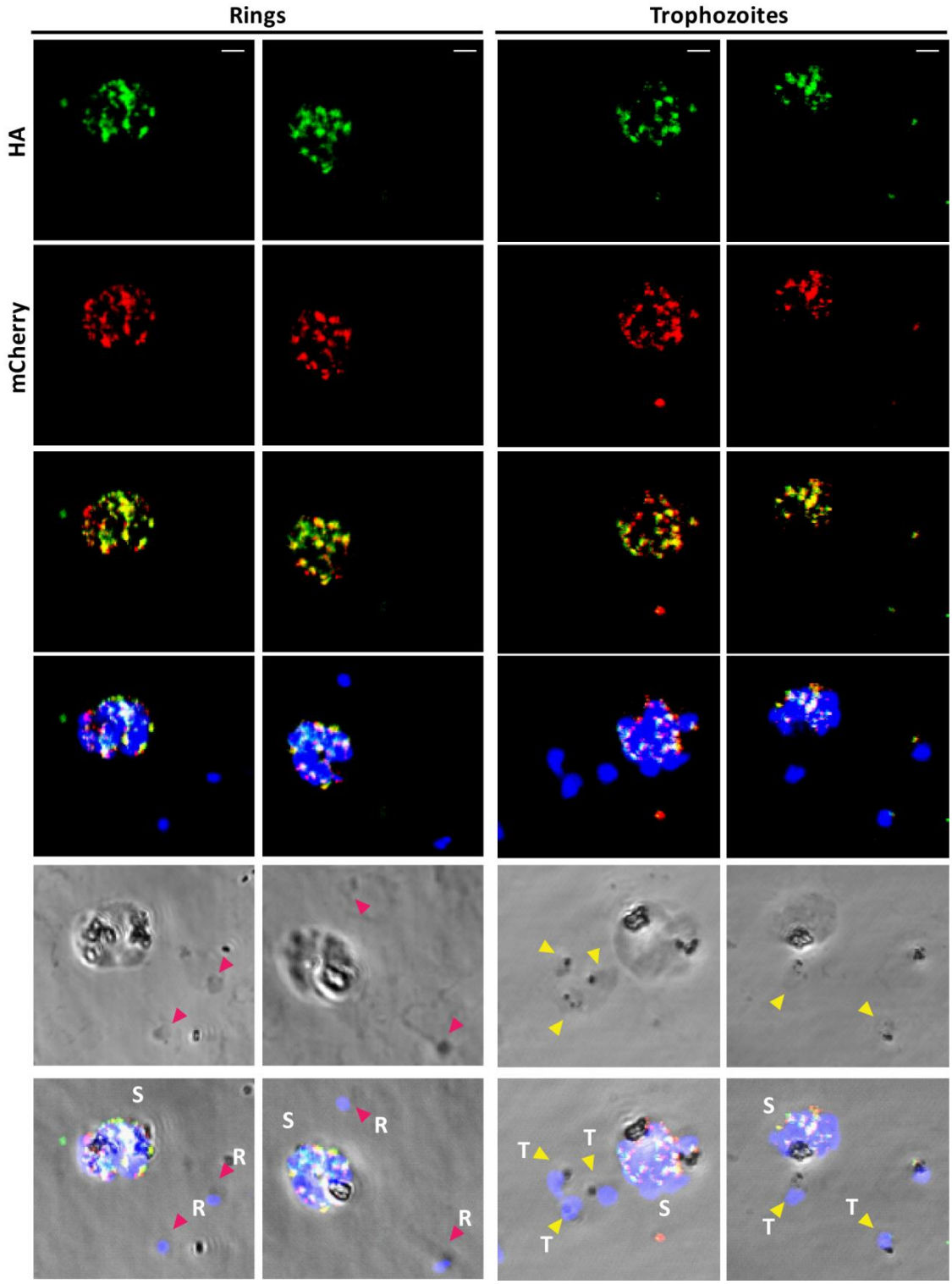


Figure 3.11 Localization of PF10_0083 during ring and trophozoite stages.

IFA was carried out to determine the intracellular localization of PF10_0083 during the ring and trophozoite stages in the PF10_0083-CHA18 cell line. IFA was performed with the anti-mCherry (red) and anti-HA antibodies (green), and DAPI stained the nuclei blue. The rings (pink arrowheads) and trophozoites (yellow arrowheads) lack both the HA (green) and mCherry (red) signal, unlike the schizonts. R: rings, T: trophozoites, S: schizonts. Scale bar = 2μm.

3.4.3 Co-localization of PF10_0083 with other nuclear proteins

Since PF10_0083 was localized to the nucleus during the schizont stage, IFA was performed to determine the co-localization of PF10_0083 with other nuclear proteins (Figure 3.12). IFA was carried out on the over-expression cell line, T996-pARL-PF10_0083-myc1 and the endogenous-tagged cell line, PF10_0083-CHA18. IFA was performed with the anti-myc and anti-HA antibodies. Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) is a well-known housekeeping gene that localizes to the cytoplasm, and is used as a control to demonstrate that PF10_0083 protein localized mainly to the nucleus. Heterochromatin protein 1 (HP1), histone 3 lysine 9 tri-methylation (H3K9Me3), and histone 3 lysine 79 tri-methylation (H3K79Me3), are highly correlated with constitutive heterochromatin [238, 262, 272], and do not co-localize with PF10_0083. This observation is consistent with the earlier finding that PF10_0083 is loosely associated with the chromatin in the nuclear fractionation experiment (section 3.3.3, Figure 3.7). Proteins that are associated with the heterochromatin, such as HP1, are tightly bound to the chromatin and are only extracted in high salt buffer [272]. Histone 3 lysine 9 acetylation (H3K9Ac), and histone 4 lysine 5 acetylation (H4K5Ac) are both associated with euchromatin [233, 273], but do not co-localize with PF10_0083. *P. falciparum* nucleoporin, Nup100, has been shown to co-localize with many proteins known to be associated with chromatin remodeling, and they co-localized to distinct nuclear compartments [268]. It was

observed that PF10_0083 did not co-localize with Nup100. Although the list of nuclear markers used in this study is not exhaustive, the results presented here strongly suggest that PF10_0083 localized to a distinct compartment in the nucleus.

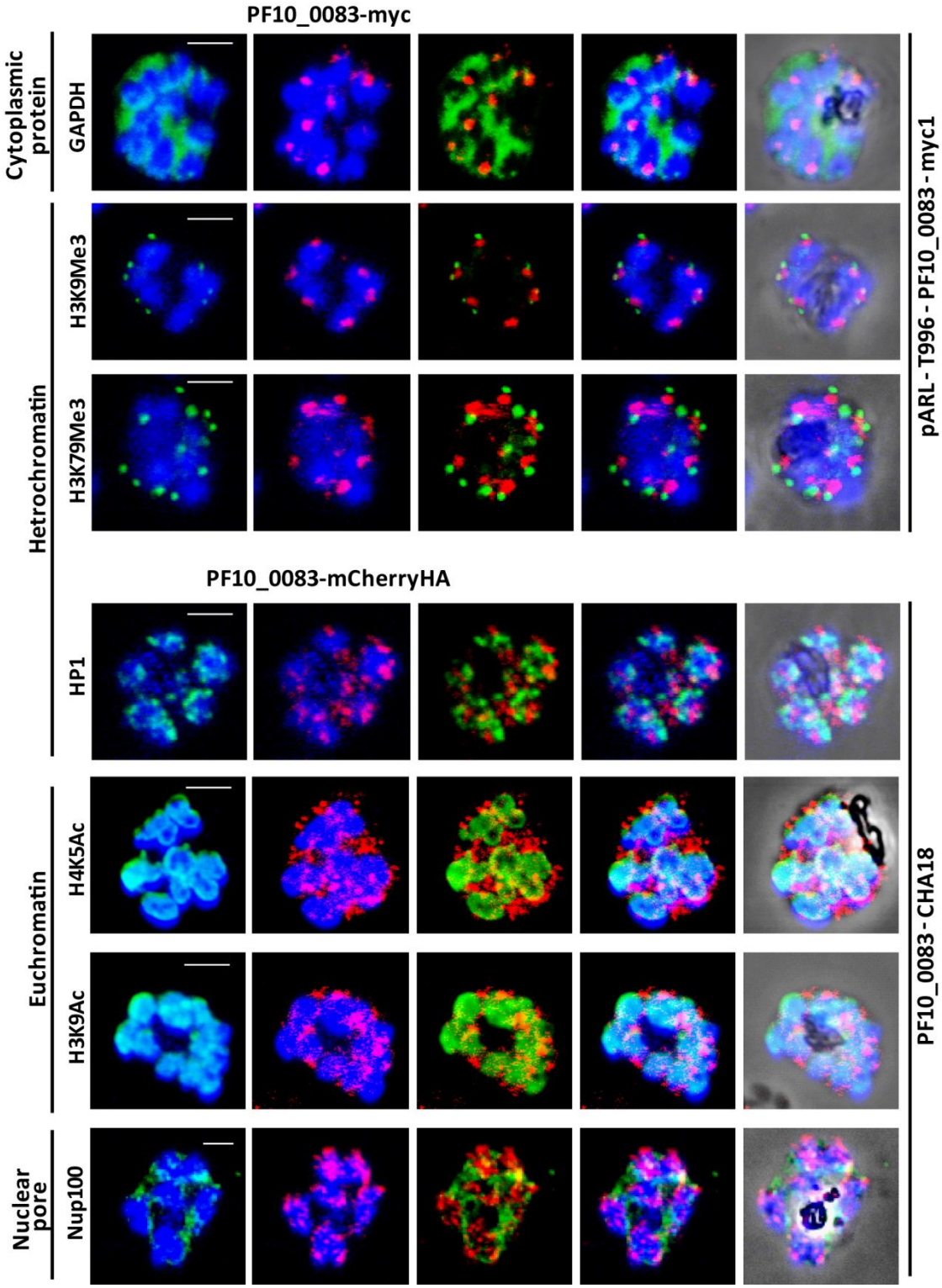


Figure 3.12 Co-localization of PF10_0083 with other nuclear proteins.
IFA was carried out to determine the co-localization of PF10_0083 with cytoplasmic and nuclear proteins. IFA was performed on the T996-pARL-PF10_0083-myc1 (over-expression) and PF10_0083-CHA18 (endogenous tagged) cell lines. IFA was performed with the anti-myc and anti-HA antibodies (red). The cytoplasmic and nuclear proteins were stained green. DAPI stained the nuclei blue. GAPDH is a cytoplasmic marker. H3K9Me3, H3K79Me3 and HP1 are heterochromatin markers. H4K5Ac and H3K9Ac are euchromatin markers. Nup100 is a nuclear pore marker. Scale bar = 2μm.

3.5 Targeted gene disruption of PF10_0083

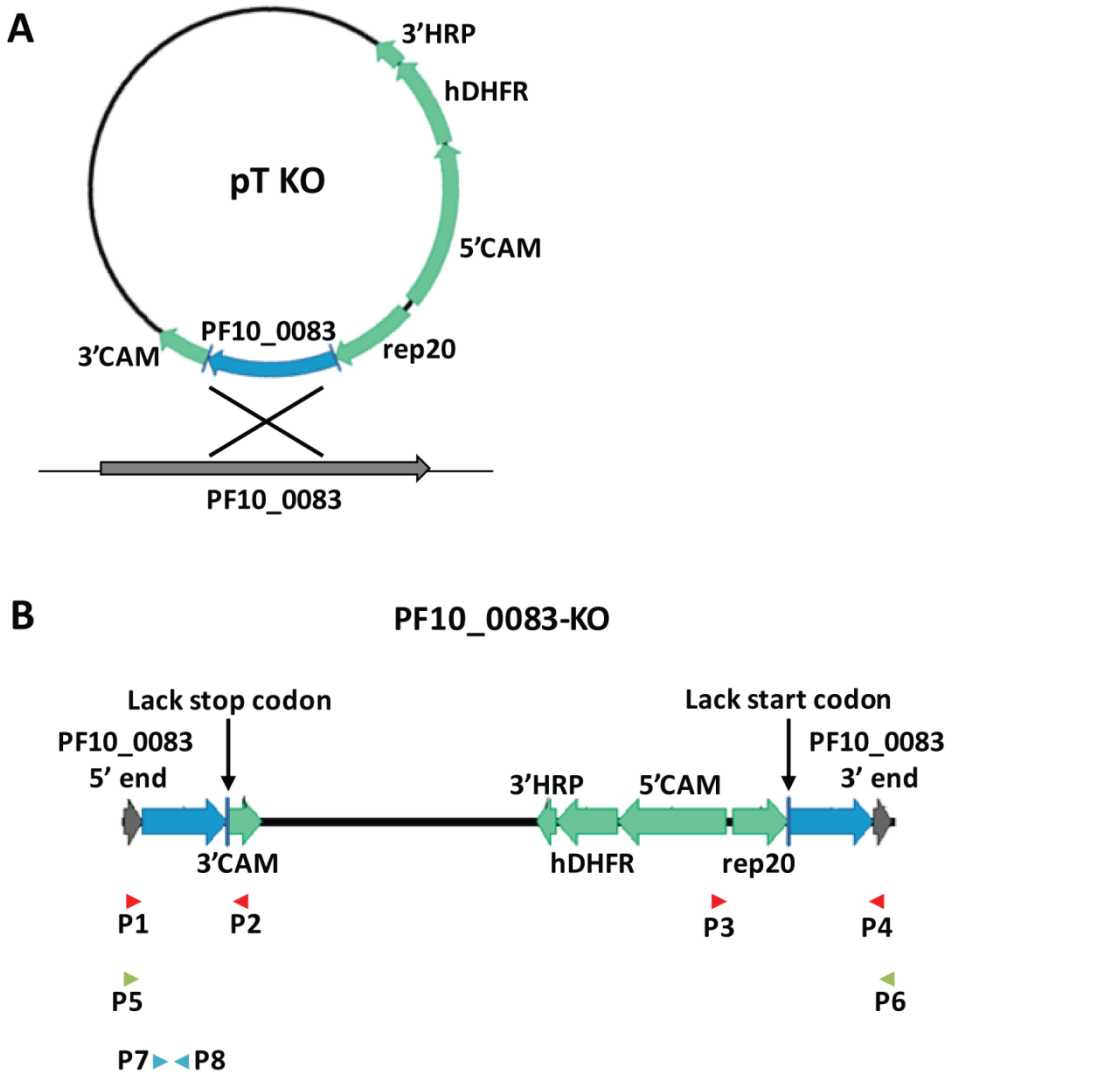
Targeted gene disruption of PF10_0083 was employed to determine the loss of function phenotype. The pTCherryHA plasmid was used to prepare the knock-out construct. A 800 bp region of PF10_0083, starting from position 161 bp to 949 bp, was cloned into the plasmid, before the 3'CAM (Figure 3.13A, blue arrow). The plasmid, now named as the pT KO plasmid, was transfected into *P. falciparum* 3D7 strain. Homologous recombination by a single crossover event occurs to introduce the hDHFR drug cassette into PF10_0083 locus and disrupt the gene at the same time (PF10_0083-KO). A truncated PF10_0083 lacking the last 200 bp region and the stop codon was generated before the hDHFR drug cassette. A truncated PF10_0083 lacking the first 200 bp region and the start codon was generated after the hDHFR drug cassette (Figure 3.13B). Parasites were cultured under WR99210 drug pressure and infected RBCs were first observed about 14-16 days after transfection. PCR was performed to confirm the disruption of PF10_0083 by means of the integration of the plasmid into the genome (Figure 3.13B, red arrowheads and Figure 3.13C, lanes 1 and 2).

WR99210 drug pressure was removed to allow the parasites to lose the episomal plasmid. However, the removal of WR99210 drug pressure did not result in all the

parasites losing the pT KO plasmid. There was a mixed population of parasites. As revealed by PCR, there was a population of parasites with PF10_0083 successfully knocked-out (Figure 3.13C, lanes 1 and 2), and another carrying the plasmid episomally in which knock-out did not occur (Figure 3.13C, lane 3). Therefore, the parasite population was cloned out to single cell by limiting dilution. In total, 15 out of 48 wells were positive for parasites (Appendix, Supplementary Figure 1). Clones that lack the intact PF10_0083 were determined by PCR (Figure 3.13B, green arrowheads and Figure 3.13C, lane 4-6). Two clones, PF10_0083-KO24 and PF10_0083-KO33, were randomly selected for further studies. Real-time PCR was carried out on cDNA which was reverse transcribed from mRNA, to determine the expression level of PF10_0083 mRNA in the knock-out cell line compared to PF10_0083-CHA18 and PF10_0083-CHA30 parasites. Real-time PCR was performed on the ring, trophozoite and schizont stages. The reference control gene used was PFL0900c (arginyl-tRNA synthetase). The control sample used was untransfected 3D7 parasites. The fold changes were calculated by the ΔΔCt method. The results indicated that the expression of PF10_0083 in PF10_0083-KO24 and PF10_0083-KO33 was much lower compared to that of the PF10_0083-CHA18 and PF10_0083-CHA30 parasites (Figure 3.13B, blue arrowheads and Figure 3.14).

It is important to note that the knock-out generated truncated PF10_0083 lacking the last 200 bp region and the stop codon or the first 200 bp region and the start codon (Figure 3.13B). The real-time PCR indicated that the mRNA levels of PF10_0083 in knock-out cell lines was much lower compared to that of the endogenous-tagged cell lines (Figure 3.14). However, we were not able to determine the protein level of the PF10_0083 knock-out lines. We have generated antibodies against the protein.

However, the antibodies generated produced unspecific binding signals (results not shown). As such, we are unable to rule out that a functional truncated protein is produced in the knock-out cell lines.



P1, P2, P3, P4: PCR to check for integration of plasmid
P5, P6: PCR to identify parasites without integration, but carrying plasmid episomally
P7, P8: RT-PCR

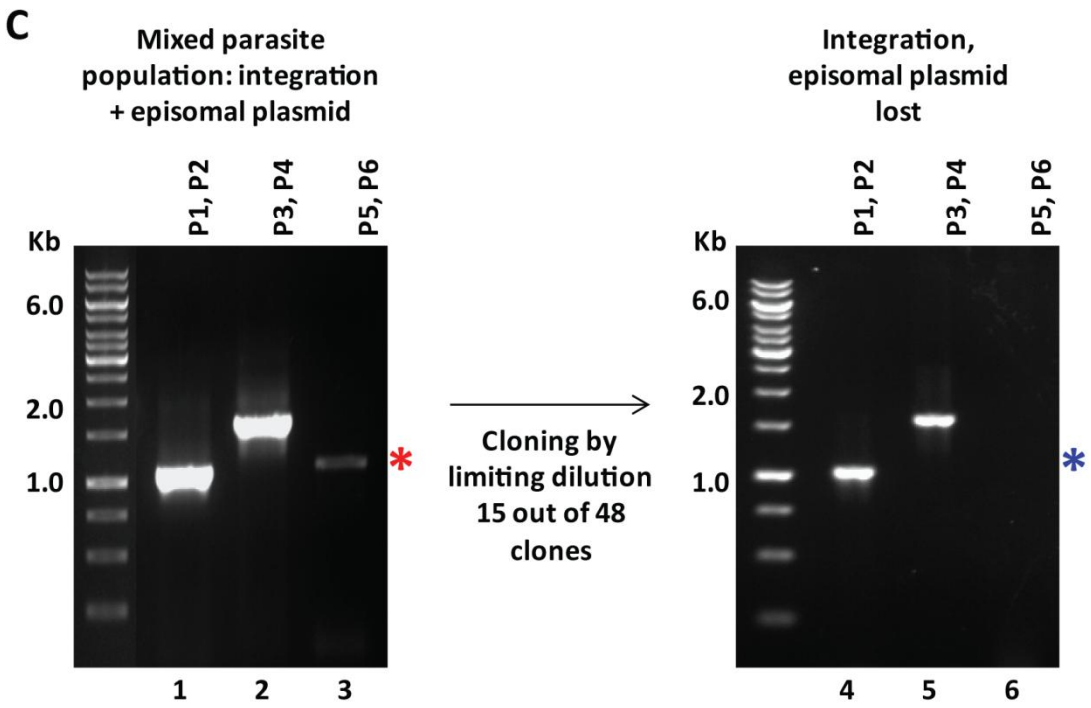


Figure 3.13 Targeted gene disruption of PF10_0083.

Targeted gene disruption by homologous integration of the plasmid DNA by a single crossover event resulted in the insertion of the resistance gene cassette (green arrows) and the disruption of the endogenous PF10_0083 gene. (A) The pTCherryHA plasmid was used to prepare the knock-out construct. A 800 bp region of PF10_0083, starting from position 161 bp to 949 bp (blue arrow), was cloned into the plasmid, before the 3'CAM (green arrow). Homologous recombination occurred at the endogenous PF10_0083 locus (grey arrow). (B) The endogenous PF10_0083 locus contains a fragment of gene that lacks the last 200 bp region and the stop codon before the drug cassette, and a fragment that lacks the first 200 bp region and the start codon after the drug cassette (blue arrows). The red arrowheads represent the position of primers for PCR to check for plasmid integration. The green arrowheads represent the positions of primers for PCR to check for parasites without integration. The blue arrowheads represent the positions of primers for RT-PCR. (C) PCR showed that targeted gene disruption of PF10_0083 (lanes 1 and 2) has been achieved. However, the parasite population still contains parasites that have the intact PF10_0083 (red asterisk). PCR was carried out to identify the parasites carrying the plasmid episomally without integration (lane 3). Cloning by limiting dilution selected for parasites that lack the intact PF10_0083 gene (lane 6, blue asterisk).

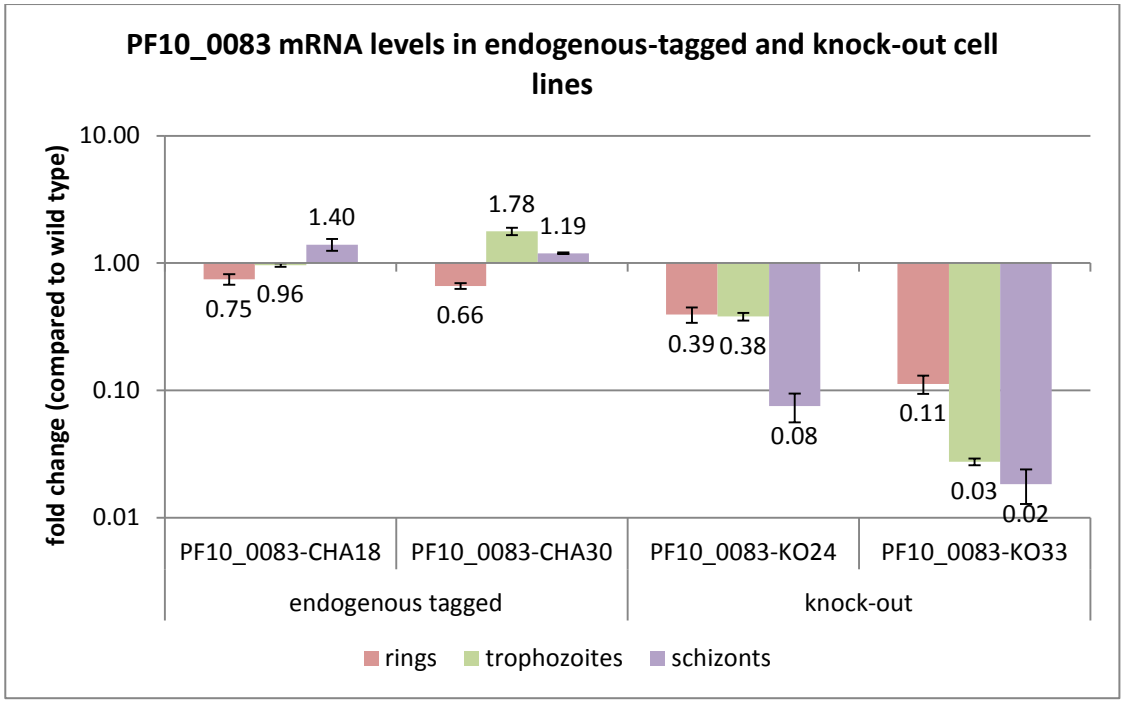


Figure 3.14 PF10_0083 mRNA levels in endogenous-tagged and knock-out cell lines. Real-time PCR determined the level of mRNA of PF10_0083 in PF10_0083-KO24 and PF10_0083-KO33 cell lines, compared to PF10_0083-CHA18 and PF10_0083-CHA30 cell lines. Real-time PCR determined the level of mRNA of PF10_0083 in ring, trophozoite, and schizont stages. Real-time PCR experiments were performed in triplicates. The reference control gene used was PFL0900c (arginyl-tRNA synthetase). The control sample used was untransfected 3D7 parasites. The ΔΔCt method was used to calculate the fold change. ΔΔCt equals to (Ct of test sample target gene - Ct of test sample reference gene) minus (Ct of control sample target gene - Ct of control sample reference gene), where Ct is the threshold cycle. The calculated fold change in PF10_0083-KO24 and PF10_0083-KO33 cell lines was lower than the endogenous level (PF10_0083-CHA18 and PF10_0083-CHA30 cell lines). Error bars represent standard deviation of the mean.

3.5.1 Growth rates of PF10_0083-KO and untransfected control parasites

Similar to the over-expression cell lines, the PF10_0083-KO parasites displayed no observable morphological difference from the untransfected control parasites by means of microscopic examination. The PF10_0083-KO parasites were observed to have a similar growth rate to the untransfected 3D7 parasites and the PF10_0083-CHA

parasites. The growth rates for a total of 20 parasitic cycles were recorded for wild-type 3D7 parasites, PF10_0083-CHA18 and PF10_0083-KO33 (Figure 3.15). The growth rate of the parasites was determined by the fold change in the parasitemia from schizont stage to the next ring stage by microscopy examination. The average growth rates of wild-type 3D7 parasites, PF10_0083-CHA18 and PF10_0083-KO33 were 3.2±1.3, 3.4±1.5 and 3.6±1.9 fold respectively. The differences in the growth rates between the wild-type 3D7 parasites and PF10_0083-CHA18, between wild-type 3D7 parasites and PF10_0083-KO33, and between PF10_0083-CHA18 and PF10_0083-KO33 were not statistically significant.

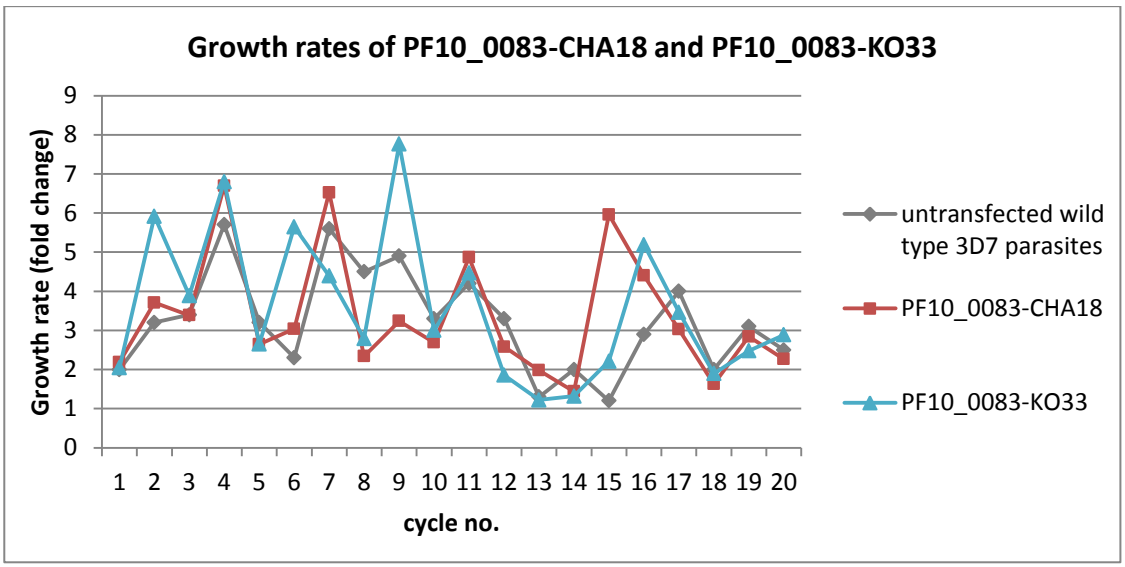


Figure 3.15 Growth rates of parasites in endogenous-tagged and knock-out cell lines. Comparison of growth rates of untransfected wild type 3D7 parasites, endogenous-tagged cell line PF10_0083-CHA18 and knock-out cell line PF10_0083-KO33 over 20 parasitic cycles. The differences in the growth rates between wild-type 3D7 parasites and PF10_0083-CHA18, between wild-type 3D7 parasites and PF10_0083-KO33, and between PF10_0083-CHA18 and PF10_0083-KO33 were not statistically significant. The growth rate of the parasites was determined by the fold change in the parasitemia from schizont stage to the next ring stage by microscopy examination.

Since PF10_0083-CHA18 and PF10_0083-KO33 cell lines were not cultured under drug pressure, the difference in the growth rates obtained from the 20 parasitic cycles is not directly comparable to the over-expression cell lines which were always grown under drug pressure. As such, we cultured the PF10_0083-CHA18 and PF10_0083-KO33 cell lines under the same drug pressure as the over-expression cell lines. At the same time, we also cultured 3D7-pARL-PF10_0083-myc3 transfectants and untransfected wild-type 3D7 parasites concurrently. The growth rates for a total of 20 parasitic cycles were recorded. The average growth rates of wild-type 3D7 parasites, 3D7-pARL-PF10_0083-myc3, PF10_0083-CHA18 and PF10_0083-KO33 were 2.9±1.1, 2.2±1.0, 3.1±1.2 and 3.0±0.8 fold respectively (Figure 3.16). The differences in the growth rates between the wild-type 3D7 parasites and PF10_0083-CHA18, between wild-type 3D7 parasites and PF10_0083-KO33, and between PF10_0083-CHA18 and PF10_0083-KO33 were not statistically significant. The difference in the growth rates between wild-type 3D7 and 3D7-pARL-PF10_0083-myc3 is statistically significant with a p-value of 0.03. The difference in the growth rates between PF10_0083-CHA18 and 3D7-pARL-PF10_0083-myc3 is also statistically significant with a p-value of 0.011. In addition, the difference in the growth rates between PF10_0083-KO33 and 3D7-pARL-PF10_0083-myc3 is also statistically significant with a p-value of 0.0001.

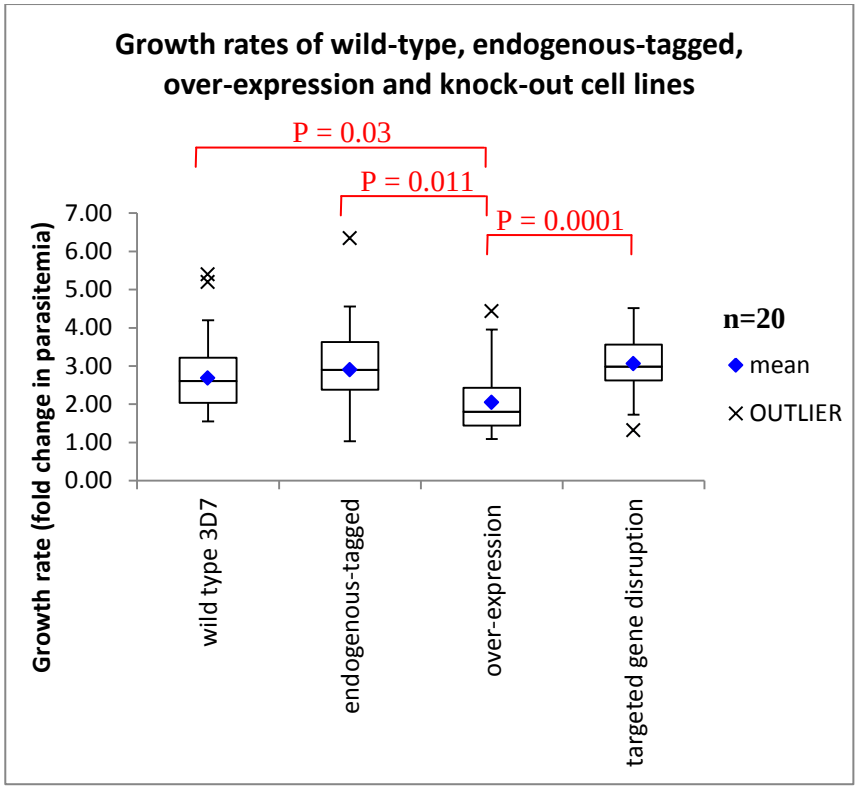


Figure 3.16 Growth rates of parasites in wild-type, endogenous-tagged, over-expression and knock-out cell lines.

Comparison of growth rates of untransfected wild type 3D7 parasites, endogenous-tagged cell line PF10_0083-CHA18, over-expression cell line 3D7-pARL-PF10_0083-myc3, and knock-out cell line PF10_0083-KO33 over 20 parasitic cycles. The differences in the growth rates between the wild-type 3D7 parasites and PF10_0083-KO33, between wild-type 3D7 parasites and PF10_0083-CHA18, and between PF10_0083-CHA18 and PF10_0083-KO33 were not statistically significant. The differences in the growth rates between wild-type 3D7 and 3D7-pARL-PF10_0083-myc3, between PF10_0083-CHA18 and 3D7-pARL-PF10_0083-myc3, and between PF10_0083-KO33 and 3D7-pARL-PF10_0083-myc3 are statistically significant. P-values are shown in red. The growth rate of the parasites was determined by the fold change in the parasitemia from schizont stage to the next ring stage by microscopy examination.

3.6 Protein expression

Since PF10_0083 is an uncharacterized protein, the availability of a protein structure would be very useful for further studies. As such, an attempt to express the full length PF10_0083 protein was made. We expressed PF10_0083 using the pET-24a and pET-

42a expression systems (Figure 3.17) in the BL21-CodonPlus-RIL competent cells. The pET24a vector contains a 6X histidine (his) tag, while the pET42 vector contains both the his and glutathione S-transferase (GST) tags. Briefly, PF10_0083 was cloned into the expression vectors in frame with the his tag. The recombinant plasmids were transformed into the BL21-CodonPlus-RIL competent cells. The addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) de-represses the *lac* operator and drives the expression of T7 polymerase. T7 polymerase binds to the T7 promoter that is upstream of PF10_0083, and drives protein expression. A total of 6 constructs, full length and 5 truncated variants were made (Figure 3.18). The expected sizes of the constructs are shown in Table 3.2.

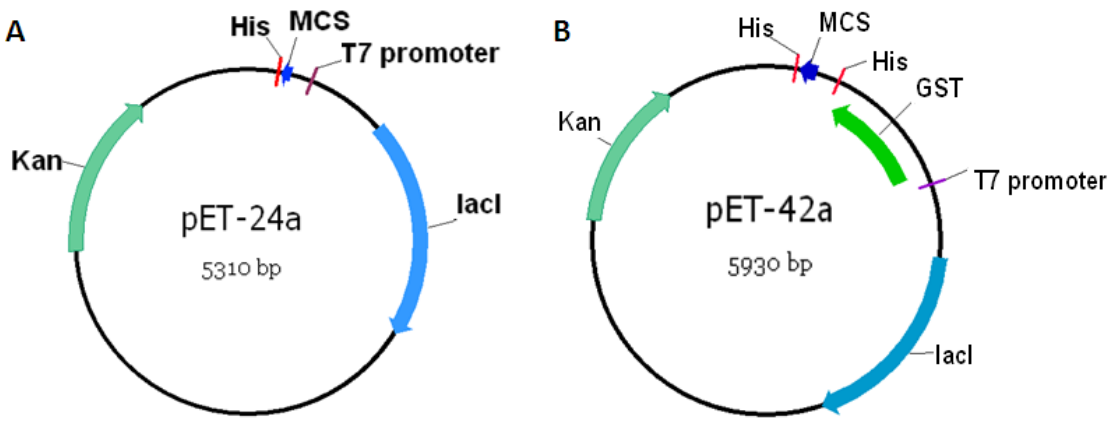


Figure 3.17 Protein expression vectors.

(A) pET-24a and (B) pET-42a expression vectors were used for PF10_0083 expression in BL21-CodonPlus-RIL *E. coli* cells. The pET24a vector contains a histidine (his) tag. The pET42a vector contains both the his and glutathione S-transferase (GST) tags. PF10_0083 was cloned into the expression vectors in frame with the his tag. The recombinant plasmids were transformed into the BL21-CodonPlus-RIL competent cells. The addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) de-represses the *lac* operator and drives the expression of T7 polymerase. T7 polymerase binds to the T7 promoter that is upstream of PF10_0083, and drives protein expression.

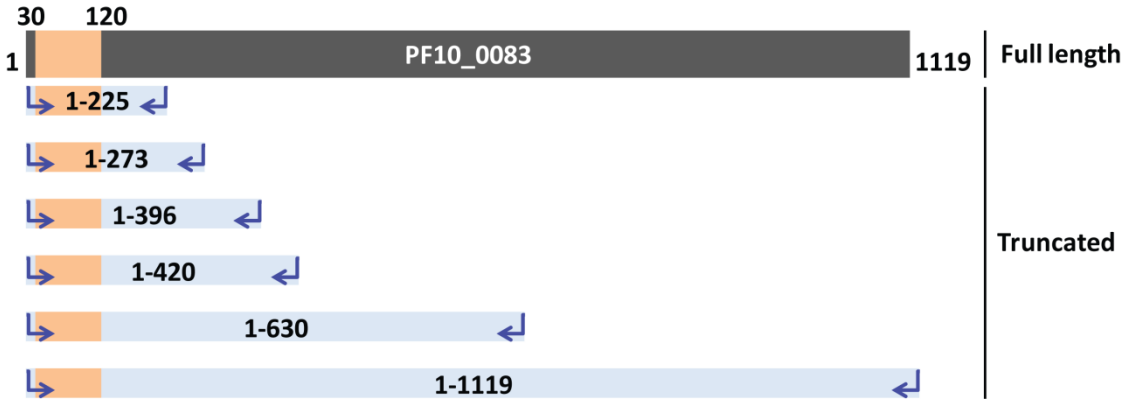


Figure 3.18 PF10_0083 protein expression constructs.

Six constructs, full length and 5 truncated variants of PF10_0083, were prepared. The positions of the primers are represented by the blue arrows. The length of each construct is shown in the blue boxes. Orange box represents the zinc finger domain between positions 30 bp and 120 bp. All 6 constructs contain the zinc finger domain.

Construct	pET-24a-PF10_0083-His (kDa)	pET-42a-GST-PF10_0083-His (kDa)
1-225	8.84	36.63
1-273	10.64	38.43
1-396	15.74	43.53
1-420	16.66	44.52
1-630	24.58	52.37
1-1119	44	71.79

Table 3.2 Expected protein sizes of the protein constructs in pET-24a and pET-42a vectors.

3.6.1 Expression of full length PF10_0083

The induction of full length PF10_0083 using the pET24a expression system in the expression host *E. coli* BL21-CodonPlus-RIL cells resulted in the expression of the protein at the expected size of ~47 kDa. Figure 3.19 illustrates the total protein sample separated on a 12% SDS-PAGE and stained by Coomassie blue. The initial induction condition used was 1 mM IPTG at 37°C for 3 hr. PF10_0083 was expressed at the

expected size of ~47 kDa compared to the uninduced control (Figure 3.18, red arrowhead). However, PF10_0083 formed inclusion bodies (Figure 3.18, solubility test: insoluble fraction, blue arrowhead). The formation of inclusion bodies suggests that the folding of PF10_0083 was improper and hence the protein accumulated as insoluble aggregates known as inclusion bodies. PF10_0083 in these inclusion bodies is misfolded and biologically inactive [274]. As such, attempts to modify the strategy for protein expression were made.

First, the induction conditions (induction time, IPTG concentration and induction temperature) were modified so that the protein solubility would improve. The first conditions tested were decreasing the induction time to 1 hr, and decreasing the IPTG concentration to 0.5 mM. However, PF10_0083 still formed inclusion bodies under both conditions. As such, the IPTG concentration was further decreased to 10 μM, 50 μM, and 100 μM. The induction temperature was also decreased to 16°C. This will allow the *E. coli* to grow at a slower rate, allowing ample time for proper protein folding. In summary, PF10_0083 protein was expressed under all these induction conditions, 0.5/1 mM IPTG at 37°C for 1 or 3 hr, and 10/50/100 μM IPTG at 16°C for 6 hr. PF10_0083 was expressed at the expected size of ~47 kDa. For the induction conditions 0.5/1 mM IPTG at 37°C for 3 hr, more protein was induced than the other induction conditions. However, in all the conditions, PF10_0083 formed inclusion bodies (Figure 3.19).

The full length PF10_0083 protein produced using the pET-24a system was misfolded and would therefore not be expected to have biological activity. As such, attempts to prepare recombinant PF10_0083 fused to a solubility enhancing protein were made.

The expression system was change to the pET-42a system. The pET-42a plasmid contains a GST tag that is known to help in the solubility of proteins. However, PF10_0083 was not expressed under any of the induction conditions tested in this study (Figure 3.20). In summary, the generation of soluble full length PF10_0083 protein under the various induction conditions was unsuccessful. The generation of recombinant PF10_0083 fused to a GST tag was also unsuccessful.

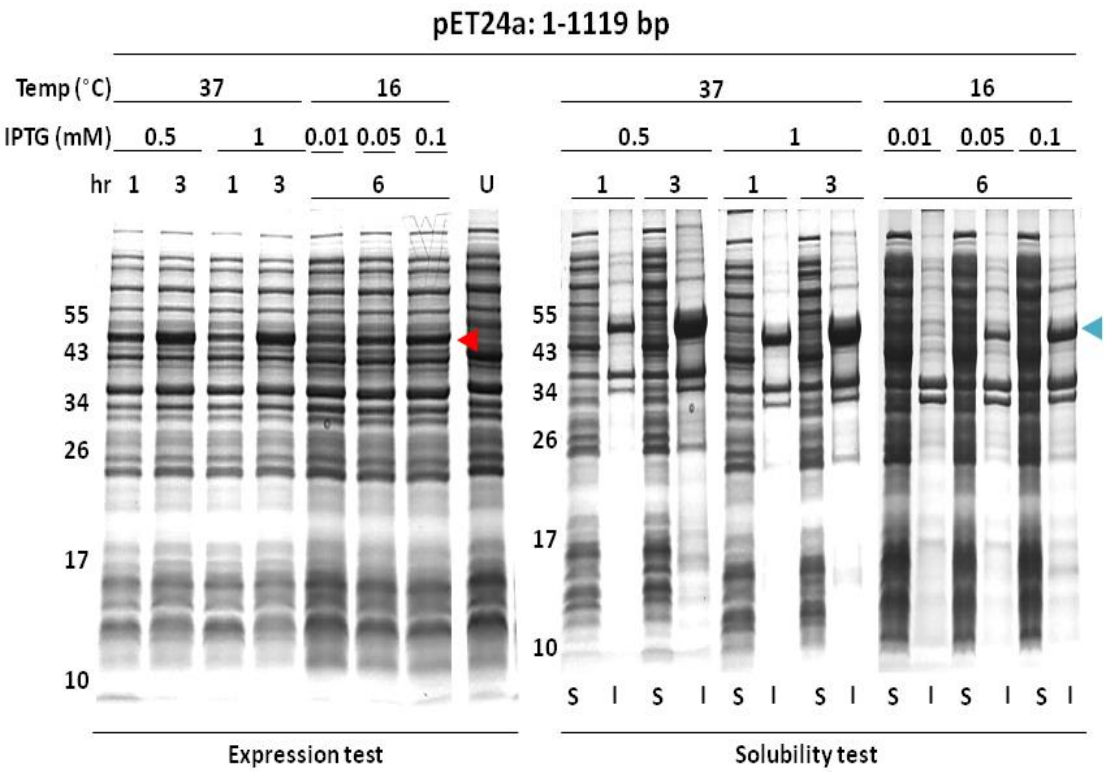


Figure 3.19 PF10_0083 (construct 1-1119 bp) protein expression.
Induction of full length PF10_0083 using the pET24a system in *E. coli* BL21-CodonPlus-RIL cells. The total protein sample was separated on a 12% SDS-PAGE and stained by Coomassie blue. The induction of full length PF10_0083 resulted in the expression of the protein at the expected size of ~47 kDa. Protein was expressed in all 7 induction conditions tested (red arrowhead). Solubility test showed that PF10_0083 was found in the inclusion bodies (insoluble fraction, blue arrowhead). U: uninduced control, S: soluble fraction, I: insoluble fraction.

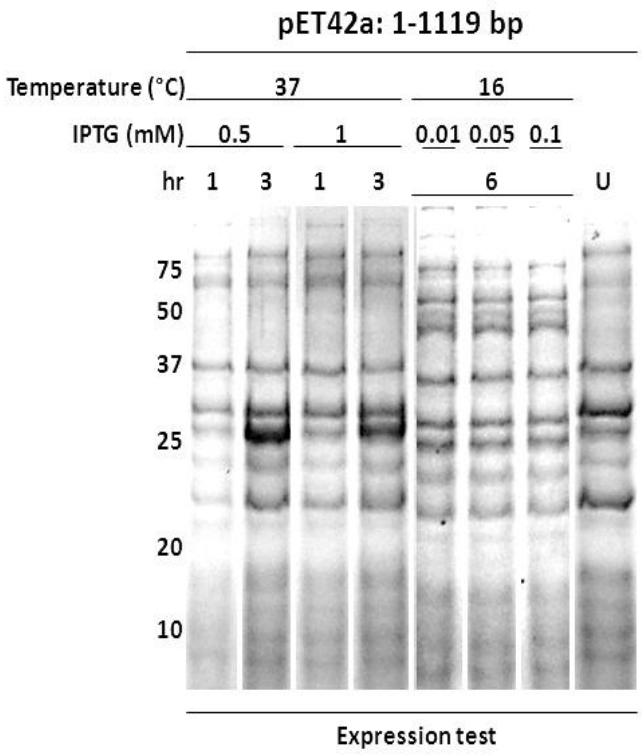
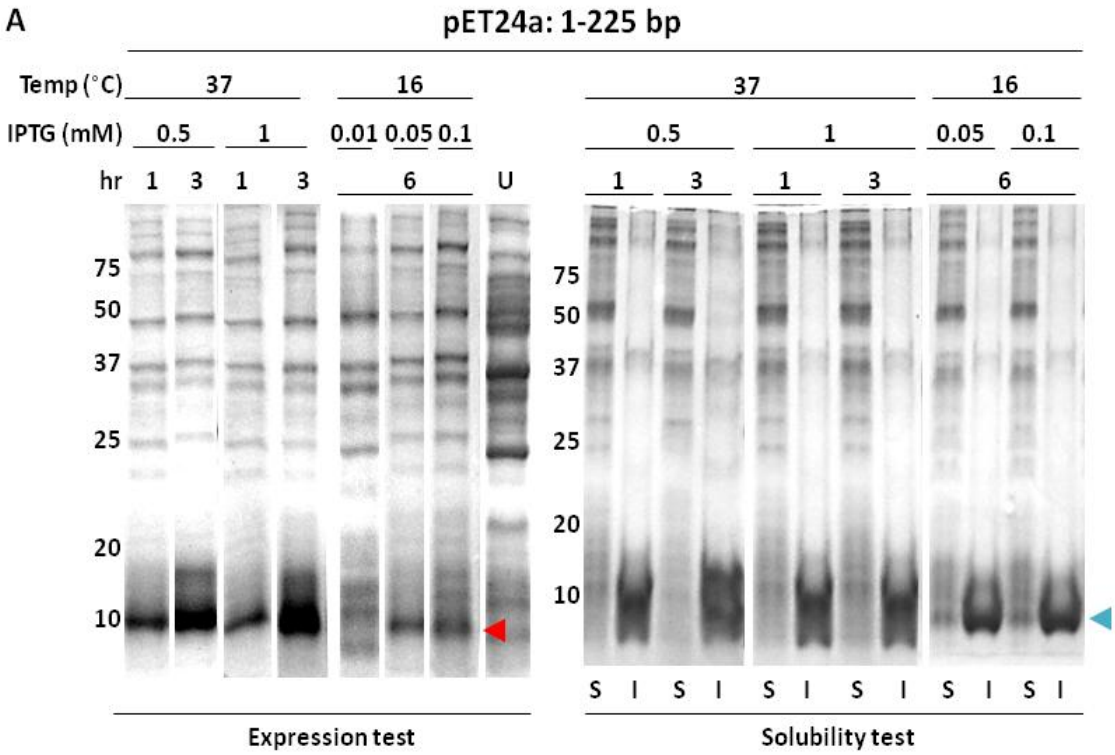


Figure 3.20 pET42a-PF10_0083 (construct 1-1119 bp) protein expression.
Induction of full length PF10_0083 using the pET42a system in *E. coli* BL21-CodonPlus-RIL cells. PF10_0083 was fused to both the his and GST tag. The total protein sample was separated on a 12% SDS-PAGE and stained by Coomassie blue. No protein was expressed under all the 7 induction conditions tested. U: uninduced control.

3.6.2 Screening of soluble variants of PF10_0083

As the full length PF10_0083 produced in both the pET24a and pET42a expression systems formed inclusion bodies, attempts to screen for soluble variants of PF10_0083 were made. A general approach to screen for soluble variants is by generating mutants of the protein of interest. Such mutants can be generated by site directed mutagenesis, random point mutation and protein truncation [275, 276]. Five truncated forms of PF10_0083 were prepared (Figure 3.18). All 5 truncated variants contain the zinc finger domain between position 30 bp and 120 bp. The truncated PF10_0083 was expressed using the pET24a expression system in the expression host *E. coli* BL21-CodonPlus-RIL cells. Construct 1-225 bp, construct 1-273 bp, and construct 1-396 bp

were expressed at the expected sizes of ~9 kDa, ~11 kDa and ~16 kDa respectively (Figure 3.21, red arrowheads). For the induction conditions, 0.5 and 1 mM IPTG at 37°C for 1 and 3 hr, more protein was induced than the other induction conditions. No protein was induced at 10 μM IPTG at 16°C for 6 hr. However, the 3 truncated variants of PF10_0083 formed inclusion bodies under all the induction conditions (Figure 3.21, solubility test: insoluble fraction, blue arrowheads).



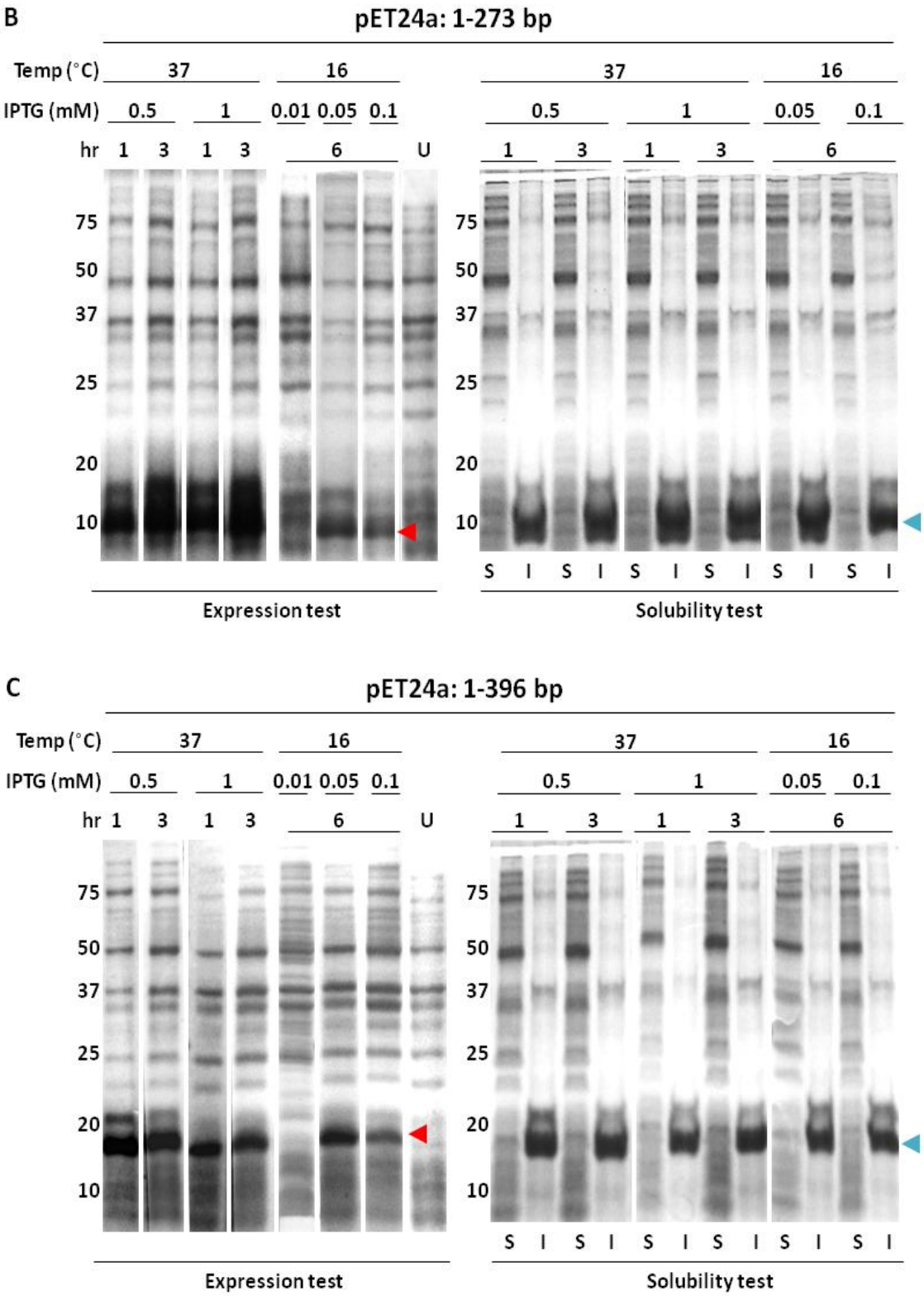


Figure 3.21 Protein expression of truncated variants of PF10_0083 in pET-24a expression system.

Protein expression and protein solubility studies of (A) construct 1-225 bp, (B) construct 1-273 bp, and (C) construct 1-396 bp. Induction of truncated variants of PF10_0083 using the pET24a system in *E. coli* BL21-CodonPlus-RIL cells. The total protein sample

was separated on a 12% SDS-PAGE and stained by Coomassie blue. The induction of truncated variants of PF10_0083, 1-225 bp, 1-273 bp, and 1-396 bp resulted in the expression of the protein at the expected sizes of ~9 kDa, ~11 kDa and ~16 kDa respectively. Protein was expressed in 6 induction conditions (red arrowheads), except for 10 μM IPTG at 16°C for 6 hr. Solubility test showed that PF10_0083 was found in the inclusion bodies (insoluble fraction, blue arrowheads). U: uninduced control, S: soluble fraction, I: insoluble fraction.

Similar to the other constructs, PF10_0083 truncated variant 1-420 bp was also expressed at the expected size of ~17 kDa under the induction conditions 0.5/1 mM IPTG at 37°C for 1/3 hr (Figure 3.22, red arrowhead). A large amount of this truncated variant formed inclusion bodies (Figure 3.22, blue arrowhead). However, a minute amount of protein was accumulated as soluble product in the bacterial cell (Figure 3.22, yellow arrowheads). As such, the induction conditions at 10/50/100 μM IPTG at 16°C for 6 hr were also tested. However, the protein produced under these conditions formed inclusion bodies. No protein was induced at 10 μM IPTG at 16°C for 6 hr. Next, we attempted to produce large amounts of construct 1-420 bp under the induction conditions 0.5/1 mM IPTG at 37°C for 1/3 hr. However, during large-scale culturing, PF10_0083 truncated variant 1-420 bp formed inclusion bodies. In summary, a truncated variant of PF10_0083 (1–420 bp) that contains the zinc finger domain was produced successfully. In addition, a small amount of this variant accumulated as soluble product in the bacterial cell. However, the production of a soluble PF10_0083 truncated variant was only successful at a small scale level. As such, more optimization work is required in order to produce large amount of the soluble PF10_0083 truncated variant.

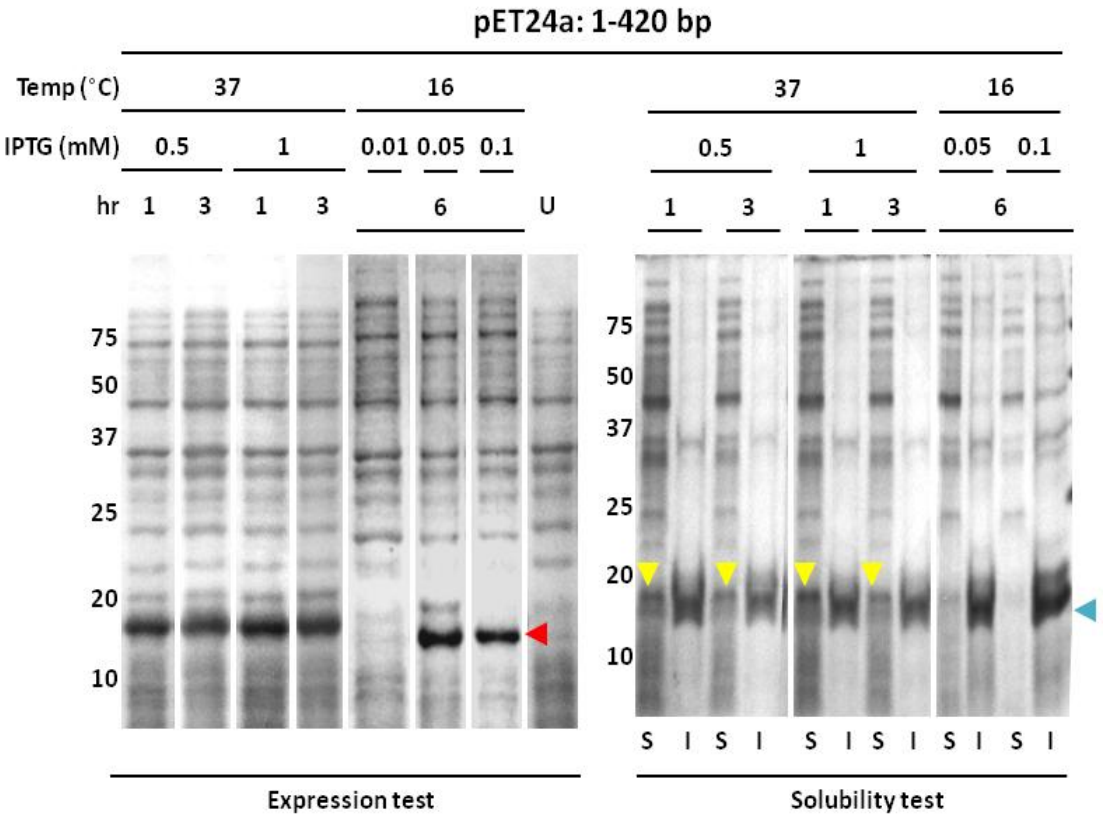


Figure 3.22 Protein expression and solubility studies of PF10_0083 truncated variant 1-420 bp in pET-24a expression system.

Induction of PF10_0083 truncated variant 1-420 bp using the pET24a system in *E. coli* BL21-CodonPlus-RIL cells. The total protein sample was separated on a 12% SDS-PAGE and stained by Coomassie blue. The induction of truncated variants of PF10_0083 resulted in the expression of the protein at the expected sizes of ~17 kDa. Protein expression was successful in 6 induction conditions (red arrowhead), except for 10 μM IPTG at 16°C for 6 hr. Solubility test showed that the majority of PF10_0083 was found in the inclusion bodies (insoluble fraction, blue arrowhead). Under the induction conditions 0.5/1 mM IPTG at 37°C for 1/3 hr, minute amount of PF10_0083 was found in the soluble fraction (yellow arrowheads). U: uninduced control, S: soluble fraction, I: insoluble fraction.

For PF10_0083 truncated variant 1-630 bp, no protein was induced under all induction conditions, 0.5/1 mM IPTG at 37°C for 1/3 hr, and 10/50/100 μM IPTG at 16°C for 6 hr (Figure 3.23). In addition, attempts to express the 5 truncated variants of PF10_0083 using the pET-42a expression system were also made. Figure 3.24

illustrates the protein expression and solubility studies of the 5 truncated variants under the induction condition 1 mM IPTG at 37°C for 3 hr. PF10_0083 truncated variants 1-225 bp, 1-273 bp, 1-396 bp, 1-420 bp and 1-630 bp were expressed at the expected sizes of ~37 kDa, ~39 kDa, ~44 kDa, ~45 kDa and ~52 kDa respectively (Figure3.24, red arrowheads). However, the protein expressed formed inclusion bodies (Figure3.24, green arrowheads). All 5 PF10_0083 truncated variants were also expressed at the expected sizes under the induction conditions 0.5 mM IPTG at 37°C for 1 or 3 hr, 1 mM IPTG at 37°C for 1 hr, and 10/50/100 μM IPTG at 16°C for 6 hr (results not shown). All the protein produced formed inclusion bodies (results not shown).

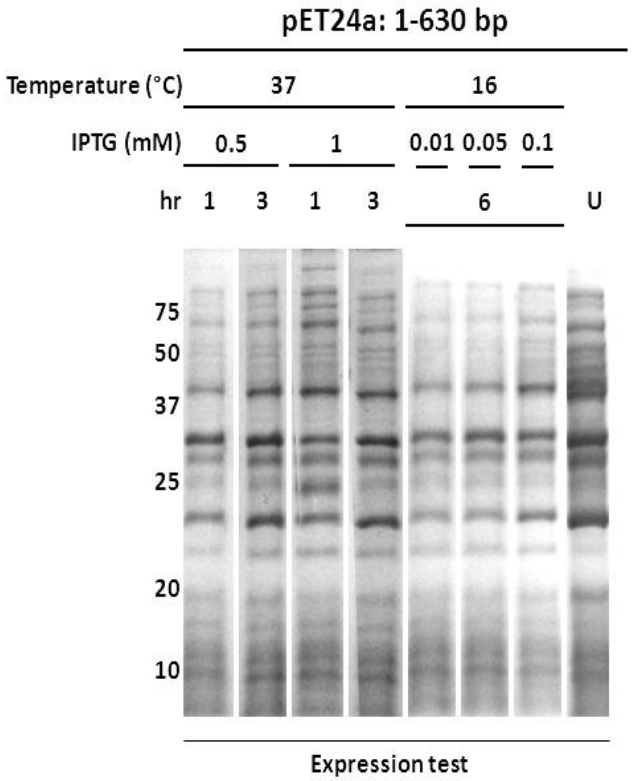


Figure 3.23 Protein expression of PF10_0083 truncated variant 1-630 bp in pET-24a expression system.

Induction of PF10_0083 truncated variant 1-630 bp using the pET24a system in *E. coli* BL21-CodonPlus-RIL cells. The total protein sample was separated on a 12% SDS-

PAGE and stained by Coomassie blue. No protein was induced under all the 7 different induction conditions. U: uninduced control.

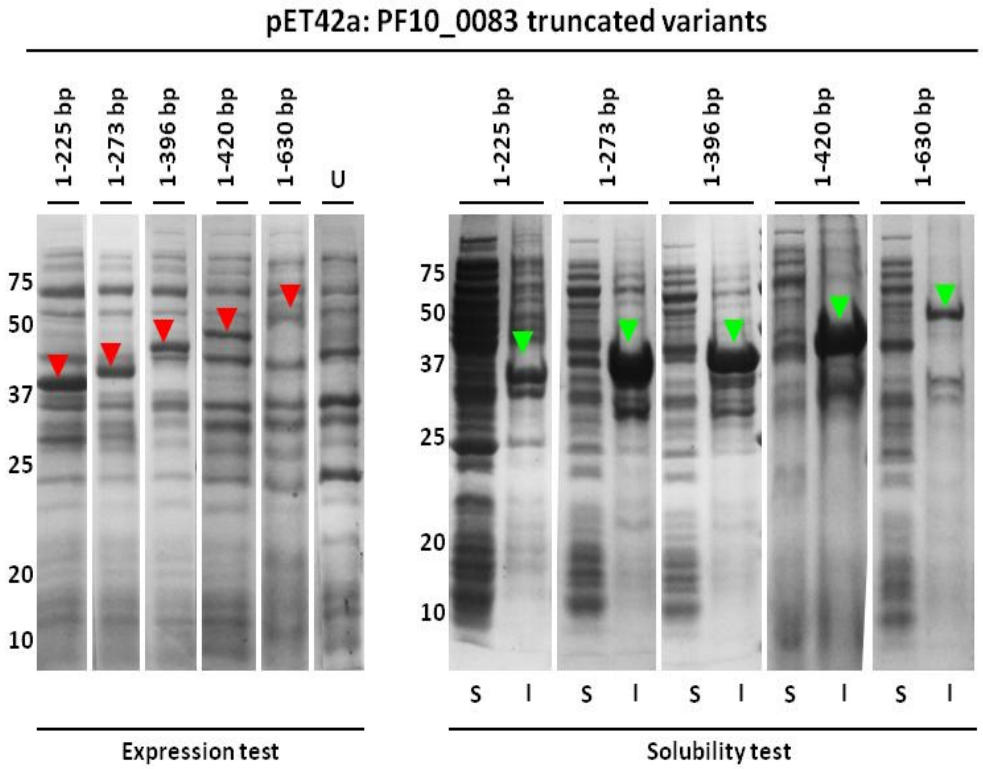


Figure 3.24 Protein expression of 5 PF10_0083 truncated variants in pET-42a expression system.

Induction of PF10_0083 truncated variants using the pET42a system in *E. coli* BL21-CodonPlus-RIL cells at 1mM IPTG at 37°C for 3 hr. The total protein sample was separated on a 12% SDS-PAGE and stained by Coomassie blue. The induction of truncated variants of PF10_0083, 1-225 bp, 1-273 bp, 1-396 bp, 1-420 bp and 1-630 bp resulted in the expression of the protein at the expected sizes of ~37 kDa, ~39 kDa, ~44 kDa, ~45 kDa and ~52 kDa respectively (red arrowheads). Solubility test showed that PF10_0083 was found in the inclusion bodies (green arrow heads). U: uninduced control, S: soluble fraction, I: insoluble fraction.

3.7 Summary

In summary, PF10_0083 over-expression, endogenous-tagged and knock-out cell lines were generated. Molecular characterization of PF10_0083 was carried out using these

cell lines. It was observed that PF10_0083 protein expression peaked during the schizont stage of the IDC in the over-expression cell lines. In addition, in the over-expression cell lines, PF10_0083 was localized to the cytoplasm during the ring and trophozoite stages, and in the nucleus during the schizont stage. In the over-expression cell lines, PF10_0083 peak protein expression coincided with its nuclear localization. In the endogenous-tagged cell lines, PF10_0083 protein expression was only detectable during the schizont stage and not in the ring and trophozoite stages. In the endogenous-tagged cell lines, PF10_0083 also localized to the nucleus during schizont stage. It was also observed that PF10_0083 over-expression cell lines had a lower parasite growth rate compared to the untransfected and transfected control cell lines. Lastly, the knock-out cell lines had similar parasite growth rate compared to the endogenous-tagged and untransfected control cell lines.

Chapter 4 The role of PF10_0083 in transcriptional regulation

4.1 Introduction

Transcriptional profiling of the over-expression cell lines was carried out. The main goal was to investigate the global transcriptional changes caused by altered PF10_0083 expression in order to identify its genomic targets. In order to study PF10_0083 occupancy throughout the genome in multiple developmental stages, ChIP-on-chip (chromatin immunoprecipitation coupled with microarray) experiments were carried out on the over-expression cell lines and the endogenous-tagged cell lines. Results from these studies were correlated with their transcriptional profiles, in order to understand the relationship between PF10_0083 occupancy and transcriptional activity. The main objective is to understand the role of PF10_0083 in gene expression and the identification of DNA regions that directly interact with PF10_0083.

4.2 Global transcriptional profiling to study the effects of PF10_0083 over-expression

To investigate the global transcriptional changes caused by PF10_0083 over-expression, the transcriptional response of the over-expression cell lines was characterized using genome-wide DNA microarrays. The main objective is to identify the putative target genes of PF10_0083. The transcriptional profiles of these cell lines obtained by genome-wide DNA microarrays were compared to the control parasites profiles (Table 4.1). An outline of the experiment is depicted in Figure 4.1. Parasites samples were harvested every 8 or 12 hours for the entire IDC (44-48 hr). Total RNA was isolated and used to synthesize cDNA for the microarray hybridization

experiments. All samples were hybridized against a common reference pool cDNA. The reference pool RNA was harvested every 8 hours for the entire IDC (44-48 hr) from 3D7 or T996 untransfected parasites. All 3D7 over-expression and control cell lines were hybridized against the 3D7 reference pool, and all T996 over-expression and control cell lines were hybridized against the T996 reference pool.

Control cell line	Over-expression cell line
Untransfected wild-type T996 parasites	T996-pARL-PF10_0083-myc1
T996-pARL-myc	T996-pARL-PF10_0083-myc2
T996-pARL-gfp	T996-pARL-PF10_0083-gfp
Untransfected wild-type 3D7 parasites	3D7-pARL-PF10_0083-myc3

Table 4.1 Summary of microarray experiments to study the effects of the over-expression of PF10_0083.
The transfected cell lines were compared to the respective control cell lines.

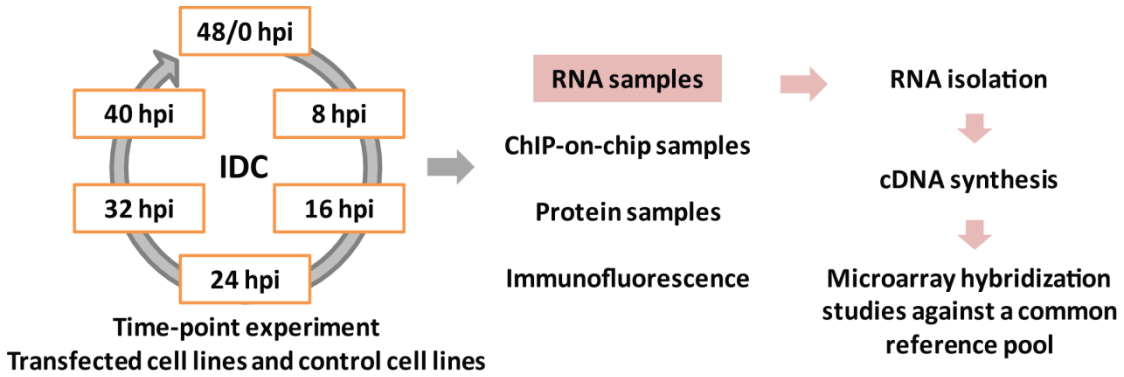


Figure 4.1 Outline of study: The transcriptional profiling of the effects of the over-expression of PF10_0083.
Parasites were harvested every 8 or 12 hours for the entire IDC, and samples for transcriptional profiling, ChIP-on-chip studies, western analysis and immunofluorescence studies were harvested. Total RNA were isolated and used to synthesize cDNA for microarray hybridization experiments. A common reference pool consisting of cDNA of T996 or 3D7 parasites from 6 time-points across the 48 hr lifecycle

was generated and both the transfectants and control time-point samples were hybridized against the common reference pool.

4.2.1 Transcriptional changes caused by PF10_0083 over-expression

Samples were harvested every 8 hours during the IDC. A total of 6 time-point samples were harvested for the T996-pARL-PF10_0083-myc1/myc2/gfp (hereafter referred to as T996-myc1, T996-myc2 and T996-gfp) parasite lines and their control cell lines (Table 4.1). For 3D7-pARL-PF10_0083-myc3 (hereafter referred to as 3D7-myc3) cell line and the control cell line, 3 samples were harvested every 12 hours during the IDC, and they represent ring, trophozoite, and schizont stage parasites. The ORFs of each gene are represented by multiple oligonucleotide probes on the microarray chip. The $\log_2$ transformed values of all the oligonucleotide probes that represent each gene were averaged. Therefore, the expression profile for each gene is represented by an average expression of all the oligonucleotide probes that represent the ORFs of each gene.

In order to identify genes that were differentially expressed, the microarray data was analyzed by calculating the Student's t-Test for every gene in the over-expression and the control cell lines. The Student's t-Test determines whether or not the difference in gene expression between the over-expression and control cell lines is statistically significant. Genes that have a p-value less than 0.05 were used for further analysis. To calculate the difference in the gene expression profiles between the over-expression and the control cell lines, the sum of each gene profile (represented by the area under the curve) was calculated. The $\log_2$ transformed values of each gene from all the time-points were summed up. The sum of each gene was calculated for both the over-expression and the control parasites. The difference in the sums between the over-

expression transfectants and the control parasites were determined. Genes are considered to be differentially expressed if the Student's t-Test p<0.05, and has at least a 2 fold difference in sums. The over-expression of PF10_0083 altered the expression of distinct sets of genes. The T996-myc1, T996-myc2 and T996-gfp and 3D7-myc3 over-expression cell lines induced a transcriptional response in 657, 1060, 90, and 85 genes respectively (Figure 4.2).

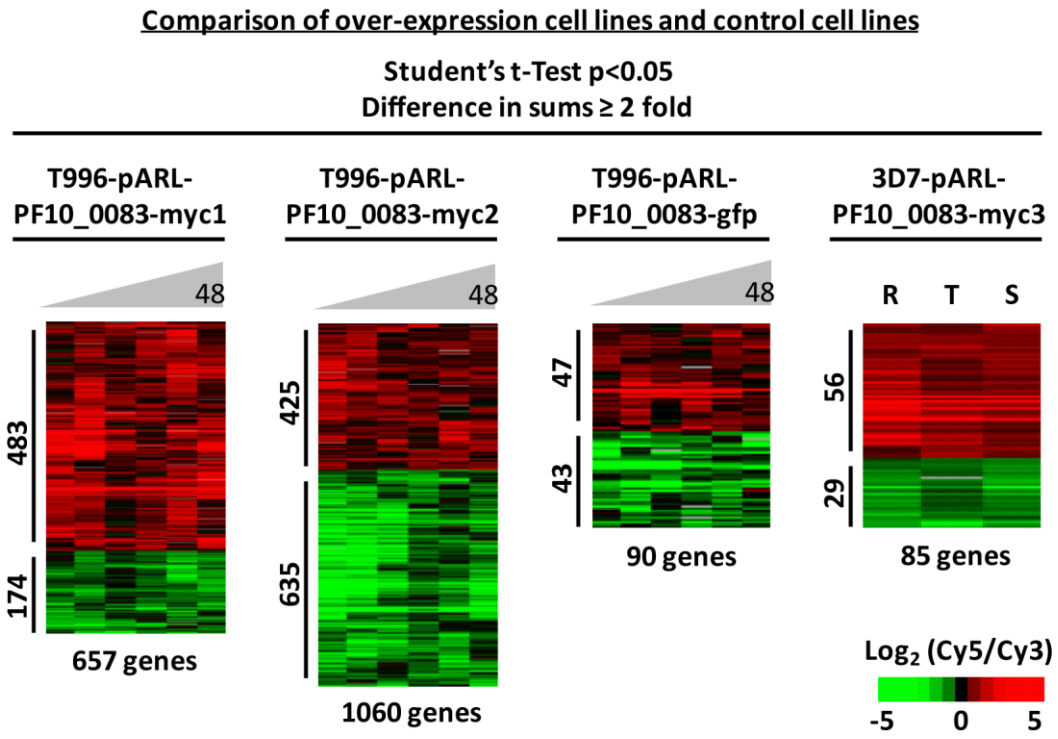


Figure 4.2 Differentially expressed genes in the over-expression cell lines compared to the control cell lines.

The Student's t-Test for every gene in the over-expression and the control cell lines was determined. The Student's t-Test determines whether or not the difference in gene expression between the over-expression and control cell lines is statistically significant. Genes that has a p-value less than 0.05 were used for further analysis. The log_2 transformed values of each gene from all the time-points in both the over-expression and the control parasites were summed up. The difference in the sums between the over-expression transfectants and the control parasites was calculated. Genes are considered to be differentially expressed if the Student's t-Test p<0.05, and has at least a 2 fold difference in sums. Hierarchical clustered genes represent the transcriptional changes

elicited by PF10_0083 over-expression. Red represents genes that had higher mRNA levels in the over-expression parasites than the control parasites while green represents genes that had lower mRNA levels in the over-expression parasites compared to control parasites. The T996-myc1, T996-myc2, T996-gfp and 3D7-myc3 over-expression cell lines induced a transcriptional response in 657, 1060, 90, and 85 genes respectively. The number of up-regulated and down-regulated genes in each cell line is shown on the left of each heat map.

T996-myc1 altered the expression in 657 genes, 483 genes were up-regulated and 174 genes were down-regulated. Functional groups involving protein motility and transport were enriched in the up-regulated genes. Functional groups involving protein translation and post-translational modification were enriched in the down-regulated genes (Appendix, Supplementary Figure 2A). The T996-myc2 cell line altered the transcriptional responses in the most number of genes. There were more genes that were down-regulated than up-regulated. Functional groups involved in transcription, host-parasites interaction and immune evasion were enriched in the 425 up-regulated genes. Functional groups involved in DNA replication, 5 zinc finger proteins and 5 CCR4 proteins were enriched amongst the 635 down-regulated genes (Appendix, Supplementary Figure 2B). The T996-gfp cell line altered the transcriptional response in 90 genes. No functional group was statistically significantly enriched amongst the 47 up-regulated genes. Functional groups involved in protein folding were enriched in the 43 down-regulated genes (Appendix, Supplementary Figure 2C). The 3D7-myc3 cell line altered the transcriptional response in the least number of genes. Functional groups involved in protein degradation were enriched in the 56 up-regulated genes. In contrast to T996-myc2, genes involved in host-parasites interactions were enriched in the 29 down-regulated genes (Appendix, Supplementary Figure 2D).

T996-myc1 and 3D7-myc3 transfectants were compared to the untransfected wild type T996/3D7 parasites. The former induced a transcriptional change in 657 genes, while the latter 85 genes. This difference in the number of genes affected could be a variation between experiments, or could be a difference in the strain of the parasites used. Since the 2 over-expression cell lines were compared to wild type T996/3D7 control cells, the transcriptional changes observed in these over-expression cell lines could be due to a selection drug induced response, and not a genuine response from PF10_0083 over-expression. As such, T996-myc2 and T996-gfp transfectants were compared to T996-pARL-myc and T996-pARL-gfp control transfectants, thus eliminating WR99210 drug induced transcriptional changes. T996-gfp induced a transcriptional change in 90 genes, 567 genes less than T996-myc1. However, T996-myc2 induced a transcriptional change in 403 more genes than T996-myc1. In addition, there were more genes that were down-regulated than up-regulated.

We also investigated the common genes that were affected between the microarrays experiments. The different cell lines were compared to one another to identify the common genes that had an altered gene expression (Figure 4.3). The overlaps between 2 out of 4 over-expression cell lines were statistically significant with a p-value less than 0.05, with the exception of the overlap between T996-gfp and 3D7-myc3 cell lines. 10 genes were common between the T996-myc1, T996-myc2, and T996-gfp cell lines. The overlap between the 3 cell lines is statistically significant with a p-value 0.0012. 10 genes were common between T996-myc1, T996-myc2, and 3D7-myc3 cell lines. The overlap between the 3 cell lines is statistically significant with a p-value 8.1 X 10^{-4}. However, no gene was commonly found to be differentially regulated in all the 4 cell lines. Figure 4.3 depicts the heat maps of the overlapping genes between 2 out of

the 4 over-expression cell lines. It was observed that the overlapping genes were not commonly up-regulated or down-regulated in both cell lines. Some genes were up-regulated in 1 cell line, but down-regulated in the other cell line. The only exception is the overlapping gene set between T996-myc1 and 3D7-myc3, where the overlapping gene set was up-regulated.

T996-myc1 and T996-myc2 have the highest number of common differentially expressed genes. 200 genes were common to both cell lines, and the overlap is statistically significant with a p-value 4.4 X 10^{-13}. 28 were up-regulated in both cell lines, but no functional group was enriched in this gene group. 25 genes were down-regulated in both cell lines, and genes involved in host-parasite interactions and replication were enriched. 141 genes were up-regulated in T996-myc1, but down-regulated in T996-myc2, and genes involved in motility were enriched. 6 genes were up-regulated in T996-myc2, but down-regulated in T996-myc1 (Appendix, Supplementary Figure 3A). There were 16 genes that were differentially expressed in both T996-myc1 and 3D7-myc3. 15 genes were up-regulated in the 2 cell lines, and genes involved in motility, folate biosynthesis, and Maurer's clefts proteins were enriched. Only 1 gene was down-regulated in both cell lines, and it is a phosphatidylinositol 4-kinase (Appendix, Supplementary Figure 3B).

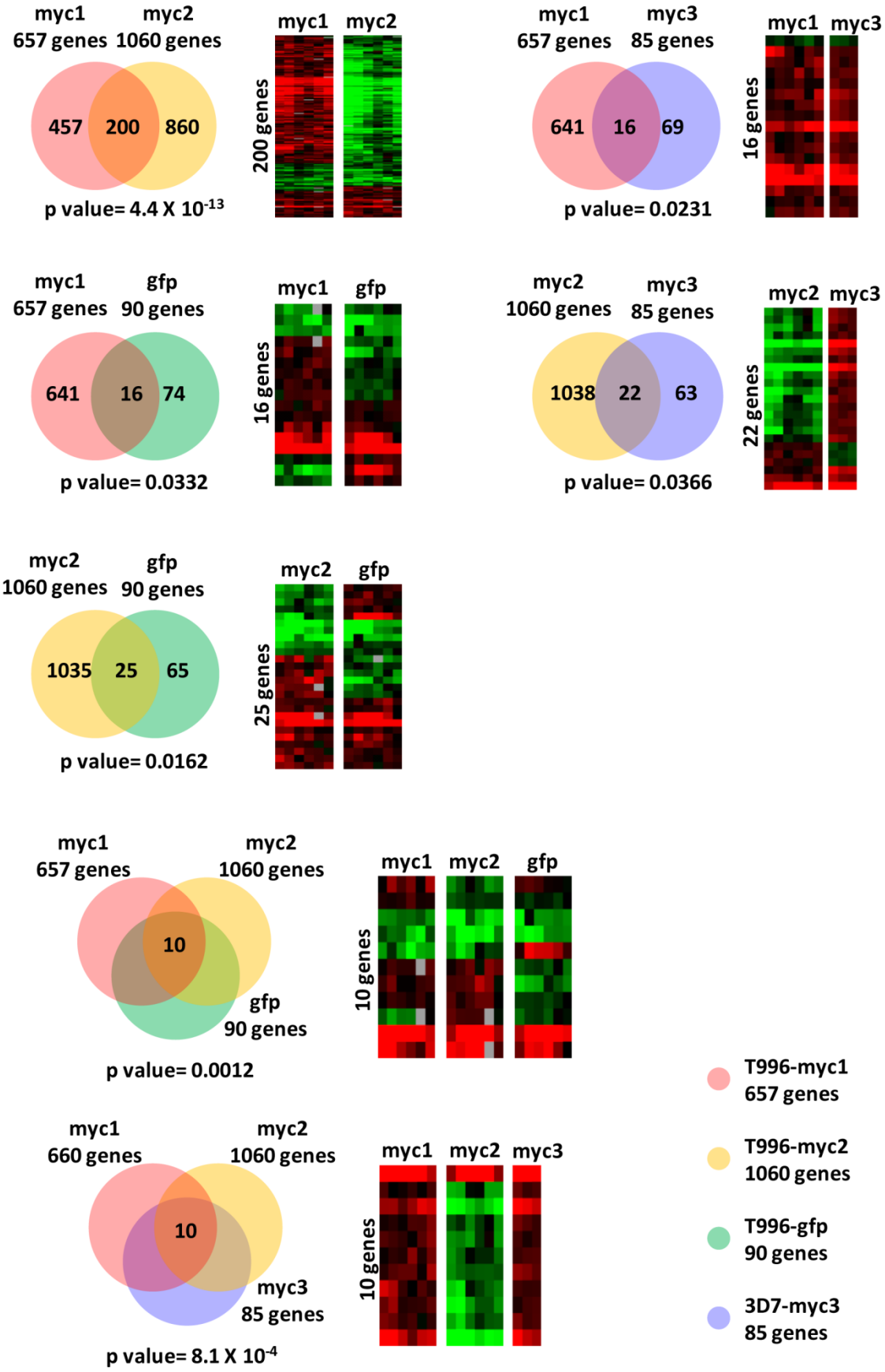


Figure 4.3 Venn diagrams illustrating the number of genes that were differentially expressed in 2 out of 4, or 3 out of 4 over-expression cell lines.

The overlapping genes between the over-expression cell lines that were statistically significant with a p-value less than 0.05 were shown. The heat maps illustrate the hierarchical clustered genes common between 2 out of 4, or 3 out of 4 over-expression cell lines. These genes were differentially expressed with a p-value <0.05 and at least a 2 fold difference in sums. Red represents genes that had higher mRNA levels in the over-expression parasites than control parasites while green represents genes that had lower mRNA levels in the over-expression parasites than the control parasites. myc1: T996-pARL-PF10_0083-myc1, myc2: T996-pARL-PF10_0083-myc2, gfp: T996-pARL-PF10_0083-gfp, myc3: 3D7-pARL-PF10_0083-myc3.

Taken together, the over-expression of PF10_0083 elicited a transcriptional response in distinct set of genes. However, it was observed that the number of genes that were transcriptionally altered in the 4 over-expression cell lines differed greatly from each other (Figure 4.2). In addition, it was observed that different sets of genes belonging to different functional groups were transcriptionally altered in the 4 over-expression cell lines (Appendix, Supplementary Figure 2). Furthermore, the same transcriptionally altered gene was up-regulated in 1 over-expression cell line but down-regulated in another over-expression cell line (Figure 3 and Appendix, Supplementary Figure 3). In summary, there is a great heterogeneity in the differentially expressed genes in the 4 over-expression cell lines. Such heterogeneity could be attributed to several reasons. First, different parasite populations (clones) could be selected during the different transfection experiments. Second, it is possible that the level of PF10_0083 in the 4 over-expression cell lines were different, which may be due to a difference in the plasmid copy number. Third, the criteria used for selecting genes that are differentially expressed was Student's t-Test $p<0.05$ between the over-expression cell line and control cell line, and has at least a 2 fold difference in sums. This criteria used might be too stringent, and as such, there were few differentially expressed genes that were common between the over-expression cell lines (Figure 4.3). Lastly, the transcriptional

changes observed in the 4 over-expression cell lines could be a general de-regulation as a result of PF10_0083 transgene over-expression. As observed in the over-expression, endogenous-tagged, knock-out, and control cell lines, over-expression of PF10_0083 is associated with a lower growth rate of the parasites (refer to section 3.3.1, Figure 3.3). However, in the knock-out and endogenous-tagged cell lines, we did not observe any impaired growth rates in the transfected cell lines (refer to section 3.5.1, Figure 3.15 and 3.16). As such, we also carried out transcriptional profiling in the endogenous-tagged cell line, PF10_0083-CHA18, and the knock-out cell line, PF10_0083-KO33.

3 samples were harvested every 12 hr during the IDC, and they represent ring, trophozoite, and schizont stage parasites. Wild-type 3D7 was used as the control cell line. The microarray data was analyzed by calculating the Student's t-Test and the difference in sums for every gene. Genes that have a p-value less than 0.05 and at least a 2 fold difference in sums were considered as differentially expressed, and were used for further analysis. Initial data analyses indicated that the endogenous tagging of PF10_0083 with mCherry-HA caused 13 genes to be differentially expressed in PF10_0083-CHA18 (Figure 4.4A). Out of these 13 genes, 2 genes were up-regulated and 11 genes were down-regulated. Functional group relating to the merozoite motility was enriched in the down-regulated gene group. The knock-out of PF10_0083 caused altered transcriptional response in 32 genes PF10_0083-KO33 (Figure 4.4B). Out of these 32 genes, 9 genes were up-regulated and 23 genes were down-regulated. In the up-regulated gene group, 2 CCR4 proteins were enriched. In the down-regulated gene group, functional group relating to the increased permeability of the host cell membrane was enriched. There were fewer genes that were transcriptionally altered in

the endogenous-tagged and knock-out cell lines compared to the over-expression cell lines. Hence, the transcriptional changes observed in the 4 over-expression cell lines could be a general de-regulation as a result of transgene over-expression.

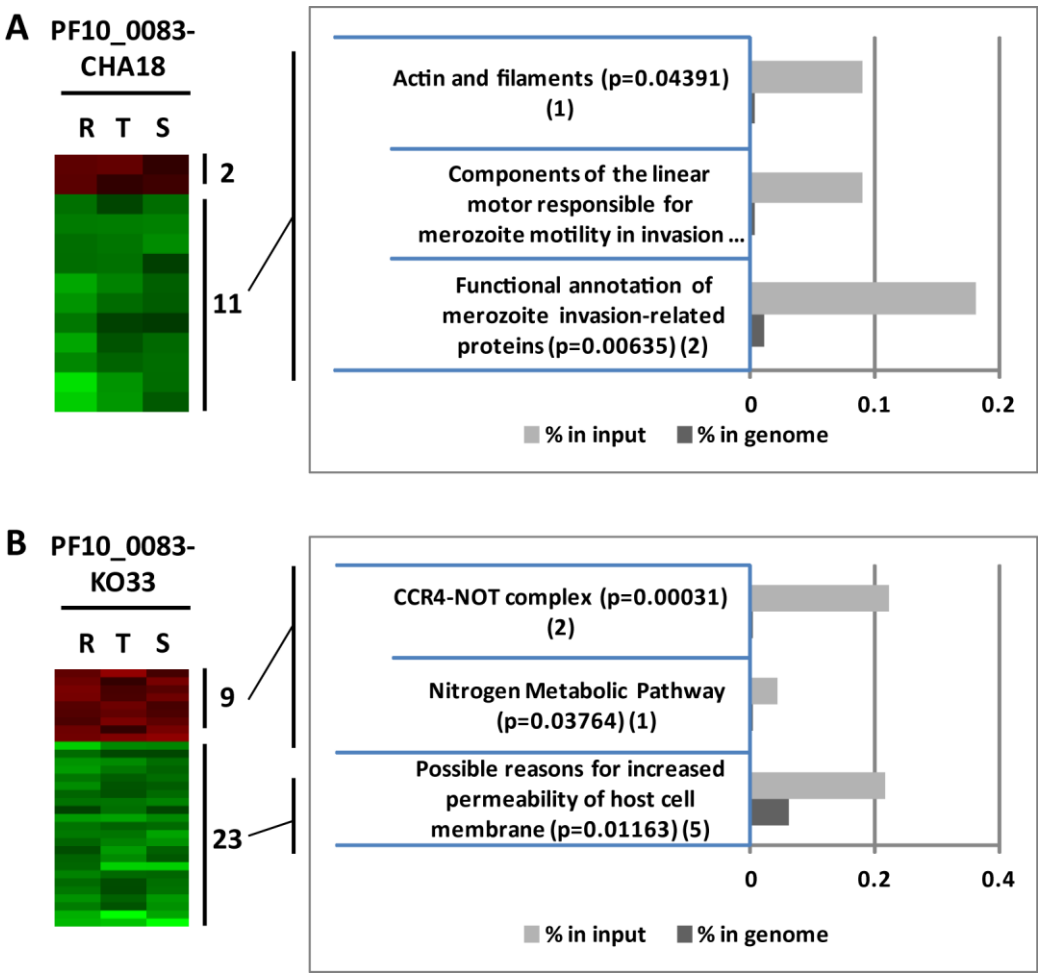


Figure 4.4 Transcriptional changes elicited by PF10_0083-CHA18 and PF10_0083-KO33. The Student's t-Test for every gene between wild type 3D7 and PF10_0083-CHA18, and between wild type 3D7 and PF10_0083-KO33 cell lines was determined. The Student's t-Test determines whether or not the difference in gene expression between the transfectants and control cell lines is statistically significant. Genes that has a p-value less than 0.05 were used for further analysis. The $\log_2$ transformed values of each gene from all the time-points in both the transfectants and the control parasites were summed up. The difference in the sums between the transfectants and the control parasites was calculated. Genes are considered to be differentially expressed if the Student's t-Test $p<0.05$, and have at least a 2 fold difference in sums. Hierarchical clustered genes of (A) PF10_0083-CHA18 and (B) PF10_0083-KO33. Red and green represents genes that had

higher and lower mRNA levels in the transfected parasites compared to control parasites. Bar graphs depict the functional enrichment (MPM: Malaria Parasites Metabolic Pathways) of genes that were transcriptionally altered. The light grey bars represent the pathway related genes in the transfectants, and the dark grey bars represent the pathway related genes in the genome. The number of genes in each pathway is shown in brackets. All pathways have a binomial p-value less than 0.05.

4.2.2 Transcriptional changes in stage specific genes

In the 4 over-expression cell lines, it was observed that a high percentage of the transcriptionally altered genes are maximally expressed during the gametocyte stage in the parasite life cycle (Figure 4.5). In T996-myc1, T996-gfp and 3D7-myc3, there were 150, 10 and 27 up-regulated genes that are maximally expressed in the gametocytes in the parasite life cycle. On the other hand, in T996-myc1, T996-gfp and 3D7-myc3, there were only 16, 6 and 2 down-regulated genes that are maximally expressed during the gametocyte stage in the parasite life cycle. In contrast, in the T996-myc1, T996-gfp and 3D7-myc3 cell lines, there were 51, 10 and 11 down-regulated genes that are most highly expressed in the early rings in the parasite life cycle. Taken together, in the T996-myc1, T996-gfp and 3D7-myc3 cell lines, the majority of the up-regulated genes are most highly expressed in the gametocytes, while the majority of the down-regulated genes are most highly expressed in the early rings. However, an exception is the T996-myc2 where the majority of the down-regulated genes are most highly expressed in the gametocytes and the majority of the up-regulated genes most highly expressed in the early rings.

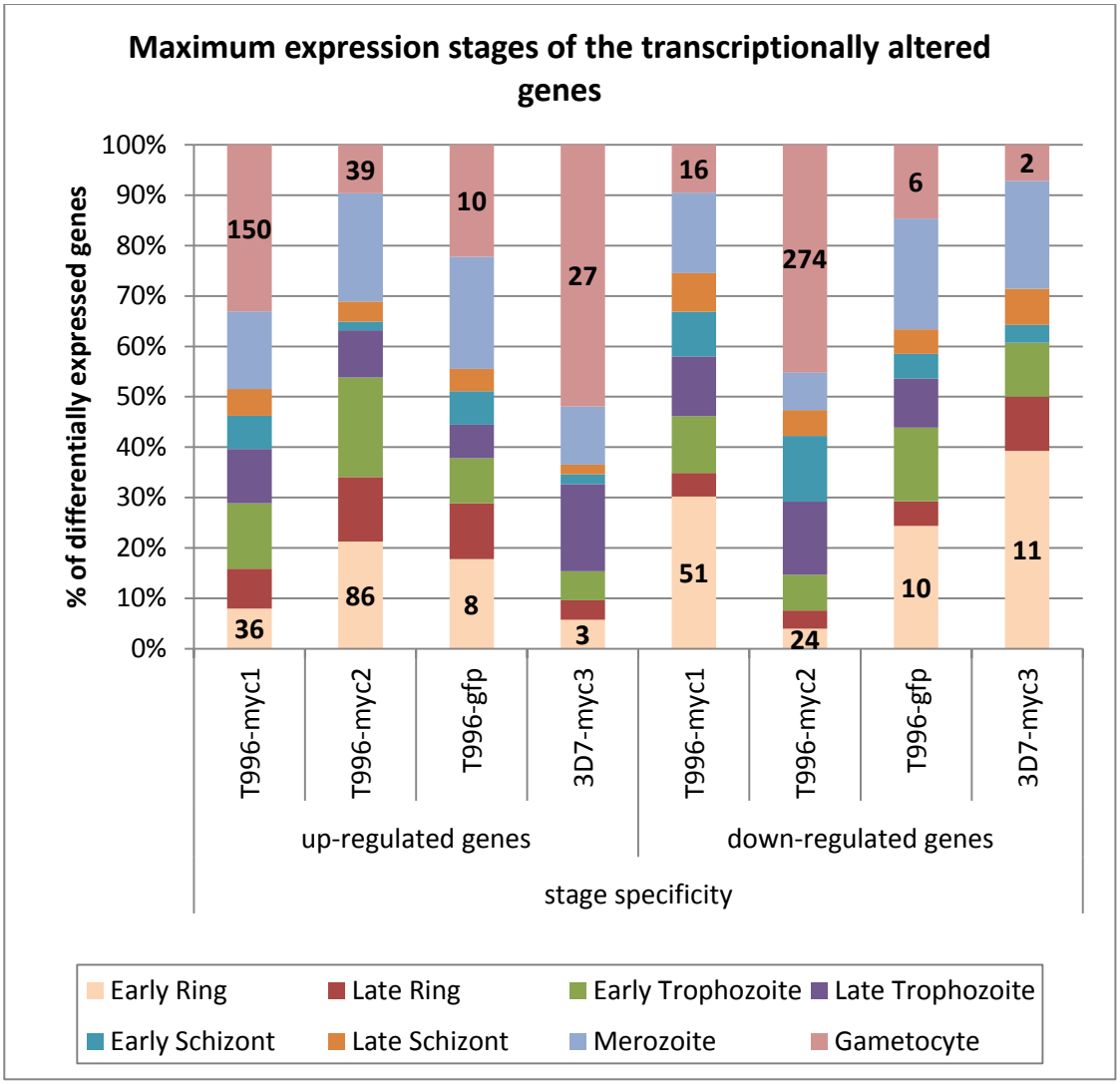


Figure 4.5 PF10_0083 over-expression affected genes classified according to their stage specificity.

Distribution of the up-regulated and down-regulated genes in the 4 over-expression cell lines, according to their normal expression profiles in the parasite life cycle. The stage specificity of each differentially expressed gene is defined as the stage during the parasite life cycle by which the gene is maximally expressed. The differentially expressed genes were then classified according to their stage specificity and plotted into this graph. T996-myc1: T996-pARL-PF10_0083-myc1, T996-myc2: T996-pARL-PF10_0083-myc2, T996-gfp: T996-pARL-PF10_0083-gfp, 3D7-myc3: 3D7-pARL-PF10_0083-myc3.

There are 10 differentially expressed genes that were common in the T996-myc1, T996-myc2, and 3D7-myc3 cell lines. All the 10 genes are known to be maximally expressed in the gametocyte and/or ookinete stages in the usual life cycle. 8 genes are

most highly expressed in the gametocytes, 1 gene in both the gametocytes and ookinete, and 1 gene in ookinete. There are another 10 differentially expressed genes that were common in T996-myc1, T996-myc2, and T996-gfp cell lines. Out of these 10 genes, 4 genes are most highly expressed in the gametocytes and 1 gene in both the ookinete and ring stages. In addition, 18 genes that were transcriptionally altered in 2 out of the 4 cell lines had been identified. 17 genes are most highly expressed in the gametocytes and 1 gene in the ookinete stage. Table 4.2 is a list of these 33 genes, and their expression levels in the 4 over-expression cell lines. Genes that were up-regulated more than 2 fold are highlighted in red, while genes that were down-regulated more than 2 fold are highlighted in green.

Out of the 33 genes (Table 4.2), 24 genes were up-regulated and 2 genes were down-regulated in T996-myc1 cell line. The sexual stage-specific protein precursor (Pfs16), gamete antigen 27/25 (Pfg27), male development gene 1 (MDV1), alpha tubulin 2, and oocyst capsule protein (Cap380) were amongst the top 10% of the 483 up-regulated genes in T996-myc1 cell line. Pfs16 and Pfg27 were previously described to be important in the early stage of gametocytogenesis [37, 38]. In addition, PF14_0588 and PF14_0708 were also found to be amongst the top 5% of the up-regulated genes. Both PF14_0588 and PF14_0708 are conserved proteins with unknown functions. They were previously identified in a proteomic study of the early stages gametocytes [277]. Out of the 33 genes, 15 genes were observed to be up-regulated and no gene was down-regulated in 3D7-myc3 cell line. 10 of these genes were also up-regulated in T996-myc1. Pfs16, alpha tubulin 2, and early transcribed membrane protein 8 (ETRAMP8), were amongst the top 10% of the 56 up-regulated genes in 3D7-myc3 cell line. Out of the 33 genes, only 3 genes were observed to be up-regulated and 4

genes were down-regulated in T996-gfp. Interestingly, the same genes that were up-regulated in these 3 cell lines were found to be down-regulated in T996-myc2. Specifically, in T996-myc2, 29 genes were down-regulated and only 3 genes were up-regulated out of the 33 genes. RESA-like protein with PHIST and DnaJ domains (PFB0090c) and another conserved protein were the only 2 genes that were up-regulated in both T996-myc1 and T996-myc2. In PF10_0083-CHA18, none of these genes show a transcriptional change greater than 2 fold. In PF10_0083-KO33, only 6-cysteine protein (P230) was observed to be up-regulated out of the 33 genes. As such, the up-regulation of these gametocyte and ookinete specific genes was possibly a result of the over-expression of PF10_0083.

It is interesting that this particular group of gametocyte and ookinete genes were up-regulated in the T996-myc1, T996-gfp and 3D7-myc3 over-expression cell lines. Although these gametocytes-related genes were up-regulated, no gametocytes were observed during the course of culturing. It is important to note that all cultures were maintained with care to ensure that there were enough nutrients and the parasites were not stressed. However, one also should take note that the early stage gametocytes are morphologically indistinguishable from the asexual trophozoite stage. As such, one cannot rule out the possibility of the presence of early stage gametocytes in the cultures.

Stage specificity	Gene name	Gene ID	T996-myc1	T996-myc2	T996-gfp	3D7-myc3
gametocyte	protein kinase, putative	MAL13P1.84	2.85	-5.72		2.01
gametocyte	early transcribed membrane protein 8 (ETRAMP8)	MAL8P1.6	3.82	-4.46		2.81
gametocyte	secreted ookinete protein, putative (PSOP1)	PF07_0089	3.66	-8.28		1.30
gametocyte	mechanosensitive ion channel protein	PF11_0092	1.63	-5.71		1.81
gametocyte	sexual stage-specific protein precursor (Pfs16)	PFD0310w	11.37	-18.53		5.54
gametocyte	alpha tubulin 2	PFD1050w	8.09	-13.29		6.60
gametocyte	hydrolase/phosphatase, putative	PFL1260w	4.67	-6.18		1.31
gametocyte	mitochondrial ribosomal protein L29/L47 precursor, putative	PFC0675c	2.67	-3.48		1.89
gametocyte + ookinete	leucine-rich repeat protein (LRR2)	PFE0455w	2.89	-3.42		1.62
okineete	nucleic acid binding protein, putative	PF11_0332	3.24	-4.91		1.16
gametocyte	biotin carboxylase subunit of acetyl CoA carboxylase, putative (ACC1)	PF14_0664	1.48	-2.22	-1.14	
gametocyte	conserved Plasmodium protein, unknown function	PFL1695c	3.60	-3.56	1.57	
gametocyte	conserved Plasmodium protein, unknown function	PFI1460c	-3.63	2.17	-2.00	
gametocyte	gametocyte erythrocyte cytosolic protein (GECO)	PFL2550w	-4.67	-9.60	-5.74	
ookinete + rings	RESA-like protein with PHIST and DnaJ domains	PFB0090c	1.98	3.59	-5.70	
gametocyte	conserved Plasmodium protein, unknown function	PFC0416w	1.54	1.10		
gametocyte	conserved Plasmodium protein, unknown function	PF14_0588	12.35	-14.17		
gametocyte	probable protein, unknown function	PF14_0708	11.23	-16.57		
gametocyte	male development gene 1 (MDV1)	PFL0795c	9.11	-16.90		
gametocyte	gamete antigen 27/25 (Pfg27)	PF13_0011	5.72	-19.00		
gametocyte	6-cysteine protein (P48/45)	PF13_0247	4.70	-9.11		
gametocyte	6-cysteine protein (P230)	PFB0405w	4.23	-5.00		

gametocyte	secreted ookinete adhesive protein (SOAP)	PF14_0040	3.91	-6.64		
gametocyte	secreted ookinete protein, putative (PSOP12)	PFE0680w	3.42	-6.06		
ookinete	oocyst capsule protein (Cap380)	PFC0905c	5.57	-4.42		
gametocyte	developmental protein, putative	PFI0300w	2.28			
gametocyte	procollagen lysine 5-dioxygenase, putative	PFI1465w		-2.62	1.58	
gametocyte	metallopeptidase, putative	PF10_0092		-3.80	1.10	
gametocyte	25 kDa ookinete surface antigen precursor (Pfs25)	PF10_0303		-2.74		1.26
gametocyte	Plasmodium exported protein, unknown function	PF14_0736		-1.45		2.19
gametocyte	CCR4-NOT transcription complex subunit 5, putative (NOT5)	PF10_0062		-3.29		1.76
gametocyte	conserved Plasmodium protein, unknown function	PF08_0080		-3.33		1.23
gametocyte	membrane skeletal protein IMC1-related	PF08_0033		-4.05		1.42

Table 4.2 List of gametocyte stage specific genes that were differentially expressed in at least 2 over-expression cell lines.

Genes that were transcriptionally altered in at least 2 out of the 4 over-expression cell lines. These genes were classified according to their stage specificity, which is defined as the stage during the parasite life cycle by which the gene is maximally expressed. The numbers represent the $\log_2$ ratio of the over-expression cell line to the control cell line. Genes with minimum 2 fold up-regulation are highlighted in red and genes with minimum 2 fold down-regulation are highlighted in green. T996-myc1: T996-pARL-PF10_0083-myc1, T996-myc2: T996-pARL-PF10_0083-myc2, T996-gfp: T996-pARL-PF10_0083-gfp, 3D7-myc3: 3D7-pARL-PF10_0083-myc3.

In summary, the 4 over-expression cell lines of PF10_0083 elicited a transcriptional response in distinct sets of genes. However, there is heterogeneity in the differentially expressed genes in the 4 over-expression cell lines. Such heterogeneity could arise from the selection of different parasite populations (clones) during the different transfection experiments, different level of PF10_0083 (difference in the plasmid copy

number) in the 4 over-expression cell lines, the criteria used for selecting genes that are differentially expressed was too stringent, or could be a general de-regulation as a result of transgene over-expression. Lastly, it was observed that the majority of the genes that were up-regulated in the T996-myc1, T996-gfp and 3D7-myc3 cell lines were maximally expressed in the gametocytes in the parasite life cycle.

4.3 Genome-wide analysis of PF10_0083 binding

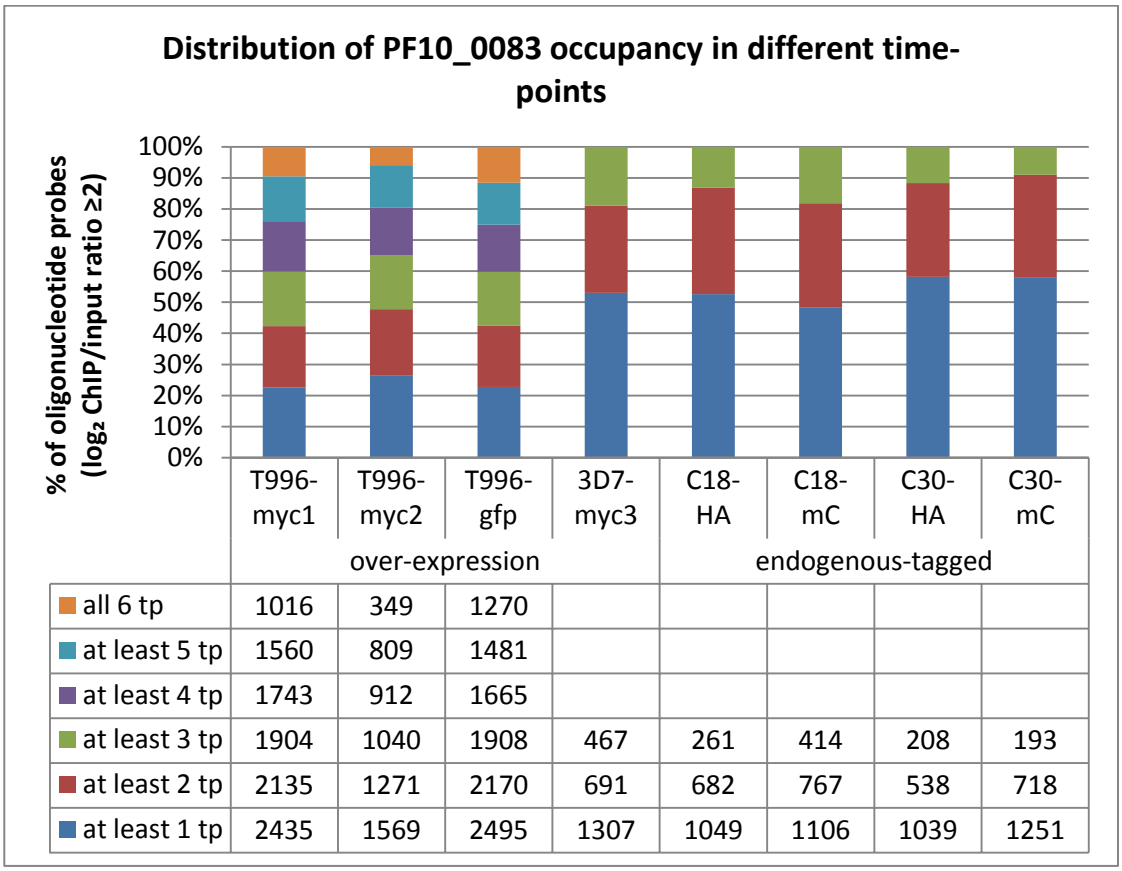
In order to study PF10_0083 occupancy throughout the genome, ChIP-on-chip experiments were carried out on the over-expression cell lines and the endogenous-tagged cell lines. The main goal was to identify DNA regions that directly interact with PF10_0083. Cross-linked parasites samples were harvested alongside the RNA samples harvested for transcriptional profiling. Samples were harvested every 8 or 12 hours for the entire IDC (44-48 hr). For T996-myc1, T996-myc2 and T996-gfp cell lines, time-points 1 and 2 are ring stages, time-points 3 and 4 are trophozoite stages, and time-points 5 and 6 are schizont stages. For 3D7-myc3, PF10_0083-CHA18 (hereafter referred to as CHA18) and PF10_0083-CHA30 (hereafter referred to as CHA30) cell lines, time-points 1, 2 and 3 represent ring, trophozoite and schizont stages respectively. Chromatin immunoprecipitation was performed on the 4 over-expression cell lines using the anti-myc or anti-gfp antibodies. For CHA18 and CHA30, experiments were performed with anti-HA and anti-mcherry antibodies. ChIP-on-chip experiments carried out on CHA18 with anti-HA are denoted as C18-HA, while experiments carried out with anti-mcherry are denoted as C18-mC. ChIP-on-chip experiments carried out on CHA30 are denoted as C30-HA and C30-mC. For every sample, DNA amplification and microarrays were carried out in triplicates. The microarray chips used contain 10,416 oligonucleotides probes representing the ORFs,

and 5,402 oligonucleotide probes representing the upstream intergenic regions (IGRs). Only oligonucleotide probes with data present in at least 2 out of the 3 triplicates were included for analysis. The data from the triplicates were averaged and used for further analysis. Oligonucleotide probes with a positive ChIP signal at least 2 fold higher than the input signal ($\log_2$ ChIP/input ratios $\geq$2) were used for further analysis.

4.3.1 PF10_0083 genome occupancy during IDC

The association of PF10_0083 with the genome in the 4 over-expression cell lines was first analyzed. Figure 4.6 depicts the number of genetic loci (represented by oligonucleotide probes with $\log_2$ ChIP/input ratios $\geq$2) that was associated with PF10_0083 in the 4 over-expression cell lines. On average, in the over-expression cell lines, PF10_0083 associated with 1951 genetic loci in at least 1 time-point. PF10_0083 associated with 40% of these genetic loci in all time-points. However, since PF10_0083 was over-expressed in these cell lines, some of these binding sites observed might be an artifact from the over-expression. As such, the genome occupancy of PF10_0083 in the endogenous-tagged cell lines was also investigated. In the endogenous-tagged cell lines, PF10_0083 was observed to be associated with 1111 genetic loci in at least 1 time-point. PF10_0083 associated with 60% of these genetic loci in at least 2 out of 3 time-points. In addition, the temporal binding pattern of PF10_0083 during the IDC was also investigated. Figure 4.7 illustrates the genome occupancy of PF10_0083 in the ring, trophozoite and schizont stages. To our surprise, we observed that the PF10_0083 associated with more genomic loci during the ring stage than in the trophozoite and schizont stages in the endogenous-tagged cell lines, even thought we were unable to detect the protein on western blot and IFA analysis. In both the over-expression and endogenous-tagged cell lines, we observed that

PF10_0083 associated with more genomic loci during the ring stage than in the trophozoite and schizont stages. PF10_0083 was associated with the least number of genetic loci in the schizont stage. Taken together, these results suggested that when PF10_0083 binds strongly to a region in the genome, it is also likely to bind in multiple time-points.



	T996-myc1	T996-myc2	T996-gfp	3D7-myc3	C18-HA	C18-mC	C30-HA	C30-mC
all 6 tp	1016	349	1270					
at least 5 tp	1560	809	1481					
at least 4 tp	1743	912	1665					
at least 3 tp	1904	1040	1908	467	261	414	208	193
at least 2 tp	2135	1271	2170	691	682	767	538	718
at least 1 tp	2435	1569	2495	1307	1049	1106	1039	1251

Figure 4.6 Association of PF10_0083 with the genome.

Chromatin immnoprecipitation was carried out on 6 time-point samples for T996-myc1, T996-myc2 and T996-gfp cell lines, and 3 time-point samples for 3D7-myc3, CHA18 and CHA30 cell lines. DNA amplification and microarrays were carried out in triplicates. Only oligonucleotide probes with data present in at least 2 out of the 3 triplicates were included for analysis. The data from the triplicates were averaged and used for further analysis. Oligonucleotide probes with a positive ChIP signal at least 2 fold higher than the input signal (log_2 ChIP/input ratios ≥2) were used for further analysis. The bar graphs depict the distribution of genetic loci (represented by oligonucleotide probes with log_2 ChIP/input ratios ≥2) associated with PF10_0083 in different number of time-points,

in the over-expression and endogenous-tagged cell lines. The table shows the number of genetic loci associated with PF10_0083 in different number of time-points. tp: time-point, T996-myc1: T996-pARL-PF10_0083-myc1, T996-myc2: T996-pARL-PF10_0083-myc2, T996-gfp: T996-pARL-PF10_0083-gfp, 3D7-myc3: 3D7-pARL-PF10_0083-myc3, C18: CHA18, C30: CHA30, HA: anti-HA, mC: anti-mCherry.

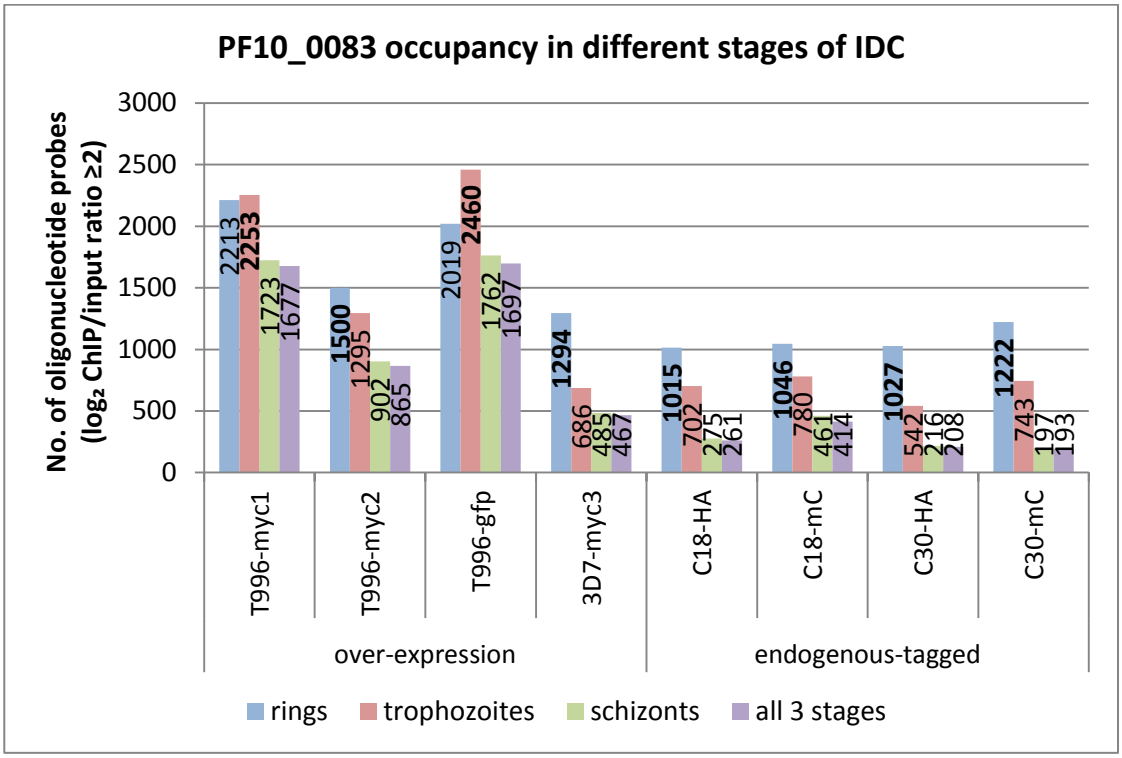
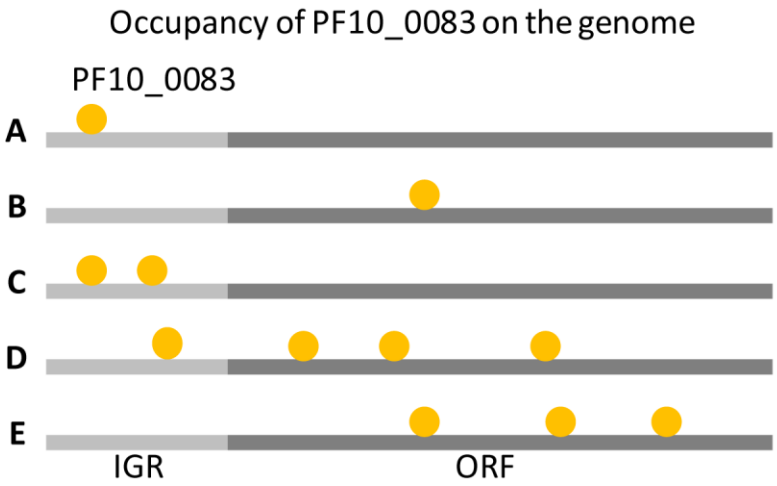


Figure 4.7 Association of PF10_0083 with the genome in different stages of IDC.
DNA amplification and microarrays were carried out in triplicates. Only oligonucleotide probes with data present in at least 2 out of the 3 triplicates were included for analysis. The data from the triplicates were averaged and used for further analysis. Oligonucleotide probes with a positive ChIP signal at least 2 fold higher than the input signal (log_2 ChIP/input ratios ≥2) were used for further analysis. The bar graphs depict the number of genetic loci (represented by oligonucleotide probes with log_2 ChIP/input ratios ≥2) associated with PF10_0083 in ring, trophozoites, and schizont stages in the over-expression and endogenous tagged cell lines. The stages with the most number of PF10_0083 associated genetic loci are highlighted in bold. T996-myc1: T996-pARL-PF10_0083-myc1, T996-myc2: T996-pARL-PF10_0083-myc2, T996-gfp: T996-pARL-PF10_0083-gfp, 3D7-myc3: 3D7-pARL-PF10_0083-myc3, C18: CHA18, C30: CHA30, HA: anti-HA, mC: anti-mCherry.

4.3.2 PF10_0083 occupancy patterns on different genomic regions

T996-myc1, T996-myc2, T996-gfp and 3D7-myc3 were associated with 2435, 1569, 2495 and 1307 genetic loci in at least 1 time-point. These genetic loci represented 2031, 1398, 2068 and 1177 genes in T996-myc1, T996-myc2, T996-gfp and 3D7-myc3 respectively. C18-HA, C18-mC, C30-HA and C30-mC were associated with 1049, 1106, 1039 and 1251 genetic loci in at least 1 time-point. These probes represented 965, 1011, 960 and 1132 genes in C18-HA, C18-mC, C30-HA and C30-mC respectively. This data suggests that for a subset of genes, PF10_0083 binds to more than one region on a single gene. As such, the binding pattern of PF10_0083 with respect to its distribution over the genetic loci was investigated (Figure 4.8). The binding pattern of PF10_0083 is categorized into 5 types. PF10_0083 was observed to bind to (A) only a single region on the gene's IGR, (B) only to a single region on the gene's ORF, (C) more than 1 region on the IGR, (D) both the IGR and ORF, and (E) more than 1 region on the ORF.

Figure 4.9 illustrates the number of genes associated with PF10_0083 through 1 binding site or through multiple binding sites on the IGRs and/or ORFs. On average, in the over-expression and endogenous tagged cell lines, 89% of the genes were associated with PF10_0083 through a single region on either the IGR or the ORF. 1% of the genes were associated with PF10_0083 through multiple regions on the IGR. 3% of the genes were associated with PF10_0083 through multiple regions on the ORF. 7% of the genes were associated with PF10_0083 through multiple regions on both the IGR and ORF.



	T996-myc1	T996-myc2	T996-gfp	3D7-myc3	C18-HA	C18-mC	C30-HA	C30-mC
A	32.1%	53.4%	33.5%	51.6%	46.8%	45.7%	47.2%	44.7%
B	50.9%	36.1%	49.5%	38.2%	44.8%	45.3%	44.7%	45.2%
C	0.4%	0.9%	0.6%	1.8%	1.5%	1.4%	1.3%	1.2%
D	10.4%	6.2%	11%	6.9%	5%	5.2%	5%	6.4%
E	6.2%	3.4%	5.5%	1.6%	2%	2.4%	2%	2.5%

Figure 4.8 The occupancy pattern of PF10_0083 on the genome.

(A) PF10_0083 binds only a single region on the gene's intergenic regions (IGR), (B) PF10_0083 binds only to a single region on the gene's open reading frames (ORF), (C) PF10_0083 binds to more than 1 region on the IGR, (D) PF10_0083 binds to both the IGR and ORF, (E) PF10_0083 binds to more than 1 region on the ORF. The percentages of genes in the over-expression and endogenous-tagged cell lines that fall into the different categories are shown at the bottom. T996-myc1: T996-pARL-PF10_0083-myc1, T996-myc2: T996-pARL-PF10_0083-myc2, T996-gfp: T996-pARL-PF10_0083-gfp, 3D7-myc3: 3D7-pARL-PF10_0083-myc3, C18: CHA18, C30: CHA30, HA: anti-HA, mC: anti-mCherry.

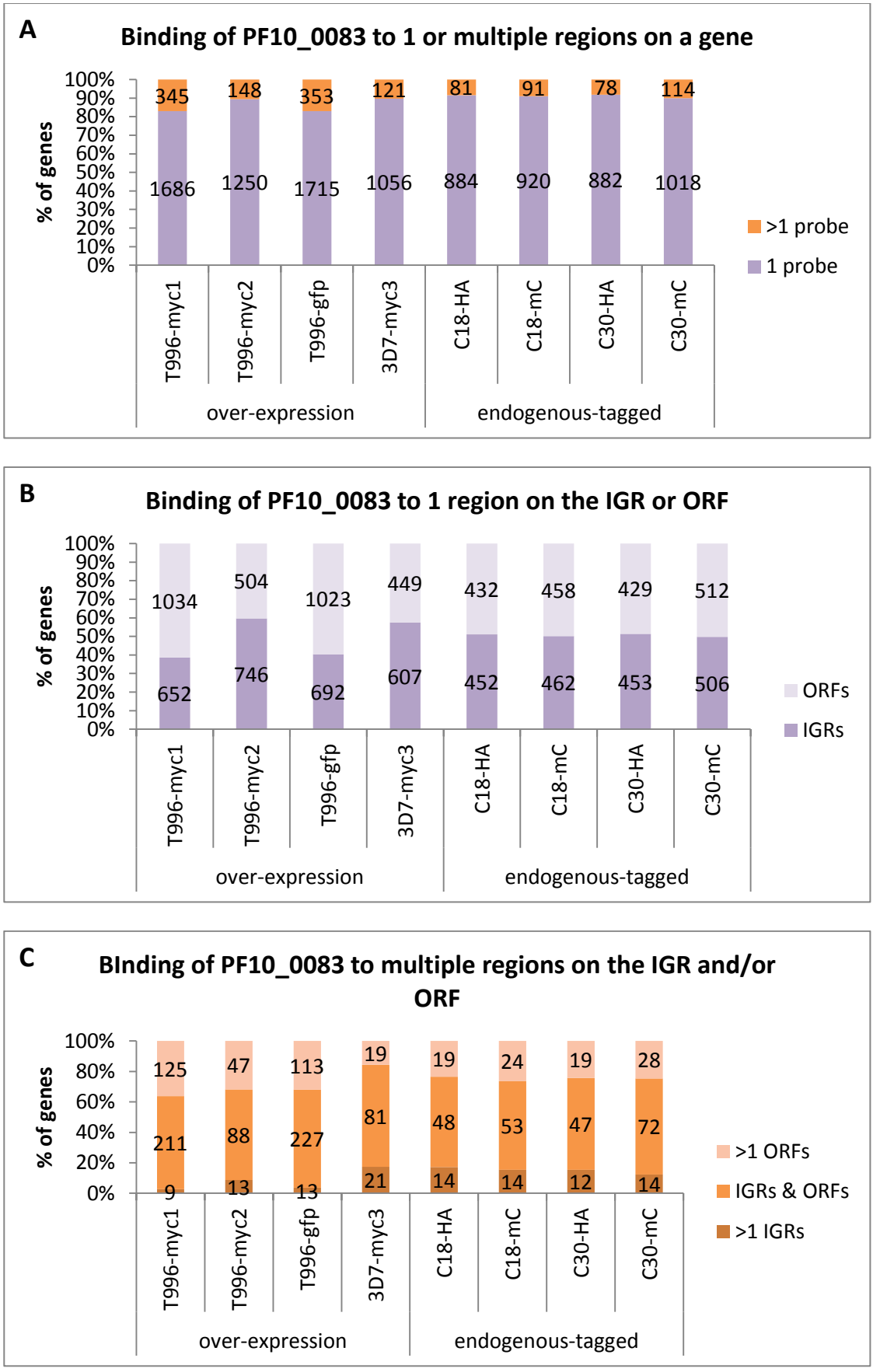


Figure 4.9 Association of PF10_0083 with different genomic regions.

DNA amplification and microarrays were carried out in triplicates. Only oligonucleotide probes with data present in at least 2 out of the 3 triplicates were included for analysis. The data from the triplicates were averaged and used for further analysis. Oligonucleotide probes with a positive ChIP signal at least 2 fold higher than the input signal (log_2 ChIP/input ratios ≥2) were used for further analysis. (A) The number of genes that were associated with PF10_0083 through a single region or multiple regions of a gene. (B) The number of genes that were associated with PF10_0083 through a single region on the IGR or ORF of a gene. (C) The number of genes that were associated with PF10_0083 through multiple regions on the IGR and/or ORF of a gene. T996-myc1: T996-pARL-PF10_0083-myc1, T996-myc2: T996-pARL-PF10_0083-myc2, T996-gfp: T996-pARL-PF10_0083-gfp, 3D7-myc3: 3D7-pARL-PF10_0083-myc3, C18: CHA18, C30: CHA30, HA: anti-HA, mC: anti-mCherry.

4.3.3 Positional enrichment of PF10_0083

In the next step, the binding positions of PF10_0083 along *P. falciparum* genes were analyzed to investigate the preferred binding positions of PF10_0083 along the genome. The microarray chips used contain 15,818 oligonucleotide probes. There is different number of oligonucleotide probes designed for different regions on the genome. As such, the number of genetic loci that was associated with PF10_0083 on different regions along the genes was normalized to the number of probes which represent that particular region on the genome. The distribution of the PF10_0083 binding positions was plotted. It was observed that there was a biased distribution of PF10_0083 binding in the IGRs in both the over-expression and endogenous tagged cell lines. There was more association of PF10_0083 with the genetic loci at the IGRs and a gradual decrease towards the ORFs (Figure 4.10). In the T996-myc1 and T996-gfp over-expression cell lines, PF10_0083 associated with 25% to 30% of the genetic loci at position -1500 bp to -5000 bp in the IGRs. In both cell lines, only approximately 15% of the genetic loci at the ORFs regions were associated with PF10_0083. In T996-myc2, PF10_0083 associated with 18% of genetic loci at position

-500 bp to -5000 bp in the IGRs. In 3D7-myc3, PF10_0083 associated with 16% of genetic loci at position -1 bp to -1500 bp. In both T996-myc2 and 3D7-myc3, PF10_0083 associated with only 7% of the genetic loci at the ORFs (Figure 4.10A). In the endogenous-tagged cell lines, C18-HA, C18-mC, C30-HA and C30-mC, PF10_0083 associated with approximately 25% of the genetic loci at position -1500 bp to -5000 bp (Figure 4.10B). The positional enrichment of PF10_0083 along the genome in rings, trophozoites and schizonts was also analyzed. It was observed that in all the over-expression and endogenous-tagged cell lines, PF10_0083 associated with the highest number of genetic loci at the IGRs in all the 3 stages (Figure 4.11). In summary, PF10_0083 preferentially binds to the IGRs than the ORFs in all the 3 stages in both the over-expression and endogenous-tagged cell lines.

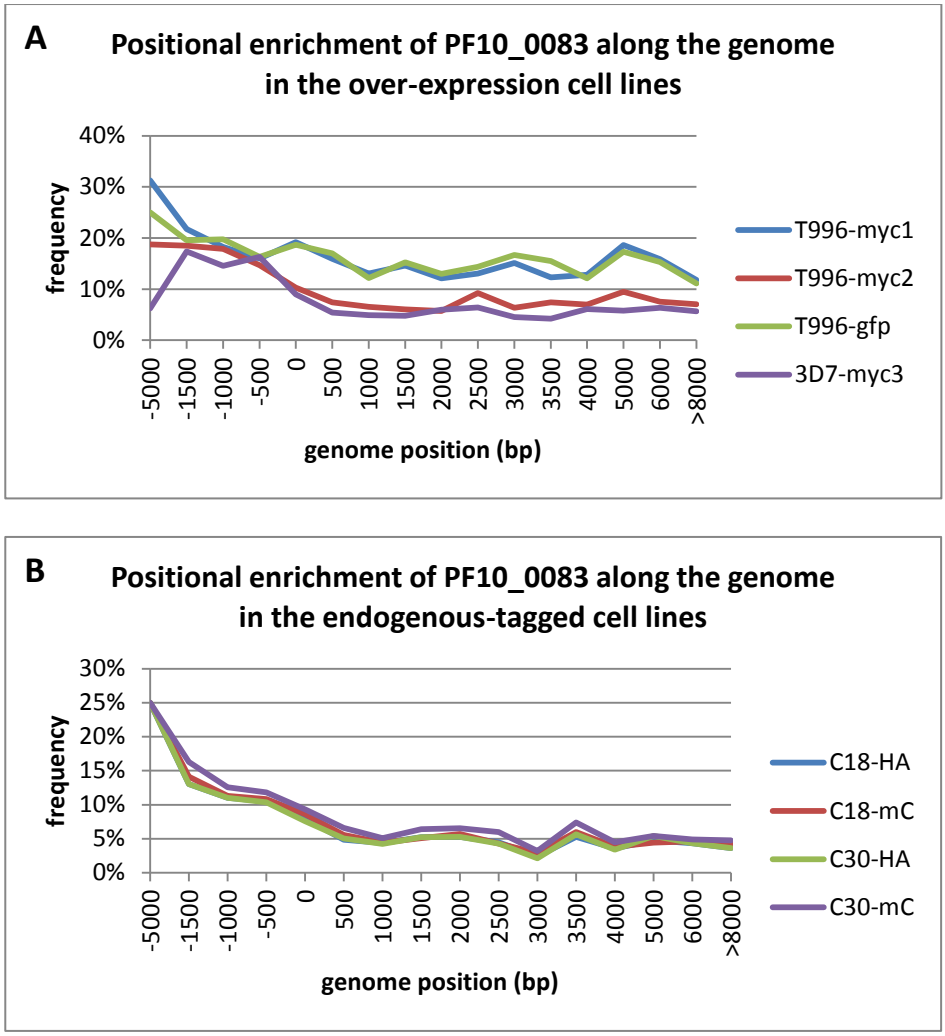


Figure 4.10 The positional enrichment of PF10_0083 along the genome.

DNA amplification and microarrays were carried out in triplicates. Only oligonucleotide probes with data present in at least 2 out of the 3 triplicates were included for analysis. The data from the triplicates were averaged and used for further analysis. Oligonucleotide probes with a positive ChIP signal at least 2 fold higher than the input signal ($\log_2$ ChIP/input ratios ≥2) were used for further analysis. The graphs depict the distribution of the PF10_0083 binding positions along the genome in (A) over-expression cell lines, and (B) endogenous-tagged cell lines. The horizontal axis is the oligonucleotide probes position relative to the ATG start codon. Position -1 bp to -5000 bp represent the IGRs, and position 0 bp to >8000 bp represent the ORFs. The vertical axis is the number of PF10_0083 associated genetic loci normalized to the number of probes which represent that particular region on the genome. T996-myc1: T996-pARL-PF10_0083-myc1, T996-myc2: T996-pARL-PF10_0083-myc2, T996-gfp: T996-pARL-PF10_0083-gfp, 3D7-myc3: 3D7-pARL-PF10_0083-myc3, C18: CHA18, C30: CHA30, HA: anti-HA, mC: anti-mCherry.

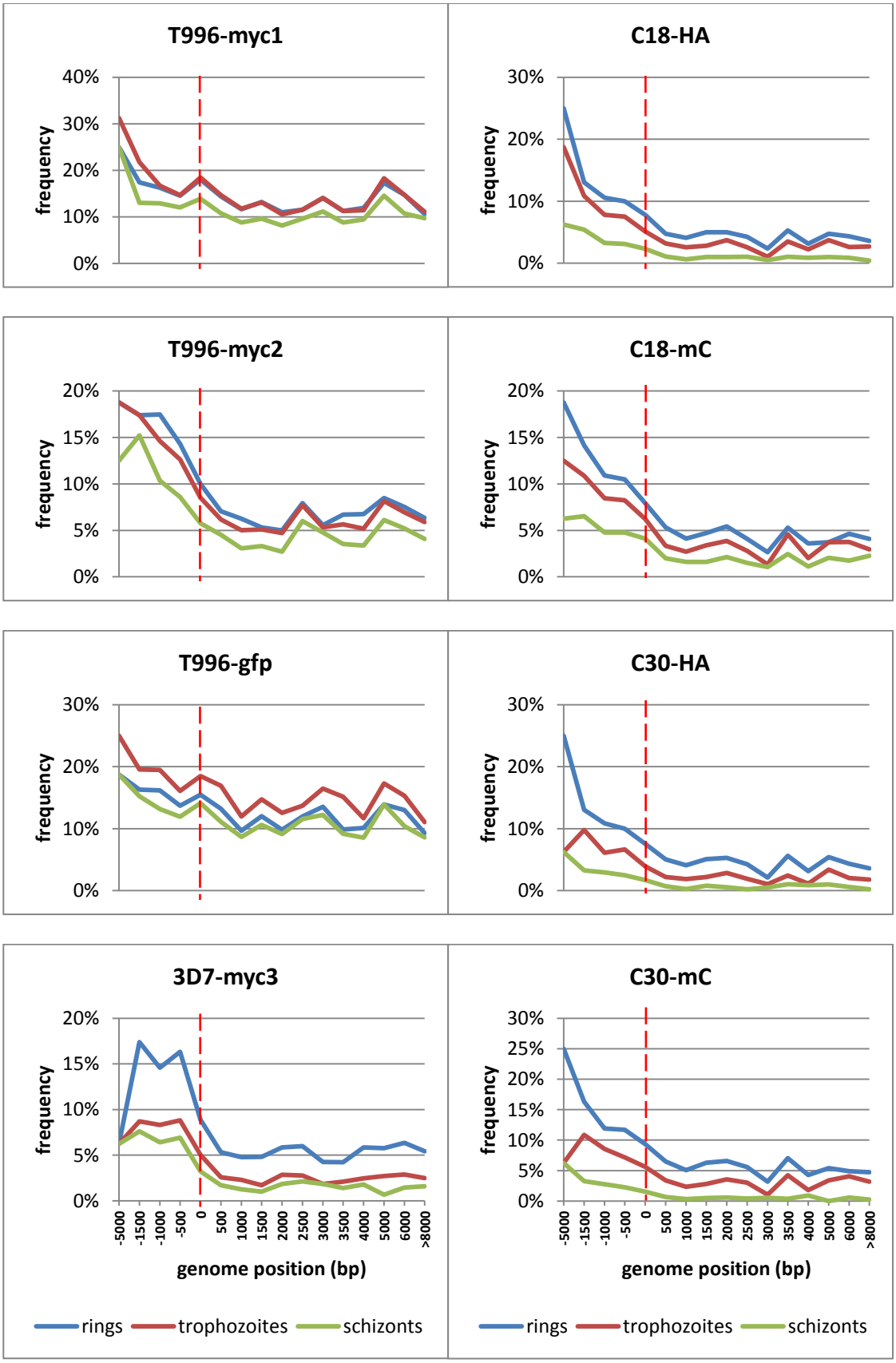


Figure 4.11 The positional enrichment of PF10_0083 along the genome in rings, trophozoites and schizonts.

DNA amplification and microarrays were carried out in triplicates. Only oligonucleotide probes with data present in at least 2 out of the 3 triplicates were included for analysis. The data from the triplicates were averaged and used for further analysis. Oligonucleotide probes with a positive ChIP signal at least 2 fold higher than the input signal (log_2 ChIP/input ratios ≥2) were used for further analysis. The graphs depict the distribution of the PF10_0083 binding positions along the genome in the different stages of the IDC. The horizontal axis is the oligonucleotide probes position relative to the ATG start codon. Position -1 bp to -5000 bp represent the IGRs, and position 0 bp to >8000 bp represent the ORFs. The vertical axis is the number of PF10_0083 associated genetic loci normalized to the number of probes which represent that particular region on the genome. T996-myc1: T996-pARL-PF10_0083-myc1, T996-myc2: T996-pARL-PF10_0083-myc2, T996-gfp: T996-pARL-PF10_0083-gfp, 3D7-myc3: 3D7-pARL-PF10_0083-myc3, C18: CHA18, C30: CHA30, HA: anti-HA, mC: anti-mCherry.

4.4 PF10_0083 binding and transcriptional regulation

In order to investigate the relationship between PF10_0083 binding and the transcriptional changes observed in the over-expression cell lines, the ChIP-on-chip data was correlated to the transcriptome data. Genes with ChIP-on-chip and transcriptome data correlated with Pearson correlation coefficient (PCC) greater than 0.4 were used for further analysis. It was observed that amongst the PF10_0083 associated genes, 185, 130, 21, and 11 genes were observed to be up-regulated in T996-myc1, T996-myc2, T996-gfp and 3D7-myc3 cell lines respectively. PF10_0083 associated with these up-regulated genes at the IGRs. On the other hand, 72, 177, 17, and 5 PF10_0083 associated genes were down-regulated in T996-myc1, T996-myc2, T996-gfp and 3D7-myc3 cell lines respectively. PF10_0083 associated with these down-regulated genes at the ORFs. Specifically, for T996-myc1, PF10_0083 displayed a positional bias at position -1 bp to -1500 bp in the up-regulated genes (Figure 4.12, red bars). In contrast, in the down-regulated genes, PF10_0083 displayed a positional bias at position 2500 bp and further downstream (Figure 4.12, green bars).

Taken together, the binding of PF10_0083 to the genes at the IGRs correlated with their up-regulation. On the contrary, the binding of PF10_0083 to the genes at the ORFs correlated with their down-regulation.

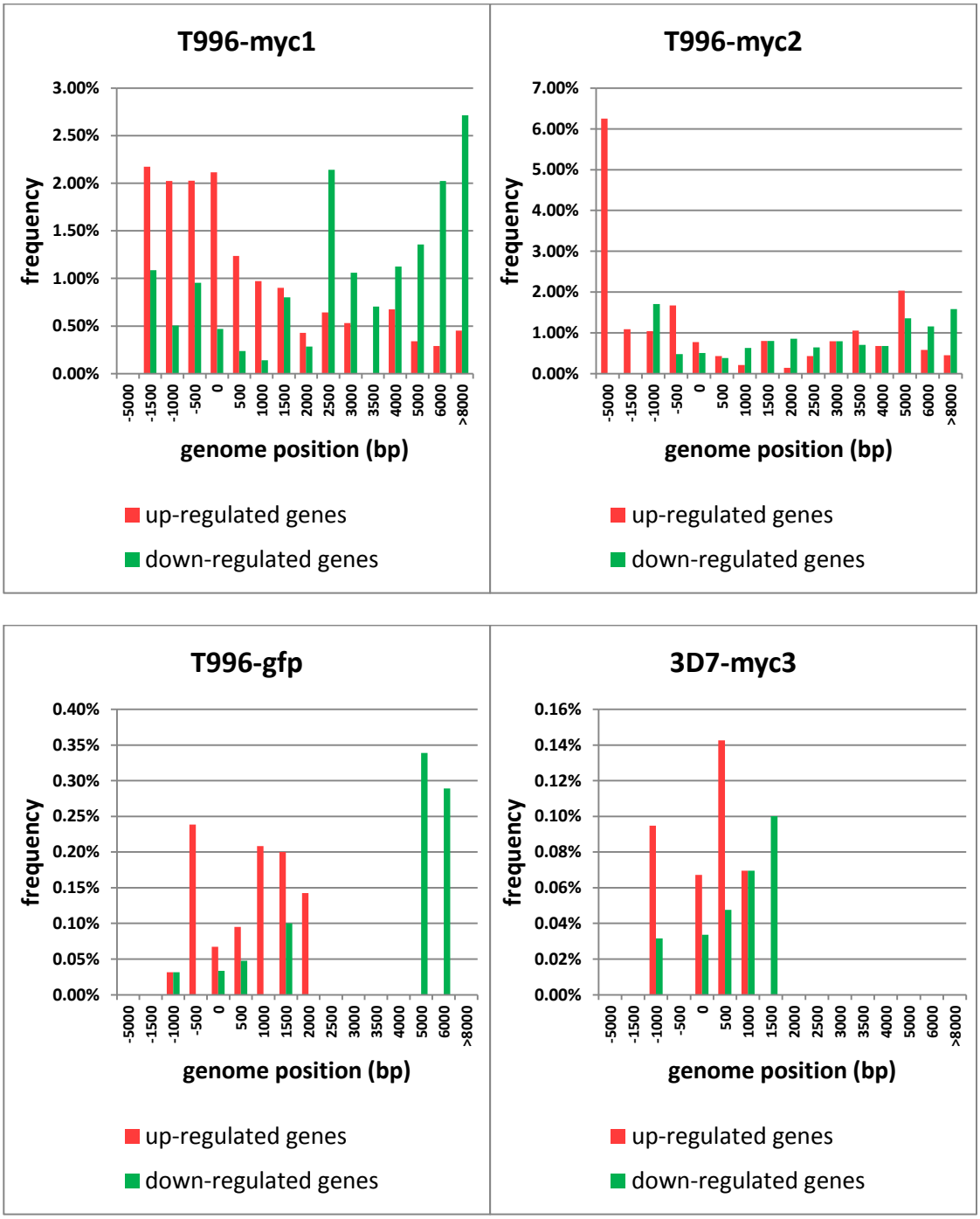


Figure 4.12 PF10_0083 binding position of the genes that were differentially expressed in the over-expression cell lines.

DNA amplification and microarrays were carried out in triplicates. Only oligonucleotide probes with data present in at least 2 out of the 3 triplicates were included for analysis. The data from the triplicates were averaged and used for further analysis. Oligonucleotide probes with a positive ChIP signal at least 2 fold higher than the input signal (log_2 ChIP/input ratios ≥2) were used for further analysis. The ChIP-on-chip data was correlated to the transcriptome data. Genes with ChIP-on-chip and transcriptome data correlated with Pearson correlation coefficient (PCC) greater than 0.4 were used for further analysis. The binding of PF10_0083 to the genes at the IGRs correlated with their up-regulation. The binding of PF10_0083 to the genes at the ORFs correlated with their down-regulation. The horizontal axis is the oligonucleotide probes position relative to the ATG start codon. Position -1 bp to -5000 bp represent the IGRs, and position 0 bp to >8000 bp represent the ORFs. The vertical axis is the number of PF10_0083 associated genetic loci normalized to the number of probes which represent that particular region on the genome. T996-myc1: T996-pARL-PF10_0083-myc1, T996-myc2: T996-pARL-PF10_0083-myc2, T996-gfp: T996-pARL-PF10_0083-gfp, 3D7-myc3: 3D7-pARL-PF10_0083-myc3.

Figure 4.13 illustrates 4 examples where the binding of PF10_0083 to the IGRs of these genes positively correlated with their up-regulation. PF10_0083 positively correlated with the up-regulation of RWD domain-containing protein, GPI (glycosylphosphatidylinositol) anchor protein, putative transcription activator, and 40S ribosomal protein, with PCC greater than 0.4. Figure 4.14 illustrates 3 examples, where the binding of PF10_0083 to the ORFs of these genes positively correlated with their down-regulation. The binding of PF10_0083 at 2 loci at the ORFs of a putative ribonuclease, a putative RNA pseudouridylate synthase, and a conserved protein, correlated with their down-regulation (PCC>0.4). Figure 4.15 illustrates 3 examples of PF10_0083 binding to both the IGRs and ORFs of a gene. The binding of PF10_0083 to hexose transporter, putative tubulin-specific chaperone and *Plasmodium* exported protein (PHISTc), was not correlated to their transcriptional changes (-0.4≤PCC≤0.4).

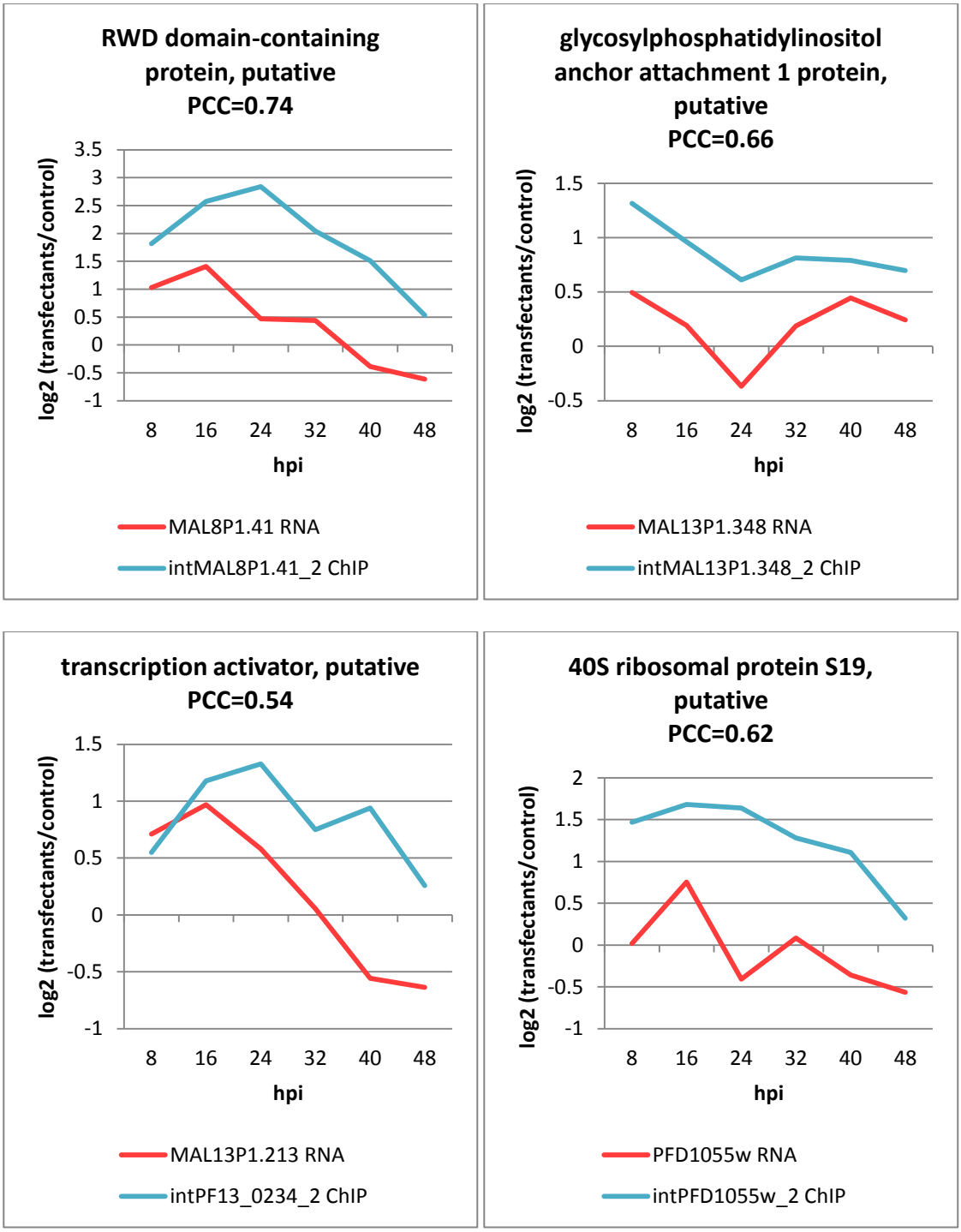


Figure 4.13 PF10_0083 binding at IGRs and RNA levels.

The binding of PF10_0083 to the IGRs of RWD domain-containing protein, GPI anchor protein, transcription activator, and 40S ribosomal protein, positively correlated with the up-regulation of these genes (PCC>0.4). PCC: Pearson correlation coefficient.

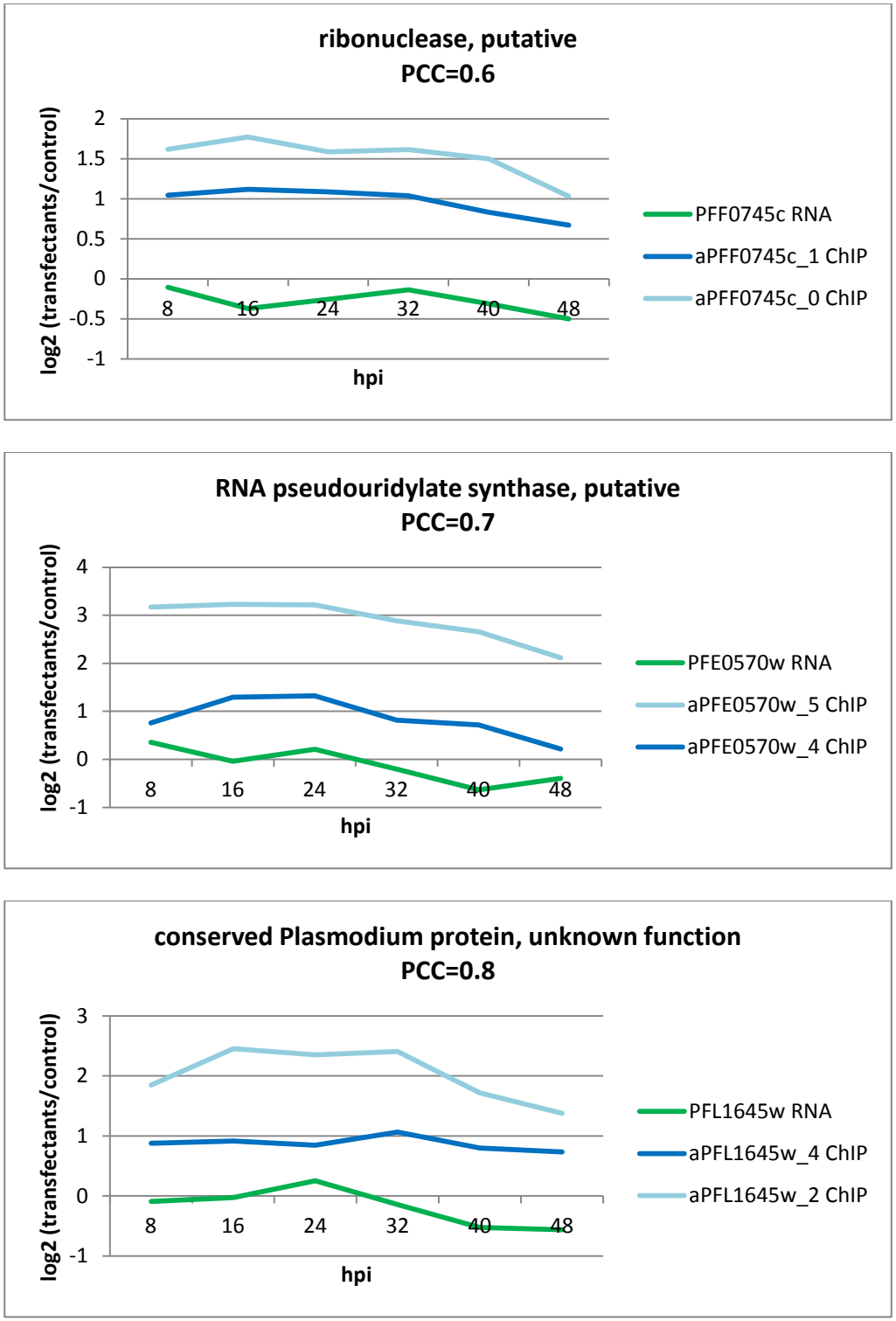
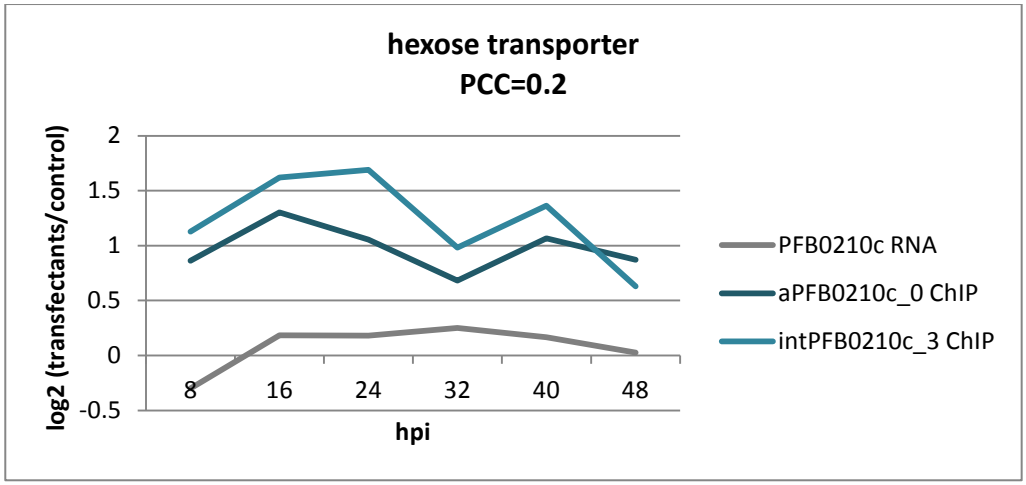
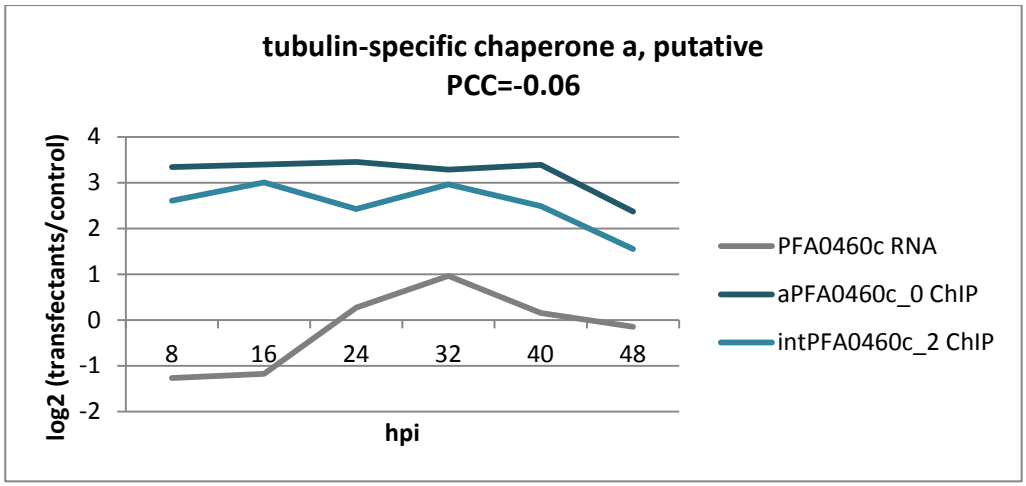


Figure 4.14 PF10_0083 binding at ORFs and RNA levels.

The binding of PF10_0083 to the ORFs of a putative ribonuclease, a putative RNA pseudouridylate synthase, and a conserved protein correlated with the down-regulation of these genes (PCC>0.4). PCC: Pearson correlation coefficient.





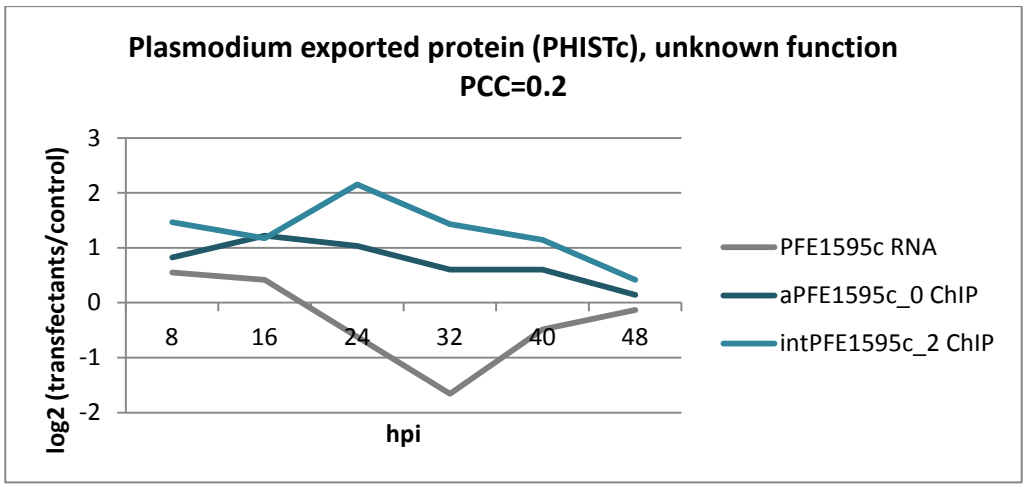


Figure 4.15 PF10_0083 binding at both the IGRs and ORFs regions, and RNA levels. **The binding of PF10_0083 to both the IGRs and ORFs of hexose transporter, putative tubulin-specific chaperone and *Plasmodium* exported protein was not correlated to the transcription changes (-0.4≤PCC≤0.4). PCC: Pearson correlation coefficient.**

4.4.1 PF10_0083 binding correlates with the up-regulation of gametocyte specific genes

From the transcriptome data, it was observed that the over-expression of PF10_0083 resulted in the up-regulation of gametocyte genes (Section 4.2.2, Figure 4.5 and Table 4.2). As such, we analyzed the genes associated with PF10_0083 and were transcriptionally altered. In T996-myc1 and T996-gfp, it was observed that the majority of the up-regulated genes associated with PF10_0083 are maximally expressed in the gametocytes in the normal parasite life cycle (Figure 4.16). For instance, in T996-myc1, 185 genes that were associated with PF10_0083 were up-regulated, out of which, 67 (36%) of the genes are maximally expressed in the gametocytes in the life cycle. Twenty-two genes (31%) out of the 72 down-regulated genes were maximally expressed in the early rings. In T996-gfp, 19% of the up-regulated genes are gametocyte-stage-specific genes. For example, in T996-myc1, the binding of PF10_0083 to the IGRs of the gametocyte genes, ETRAMP 10.3, secreted ookinete adhesive protein (SOAP), tubulin-tyrosine ligase, and nucleoside transporter positively correlated with their up-regulation with PCC>0.4 (Figure 4.17).

However, in T996-myc2, the opposite was observed. PF10_0083 associated with 130 up-regulated genes. Thirty (23%) of these genes are early-rings-stage-specific. PF10_0083 associated with 177 down-regulated genes, out of which, 68 genes (38%) are maximally expressed in the gametocytes in the parasite life cycle (Figure 4.16). The sexual stage-specific protein precursor (Pfs16) was up-regulated in T996-myc1 but down-regulated in T996-myc2. In both cell lines, the ChIP-on-chip profiles were similar. The binding of PF10_0083 to Pfs16 correlated with its up-regulation in T996-myc1, and its down-regulation in T996-myc2 (Figure 4.18).

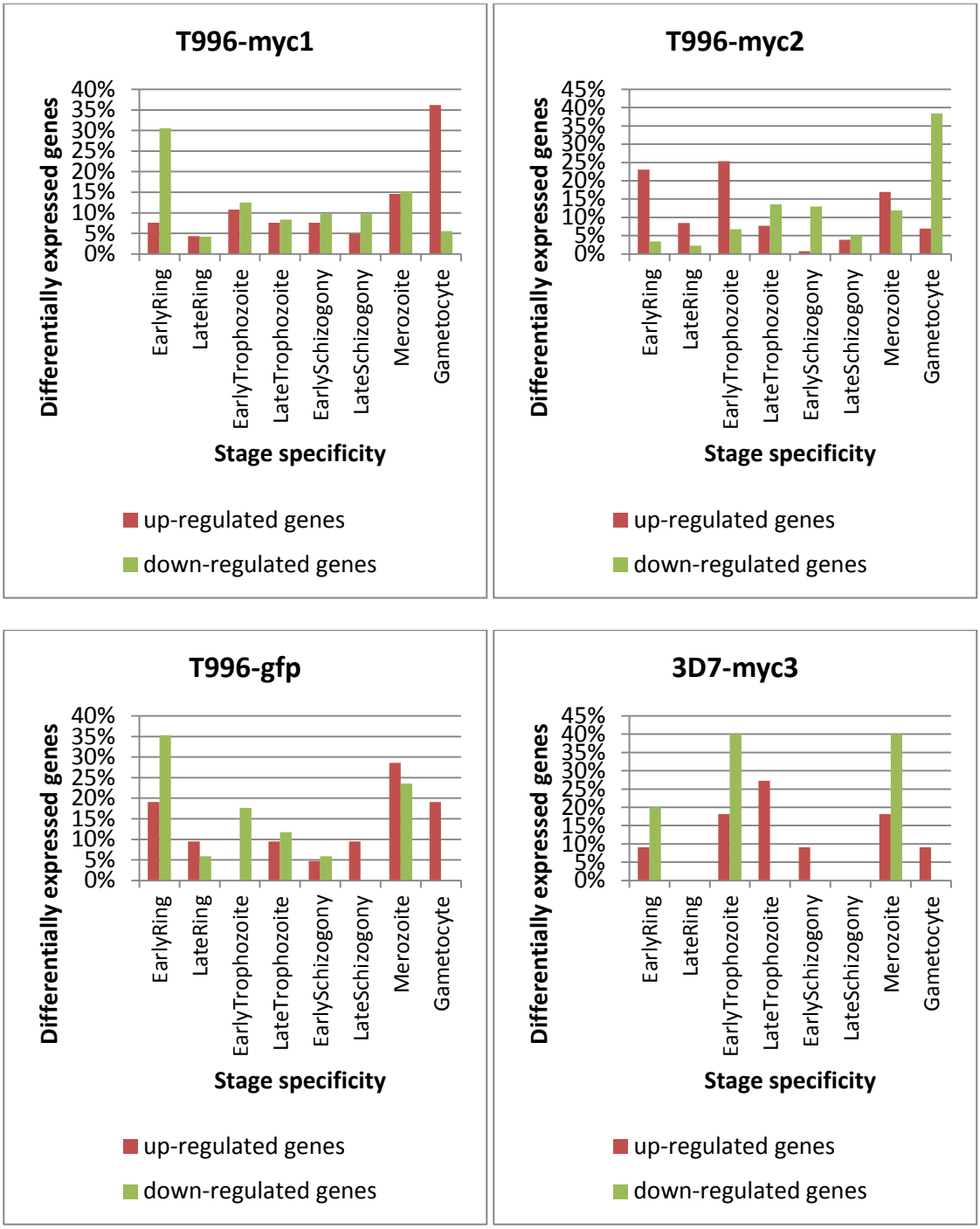


Figure 4.16 Distribution of PF10_0083 associated genes according to their stage specificity.

Distribution of the up-regulated and down-regulated genes that were associated with PF10_0083 in the 4 over-expression cell lines, according to their normal expression profiles in the parasite life cycle. The stage specificity of each differentially expressed gene is defined as the stage during the parasite life cycle by which the gene is maximally expressed. The differentially expressed genes were then classified according to their stage specificity and plotted into this graph. T996-myc1: T996-pARL-PF10_0083-myc1, T996-

myc2: T996-pARL-PF10_0083-myc2, T996-gfp: T996-pARL-PF10_0083-gfp, 3D7-myc3: 3D7-pARL-PF10_0083-myc3.

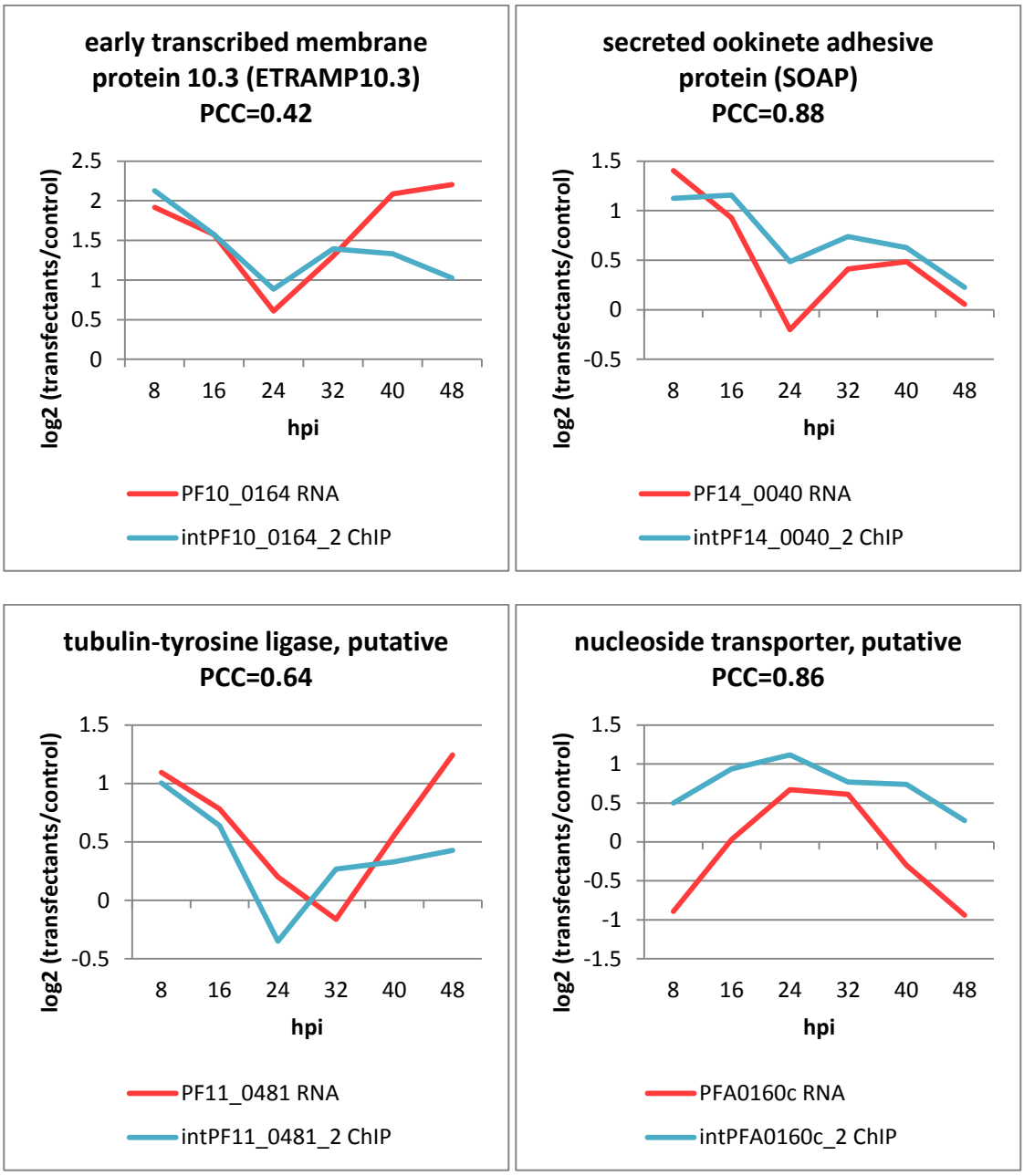


Figure 4.17 The binding of PF10_0083 to gametocyte stage specific genes correlated with their up-regulation.

The binding of PF10_0083 to the IGRs of the gametocyte genes, ETRAMP10.3, SOAP, tubulin-tyrosine ligase and nucleoside transporter in the T996-myc1 cell line, positively correlated with their up-regulation. PCC: Pearson correlation coefficient.

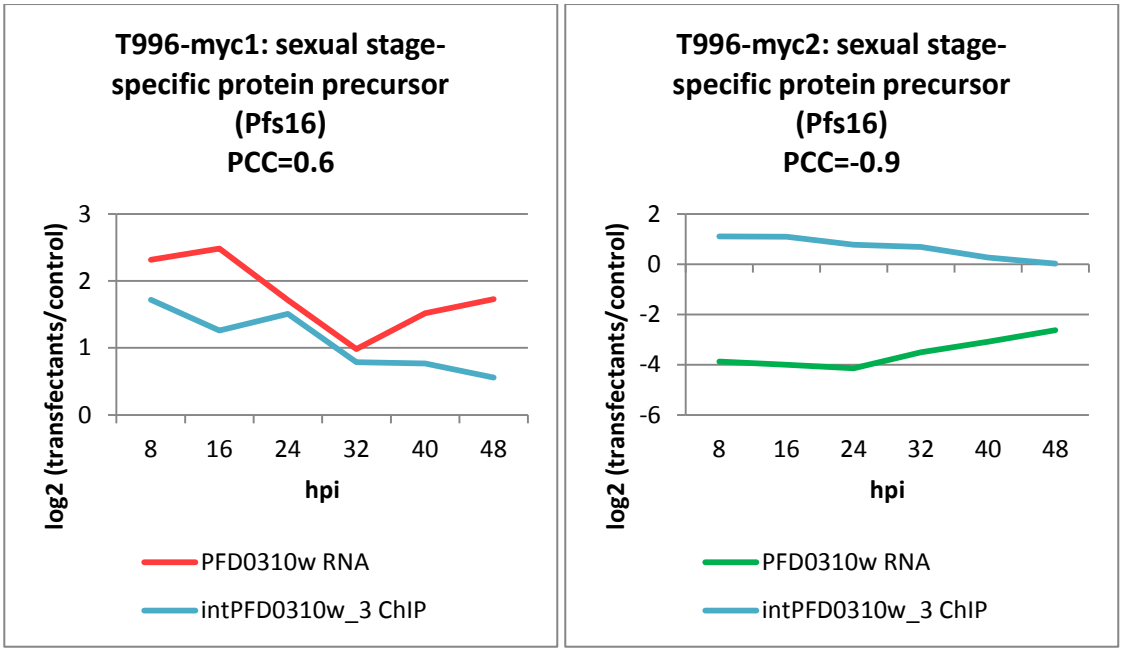


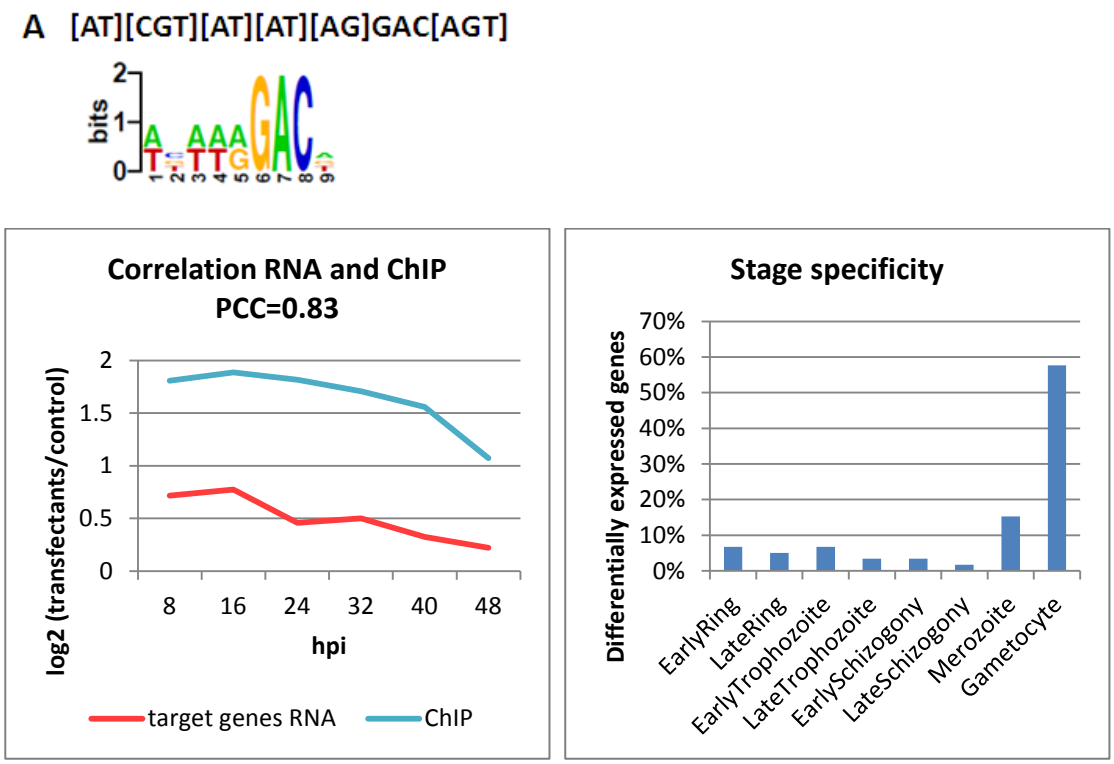
Figure 4.18 PF10_0083 binding and RNA levels of Pfs16.

The binding of PF10_0083 to Pfs16 resulted in its up-regulation in T996-myc1 and its down-regulation in T996-myc2. T996-myc1: T996-pARL-PF10_0083-myc1, T996-myc2: T996-pARL-PF10_0083-myc2. PCC: Pearson correlation coefficient.

4.4.2 PF10_0083 binds specific DNA motifs

We determined the potential DNA sequence motifs that PF10_0083 binds to using the FIRE program [145]. We predicted 3 DNA motifs in the genes where the binding of PF10_0083 (ChIP-on-chip profiles) and the transcriptome profiles were positively correlated (PCC>0.4). The first motif, [AT][CGT][AT][AT][AG]GAC-[AGT], was found in 63 genes. The RNA profiles and ChIP-on-chip profiles of these genes were averaged and plotted. The RNA profiles and ChIP-on-chip profiles were positively correlated with a PCC of 0.83. In addition, we observed that 34 of these genes are gametocyte stage specific genes (Figure 4.19A). The second motif, [AT][CG][AT]CA[CG][AGT]C[ACG], was found in 29 genes. Eighteen of these genes are gametocytes genes. The average RNA profiles and ChIP-on-chip profiles were positively correlated with a PCC of 0.88 (Figure 4.19B). The third motif,

[AT][AG][AT]AT[ACG]GC[ACT], was found in 60 genes. The average RNA profiles and ChIP-on-chip profiles were also positively correlated with a PCC of 0.93. Twenty-four of these genes are gametocyte genes (Figure 4.19C). Eight genes were found to contain all 3 motifs, while 28 genes contain 2 of the 3 motifs (Figure 4.20). For example, Pfs16 was found to contain all 3 motifs (Figure 4.18), while SOAP contains the first and third motifs (Figure 4.17). ETRAMP 10.3 and nucleoside transporter both contain the third motif, while tubulin-tyrosine ligase contains only the first motif (Figure 4.17). These genes contain the predicted motif(s), and PF10_0083 bound to their IGRs regions and was positively correlated with their transcriptional up-regulation.



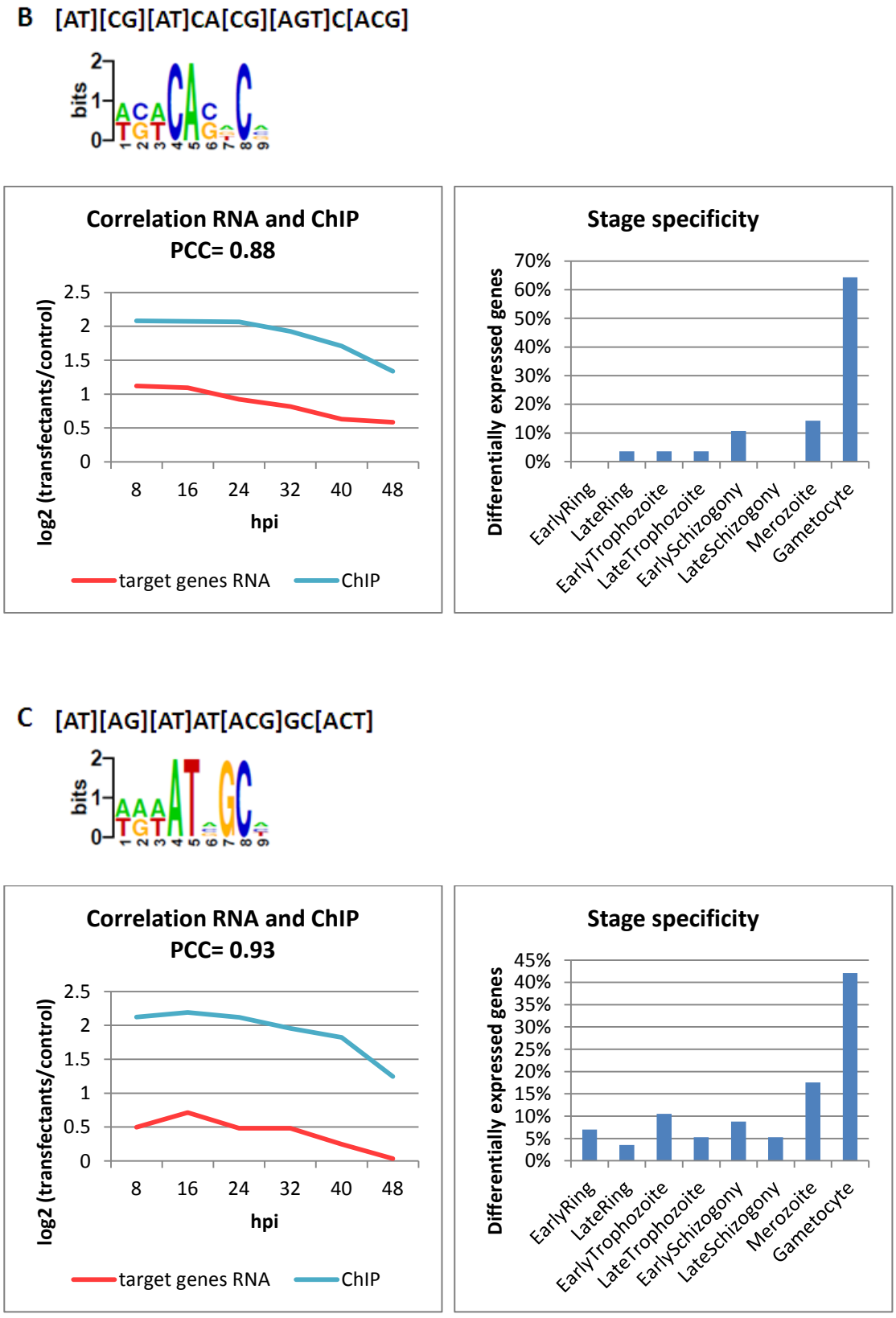


Figure 4.19 PF10_0083 binds specific DNA sequence motifs.

3 DNA motifs were predicted (A) [AT][CGT][AT][AT][AG]GAC[AGT], (B) [AT][CG][AT]CA-[CG][AGT]C[ACG], and (C) [AT][AG][AT]AT[ACG]GC[ACT]. The average transcriptome and ChIP-on-chip profiles of the genes containing these motif(s)

were positively correlated (PCC>0.4). On average, 55% of the genes that contain these motifs are gametocyte stage specific genes. PCC: Pearson correlation coefficient.

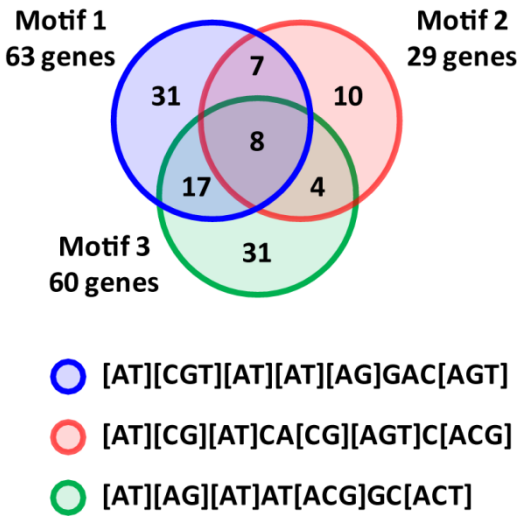


Figure 4.20 Differentially expressed genes containing multiple DNA sequence motifs. 8 genes were found to contain all 3 predicted DNA motifs. 28 genes were found to contain 2 out of the 3 predicted DNA motifs.

4.5 Summary

Taken together, PF10_0083 binds to multiple stages during the IDC. PF10_0083 was found to bind to both IGRs and ORFs. However, PF10_0083 displayed a positional bias for the IGRs. In addition, the binding of PF10_0083 to the IGRs correlated with transcriptional up-regulation, while the binding of PF10_0083 to the ORFs correlated with transcriptional down-regulation. In addition, PF10_0083 binds to a group of genes which are most highly expressed in the gametocytes in the usual life cycle. It was observed that the binding of PF10_0083 to these genes correlated with their transcriptional up-regulation in T996-myc1. In addition, 3 DNA sequence motifs that PF10_0083 potentially binds to amongst the up-regulated genes were identified.

Chapter 5 The role of PF10_0083 in gametocyte development

5.1 Introduction

From the transcriptome data, it was observed that the over-expression of PF10_0083 induced transcriptional changes in a subset of genes that are known to be highly expressed in the gametocyte stage of the parasite life cycle (Section 4.2.2, Figure 4.5 and Table 4.2). From the ChIP-on-chip data, it was observed that PF10_0083 associated with a group of genes which are most highly expressed in the gametocytes in the parasite life cycle (Section 4.4.1, Figure 4.16). The binding of PF10_0083 to these genes positively correlated with their transcriptional up-regulation (Section 4.4.1, Figure 4.17). As such, we investigated the role of PF10_0083 in gametocyte development. We tested the effects of the over-expression and knock-out of PF10_0083 on gametocyte production and development.

5.2 Gametocyte cultivation

5.2.1 Validate the ability of 3D7 parasites to develop into gametocytes

In order to induce asexual stages parasites to differentiate into gametocytes, we first tested the ability of the parasites to produce gametocytes. The protocol used is commonly known as the "crash" method and is aimed at producing high number of asynchronous gametocytes [278]. Wild type 3D7 strain parasites were used. A starter culture of healthy and synchronized ring stage parasites at 3% to 5% parasitemia was prepared. Culture media was changed every day, but the supply of fresh blood was stopped. After 5 to 7 days, a crash of the asexual parasites was observed. Stressed

schizonts which appeared small, with fewer merozoites and poorly defined nuclei started to appear (Figure 5.1). In addition, gametocytes were also observed. For the next 5 to 7 days, gametocytes number increased, and few late stage gametocytes (stages IV and V) were also observed. The parasitemia of the culture was calculated every day. The total number of asexual and sexual parasites over the total number of RBCs is defined as the total parasitemia. The total number of gametocytes over the total number of RBCs is defined as the gametocytemia. The percentage of gametocytes is defined by the gametocytemia over the total parasitemia. As illustrated in Figure 5.2, on day 4, gametocytes were first observed, and between day 8 and day 12, up to 23% of the total parasitemia were gametocytes. However, the asexual stages were not completely eliminated. In addition, the gametocytes were not synchronized. From day 5 to day 14, both early (stages II and III) and late stages (stages IV and V) gametocytes were observed.

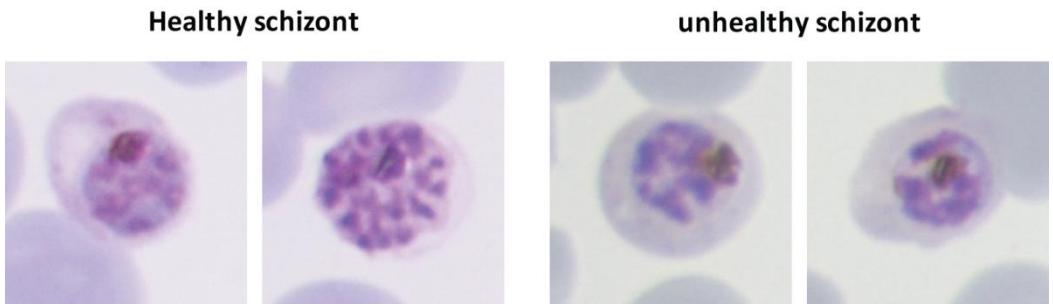


Figure 5.1 Microscope images of healthy and unhealthy schizonts.
Giemsa stained images of healthy and unhealthy schizonts. Healthy schizonts have clearly defined nuclei. Unhealthy schizonts appeared small and have undefined nuclei.

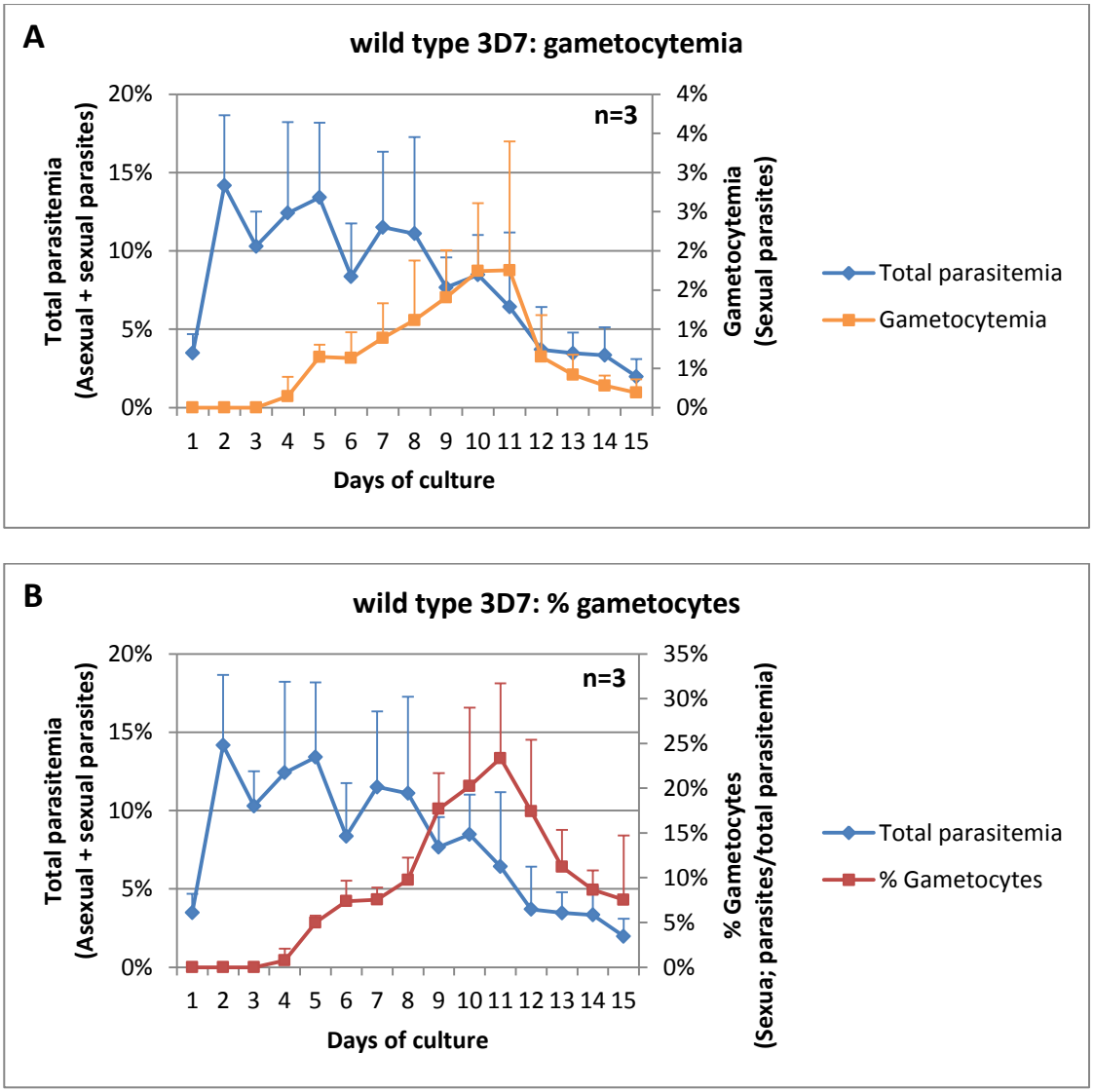


Figure 5.2 Validate the ability of the 3D7 parasites to produce gametocytes using the "crash" method.

The "crash method" was employed to produce large number of asynchronous gametocytes. A starter culture of 3% to 5% ring stage parasites was used. At day 5 to day 7, a crash of the asexual parasites was observed, and gametocytes were observed. For the next 5 to 7 days, gametocytes number increased. Between day 8 and day 12, up to 23% of the total parasitemia were gametocytes. (A) The total parasitemia is defined as the total number of asexual and sexual parasites over the total number of RBCs counted. Gametocytemia is defined as the total number of gametocytes over the total number of RBCs. (B) The percentage of gametocytes over the total parasitemia was also calculated. The graphs represent results from 3 experiments. Error bars represent standard deviation of the mean.

5.2.2 Culturing synchronous gametocytes

Since it was confirmed that wild type 3D7 strain parasites are capable of gametocyte development, we tested another protocol (developed by Fivelman et al.) which aims at producing synchronized stages of gametocytes [279]. Figure 5.3 is a flowchart describing the protocol. For this method, it is important to determine the invasion efficiency of merozoites as assessed by the formation of new rings. The number of new rings formed during every invasion will differ with different batches of blood used. Hence, we carefully determined the invasion efficiency with every batch of blood before attempting this gametocyte induction protocol. For instance, the parasites gave an invasion efficiency of 5 fold in a batch of blood. On day -4, a starter culture of healthy and highly synchronized ring stage parasites at 1.5% to 2% parasitemia was prepared. On day -2, this 1.5% to 2% rings culture will produce a 8% to 10% parasitemia during the next invasion since the invasion efficiency of the parasites in this batch of blood is 5 fold.

To apply stress to these parasites, not all the old culture media was removed. Instead, only half the old media was removed, and an equal volume of fresh media was provided. Stressed schizont parasites were observed on the next day (day -1). The stressed schizont parasites were diluted to 1.5% to 2% parasitemia. The same batch of blood used for preparing the starter culture was used for the dilution of the parasites, so as not to affect the invasion efficiency of the merozoites to form new rings. One third volume of old media and two third volumes of fresh media were provided to the parasite culture. On the following day, parasitemia was checked to ensure that 8% to 10% new rings were formed (day 0).

To enrich and isolate early stage gametocytes, culture medium was changed normally on day 0 and treated with 5% sorbitol to remove late asexual trophozoite and schizont stages on day 1. Early stage gametocytes will be enriched between day 2 and day 4. To enrich and isolate late stage gametocytes, culture media was supplemented with 50 mM N-acetylglucosamine (GlcNAc) from day 1 to day 5. GlcNAc is toxic to trophozoites and schizonts but does not affect gametocyte development. Late stage gametocytes will be ready for harvest on day 6. Unfortunately, despite several attempts, we were unsuccessful at inducing synchronized gametocytes using this protocol. The sorbitol treatment resulted in the crash of the cultures, and we did not observe gametocytes. The exposure of the parasites to GlcNAc resulted in a mixture of early and late stage gametocytes.

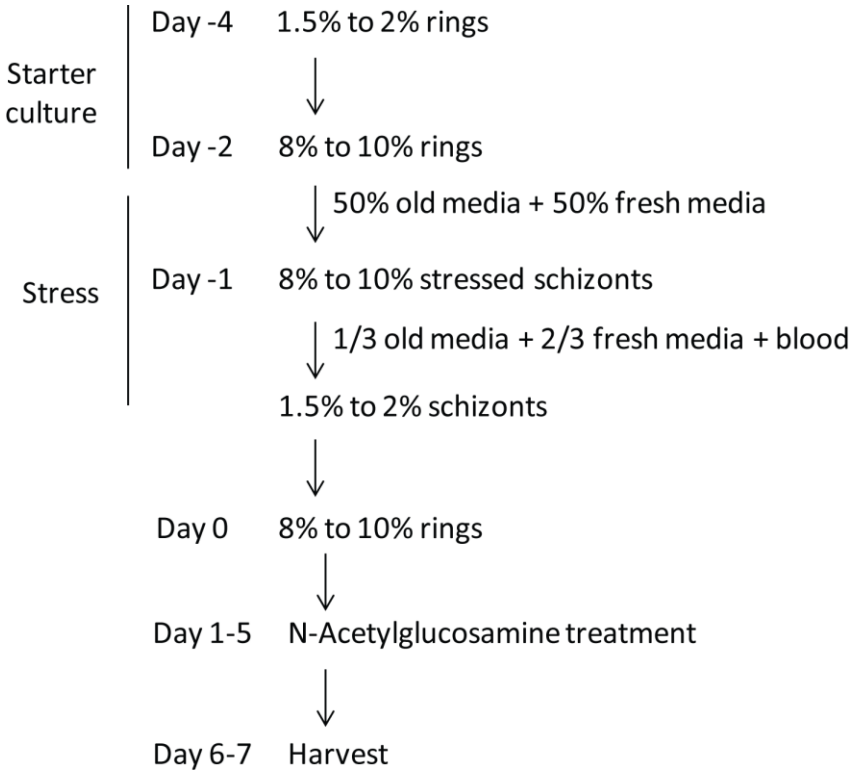
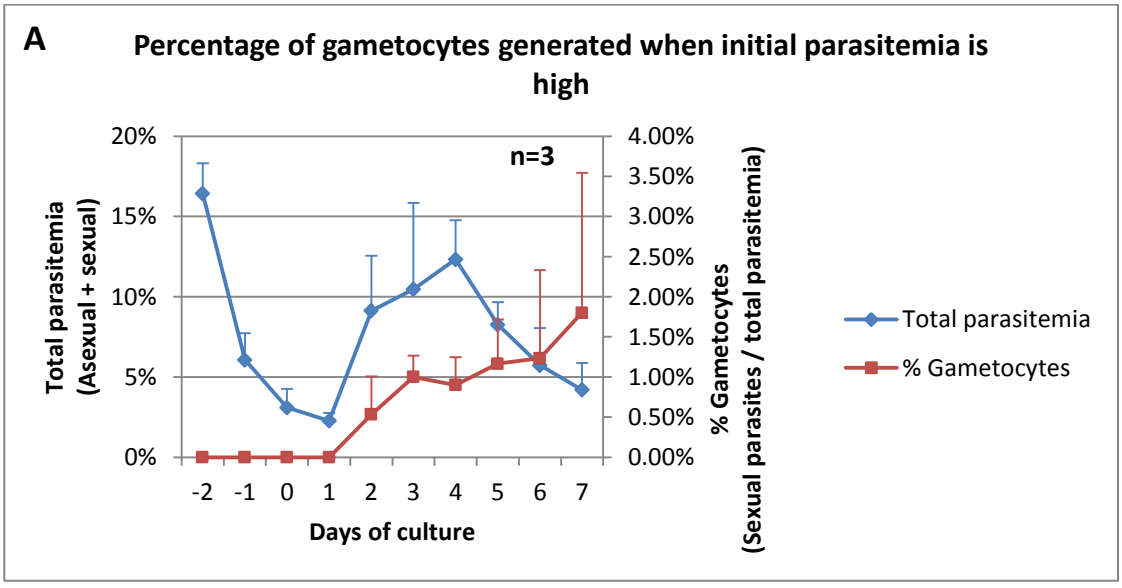


Figure 5.3 Flowchart describing the steps to culture synchronized gametocytes.
On day -4, a starter culture of healthy and highly synchronized ring stage parasites at 1.5% to 2% parasitemia was prepared for gametocyte induction. On day -2, this 1.5% to 2%

rings culture will produce a 8% to 10% parasitemia during the next invasion if the invasion efficiency of the parasites in the batch of blood used is 5 fold. From day -2 to day -1, parasites were cultured under stress conditions, where old culture media was supplied together with fresh culture media. From day 1 to day 5, culture media was supplemented with 50 mM N-acetylglucosamine (GlcNAc). On day 6, gametocytes will be ready for harvest.

It is important to keep the parasitemia on day -2 between 8% to 10% for every gametocyte induction experiment. This is because when the parasites are stressed on day -2, there will be too many cell deaths on day -1 when the parasitemia on day -2 is too high. As a result, the percentage of gametocytes produced will be affected. As illustrated in Figure 5.4A, on day -2, approximately 16% total parasitemia were observed. On day -1, the total parasitemia dropped to approximately 6%. From day 2 to day 7, the percentage of gametocytes produced were only between 0.5% to 1.8% of the total parasitemia. On the other hand, when the total parasitemia on day -2 was kept at approximately 9.3%, the parasitemia remained at approximately 9.6% on day -1. From day 2 to day 7, the percentage of gametocytes produced were between 0.7% to 22% of the total parasitemia (Figure 5.4B).



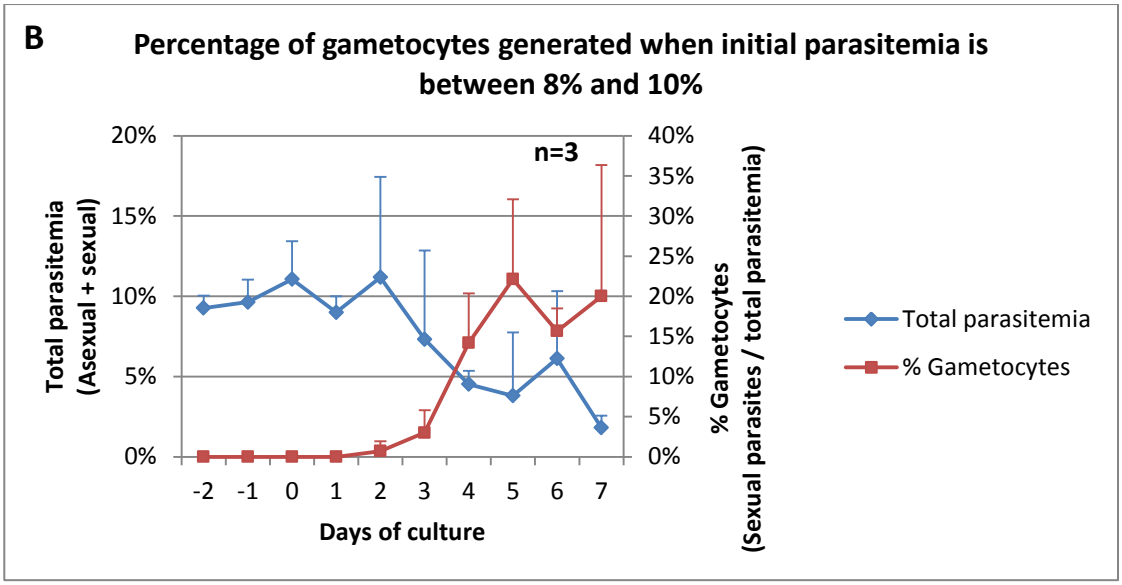


Figure 5.4 Initial parasitemia during gametocyte induction experiment affects the percentage of gametocytes generated.

(A) On day -2, approximately 16% total parasitemia were observed. On day -1, the total parasitemia dropped to approximately 6%. From day 2 to day 7, the percentage of gametocytes produced were between 0.5% to 1.8% of the total parasitemia. (B) On day -2, approximately 9.3% total parasitemia were observed. The parasitemia remained at approximately 9.6% on day -1. From day 2 to day 7, the percentage of gametocytes produced were between 0.7% to 22% of the total parasitemia. The total parasitemia is defined as the total number of asexual and sexual parasites over the total number of RBCs counted. The total number of gametocytes over the total number of RBCs is defined as the gametocytemia. The percentage of gametocytes is defined by the gametocytemia over the total parasitemia. The graphs represent results from 3 experiments. Error bars represent standard deviation of the mean.

Figure 5.5 illustrates the progression of the ring, trophozoite and schizont asexual stage parasites and sexual stage parasites development during gametocyte induction. On day -1, approximately 2% stressed schizonts were prepared. On day 0, approximately 8% to 10% rings were observed. On day 1, GlcNAc was added to the culture media, and asexual parasites began to decrease on day 3. Gametocytes were first observed on day 2. On day 6, the maximum number of gametocytes was observed. Approximately 21% of the total number of parasites were gametocytes. On day 6 and

day 7, the number of gametocytes started to decrease. Just before the crash of the asexual parasites, the gametocytes started to increase. The crash of the asexual parasites coincided with the maximum number of gametocytes. However, due to the crash of asexual parasites, there was RBC lysis and the presence of dead asexual stage parasites. These factors resulted in a decreased number of gametocytes. Hence, for further studies of the gametocytes, the day of harvest is important to maximize sample collection.

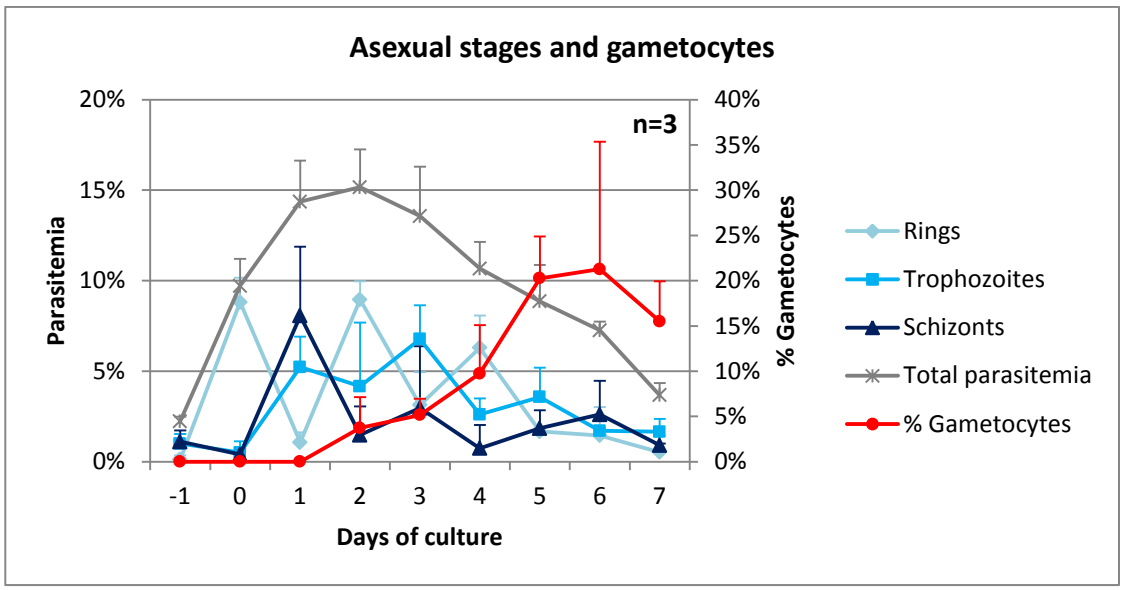


Figure 5.5 Percentages of rings, trophozoites and schizonts asexual stages and gametocytes during each day of the gametocyte induction experiment.

On day -1, approximately 2% stressed schizonts were prepared. On day 0, approximately 8% to 10% rings were observed. On day 1, GlcNAc was added to the culture media. On day 3, asexual parasites began to decrease. On day 2, gametocytes were first observed. On day 6, approximately 21% of the total parasitemia were gametocytes. The total parasitemia is defined as the total number of asexual and sexual parasites over the total number of RBCs counted. The total number of gametocytes over the total number of RBCs is defined as the gametocytemia. The percentage of gametocytes is defined by the gametocytemia over the total parasitemia. The graphs represent results from 3 experiments. Error bars represent standard deviation of the mean.

Gametocyte induction was carried out on wild type 3D7 parasites, PF10_0083-CHA18 endogenous-tagged cell line, 3D7-myc3 over-expression cell line, and PF10_0083-KO33 knock-out cell line. On average, on day 3, early stage gametocytes (stage II-III) were observed. Between day 4 and day 5, late stage gametocytes (Stage IV-V) were observed. All 4 cell lines were induced at the same time for all induction experiments. This was done to minimized experiment to experiment variations. In total, 15 gametocyte induction experiments were carried out, out of which gametocytes did not develop in 3 experiments. The parasite counts from the 12 experiments were calculated and averaged.

5.3 Gametocyte induction

5.3.1 3D7 and PF10_0083-CHA18 control cells

Since 3D7-myc3, PF10_0083-CHA18 and PF10_0083-KO33 cell lines transfection were carried out in 3D7 strain parasites, the wild type 3D7 strain parasites were used as a control for all experiments. On average, it was observed that the total parasitemia started to decrease starting from day 3 and day 4, and the total parasitemia decreased to 1.7% on day 7 of the experiment. The decrease of the total parasitemia coincided with the appearance of the gametocytes. On average, we started to observe gametocytes from day 2 and day 3 onwards. On day 5, the number of gametocytes peaked. Approximately 18% of the total number of parasites were gametocytes (Figure 5.6A). Although we were able to induce a good percentage of parasites to differentiate into gametocytes, we were not able to obtain synchronized stages of gametocytes. Throughout the course of the experiments, there was a mixture of early and late stages gametocytes. On average, early stage gametocytes peaked during day 3 and day 5. We started to observe late stage gametocytes starting from day 4 to day 5. On day 7 of the

experiments, it was observed that approximately 65% of the gametocytes were early stage, and 35% of them were late stage (Figure 5.6B).

Since 3D7-myc3 was under continuous WR99210 drug pressure, this cell line was not directly comparable to the wild type 3D7 parasites. PF10_0083-CHA18 cell line is PF10_0083 tagged at the endogenous locus, and contains the hDHFR drug cassette. Therefore, we were able to culture this cell line under continuous WR99210 drug pressure. As such, PF10_0083-CHA18 was also used as a control along with wild type 3D7. Similar to the 3D7 parasites, we started to observe gametocytes from day 2 and 3, and the number of gametocytes peaked on day 6. Approximately 17% of the total number of parasites were gametocytes (Figure 5.7A). Similar to the 3D7 parasites, there was a mixture of early and late stages gametocytes. Early stage gametocytes peaked between day 3 and day 5. We started observing late stage gametocytes starting from day 4 and day 5. Similar to the wild type 3D7, we observed that approximately 60% of the gametocytes were early stage, and 40% of them were late stage on day 7 of the experiments (Figure 5.7B). The difference in the number of gametocytes produced in the 2 cell lines is not statistically significant.

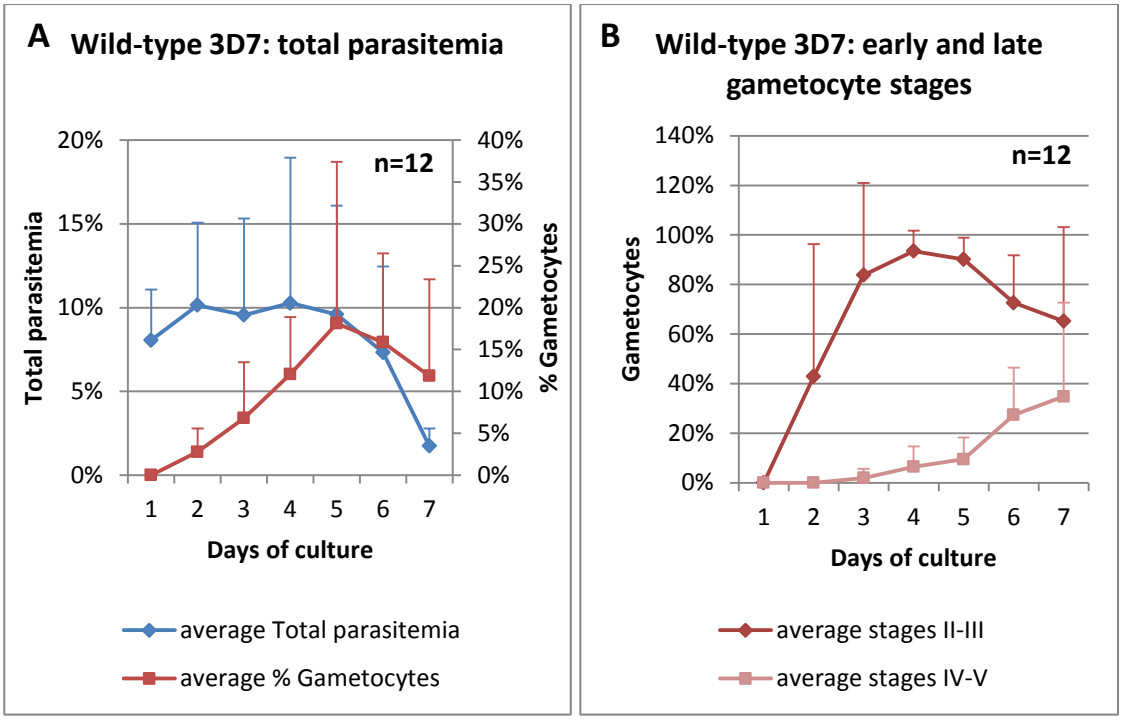


Figure 5.6 Gametocyte induction in wild type 3D7 control parasites.

(A) The total parasitemia and percentage of gametocytes, averaged from 12 experiments. The maximum number of gametocytes produced was 18% of the total parasitemia on day 5 of the gametocyte induction. (B) The percentages of the early and late stages gametocytes, averaged from 12 experiments. On average, 65% of the total gametocytes were early stage (stages II and III), and 35% of the gametocytes were late stage (stages IV and V). The total parasitemia is defined as the total number of asexual and sexual parasites over the total number of RBCs counted. The total number of gametocytes over the total number of RBCs is defined as the gametocytemia. The percentage of gametocytes is defined by the gametocytemia over the total parasitemia. Error bars represent standard deviation of the mean.

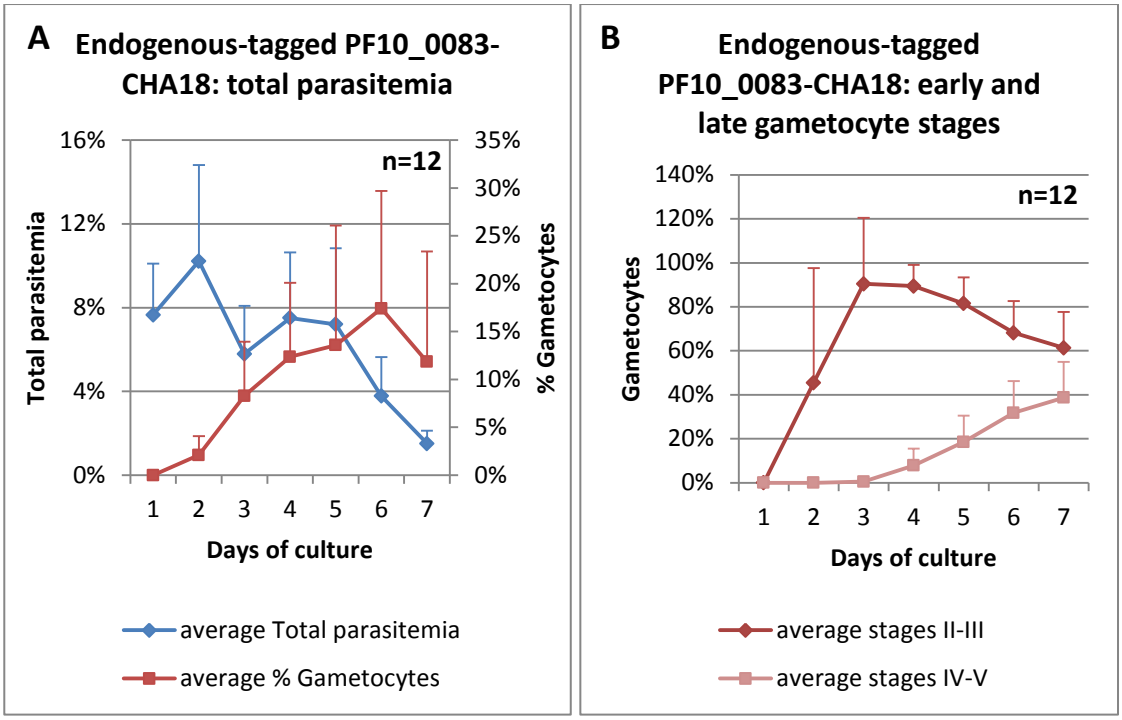


Figure 5.7 Gametocyte induction in the endogenous-tagged PF10_0083-CHA18 control parasites.

(A) The total parasitemia and percentage of gametocytes, averaged from 12 experiments. The maximum number of gametocytes formed was 17% of the total parasitemia on day 6 of the gametocyte induction. (B) The percentages of the early and late stages gametocytes, averaged from 12 experiments. On average, 60% of the total gametocytes were early stage (stages II and III), and 40% of the gametocytes were late stage (stages IV and V). The total parasitemia is defined as the total number of asexual and sexual parasites over the total number of RBCs counted. The total number of gametocytes over the total number of RBCs is defined as the gametocytemia. The percentage of gametocytes is defined by the gametocytemia over the total parasitemia. Error bars represent standard deviation of the mean.

5.3.2 PF10_0083 over-expression enhances gametocyte production

In the 3D7-myc3 parasites, the total parasitemia decreased gradually starting from day 3 to day 5, and the total parasitemia decreased to approximately 1% on day 7 of the experiment. We started to observe gametocytes from day 2 onwards. On day 5, the number of gametocytes peaked. In contrast to wild-type 3D7 and PF10_0083-CHA18,

there were more gametocytes during day 5 of the experiment. Around 29% of the total number of parasites were gametocytes (Figure 5.8A). The difference in the number of gametocytes produced from day 2 to day 7 between wild-type 3D7 and 3D7-myc3 is statistically significant with a p-value 8.5 X 10^{-4} (paired Student's t-test). The difference in the number of gametocytes produced from day 2 to day 7 between PF10_0083-CHA18 and 3D7-myc3 is also statistically significant with a p-value 2.7 X 10^{-3} (paired Student's t-test). Similar to the control cell lines, early stages gametocytes peaked between day 3 and day 5, and we started to observe late stages gametocytes starting from day 4 and day 5. Although there were more gametocytes in the over-expression cell line, the number of early stage gametocytes was much higher than the late stage gametocytes. Approximately 80% of the gametocytes were early stage, and only 20% of the gametocytes were late stage on day 7 of the experiments (Figure 5.8B).

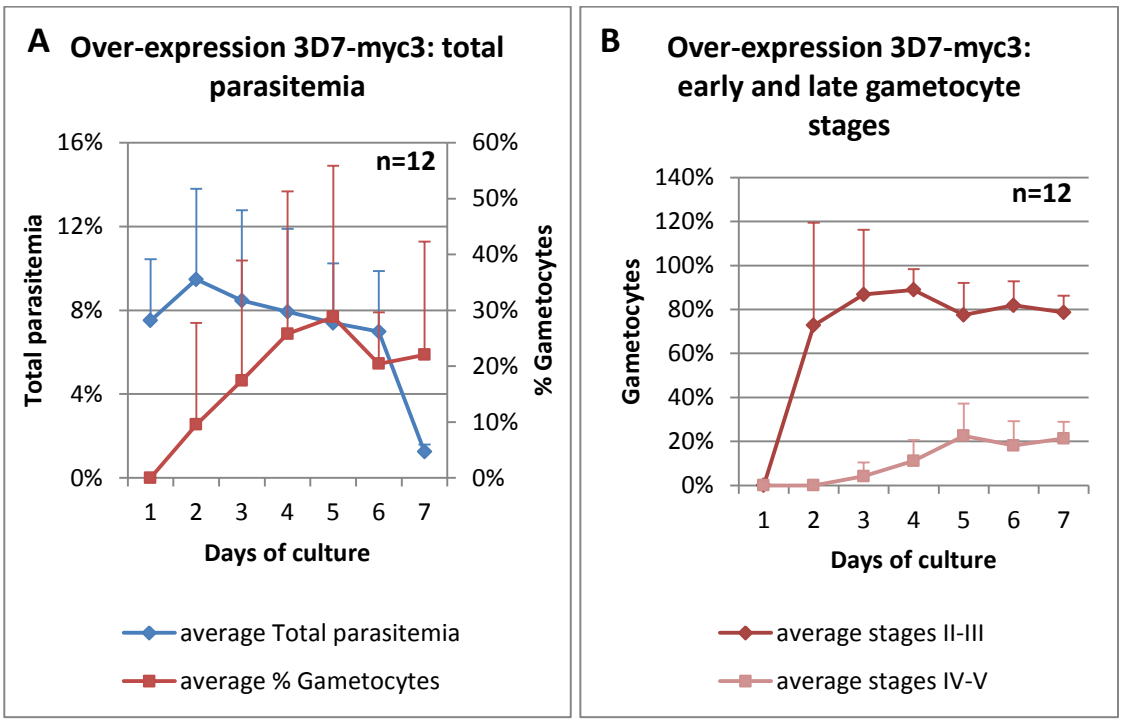


Figure 5.8 Gametocyte induction in 3D7-myc3 over-expression parasites.

(A) The total parasitemia and percentage of gametocytes, averaged from 12 experiments. The maximum number of gametocytes formed was 29% of the total parasitemia on day 5 of the gametocyte induction. (B) The percentages of the early and late stages gametocytes, averaged from 12 experiments. On average, 80% of the total gametocytes were early stage (stages II and III), while only 20% of the gametocytes were late stage (stages IV and V). The total parasitemia is defined as the total number of asexual and sexual parasites over the total number of RBCs counted. The total number of gametocytes over the total number of RBCs is defined as the gametocytemia. The percentage of gametocytes is defined by the gametocytemia over the total parasitemia. Error bars represent standard deviation of the mean.

5.3.3 PF10_0083 knock-out decreases gametocyte production

The PF10_0083-KO33 parasites also contain the hDHFR drug cassette. Therefore, we also cultured this cell line under continuous WR99210 drug pressure. We started to observe gametocytes from day 2 onwards, and on day 4, the number of gametocytes peaked. Contrary to wild-type 3D7, PF10_0083-CHA18 and 3D7-myc3, there were fewer gametocytes during the course of the experiment. Only up to 9.4% of the total number of parasites were gametocytes (Figure 5.9A). The difference in the number of gametocytes produced from day 2 to day 7 between wild-type 3D7 and PF10_0083-KO33 is statistically significant with a p-value 0.0032 (paired Student's t-test). The difference in the number of gametocytes produced from day 2 to day 7 between PF10_0083-CHA18 and PF10_0083-KO33 is also statistically significant with a p-value 0.0031 (paired Student's t-test). The difference in the number of gametocytes produced from day 2 to day 7 between 3D7-myc3 and PF10_0083-KO33 is also statistically significant with a p-value 5.1×10^{-4} (paired Student's t-test). On day 4, the early stage gametocytes peaked, and we started to observe late stage gametocytes. The percentage of early and late stages gametocytes was similar to the 3D7-myc3 over-expression cell line. Approximately 80% of the gametocytes were early stage, and

only 20% of the gametocytes were late stage on day 7 of the experiments (Figure 5.9B).

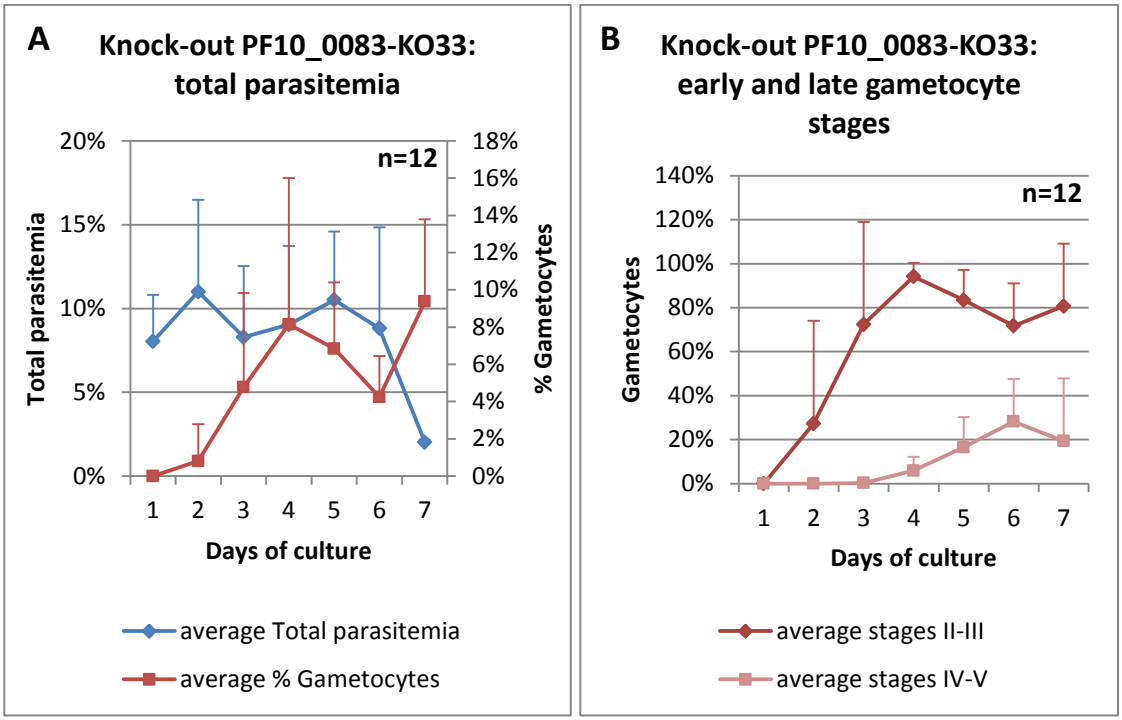


Figure 5.9 Gametocyte induction in PF10_0083-KO33 knock-out parasites.
(A) The total parasitemia and percentage of gametocytes, averaged from 12 experiments. The total parasitemia remained high from day 1 to day 6 and decreased to 2% on day 7 of the experiment. The maximum number of gametocytes formed was only 9.4% of the total parasitemia. (B) The percentages of the early and late stages gametocytes, averaged from 12 experiments. On average, 80% of the total gametocytes were early stage (stages II and III), and 20% of the gametocytes were late stage (stages IV and V). The total parasitemia is defined as the total number of asexual and sexual parasites over the total number of RBCs counted. The total number of gametocytes over the total number of RBCs is defined as the gametocytemia. The percentage of gametocytes is defined by the gametocytemia over the total parasitemia. Error bars represent standard deviation of the mean.

5.3.4 Gametocyte production in other clones of the endogenous-tagged and knock-out cell lines

The continuous culturing of *P. falciparum* under *in vitro* conditions has been associated with the loss of the ability to produce gametocytes [280, 281]. For this reason, we have performed another gametocyte induction on another clone of the endogenous-tagged cell line and the knock-out cell line. We have used the PF10_0083-CHA30 (endogenous-tagged) and PF10_0083-KO24 (knock-out) cell lines alongside wild-type 3D7, PF10_0083-CHA18, over-expression parasites 3D7-myc3 and PF10_0083-KO33 cell lines. It is important to note that PF10_0083-CHA18, PF10_0083-CHA30, PF10_0083-KO24 and PF10_0083-KO33 have been kept in culture for the same amount of time. Five gametocyte induction experiments were carried out. The results are summarized in Figure 5.10.

In wild-type 3D7, PF10_0083-CHA18 and PF10_0083-CHA30, 19.7%, 19.5% and 15.1% of the total number of parasites were gametocytes. In all the 3 cell lines, the number of gametocytes peaked on day 6. The difference in the number of gametocytes produced between wild-type 3D7 and PF10_0083-CHA18, between wild-type 3D7 and PF10_0083-CHA30, and between PF10_0083-CHA18 and PF10_0083-CHA30 is not statistically significant. In contrast, in the over-expression cell line 3D7-myc3, 27.6% of the total number of parasites were gametocytes. The difference in the number of gametocytes produced from day 1 to day 7 between 3D7-myc3 and wild-type 3D7, between 3D7-myc3 and PF10_0083-CHA18, and between 3D7-myc3 and PF10_0083-CHA30 is statistically significant with p-values 0.0056, 0.0057 and 0.0009 respectively (paired Student's t-test). On the other hand, there were fewer gametocytes in the 2 knock-out cell lines. In PF10_0083-KO24 and PF10_0083-KO33, we

observed that only 5% and 5.6% of the total number of parasites were gametocytes. The difference in the number of gametocytes produced between PF10_0083-KO24 and PF10_0083-KO33 is not statistically significant. The difference in the number of gametocytes produced from day 1 to day 7 between PF10_0083-KO24 and 3D7, between PF10_0083-KO24 and PF10_0083-CHA18, between PF10_0083-KO24 and PF10_0083-CHA30, and between PF10_0083-KO24 and 3D7-myc3 is statistically significant with a p-value <0.05 (paired Student's t-test). Similarly, the difference in the number of gametocytes produced from day 1 to day 7 between PF10_0083-KO33 and 3D7, between PF10_0083-KO33 and PF10_0083-CHA18, between PF10_0083-KO33 and PF10_0083-CHA30, and between PF10_0083-KO33 and 3D7-myc3 is statistically significant with a p-value <0.05 (paired Student's t-test). Since PF10_0083-CHA18, PF10_0083-CHA30, PF10_0083-KO24 and PF10_0083-KO33 have been kept in culture for the same amount of time, the results from these experiments strongly suggest that the decreased gametocyte production observed in the knock-out cell lines is a genuine effect resulting from the knock-out of PF10_0083, and not due to a spontaneous mutation that arises from the continuous *in vitro* culturing of the parasites.

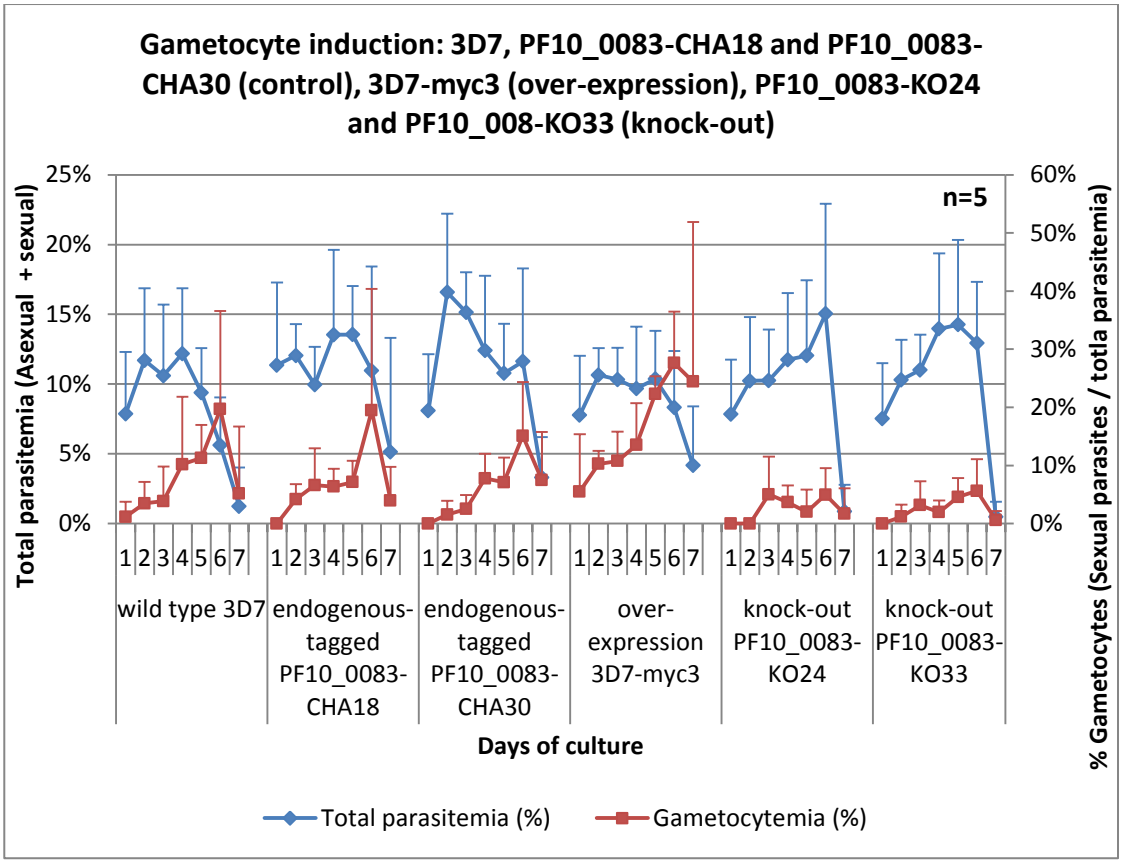


Figure 5.10 Gametocyte induction in 3D7, PF10_0083-CHA18, PF10_0083-CHA30 (endogenous-tagged), 3D7-myc3 (over-expression), PF10_0083-KO24 and PF10_0083-KO33 (knock-out) parasites.

The total parasitemia and percentage of gametocytes, averaged from 5 experiments. In 3D7, PF10_0083-CHA18 and PF10_0083-CHA30, 19.7%, 19.5% and 15.1% of the total number of parasites were gametocytes. In 3D7-myc3, 27.6% of the total number of parasites were gametocytes. In PF10_0083-KO24 and PF10_0083-KO33, 5% and 5.6% of the total number of parasites were gametocytes. The difference in the number of gametocytes produced between wild-type 3D7 and PF10_0083-CHA18, between wild-type 3D7 and PF10_0083-CHA30, and between PF10_0083-CHA18 and PF10_0083-CHA30 is not statistically significant. The difference in the number of gametocytes produced from day 1 to day 7 between 3D7-myc3 and wild-type 3D7, between 3D7-myc3 and PF10_0083-CHA18, and between 3D7-myc3 and PF10_0083-CHA30 is statistically significant with p-values 0.0056, 0.0057 and 0.0009 respectively (paired Student's t-test). The difference in the number of gametocytes produced from day 1 to day 7 between PF10_0083-KO24 and 3D7, between PF10_0083-KO24 and PF10_0083-CHA18, between PF10_0083-KO24 and PF10_0083-CHA30, and between PF10_0083-KO24 and 3D7-myc3 is statistically significant with a p-value <0.05 (paired Student's t-test). The difference in the number of gametocytes produced from day 1 to day 7 between

PF10_0083-KO33 and 3D7, between PF10_0083-KO33 and PF10_0083-CHA18, between PF10_0083-KO33 and PF10_0083-CHA30, and between PF10_0083-KO33 and 3D7-myc3 is also statistically significant with a p-value <0.05 (paired Student's t-test). The total parasitemia is defined as the total number of asexual and sexual parasites over the total number of RBCs counted. The total number of gametocytes over the total number of RBCs is defined as the gametocytemia. The percentage of gametocytes is defined by the gametocytemia over the total parasitemia. Error bars represent standard deviation of the mean.

5.3.5 Gametocyte induction by pyrimethamine treatment

To further validate the findings from above, we have performed another gametocyte induction by pyrimethamine treatment [282]. 1.5% to 2% wild type 3D7 ring stage parasites were prepared as starter culture (day -2). On the next day (day -1), the parasites were treated with pyrimethamine for the next 48 hours. After 48 hours of treatment (day 1), the parasites were washed twice in normal culture media to remove the pyrimethamine. From day 1 to day 7, the total parasitemia and percentage of gametocytes produced were recorded. The parasites were subjected to 3 concentrations of pyrimethamine treatment, 0.05 ng/μl, 0.1 ng/μl and 0.2 ng/μl (Figure 5.11). For 0.05 ng/μl pyrimethamine treatment, we observed that no gametocyte was generated after 7 days of culture. The total parasitemia remained at approximately 9.6% on day 7. For 0.1 ng/μl pyrimethamine treatment, the total parasitemia started to decrease on day 3, and dropped to approximately 7.9% on day 7. We started to observe gametocytes from day 3. The percentage of gametocytes peaked on day 5 at approximately 8.3%. However, the percentage of gametocytes decreased to 6% on day 6 and 1.6% day 7. For 0.2 ng/μl pyrimethamine treatment, the total parasitemia started to decrease on day 3, and dropped to approximately 0.9% on day 5. On day 6, all the parasites died. No gametocyte was observed from day 1 to day 5.

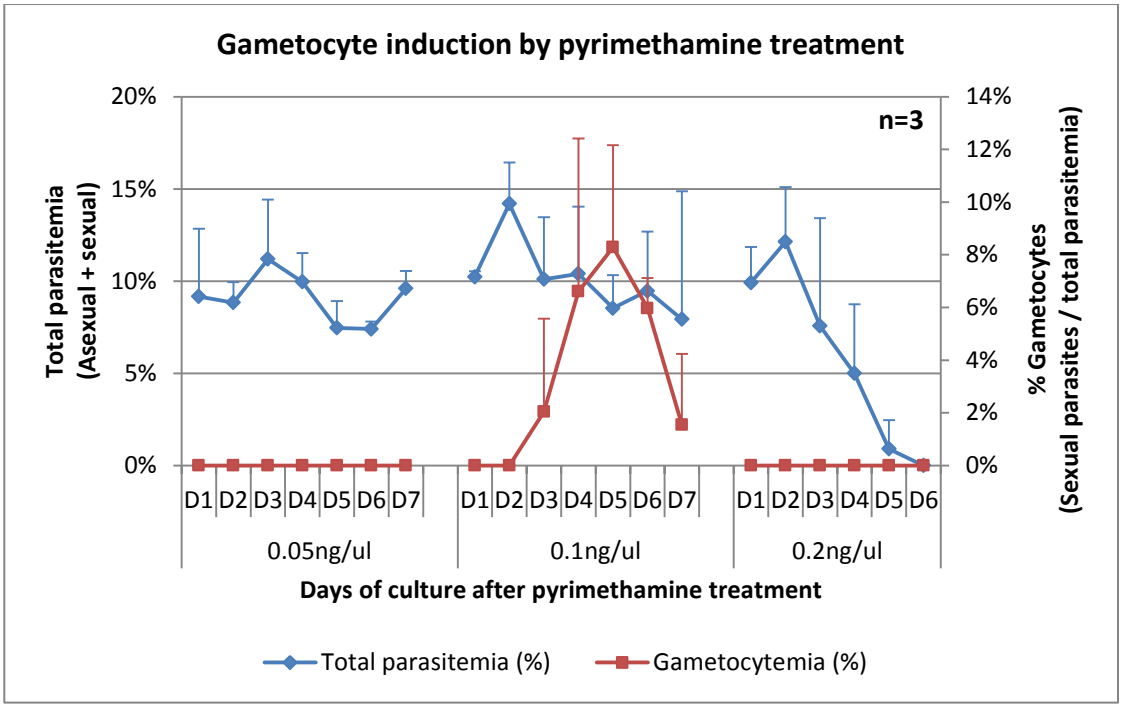


Figure 5.11 Gametocyte induction by pyrimethamine treatment.

1.5% to 2% wild type 3D7 parasites were prepared as starter culture. On the next day, the parasites were treated with pyrimethamine for the next 48 hours. After 48 hours of treatment, the parasites were washed twice in normal culture media to remove the pyrimethamine. The total parasitemia and percentage of gametocytes produced were recorded for the next 7 days. The parasites were subjected to 3 concentrations of pyrimethamine, 0.05 ng/µl, 0.1 ng/µl and 0.2 ng/µl. For 0.05 ng/µl pyrimethamine treatment, no gametocyte was generated after 7 days of culture. The total parasitemia remained at approximately 9.6% on day 7. For 0.1 ng/µl pyrimethamine treatment, the total parasitemia started to decrease on day 3, and dropped to approximately 7.9% on day 7. Gametocytes were observed from day 3. The percentage of gametocytes peaked on day 5 at approximately 8.3%. The percentage of gametocytes decreased to 6% on day 6 and 1.6% day 7. For 0.2 ng/µl pyrimethamine treatment, the total parasitemia started to decrease on day 3, and dropped to approximately 0.9% on day 5. On day 6, all the parasites died. No gametocyte was observed from day 1 to day 5. The total parasitemia is defined as the total number of asexual and sexual parasites over the total number of RBCs counted. The total number of gametocytes over the total number of RBCs is defined as the gametocytemia. The percentage of gametocytes is defined by the gametocytemia over the total parasitemia. The total parasitemia and percentage of gametocytes were averaged from 3 experiments. Error bars represent standard deviation of the mean.

Gametocyte induction with 0.2 ng/μl pyrimethamine was carried out on wild type 3D7 parasites, PF10_0083-CHA18 and PF10_0083-CHA30 endogenous-tagged cell line, 3D7-myc3 over-expression cell line, and PF10_0083-KO24 and PF10_0083-KO33 knock-out cell line. Six gametocyte induction experiments were carried out. The results are summarized in Figure 5.12.

In wild-type 3D7, we started to observe gametocytes from day 3. The number of gametocytes peaked on day 5 at 7.1%. In PF10_0083-CHA18 and PF10_0083-CHA30, gametocytes were observed from day 2. The number of gametocytes peaked on day 4 at approximately 6.4% and 7.3% in PF10_0083-CHA18 and PF10_0083-CHA30, respectively. The difference in the number of gametocytes produced between wild-type 3D7 and PF10_0083-CHA18, between wild-type 3D7 and PF10_0083-CHA30, and between PF10_0083-CHA18 and PF10_0083-CHA30 is not statistically significant. On the other hand, in the over-expression cell line 3D7-myc3, the number of gametocytes peaked on day 7 at approximately 19.9%. The difference in the number of gametocytes produced from day 2 to day 7 between 3D7-myc3 and wild-type 3D7, between 3D7-myc3 and PF10_0083-CHA18, and between 3D7-myc3 and PF10_0083-CHA30 is statistically significant with p-values 0.01, 0.008 and 0.01 respectively (paired Student's t-test). There were fewer gametocytes in the 2 knock-out cell lines. In PF10_0083-KO24, the number of gametocytes peaked on day 7 at approximately 5%. In PF10_0083-KO33, the number of gametocytes peaked on day 4 at approximately 5.2%. The difference in the number of gametocytes produced between PF10_0083-KO33 and PF10_0083-KO24 is not statistically significant. The difference in the number of gametocytes produced from day 2 to day 7 between PF10_0083-KO24 and 3D7, between PF10_0083-KO24 and PF10_0083-CHA18, between

PF10_0083-KO24 and PF10_0083-CHA30, and between PF10_0083-KO24 and 3D7-myc3 is statistically significant with a p-value <0.05 (paired Student's t-test). Similarly, the difference in the number of gametocytes produced from day 1 to day 7 between PF10_0083-KO33 and 3D7, between PF10_0083-KO33 and PF10_0083-CHA18, between PF10_0083-KO33 and PF10_0083-CHA30, and between PF10_0083-KO33 and 3D7-myc3 is statistically significant with a p-value <0.05 (paired Student's t-test). The results from these experiments are in agreement with the above mentioned results that the over-expression of PF10_0083 enhanced gametocyte production, while the knock-out of PF10_0083 decreased gametocyte production.

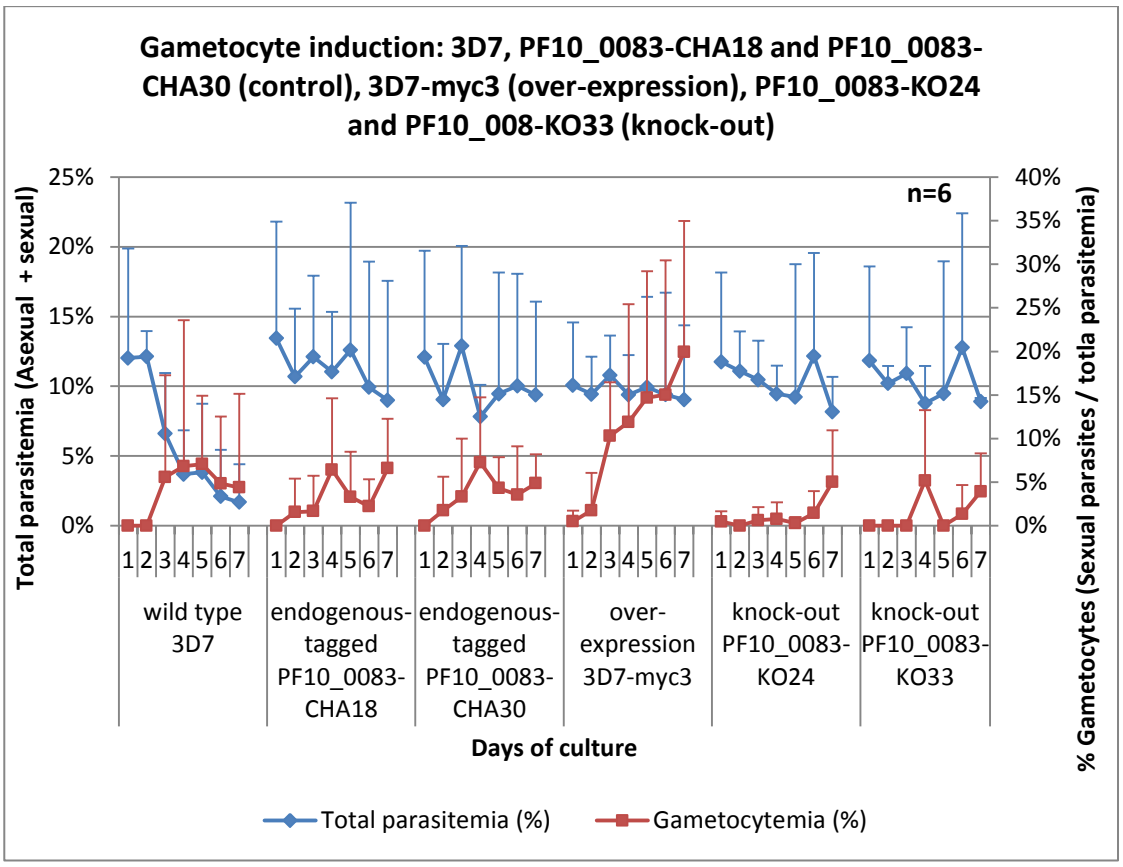


Figure 5.12 Gametocyte induction by pyrimethamine treatment in 3D7, PF10_0083-CHA18, PF10_0083-CHA30 (endogenous-tagged), 3D7-myc3 (over-expression), PF10_0083-KO24 and PF10_0083-KO33 (knock-out) parasites.

The total parasitemia and percentage of gametocytes were averaged from 6 experiments. In 3D7, PF10_0083-CHA18 and PF10_0083-CHA30, the number of gametocytes peaked at approximately 7.1%, 6.4% and 7.3% respectively. In the over-expression cell line 3D7-myc3, the number of gametocytes peaked at approximately 19.9%. In PF10_0083-KO24 and PF10_0083-KO33, the number of gametocytes peaked at 5% and 5.2% respectively. The difference in the number of gametocytes produced between wild-type 3D7 and PF10_0083-CHA18, between wild-type 3D7 and PF10_0083-CHA30, and between PF10_0083-CHA18 and PF10_0083-CHA30 is not statistically significant. The difference in the number of gametocytes produced from day 2 to day 7 between 3D7-myc3 and wild-type 3D7, between 3D7-myc3 and PF10_0083-CHA18, and between 3D7-myc3 and PF10_0083-CHA30 is statistically significant with p-values 0.01, 0.008 and 0.01 respectively (paired Student's t-test). The difference in the number of gametocytes produced between PF10_0083-KO33 and PF10_0083-KO24 is not statistically significant. The difference in the number of gametocytes produced from day 2 to day 7 between PF10_0083-KO24 and 3D7, between PF10_0083-KO24 and PF10_0083-CHA18, between PF10_0083-KO24 and PF10_0083-CHA30, and between PF10_0083-KO24 and 3D7-myc3 is statistically significant with a p-value <0.05 (paired Student's t-test). The difference in the number of gametocytes produced from day 1 to day 7 between PF10_0083-KO33 and 3D7, between PF10_0083-KO33 and PF10_0083-CHA18, between PF10_0083-KO33 and PF10_0083-CHA30, and between PF10_0083-KO33 and 3D7-myc3 is also statistically significant with a p-value <0.05 (paired Student's t-test). The total parasitemia is defined as the total number of asexual and sexual parasites over the total number of RBCs counted. The total number of gametocytes over the total number of RBCs is defined as the gametocytemia. The percentage of gametocytes is defined by the gametocytemia over the total parasitemia. Error bars represent standard deviation of the mean.

5.4 PF10_0083 promotes the differentiation of asexual parasites to sexual parasites

Taken together, gametocyte induction was successful in all the 4 cell lines. In the two control cell lines, wild-type 3D7 and PF10_0083-CHA18, up to 18% and 17% of the total parasitemia were gametocytes. In 3D7-myc3 over-expression cell line, up to 29% of the total parasitemia were gametocytes. On the other hand, the knock-out of

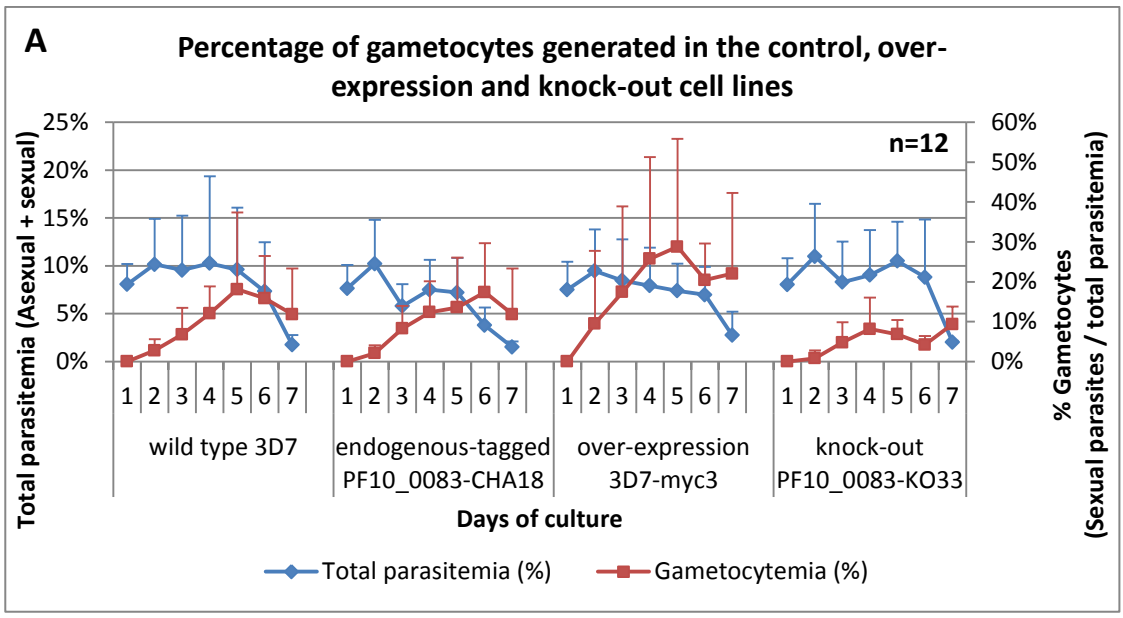
PF10_0083 resulted in less gametocytes being produced. In PF10_0083-KO33, the maximum number of gametocytes produced was only 9.4% of the total parasitemia. The difference in the number of gametocytes produced between the 3D7 and PF10_0083-CHA18 control cell lines is not statistically significant. In contrast, the difference in the number of gametocytes produced from day 2 to day 7 between 3D7 control and 3D7-myc3 over-expression cell lines is statistically significant with a p-value 8.5×10^{-4} (paired Student's t-test). The difference in the number of gametocytes produced from day 2 to day 7 between PF10_0083-CHA18 control and 3D7-myc3 over-expression cell lines is also statistically significant with a p-value 2.7×10^{-3} (paired Student's t-test). The difference in the number of gametocytes produced from day 2 to day 7 between 3D7-myc3 over-expression and PF10_0083-KO33 knock-out cell lines is also statistically significant with a p-value 5.1×10^{-4} (paired Student's t-test) (Figure 5.13A). These results were confirmed in another clone of endogenous-tagged and knock-out cell lines, PF10_0083-CHA30 and PF10_0083-KO24 (section 5.3.4, Figure 5.10). We also validated the results by generating gametocytes using pyrimethamine treatment (section 5.3.5, Figure 5.12). Together, these results showed that the over-expression of PF10_0083 enhanced gametocyte production, while the knock-out of PF10_0083 decreased gametocyte production.

In addition, we observed that in all the cell lines, gametocytes were first observed starting from day 2 and day 3. In the 3D7 and PF10_0083-CHA18 control cell lines, the percentage of gametocytes observed on day 2 was around 2% to 3%. Approximately 7% to 8% gametocytes were observed on day 3. By day 5 and day 6, the percentage of gametocytes increased to approximately 17% to 18%. In contrast, in the 3D7-myc3 over-expression cell line, 9.5% gametocytes was already observed on

day 2, and increased to 17.4% on day 3. By day 4 and day 5, the percentage of gametocytes peaked at around 29%. In PF10_0083-KO33 knock-out cell line, around 0.8% gametocytes was observed on day 2, increased to 5% on day 3 and 8% on day 4, and peaked at only 9.4% on day 7 (Figure 5.13A).

Furthermore, it was also observed that the percentages of early and late stages gametocytes between 3D7 and PF10_0083-CHA18 control cell lines were similar. On average, approximately 60% to 65% of the total gametocytes were stages II and III gametocytes, and 35% to 40% of the total gametocytes were stages IV and V. In both the control cell lines, the number of early stage gametocytes peaked on day 3 and day 4 at approximately 90%. After which, the number of early stage gametocytes decreased to around 60% to 65% on day 7. As the early stage gametocytes decreased from day 3 to day 7, the late stage gametocytes increased and peaked on day 7. In contrast to the control cell lines, in 3D7-myc3 over-expression cell line, the percentages of early and late stages gametocytes were around 80% and 20% respectively. Similar to the control cells, the number of early stage gametocytes peaked between day 3 and day 4 at approximately 87%. However, the early stage gametocytes then decreased slightly to about 80% and remained constant from day 5 to day 7. These findings suggested that throughout the 7 days of experiment, there were new gametocytes forming, thus constant high proportion of stages II and III gametocytes was observed in the 3D7-myc3 over-expression cell line. For the PF10_0083-KO33 knock-out cell line, even though the total gametocyte numbers were less than the over-expression and control cell lines, the proportion of the early and late stages gametocytes was similar to the over-expression cell line (Figure 5.13B).

In all the 4 cell lines, stages II to V gametocytes were observed. The morphologies of the gametocytes in the 4 cell lines were similar. Together, these findings suggested that the maturation of gametocytes from stage I to V was not affected by the over-expression or knock-out of PF10_0083. It is highly possible that the over-expression of PF10_0083 promoted the differentiation of asexual parasites into gametocytes, while the knock-out of PF10_0083 impeded the differentiation of asexual parasites into gametocytes.



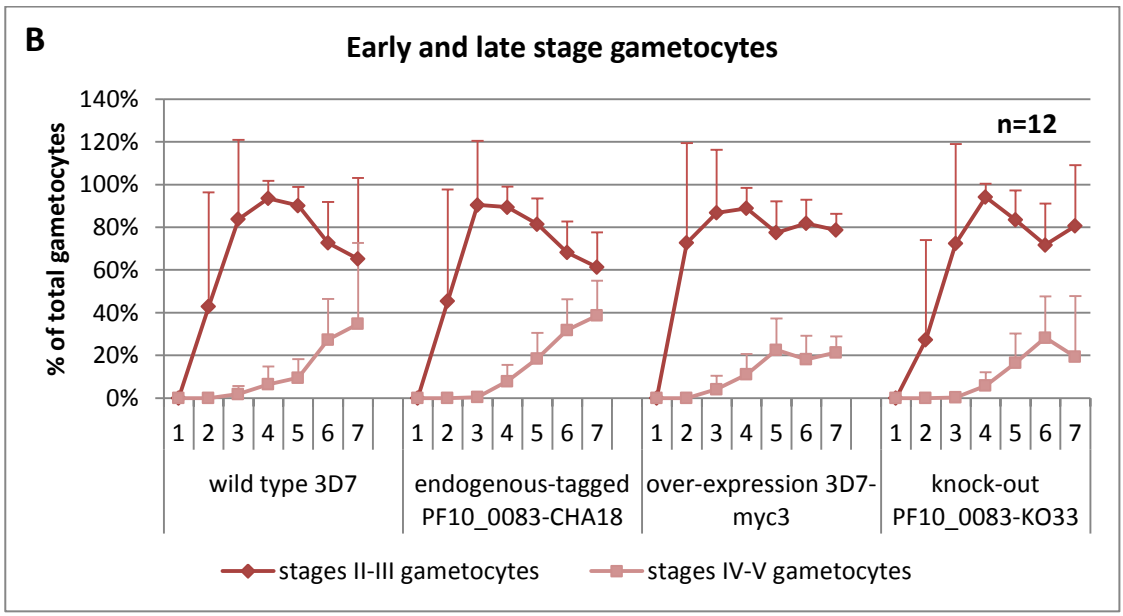


Figure 5.13 Summary of the results of gametocyte induction in 3D7, PF10_0083-CHA18, 3D7-myc3 (over-expression) and PF10_0083-KO33 (knock-out) parasites.

(A) The total parasitemia and percentage of gametocytes produced, averaged from 12 experiments. The difference in the number of gametocytes produced in wild-type 3D7 and PF10_0083-CHA18 is not statistically significant. 3D7-myc3 over-expression cell line generated the most number of gametocytes, while PF10_0083-KO33 knock-out cell line generated the least number of gametocytes. The difference in the number of gametocytes produced from day 2 to day 7 between wild-type 3D7 and 3D7-myc3 is statistically significant with a p-value 8.5 X 10^{-4} (paired Student's t-test). The difference in the number of gametocytes produced from day 2 to day 7 between PF10_0083-CHA18 and 3D7-myc3 is also statistically significant with a p-value 2.7 X 10^{-3}. The difference in the number of gametocytes produced from day 2 to day 7 between wild-type 3D7 and PF10_0083-KO33 is statistically significant with a p-value 0.0032. The difference in the number of gametocytes produced from day 2 to day 7 between PF10_0083-CHA18 and PF10_0083-KO33 is also statistically significant with a p-value 0.0031. The difference in the number of gametocytes produced from day 2 to day 7 between 3D7-myc3 and PF10_0083-KO33 is also statistically significant with a p-value 5.1 X 10^{-4}. The total parasitemia is defined as the total number of asexual and sexual parasites over the total number of RBCs counted. The total number of gametocytes over the total number of RBCs is defined as the gametocytemia. The percentage of gametocytes is defined by the gametocytemia over the total parasitemia. (B) The percentages of the early and late stages gametocytes, averaged from 12 experiments. In 3D7 and PF10_0083-CHA18 control cell lines, 60% of the total gametocytes were early stage (stages II and III), and 40% of the gametocytes were late stage (stages IV and V). In 3D7-myc3 over-expression and PF10_0083-KO33 knock-out cell lines, approximately 80% of the total gametocytes were early stage (stages II and III), and only 20% of the gametocytes were late stage (stages IV and V). Error bars represent standard deviation of the mean.

5.5 PF10_0083 expression level increases in gametocytes

Since the over-expression of PF10_0083 resulted in enhanced gametocytes production, we hypothesized that the expression levels of PF10_0083 would increase in the parasites upon gametocyte induction. As such, we checked the expression levels of PF10_0083 using real-time PCR. We analyzed the mRNA levels of PF10_0083 and sexual stage-specific protein precursor (Pfs16), a gametocyte marker, in the

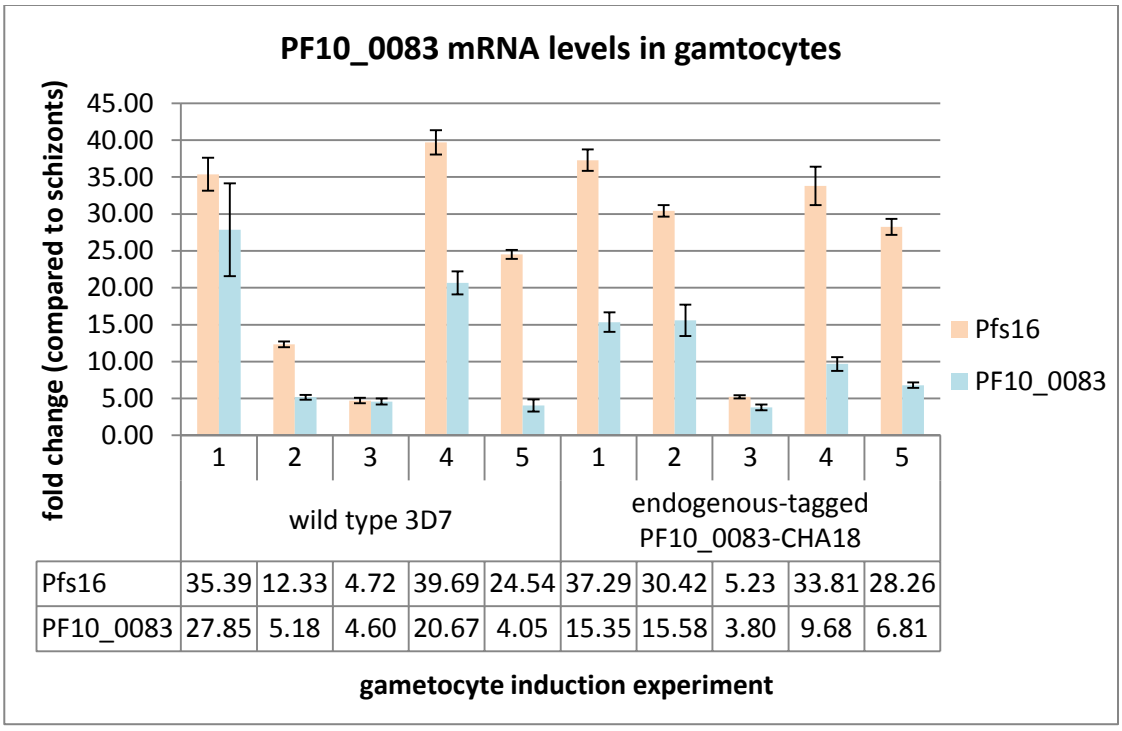
gametocyte cultures and asexual cultures. From the 12 sets of gametocyte induction cultures, we carried out real-time PCR experiments on 5 sets of experiments, G3, G4, G5, G7 and G9. Table 5.1 shows the percentage of gametocytes present in the culture during sampling for real-time PCR, and the day of gametocyte induction experiment during sampling.

Experiment	G3		G4		G5		G7		G9	
Day of harvest	6		7		5		3		4	
Cell line	**3D7**	**CHA 18**	**3D7**	**CHA 18**	**3D7**	**CHA 18**	**3D7**	**CHA 18**	**3D7**	**CHA 18**
Total parasitemia (%)	7	6	7	1	12	11	4	3	2	4
% Gametocytes	9	12	4	4	12	6	17	11	12	18

Table 5.1 Gametocyte samples harvested for real-time-PCR.
The total parasitemia and the percentage of gametocytes in wild type 3D7 and endogenous-tagged PF10_0083-CHA18 (CHA18) control cell lines, and the day of experiments, during sampling. Samples of G3, G4, G5, G7 and G9 were harvested on day 6, day 7, day 5, day 3 and day 4 respectively.

Real-time PCR was performed on the gametocyte stages of the wild type 3D7 and PF10_0083-CHA18 control cell line. Real-time PCR was carried out on cDNA which was reverse transcribed from mRNA. The control sample used was 3D7 schizont stage parasites. The reference control gene used was PFL0900c (arginyl-tRNA synthetase). PFL0900c has a similar expression level in both the gametocytes and schizonts. For each sample, real-time PCR was performed 3 times and the readings were averaged. The fold changes were calculated by the ΔΔCt method. In all the 5 experiments, in both the 3D7 and PF10_0083-CHA18 control cell lines, the gametocyte marker Pfs16 was up-regulated in the gametocytes compared to the schizonts. PF10_0083 was also

up-regulated in the gametocytes compared to the schizonts. In the 3D7 control parasites, PF10_0083 was up-regulated by approximately 5 fold in G4, G5 and G9, 28 fold in G3 and 21 fold in G7. In endogenous-tagged PF10_0083-CHA18 cell line, PF10_0083 was up-regulated by 15 fold in both G3 and G4, 4 fold in G5, 10 fold in G7, and 7 fold in G9. The levels of Pfs16 were always higher than PF10_0083. In G3, G7 and G9, the levels of the Pfs16 and PF10_0083 between 3D7 and CHA18 were comparable. In G4, both the levels of Pfs16 and PF10_0083 were higher in CHA18 than in 3D7. This is probably due to the presence of more asexual parasites in the 3D7 cultures. In G5, both the levels of Pfs16 and PF10_0083 in 3D7 and CHA18 were lower than the samples in G3, G4, G7 and G9. This is also likely because more asexual parasites were present in the cultures in G5 experiment (Figure 5.14). In summary, PF10_003 was up-regulated in the gametocytes.



	1	2	3	4	5	1	2	3	4	5
Pfs16	35.39	12.33	4.72	39.69	24.54	37.29	30.42	5.23	33.81	28.26
PF10_0083	27.85	5.18	4.60	20.67	4.05	15.35	15.58	3.80	9.68	6.81

Figure 5.14 PF10_0083 mRNA levels increased in the gametocyte cultures.
Real-time PCR was carried out to determine the mRNA levels of PF10_0083 and Pfs16, a gametocyte marker, in wild type 3D7 and endogenous-tagged PF10_0083-CHA18 control

cells, from 5 gametocyte induction experiments. Real-time PCR experiments were performed in triplicates. The readings were averaged. The control sample used was 3D7 schizont stage parasites. The reference control gene used was PFL0900c (arginyl-tRNA synthetase). The ΔΔCt method was used to calculate the fold change. ΔΔCt equals to (Ct of test sample target gene - Ct of test sample reference gene) minus (Ct of control sample target gene - Ct of control sample reference gene), where Ct is the threshold cycle. In all the 5 experiments, in both the 3D7 and PF10_0083-CHA18 control cell lines, Pfs16 and PF10_0083 were up-regulated in the gametocytes compared to the schizonts. Error bars represent standard deviation of the mean.

We also performed real-time PCR experiments on gametocytes generated from wild type 3D7, PF10_0083-CHA18 and PF10_0083-CHA30 endogenous-tagged cell lines by pyrimethamine treatment. We analyzed the expression levels of PF10_0083 in the gametocyte cultures and asexual cultures. Real-time PCR was carried out on cDNA which was reverse transcribed from mRNA. The control sample used was 3D7 schizont stage parasites. The reference control gene used was PFL0900c (arginyl-tRNA synthetase). For each sample, real-time PCR was performed 3 times and the readings were averaged. The fold changes were calculated by the ΔΔCt method. Table 5.2 shows the percentage of gametocytes present in the culture during sampling for real-time PCR, and the day of gametocyte induction experiment during sampling. In all the 3 cell lines, PF10_0083 was up-regulated by approximately 3 fold (Figure 5.15).

Experiment	G28		
Day of harvest	7		
Cell line	**Wild type 3D7**	**PF10_0083-CHA18**	**PF10_0083-CHA30**
Total parasitemia (%)	5	5.7	9
% Gametocytes	7.6	5.85	6.58

Table 5.2 Gametocyte samples generated by pyrimethamine treatment harvested for real-time-PCR.

The total parasitemia and the percentage of gametocytes in wild type 3D7 and endogenous-tagged PF10_0083-CHA18 and PF10_0083-CHA30 control cell lines, and the day of experiments, during sampling.

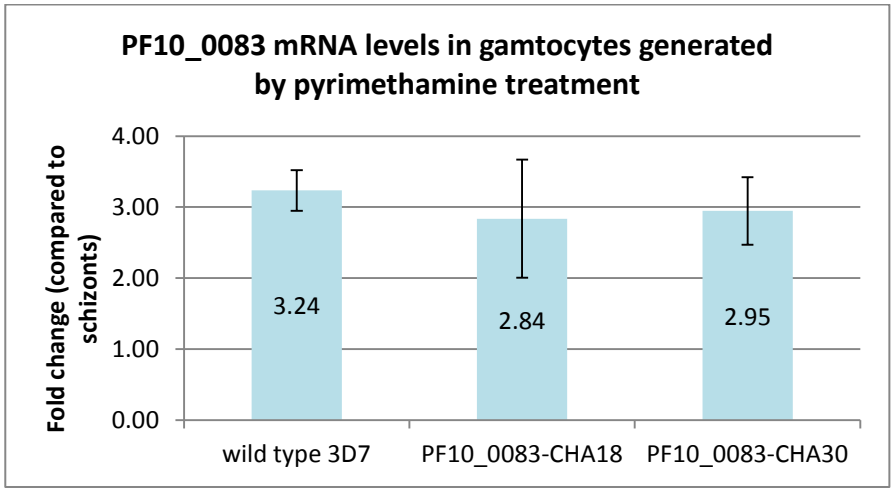


Figure 5.15 PF10_0083 mRNA levels increased in the gametocytes generated by pyrimethamine treatment.

Real-time PCR was carried out to determine the mRNA levels of PF10_0083 in wild type 3D7 and endogenous-tagged PF10_0083-CHA18 and PF10_0083-CHA30 control cells. Real-time PCR experiments were performed in triplicates. The readings were averaged. The control sample used was 3D7 schizont stage parasites. The reference control gene used was PFL0900c (arginyl-tRNA synthetase). The ΔΔCt method was used to calculate the fold change. ΔΔCt equals to (Ct of test sample target gene - Ct of test sample reference gene) minus (Ct of control sample target gene - Ct of control sample reference gene), where Ct is the threshold cycle. In all the 3 cell lines, PF10_0083 were up-regulated in the gametocytes compared to the schizonts. Error bars represent standard deviation of the mean.

5.6 Female to male gametocyte ratio not affected

The mature female gametocytes are characterized by a blue staining cytoplasm. The mature male gametocytes are characterized by a pink staining cytoplasm. In addition, mature female gametocytes are more elongated than the mature male gametocytes. Mature female gametocytes also have a more marked axial curve [17]. We quantified the number of male and female gametocytes on day 7 of gametocyte induction from 3

experiments in wild type 3D7 parasites, PF10_0083-CHA18 endogenous-tagged cell line, 3D7-myc3 over-expression cell line, and PF10_0083-KO33 knock-out cell line. In wild type 3D7, there were approximately 72% female and 28% male gametocytes. In PF10_0083-CHA18, there were approximately 79% female and 21% male gametocytes. In 3D7-myc3, there were approximately 79% female and 21% male gametocytes. In PF10_0083-KO33, there were approximately 76% female and 24% male gametocytes (Figure 5.16). The ratios of female to male gametocytes are summarized in Table 5.3. In wild type 3D7, PF10_0083-CHA18, 3D7-myc3, and PF10_0083-KO33, the average ratios of female to male gametocytes were 2.61:1, 4.38:1, 3.85:1 and 3.81:1 respectively. The difference in the ratios of female to male gametocytes in the 4 cell lines is not statistically significant. We observed that the ratios of female to male gametocytes in the 4 cell lines were between 2.61:1 and 4.38:1. It was previously reported that the sex ratio in *P. falciparum* is female bias at approximately 3-5 female gametocytes to every 1 male gametocytes [283]. As such, the sex ratio is not affected by the over-expression or knock-out of PF10_0083.

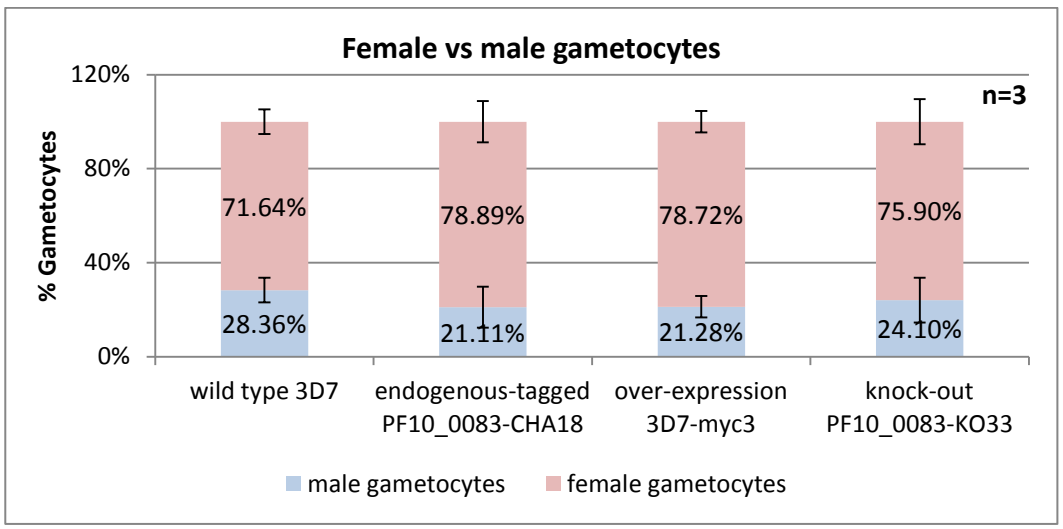


Figure 5.16 Percentages of female and male gametocytes in 3D7, PF10_0083-CHA18, 3D7-myc3 (over-expression) and PF10_0083-KO33 (knock-out) parasites.

The number of male and female gametocytes on day 7 of gametocyte induction from 3 experiments was quantified. In wild type 3D7, there were approximately 72% and 28% female and male gametocytes. In PF10_0083-CHA18, there were approximately 79% and 21% female and male gametocytes. In 3D7-myc3, there were approximately 79% and 21% female and male gametocytes. In PF10_0083-KO33, there were approximately 76% and 24% female and male gametocytes. Error bars represent standard deviation of the mean.

	Ratio of female to male gametocytes			
	Count 1	Count 2	Count 3	**Average**
Wild type 3D7	2.00:1	3.38:1	2.46:1	2.61:1
Endogenous-tagged PF10_0083-CHA18	7.00:1	3.8:1	2.33:1	4.38:1
Over-expression 3D7-myc3	2.88:1	3.67:1	5.00:1	3.85:1
Knock-out PF10_0083-KO33	2.42:1	6.67:1	2.33:1	3.81:1

Table 5.3 The ratio of female to male gametocytes in 3D7, PF10_0083-CHA18 (endogenous-tagged), 3D7-myc3 (over-expression) and PF10_0083-KO33 (knock-out) parasites.

The number of male and female gametocytes on day 7 of gametocyte induction from 3 experiments was quantified. The ratios of female to male gametocytes were calculated and averaged. In wild type 3D7, PF10_0083-CHA18, 3D7-myc3, and PF10_0083-KO33, the average ratios of female to male gametocytes were 2.61:1, 4.38:1, 3.85:1 and 3.81:1 respectively. The difference in the ratios of female to male gametocytes in the 4 cell lines is not statistically significant.

5.7 PF10_0083 expressed only in a small population of schizonts

Only a sub-population of parasites exits asexual multiplication and differentiates into gametocytes. The commitment of asexual parasites to male or female gametocyte development was found to be determined before schizonts maturation and merozoites release [29, 30]. As such, we investigated the number of schizonts that expressed

PF10_0083. We showed that PF10_0083 localized only to the nucleus during the schizont stage (Chapter 3, Figure 3.6 and Figure 3.10). In addition, it was observed that not all the schizonts have fluorescence signal detected. Figure 5.17 illustrates that only some schizonts expressed PF10_0083 (with red fluorescence signals). On the other hand, H3K9Me3 (green fluorescence signals) was detected in every schizont. We counted the number of schizonts that expressed PF10_0083-myc or PF10_0083-CHA in the 3D7-myc3 (over-expression) and PF10_0083-CHA18 (endogenous tagged) cell lines (Table 5.4). We performed 3 independent immunofluorescence experiments with smears prepared on 3 different days. On average, we counted 514 schizonts from each experiment in the 3D7-myc3 over-expression cell line. There were 18 schizonts with positive fluorescence signals. 3.5% of the total number of schizonts counted expressed PF10_0083. In PF10_0083-CHA18 cell line, we counted an average of 757 schizonts from each experiment. Out of which, 13 schizonts, which represented 1.72% of the total number of schizonts, expressed PF10_0083.

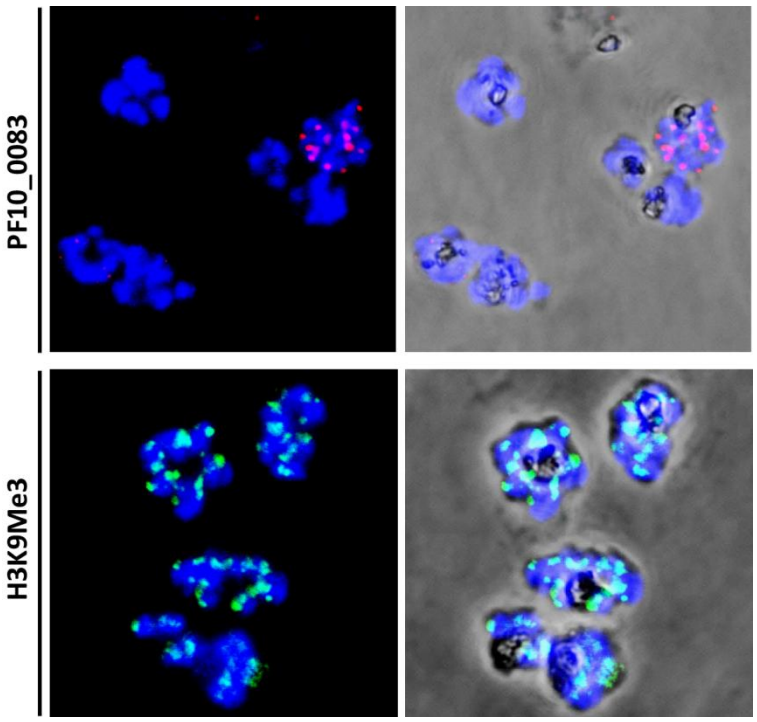


Figure 5.17 PF10_0083 localization in a sub-population of schizonts.
Immunofluorescence assay (IFA) was carried out to determine the intracellular localization of PF10_0083 and H3K9Me3 in the schizonts in the 3D7-myc3 over-expression cell line. IFA was performed with anti-myc antibodies (red) or anti-H3K9Me3 (green), and DAPI stained the nuclei blue. Only a small number of schizonts expressed PF10_0083. Fluorescence signals were detected mainly in the nuclei. On the other hand, H3K9Me3 (green fluorescence signals) was detected in every schizont.

	Over-expression: 3D7-myc3				**Endogenous-tagged: PF10_0083-CHA18**			
	Count 1	Count 2	Count 3	Ave	Count 1	Count 2	Count 3	Ave
Total count	482	499	560	513.7	730	742	800	757.3
Total no. of positive schizonts	16	18	20	18	12	13	14	13
% of positive schizonts	3.32%	3.61%	3.57%	3.50%	1.64%	1.75%	1.75%	1.72%

Table 5.4 The number of schizont stage parasites expressing PF10_0083.
Three independent immunofluorescence experiments were performed with smears prepared on 3 different days. Parasite count in the over-expression cell line 3D7-myc3, and endogenous-tagged cell line PF10_0083-CHA18. The total count of the parasites refers to the total number of schizonts counted. The total number of positive cells refers to the total number of schizonts expressing PF10_0083 (with fluorescence signal) counted. The percentage of schizonts expressing PF10_0083 over the total number of schizonts counted was calculated.

5.8 PF10_0083 localized to the cytoplasm in the gametocytes

An immunofluorescence assay was carried out to determine the localization of PF10_0083 in the gametocytes in the over-expression 3D7-myc3 and endogenous-tagged PF10_0083-CHA18 cell lines. The gametocyte markers, 25 kDa ookinete surface antigen precursor (Pfs25), and gamete antigen 25/27 (Pfg27) were used to identify the gametocytes. Pfs25 is a late gametocyte marker, and is localized to the

parasite surface. Pfg27 is a cytoplasmic protein and is expressed from stage II to stage V gametocytes. Anti-myc stained PF10_0083 red, while anti-HA stained PF10_0083 green. In both 3D7-myc3 and PF10_0083-CHA18, the staining pattern of PF10_0083 in the gametocytes resembled that of Pfg27. As such, we conclude that the majority of PF10_0083 was found to be localized to the cytoplasm in all gametocyte stages (Figure 5.18). However, we observed that PF10_0083 localization in some stage IV and stage V gametocytes (Figure 5.18, seventh and ninth rows) resembled that of the Pfs25 (Figure 5.18, second row). Since Pfs25 is a parasite surface marker, a small amount of PF10_0083 possibly localized to the parasite surface as well. In addition, we also observed that a minute amount of PF10_0083 was localized to the nucleus (Figure 5.18, third, fifth and tenth rows).

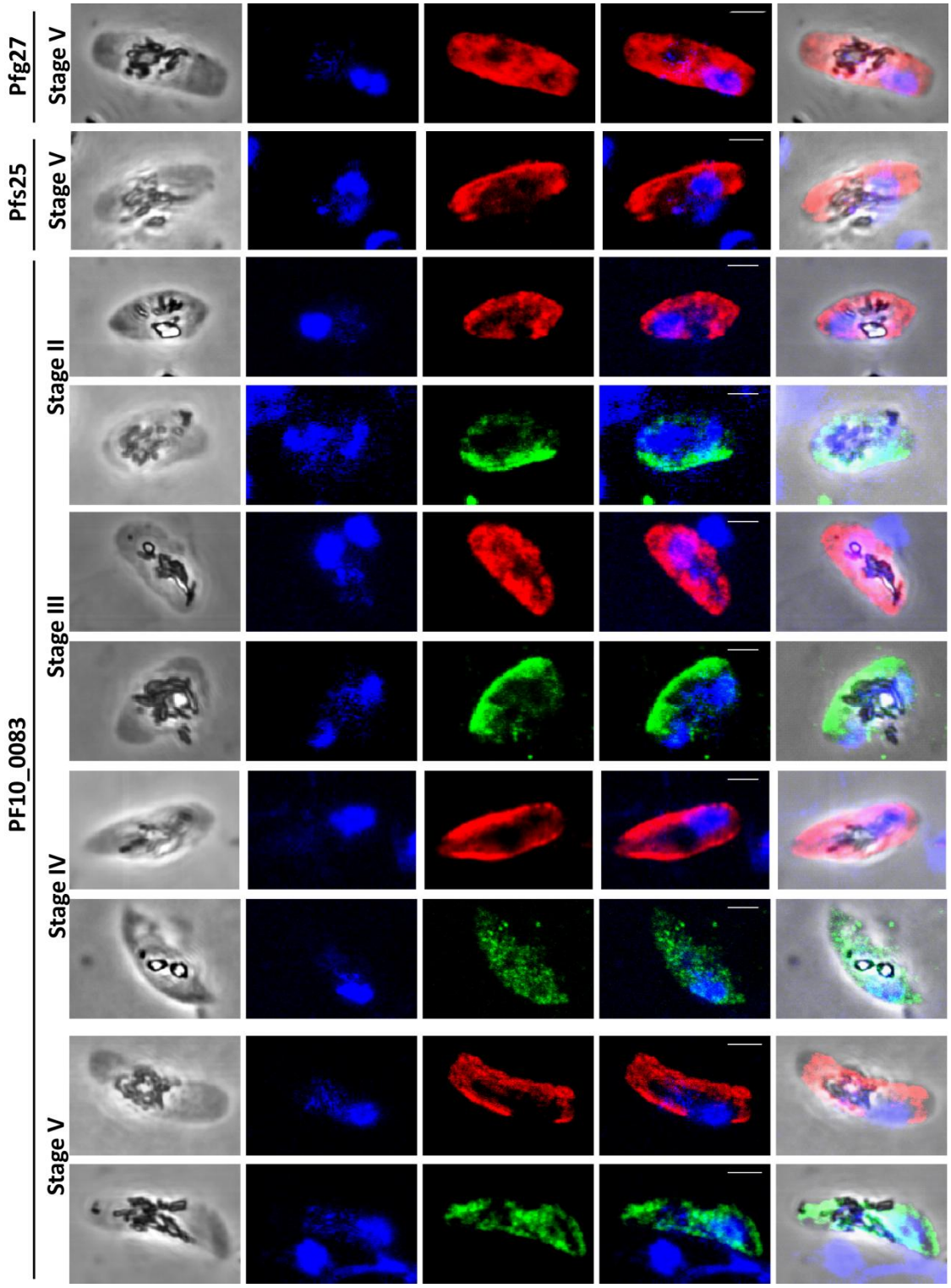


Figure 5.18 PF10_0083 localization in the gametocytes.

Immunofluorescence assay (IFA) was carried out to determine the intracellular localization of PF10_0083, gamete antigen 25/27 (Pfg27) and 25 kDa ookinete surface antigen precursor (Pfs25) in the gametocytes in the 3D7-myc3 over-expression and PF10_0083-CHA18 endogenous-tagged cell lines. IFA was performed with anti-myc (red), anti-HA (green), anti-Pfg27 (red) and anti-Pfs25 (red) antibodies, and DAPI stained the

nuclei blue. Pfg27 and Pfs25 are gametocyte markers. Pfs25 localized to the parasite surface in stage V gametocytes. Pfg27 localized to the cytoplasm in all stages of gametocytes. The staining pattern of PF10_0083 in the gametocytes resembled that of Pfg27. The majority of PF10_0083 was found to be localized to the cytoplasm in all gametocyte stages. PF10_0083 localization in some stage IV and stage V gametocytes (seventh and ninth rows) resembled that of the Pfs25 (second row). A small amount of PF10_0083 possibly localized to the parasite surface. A small amount of PF10_0083 also localized to the nucleus (third, fifth and tenth rows). Scale bar = 2μm.

5.9 Global transcriptional profiling to study the effects of gametocyte induction

To investigate the transcriptional changes caused by gametocyte induction in the control, over-expression and knock-out cell lines, samples were harvested for transcriptional profiling using genome-wide DNA microarrays. Gametocyte samples were harvested on day 6 of gametocyte induction experiment G7. Control schizont stage parasites were also harvested from uninduced control, over-expression and knock-out cell lines. Total RNA was isolated and used to synthesize cDNA for the microarray hybridization experiments. All samples were hybridized against a common reference pool cDNA. The reference pool RNA was harvested every 8 hours for the entire IDC from 3D7 untransfected parasites. The transcriptional profiles of the gametocyte samples obtained by genome-wide DNA microarrays were compared to the uninduced parasites profiles. In order to identify genes that were differentially expressed, the microarray data was analyzed by calculating the Student's t-Test for every gene in the gametocytes and the uninduced parasites. Genes are considered to be differentially expressed if the Student's t-Test $p<0.05$, and has at least a 2 fold difference in the expression between the gametocytes and uninduced parasites.

Table 5.5 shows the percentage of gametocytes present in the culture during sampling for transcriptional profiling. In all the 4 cell lines, wild type 3D7, 3D7-myc3 over-expression, endogenous-tagged PF10_0083-CHA18 and PF10_0083-KO33 knock-out, known sexual stage genes, including Pfs16, Pfs25, Pfs28, alpha tubulin 2 and male development gene 1, were up-regulated in the gametocytes (Figure 5.19). In addition to the gametocyte genes, genes involved in DNA replication, such as DNA polymerase and helicase, expressed in late trophozoites and schizonts, were also up-regulated in the gametocyte cultures. Genes that encode for tricarboxylic acid (TCA) cycle enzymes, such as malate dehydrogenase and pyruvate dehydrogenase, were also up-regulated in the gametocytes. General cellular processes in the asexual stages, such as glycolysis (hexokinase, phosphoglycerate kinase and 6-phosphofructokinase etc.), protein synthesis (translation initiation and elongation factors) and hemoglobin digestion (plasmepsin and falcipain) are normally up-regulated in the rings and trophozoites. These genes were found to be down-regulated in the gametocytes.

Experiment	G4			
Day of harvest	6			
Cell line	**Wild type 3D7**	**Endogenous-tagged PF10_0083-CHA18**	**Over-expression 3D7-myc3**	**Knock-out PF10_0083-KO33**
Total parasitemia (%)	3.4	3.2	3.3	3.4
% Gametocytes	16.7	13.8	33.3	10.5

Table 5.5 Gametocyte samples harvested for transcriptional profiling.
Transcriptional profiling was carried out on wild type 3D7 and endogenous-tagged PF10_0083-CHA18 control cell lines, 3D7-myc3 over-expression and PF10_0083-KO33 knock-out cell lines. The total parasitemia and the percentage of gametocytes in the 4 cell lines and the day of experiment during sampling are shown in the table. In 3D7, PF10_0083-CHA18, 3D7-myc3 and PF10_0083-KO33, the total parasitemia was around

3.2% to 3.4%. There were 16.7%, 13.8%, 33.3% and 10.5% gametocytes in 3D7, PF10_0083-CHA18, 3D7-myc3 and PF10_0083-KO33 cell lines respectively.

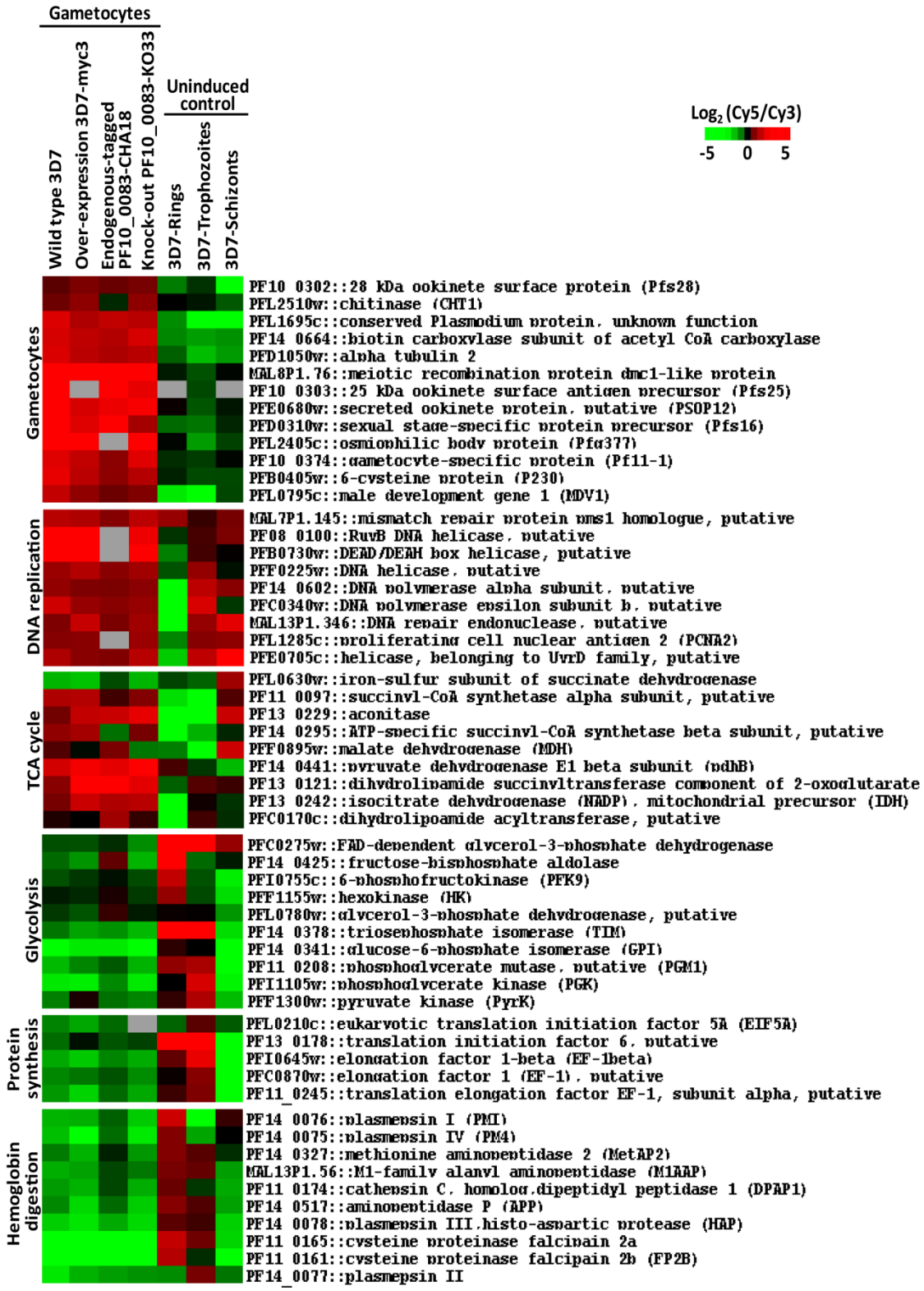


Figure 5.19 Differentially expressed genes in the gametocytes compared to the control schizont parasites.

Transcriptional profiling was carried out on wild type 3D7, 3D7-myc3 over-expression, endogenous-tagged PF10_0083-CHA18 and PF10_0083-KO33 knock-out cell lines. Gametocyte samples were harvested on day 6 of the experiment. Schizont stage parasites were also harvested from uninduced wild type 3D7, 3D7-myc3 over-expression, endogenous-tagged PF10_0083-CHA18 and PF10_0083-KO33 knock-out cell lines. The Student's t-Test for every gene in the gametocytes and the uninduced control cell lines was determined. The Student's t-Test determines whether or not the difference in gene expression between the gametocytes and the uninduced cell lines is statistically significant. Genes are considered to be differentially expressed if the Student's t-Test $p<0.05$, and has at least a 2 fold difference in expression between the gametocytes and uninduced control parasites. Hierarchical clustered genes represent the transcriptional changes elicited from gametocyte induction. Red represents genes that had higher mRNA levels in the gametocytes than the control parasites while green represents genes that had lower mRNA levels in the gametocytes compared to control parasites. Gametocyte genes, such as Pfs16, Pfs25, Pfs28, alpha tubulin 2 and MDV1, were up-regulated in the gametocytes in the 4 cell lines, wild type 3D7, 3D7-myc3 over-expression, endogenous-tagged PF10_0083-CHA18 and PF10_0083-KO33 knock-out. Genes involved in DNA replication were up-regulated in the gametocytes, while genes involved in general cellular processes, such as glycolysis, protein synthesis and hemoglobin digestion, were down-regulated in the gametocytes.

5.10 Summary

In summary, we have successfully induced gametocyte production in 6 cell lines, wild type 3D7, 3D7-myc3 over-expression, PF10_0083-CHA18 (endogenous-tagged), PF10_0083-CHA30 (endogenous-tagged), PF10_0083-KO24 (knock-out) and PF10_0083-KO33 (knock-out). We have successfully induced gametocyte production by nutritional stress and pyrimethamine treatment. We demonstrated that the over-expression of PF10_0083 resulted in enhanced gametocyte production, approximately 10% to 15% more than the control cells. In addition, we demonstrated that the mRNA levels of PF10_0083 increased in the gametocytes in the wild-type and endogenous-tagged control cell lines. In contrast, the knock-out of PF10_0083 resulted in

decreased gametocyte production, approximately half of the gametocyte number in the control cell lines. The impaired gametocyte production in both PF10_0083-KO24 and PF10_0083-KO33 knock-out clones suggests that the decreased gametocyte production is a genuine effect resulting from the knock-out of PF10_0083, and not due to a spontaneous mutation that arises from the continuous *in vitro* culturing of the parasites. However, the knock-out of PF10_0083 did not totally abrogate the gametocyte production. It is apparent that there is more than one factor contributing to gametocyte production. It was also observed that the development of gametocytes from stage I to stage V was normal in 3D7-myc3, PF10_0083KO24 and PF10_0083-KO33 compared to the 3D7, PF10_0083-CHA18 and PF10_0083-CHA30 control cell lines. The ratio of female to male gametocytes was not affected by the over-expression or knock-out of PF10_0083. Interestingly, PF10_0083 was found to be localized to a small percentage of schizonts. Taken together, these results suggest that PF10_0083 probably plays a more active role in the differentiation of asexual parasites into sexual parasites, and not in the maturation of the gametocytes.

Chapter 6 Discussion and conclusion

P. falciparum has a complex life cycle that involves multiple hosts and distinct morphological forms. In addition, *P. falciparum* displays a highly regulated transcriptome. Together, these processes indicate the importance of transcriptional regulation in *Plasmodium* parasites. However, relatively little is known about gene regulation in *Plasmodium*. To date, very few STFs and DNA regulatory motifs have been identified in *Plasmodium*. Amongst the STFs identified, there is a class of ZnF family proteins. In particular, the CCCH-type ZnF proteins are highly over-represented in the *P. falciparum* genome. As such, PF10_0083, a CCCH type ZnF, was characterized.

6.1 PF10_0083 in the asexual stages

6.1.1 Regulation of PF10_0083 protein levels

Four transgenic *P. falciparum* cell lines with PF10_0083 over-expression were prepared. PF10_0083 expression was under the control of a constitutive promoter. Real-time PCR indicated that the mRNA levels of PF10_0083 in rings, trophozoites and schizonts were similar (section 3.3.2.1, Figure 3.4). PF10_0083 was found to be expressed throughout the entire IDC at the mRNA level. PF10_0083 was also found to be expressed throughout the entire IDC at the protein level (section 3.3.2.2, Figure 3.5). In contrast to the mRNA levels, the peak protein levels were at the 32 and 40 hpi of IDC. However, in the endogenous-tagged cell line, PF10_0083 was only detected in the schizont stages but undetectable in the ring and trophozoite stages at the protein level (section 3.4.2, Figure 3.9). As such, we concluded that the protein levels of

PF10_0083 is the lowest in the ring and trophozoite stages, and peaks during the schizont stage in both the over-expression and the endogenous-tagged cell lines.

The peak protein levels coincide with the DNA replication timings of the parasites during late trophozoite and schizont stages. PF10_0083 might be involved in specialized functions or parasite growth and development. The level of PF10_0083 at different stages in the IDC is likely to be developmentally regulated. In eukaryotes, protein kinases and its inhibitors that function to regulate TFs during cell cycle progression are known to have high protein turnover rates, and the disruption of their level causes cancer [284]. For example, β-catenin, which is part of the Wnt signaling pathway, is normally targeted for degradation. Wnt signaling stabilizes β-catenin and allows it to enter the nucleus and assist in the transcription of Wnt-responsive genes. Abnormal regulation of β-catenin de-regulates embryonic development and causes carcinogenesis [285]. Since PF10_0083 is a putative TF, its endogenous steady state level must be tightly regulated and its over-expression might be unfavourable to the parasites. This is further exemplified by the observation that the over-expression cell lines displayed a lower growth rate when compared to the control, endogenous-tagged and knock-out cell lines (section 3.3.1, Figure 3.3 and section 3.5.1, Figure 3.16). It is likely that high level of PF10_0083 is unfavourable during asexual development. Consequently, the parasites regulate the protein level of the over-expressed PF10_0083 in order to develop normally.

6.1.2 PF10_0083 localized to both the cytoplasm and nucleus in the asexual stages

In addition, in the over-expression cell lines, it was observed that PF10_0083 was localized mainly to the cytoplasm in the rings, but gets transported into the nucleus during trophozoite stage, and was enriched in the nucleus in the schizonts (section 3.3.3, Figure 3.6). In eukaryotes, many TFs are not constitutively localized to the nuclei, but require appropriate activating signals in order to relocate from the cytoplasm to the nucleus [286]. For example, the SWI5 is a cell cycle regulated zinc finger protein which functions to activate several G_1-specific genes. SWI5 is produced throughout the cell cycle, but only relocates to the nucleus when it is dephosphorylated by CDC28 kinase during anaphase, and is rapidly degraded during the G_1 phase [287]. In plants, a CCCH type ZnF, AtTZF1, was demonstrated to be able to shuttle between the cytoplasm and nucleus. AtTZF1 was shown to bind to mRNA in the cytoplasm and repress mRNA translation. AtTZF1 was also shown to bind to DNA in the nucleus, and is involved in both DNA and/or RNA regulation [288]. Therefore, in the over-expression cell lines, the localization of PF10_0083 both in the cytoplasm and nucleus during the early stages and only in the nucleus in the late stages could be due to a developmental trigger that shifts PF10_0083 into the nucleus at a precise time during the 48-hour IDC. However, in the endogenous-tagged cell line, PF10_0083 was only detectable in the schizont stage but undetectable during the ring and trophozoite stages by IFA (section 3.4.2, Figure 3.10 and Figure 3.11). As such, we are unable to draw the same conclusion in the endogenous-tagged cell line. It is also possible that the translocation of PF10_0083 from the cytoplasm to the nucleus is just a normal transfer that must occur after the nuclear protein has translated in the cytoplasm. Therefore, more studies are required before we can confirm whether the transport of PF10_0083 into the nucleus during schizont stage is due to a developmental trigger.

6.1.3 PF10_0083 lacks nuclear localization signal

A protein can get transported into the nucleus if it contains a nuclear localization signal (NLS), binds to another protein that contains a NLS, or it is small enough to overcome the diffusion limits of nuclear pore complexes [289]. In eukaryotes, many proteins that lack NLS but localize to the nucleus have been identified. For example, actin forms a complex with cofilin and enters nucleus via the nuclear pore complexes [290]. There are also factors that function to promote and inhibit nuclear translocation. For example, NF-κB is prevented from nuclear uptake by binding to a group of IκB inhibitors. The phosphorylation of IκB frees NFκB to be translocated to the nucleus [291]. It is interesting to note that PF10_0083 lacks a NLS. It is of interest to investigate how PF10_0083 gets translocated into the nucleus. Therefore, further studies to identify potential binding partners of PF10_0083 are required. We carried out protein co-immunoprecipitation using an antibody targeted against the tagged PF10_0083 in the over-expression cell lines. However, we were unable to pull down even the targeted protein (Appendix, Supplementary Figure 4). As such, attempts to express the full length PF10_0083 protein were made. Unfortunately, we were unable to produce large quantities of soluble full length protein (section 3.6). As a result, we are still unable to carry out protein co-immunoprecipitation using a recombinant protein. Further optimizations of these experiments are required.

6.1.4 PF10_0083 localized to distinct compartments in the nuclear periphery in the late asexual stages

Nuclear compartmentalization has also been implicated in gene regulation (refer to section 1.12). Euchromatin and heterochromatin are organized into different compartments and are replicated during different phases of the cell cycle. In

Saccharomyces cerevisiae, the silenced telomeres are localized to the nuclear periphery [292]. Compartments which contain proteins that block the spread of heterochromatin are also located at the nuclear periphery [293]. In eukaryotes, nuclear foci with active transcription machineries are also localized to the nuclear periphery [247-250]. In *Plasmodium*, repressive and activating histone modifications are localized to distinct compartments within the nucleus [266, 267]. Similar to other eukaryotes, active epigenetic marks are mostly found in the nuclear periphery in *P. falciparum*. We observed that PF10_0083 localized to distinct compartments near the nuclear periphery. As such, co-localization studies of PF10_0083 with known heterochromatin and euchromatin markers during schizont stage were carried out (section 3.4.3, Figure 3.12). It was observed that PF10_0083 does not co-localize with the heterochromatin marks, including HP1, H3K9Me3 and H3K79Me3. Although PF10_0083 is localized to the nuclear periphery, it was also not co-localized with known euchromatin marks, including H3K9Ac and H4K5Ac. Although the list of nuclear markers used in this study is not exhaustive, the results presented here strongly suggest that PF10_0083 localized to a unique and distinct compartment in the nucleus during schizont stage, which is neither silent nor actively transcribed. More studies are required in order to fully understand the implications of PF10_0083 within the nucleus. Specifically, the nuclear compartments which PF10_0083 resides in the ring and trophozoite stages remain to be determined. In addition, given the long 48 hours asexual life cycle, the precise timing of PF10_0083 compartmentalization might be crucial. As such, co-localization studies with the heterochromatin and euchromatin markers can also be done on young-, mid- and late schizonts. Since PF10_0083 is likely to be involved in the commitment of the asexual parasites to gametocytogenesis,

it would be of interest to determine the localization of PF10_0083 after the parasites are subjected to stress and before differentiating into gametocyte.

6.2 PF10_0083 binds DNA

6.2.1 PF10_0083 binds preferentially to the IGRs, and IGR binding is positively correlated with transcriptional activity

We performed DNA (section 4.3) and RNA (results not shown) ChIP-on-chip experiments using the transgenic parasite cell lines. Unlike most characterized CCCH ZnF proteins, PF10_0083 was found to bind DNA. On the other hand, we did not find evidence of PF10_0083 binding to RNA (results not shown). In the over-expression cell lines, PF10_0083 was found to associate with the genome in multiple stages during the IDC. PF10_0083 was found to be associated with more genetic loci during the early stages of the IDC. PF10_0083 was found to be associated with the least number of genetic loci during the late stages of the IDC (section 4.3.1, Figure 4.6 and Figure 4.7). It is interesting that although PF10_0083 was enriched in the late asexual stages, it was observed that PF10_0083 was associated to the genome most weakly during the schizont stage. To our surprise, we observed that PF10_0083 associated with more genomic loci during the ring stage than in the trophozoite and schizont stages in the endogenous-tagged cell lines, even thought we were unable to detect the protein by western blot and IFA analysis. It has been previously reported that in the early stage parasites, DNA is associated with lower nucleosome occupancy and is therefore more accessible to TFs binding. On the other hand, in the schizont stage parasites, DNA is associated with higher nucleosome occupancy and is therefore less accessible to TFs binding [294]. As such, we speculate that the higher number of

genetic loci associated with PF10_0083 in the rings might be a result of lower nucleosome occupancy and greater accessibility to TFs in general.

It was also observed that PF10_0083 binds to both the IGRs and ORFs regions on the genome. However, PF10_0083 displayed a positional bias for the IGRs than the ORFs (section 4.3.3, Figure 4.10 and Figure 4.11). In addition, the binding of PF10_0083 to the IGRs was positively correlated with transcriptional activity. On the other hand, the binding of PF10_0083 to multiple regions on the ORFs was correlated with transcriptional repression (section 4.4, Figure 4.12). In addition, we found that PF10_0083 binds to the IGRs of a group of genes which are normally most highly expressed in the gametocytes during the parasite life cycle (section 4.4.1, Figure 4.16). The binding of PF10_0083 to the IGRs of these gametocyte genes (ETRAMP 10.3, SOAP, tubulin-tyrosine ligase, nucleoside transporter, and Pfs16) correlated with their transcriptional up-regulation (section 4.4.1, Figure 4.17 and Figure 4.18). In addition, we identified 3 DNA sequence motifs that PF10_0083 binds to (section 4.4.2, Figure 4.19). These motifs are enriched in the 1 kb region upstream of the genes that were transcriptionally up-regulated. The RNA profiles of these genes were positively correlated to their corresponding ChIP-on-chip profiles. However, the binding of PF10_0083 to these sequences require further validation such as electrophoretic mobility shift assay (EMSA).

In summary, PF10_0083 binds to the ORFs of metabolic genes and correlated with the down-regulation of these genes, seemingly preparing the parasite for quiescent gametocyte state. At the same time, PF10_0083 binds to the IGRs of the gametocyte

genes and correlated with their up-regulation, preparing the parasite for gametocyte differentiation.

6.2.2 The binding of PF10_0083 to non-regulatory targets

The eukaryotic model of transcriptional regulation hypothesizes that TFs bind to DNA regulatory sequences, and recruit the RNA polymerase complex via direct interaction with the RNA polymerase or via indirect interaction with co-factors or other regulatory proteins. This eukaryotic model suggest that the binding of TFs directly correlates with gene transcription. For example, in yeast, the repressor-activator protein 1 (Rap1) was found to bind specifically to the promoter regions and not the ORFs regions, although both regions contain the DNA motif that Rap1 recognizes [295]. In another study, the yeast Ste12 TF was found to bind function-specific genes during different developmental conditions, and regulates mating and filamentous growth [296]. In *C. elegans*, the PHA-4 directly binds to and regulates the timing of expression of the pharyngeal genes at different developmental stages [297]. In humans, a zinc finger protein, the neuron-restrictive silencer factor (NRSF), was found to both activate and repress transcription of the genes that it directly binds to [298].

It was observed from the ChIP-on-chip studies that PF10_0083 binds to a large number of sites on the genome. The widespread binding of TFs has also been described in the other organisms. In human, the Myc TF was found to bind to both specific and non-specific DNA sequences in a large number of genes. Amongst these genes, only a few genes were proven to be Myc-direct targets for transcription [299]. Another human protein, interferon-γ responsive TF, STAT1, was found to bind to 41,000 genome sites in interferon-γ stimulated cells. It was found that there is no direct

relationship between binding at most of these sites and transcriptional activity [300]. In *Drosophila melanogaster*, genome-wide ChIP-on-chip studies have been carried out on embryo dorso-ventral determinants, Dorsal, Twist, and Snail. These studies identified thousands more genes than the expected target genes. In addition, these genes were found to contain highly conserved motifs. However, the biological significance of TFs binding to these genes is still unclear [301-303].

It was suggested that the widespread binding of TF at these unspecific sites are random and serves no biological functions [303]. As such, several explanations and models have been proposed for the binding of genes that the TF does not directly regulate [300]. The first model proposed that the binding of TFs to non-regulatory sites is used as a method to control the concentration and availability of TFs. The bound TFs serve as a reservoir that can be mobilized in a short time [304]. This model was supported by the observation that most of the regions not directly regulated by the TFs were bound at low affinity [303]. This model was further supported by studies that demonstrated the binding of functionally-unrelated TFs to the same sites [305]. It has also been found that such binding resulted in low expression levels that are sufficient for its biological functions [306]. The next model proposed that the binding of TF to these non-specific sites acts to regulate gene transcription at a distance. Changes of the chromosome structure bring the TF and its target gene to close proximity. This model was supported by a study that demonstrated that the formation of the chromosome loop is responsible for the physical interaction between β-globin locus control region and the β-globin promoter. This interaction resulted in β-major globin gene transcription. The hematopoietic transcription factor GATA-1 and its cofactor FOG-1 act as anchors to facilitate this interaction [307]. This model was supported by another

study that demonstrated that chromosome looping brought about the interaction of Wnt responsive enhancers and the T cell factor family of transcription factors to promote Myc transcription [308]. Some TFs and regulatory proteins have also been suggested to affect gene accessibility by re-organizing chromatin [309].

PF10_0083 binds to a large number of sites on the genome. However, the binding of PF10_0083 at the genome is both positively and negatively correlated with transcription up- and down-regulation. In addition, there was also binding of PF10_0083 at the genome that had no correlation with transcription up- or down-regulation. As such, the functional relevance of wide-spread PF10_0083 binding at the genome is unknown. Further studies are required in order to completely understand the functional relevance of PF10_0083 binding to these non-regulatory targets.

6.3 PF10_0083 over-expression promotes gametocyte production

6.3.1 PF10_0083 over-expression resulted in the up-regulation of gametocyte-specific genes

Transcriptome analysis of the over-expression cell lines was carried out (section 4.2). General cellular processes such as DNA replication, transcription and protein translation were differentially up-regulated or down-regulated in the different cell lines. Comparison of the transcriptome profiles of the 4 over-expression cell lines failed to produce a reproducible set of transcriptionally altered genes. We speculated that different parasite populations (clones) could be selected during the different transfection experiments. It is also possible that the level of PF10_0083 in the 4 over-expression cell lines were different, which may be due to a difference in the plasmid copy number. In addition, the criteria used for selecting genes that are differentially

expressed might be too stringent. Lastly, the transcriptional changes observed in the 4 over-expression cell lines could be a general de-regulation as a result of PF10_0083 transgene over-expression.

In the T996-myc1, T996-gfp and 3D7-myc3 cell lines, the majority of the up-regulated genes are most highly expressed in the gametocytes, while the majority of the down-regulated genes are most highly expressed in the early rings. However, an exception is the T996-myc2 where the majority of the down-regulated genes are most highly expressed in the gametocytes and the majority of the up-regulated genes most highly expressed in the early rings (section 4.2.2, Figure 4.5 and Table 4.2). In PF10_0083-CHA18, none of these gametocyte specific genes show a transcriptional change greater than 2 fold. In PF10_0083-KO33, only 6-cysteine protein (P230) was observed to be up-regulated out of these gametocyte specific genes. As such, the up-regulation of these gametocyte specific genes was possibly a result of the over-expression of PF10_0083. The transcriptional profiling of the 4 over-expression cell lines was performed only once. As such, we propose to carry out the transcriptional profiling of the 4 over-expression cell lines again, especially T996-myc2, to confirm the findings. In addition, we propose to carry out the transcriptional profiling on parasites that have been stressed to induce gametocyte production. It would be of interest to find out when PF10_0083 activate the switch of these gametocyte specific genes, and which genes gets up-regulated first after gametocyte induction.

In the T996-myc1, T996-gfp and 3D7-myc3 cell lines, a high percentage of genes that are normally most highly expressed in the gametocyte stage during the parasite life cycle. Although these gametocyte-related genes were transcriptionally altered, no

gametocytes were observed in the parasite cultures. Instead, it was observed that the over-expression of PF10_0083 resulted in a lower growth rate of the parasites. On the other hand, the knock-out of PF10_0083 did not have any observable effects on the parasites. The endogenous-tagged cell line was not observed to have an impaired growth rate too (section 3.3.1, Figure 3.3, and section 3.5.1, Figure 3.16). As such, the up-regulation of the gametocyte genes could also be a general stress response and secondary effect resulting from the over-expression of a trans-gene from a heterologous promoter.

6.3.2 PF10_0083 over-expression promotes differentiation of asexual blood stages parasites to gametocytes

The up-regulation of gametocyte genes in the over-expression cell lines also suggested that PF10_0083 is likely not essential in the asexual blood stages. Instead, PF10_0083 more likely affected parasite development during the sexual stages, and/or during the transition from asexual to sexual stages. As such, gametocyte induction experiments were performed. We subjected the parasites to stress conditions (section 5.2) to induce gametocyte production.

We observed successful gametocyte production in all the cell lines (wild-type, endogenous-tagged, over-expression and knock-out). We observed that the wild-type 3D7 and 2 clones of endogenous-tagged cell lines produced the same percentage of gametocytes (section 5.3.1, Figure 5.6 and Figure 5.7, and section 5.3.4, Figure 5.10). In contrast, the over-expression of PF10_0083 resulted in more gametocyte production (section 5.3.2, Figure 5.8). On the other hand, the knock-out of PF10_0083 resulted in less gametocyte production compared to the control cell lines (section 5.3.3, Figure 5.9,

and section 5.3.4, Figure 5.10). These results were further confirmed by gametocyte induction experiments by pyrimethamine treatment (section 5.3.5, Figure 5.12).

In addition, 2 clones of endogenous-tagged cell lines and 2 clones of knock-out cell lines, kept in culture for the same amount of time were used for gametocyte induction (section 5.3.4, Figure 5.10). It was observed that the 2 clones of endogenous-tagged cell lines produced the same percentage of gametocytes compared to wild type cell line. In addition, the 2 clones of knock-out cell lines produced less gametocytes compared to the wild type and endogenous-tagged cell lines. Together, these results suggested that the decreased gametocyte production in the knock-out cell lines was caused by the knock-out of PF10_0083, and not a result of other spontaneous mutations that arise from long term asexual propagation. We also observed that the mRNA levels of PF10_0083 increased in the gametocyte cultures of the wild type and endogenous-tagged cell lines (section 5.5, Figures 5.14 and 5.15). These results again suggested that the differentiation of the asexual stages to gametocytes is associated with an increase in PF10_0083 level.

We also observed the progression of stages II to V of the gametocytes in the cultures. The maturation of gametocytes from stage I to V was not affected by the over-expression or knock-out of PF10_0083 (section 5.4, Figure 5.13B). The morphologies of the gametocytes in the over-expression and knock-out cell lines were no different from the wild type and endogenous-tagged cell lines. However, in the PF10_0083 over-expression cell line, the percentages of early stage gametocytes produced were more than the control cell lines, while the percentages of late stage gametocytes produced were less than the control cell lines. We speculate that this is likely because

there were new gametocytes being produced throughout the gametocyte induction experiment, hence there was a constant high proportion of stages II and III gametocytes. It was observed in the knock-out cell line that, on average, there were also a higher percentage of early stage gametocytes when compared to the control cell lines during the last day of the gametocyte induction experiment. We speculate that this is likely because the differentiation of asexual parasites into sexual parasites might be slowed down in the absence of PF10_0083. From these results, we concluded that the over-expression of PF10_0083 promoted the production of gametocytes under stress conditions, while the knock-out of PF10_0083 impeded the differentiation of asexual parasites into gametocytes. We also observed that the sex ratio of the gametocytes in the wild-type, endogenous-tagged, over-expression and knock-out cell lines was not affected by the over-expression and knock-out of PF10_0083 (section 5.6, Figure 5.16 and Table 5.3).

Together, we demonstrated that PF10_0083 over-expression promoted the differentiation of asexual blood stage parasites to sexual stage parasites. Hence, a higher percentage of gametocytes were produced in the over-expression cell line than the control cell lines. These findings suggested that PF10_0083 plays a role in gametocytogenesis.

6.3.3 Transcriptional alteration of sexual gametocyte genes in the gametocyte cultures

Transcriptome analysis demonstrated that a high percentage of genes that are normally most highly expressed in the gametocyte stage were affected by PF10_0083 over-expression. Gametocyte genes such as Pfs16, male development gene 1, and alpha

tubulin 2 were amongst the top 10% most up-regulated genes in the over-expression cell lines. We also performed transcriptome analysis on the gametocyte samples. We showed that genes involved in DNA replication and TCA cycle enzymes were up-regulated in the gametocytes. On the other hand, the general cellular processes, glycolysis, protein synthesis and hemoglobin digestion were down-regulated in the gametocytes (section 5.9, Figure 5.19). The asexual parasites replicate in the blood, which is a hemoglobin rich environment. As such, in the sexual parasites, the genes encoding for proteins involved in hemoglobin digestion were down-regulated. In *Plasmodium*, the mitochondria in the asexual and sexual stages are markedly different. There are more mitochondria in the gametocytes than in the asexual stages. The mitochondria in the gametocytes also have more cristae. In addition, the gametocytes are particularly sensitive to inhibitors targeting the electron transport system [310]. The asexual parasites mainly rely on glucose metabolism for energy generation. On the other hand, the sexual stages, being in an oxygen-rich environment in the mosquito, use the TCA cycle to generate energy [311]. Hence, this information suggests the reason for the down-regulation of these transcripts in the gametocytes.

It is important to identify the genes that are expressed during the transition of asexual to sexual stages. It is also important to identify markers that are expressed in young gametocytes before they are morphologically distinguishable from the asexual stages. We can better understand the function and role of PF10_0083 if we can separate early and late stage gametocytes. As such, we need to further optimize the protocols relating to gametocytes culturing and purification.

6.4 Potential role of PF10_0083

6.4.1 PF10_0083 expression in a sub-population of schizonts

A sub-population of parasites spontaneously exits asexual multiplication and differentiates into gametocytes. The commitment of asexual parasites to gametocyte development was found to be determined before schizonts maturation and merozoites release [29, 30]. We observed that PF10_0083 was only expressed in a small population of schizonts in both the over-expression and endogenous-tagged cell lines (section 5.7, Figure 5.17 and Table 5.4). In addition, there were more parasites expressing PF10_0083 in the over-expressing cell line compared to the endogenous-tagged cell line. The localization of PF10_0083 in the nucleus during schizont stage was observed regardless of whether the *in vitro* cultures were subjected to stress to induce gametocyte production. In addition, the over-expression of PF10_0083 resulted in a lower growth rate of the parasites compared to the endogenous-tagged cell line, suggesting that the population of parasites expressing PF10_0083 might be committed to sexual development. Taken together, these findings suggest that PF10_0083 might be involved in the commitment of gametocytogenesis in the sub-population of schizonts expressing PF10_0083.

However, although the knock-out of PF10_0083 impaired gametocytogenesis, the gametocytes were able to develop normally. In addition, the knock-out of PF10_0083 did not result in a higher growth rate of the asexual stages. These results suggested that PF10_0083 might not be indispensable, but might play important accessory roles in the initiation of gametocytogenesis. It is likely that PF10_0083 is part of a regulatory complex in the initiation of sexual development. As such, it is also important to identify proteins that complement each other in gametocytogenesis. In eukaryotes, cell

cycle regulation often requires a complex of protein kinases. For example, in yeast, FUS3 activates FAR1 to bind to CDC28-CLN2 kinase, causing cell cycle arrest [312, 313]. Epigenetic regulation of chromatin structure also requires regulatory proteins and transcription factors to work together. For example, Rfm1 binds to histone deacetylase Hst1, and DNA-binding protein Sum1, to repress sporulation genes in yeast [314]. As such, protein co-immunoprecipitaion experiments to identify potential interacting partners of PF10_0083 during gametocytogenesis are required.

6.4.2 PF10_0083 in gametocytes

We also observed that PF10_0083 was expressed in all the stages during gametocyte development (section 5.8, Figure 5.18). The majority of PF10_0083 was found to be localized to the cytoplasm in all stages of gametocytes. A minute amount of PF10_0083 was localized to the gametocyte surface and nucleus. In eukaryotes, the regulation of TFs levels is important during embryogenesis, cell development and cell aging. The down-regulation of TFs can be achieved at both the transcriptional and post-translational level. TFs can be targeted for degradation or sequestered into inactive complexes. In *Drosophila melanogaster*, the down-regulation of TFs is an important mechanism to regulate TFs levels during dorso-ventral patterning in embryos. For example, CF2, a ZnF TF that functions during dorso-ventral patterning is down-regulated in order to release the repression of its downstream target, rhomboid. The Egfr signaling cascade down-regulates CF2 by targeting it for cytoplasmic localization and subsequently degradation [315]. In human, the Sp1 TF was demonstrated to be down-regulated by proteasomal degradation during cell aging and DNA damage [316]. The Klf4 TF was demonstrated to be down-regulated at the transcriptional level in order for T cell lineage commitment to proceed [317]. It is

possible that the localization of PF10_0083 to the cytoplasm in the gametocytes may be a mechanism to down-regulate this TF.

We also observed that a minute amount of PF10_0083 localized to the gametocyte surface. This observation was made because PF10_0083 displayed a similar staining pattern as that of Pfs25, a parasite surface marker. However, the staining pattern of Pfs25 resembled that of a cytoplasmic staining. In addition, not all parasites displayed PF10_0083 surface localization. As such, there is a need to validate these results. The IFA protocol can be further optimized in order to be able to clearly distinguish between parasite surface and cytoplasm staining. For example, the fixation conditions with methanol, acetone and paraformaldehyde can be tested and compared [318]. Different fixation and/or permeabilization conditions were reported to affect the preservation of the cell structure [319]. In addition, the antibodies staining conditions can be optimized to enhance antibody staining.

In eukaryotes, many TFs precursors are localized to the membrane. These membrane-bound TFs are activated or cleaved to be released to migrate into the nucleus. The release of the active TFs occurs in response to intracellular or environmental stimuli. For example, membrane-bound CREB3L1 is activated by proteolysis in response to virus induced stress. CREB3L1 enters the nucleus and transcribes genes coding for cell cycle inhibitors to block the proliferation of virus infected cells [320]. In plants, membrane-bound TFs are important in regulating cell division [321]. Notch is a membrane-bound TF that is activated by proteolytic cleavage. Notch targets many genes that function in developmental pathways [322]. Myelin Regulatory Factor (*Myrf*) is a putative TF that lack a NLS. Recently, it was shown that Myrf associates with the

membrane and is activated by proteolytic cleavage. The active Myrf enters the nucleus and transcribes genes involved in oligodendrocyte differentiation and central nervous system myelination [323]. At this point, it is not clear whether PF10_0083 act as a membrane-bound TF role in the gametocytes. PF10_0083 contains no NLS domain and no trans-membrane domain as well. In addition, we have no evidence of proteolysis cleavage of PF10_0083.

In addition, the staining of PF10_0083 in all the stages of gametocyte development suggests that PF10_0083 might play other roles after gametocyte development initiation. We observed that the development of gametocytes from stage II to stage V was not affected by the over-expression or knock-out of PF10_0083. However, we cannot rule out the possibility that PF10_0083 might play an accessory role during gametocyte maturation, but is dispensable during the process. Furthermore, we observed that PF10_0083 levels remained high in the late stage gametocyte. This result suggests that PF10_0083 might play other roles after gametocyte maturation. In *Plasmodium*, it was reported that the mRNA transcripts of 9 genes in the gametocytes were only translated during the ookinete stage in the mosquito vector [46, 202]. Therefore, PF10_0083 might have other functional roles after the gametocytes have matured. As such, the roles of PF10_0083 after gametocyte development, specifically in the mosquito host, remain to be determined.

Since we were able to detect PF10_0083 expression in schizonts and early in the gametocytes, the potential of using this protein as a marker in early gametocytes can be explored. In addition, the localization of PF10_0083 to the parasite surface suggested the possibility of developing antibodies against this protein. Much work has

also been done to study the sexual stage specific markers and surface proteins. Pfs16, Pfs230 and Pfs48/45 are expressed on the gametocyte surface while Pfg27 is localized in the cytoplasm. Pfs25 and Pfs28 are expressed on ookinete surface. Many of these surface proteins on the gametocytes and gametes play important roles in cell-cell interaction and signaling during fertilization, motility, adhesion and invasion inside the mosquito host. Disruption of these proteins caused reduced fertilization, reduced oocyst formation and infectivity, reduced sporozoites formation, and reduced transmission to mosquito. As such, many studies have been done on developing transmission intervention strategies against known surface proteins [324]. The role of PF10_0083 in the gametocytes remains to be determined. In addition, the role of PF10_0083 after gametocytogenesis, in the gametes and inside mosquito can be explored as well.

6.5 Future directions

We observed that PF10_0083 was localized to the nucleus in the schizonts. However, PF10_0083 lacks NLS, and might be transported into the nucleus with the help of another protein that contains a NLS. Therefore, studies such as protein co-immunoprecipitaion should be carried out to elucidate the mechanism by which PF10_0083 gets transported into the nucleus. In addition, the unique nuclear compartment that PF10_0083 resides during the schizont stage requires additional validation. The nuclear compartments which PF10_0083 resides in the rings and trophozoites need to be determined too. As such, we propose to performed co-localization studies with the heterochromatin and euchromatin markers on ring, trophozoite, young-, mid- and late schizont stages. We identified 3 DNA sequence motifs that PF10_0083 potentially binds to. The binding of PF10_0083 to these

sequences require further validation such as electrophoretic mobility shift assay (EMSA). We speculated that PF10_0083 is involved in the differentiation of asexual blood stages parasites to sexual stages parasites. However, much work needs to be done in order to validate the findings presented in this thesis. PF10_0083 appears to be dispensable during the differentiation of asexual parasites to gametocytes. It is likely that PF10_0083 is part of a regulatory complex that functions during the initiation of sexual development. As such, protein co-immunoprecipitaion experiments to identify the potential interacting partners of PF10_0083 during gametocytogenesis are required. We observed that PF10_0083 was only expressed in a sub-population of schizonts, and proposed that these schizonts are committed to sexual development. As such, we propose to characterize PF10_0083 after subjecting the parasites to stress. In particular, we propose to study the localization of PF10_0083 in the asexual stages after stress and before differentiating into gametocyte. In addition, we showed that the level of PF10_0083 increased in the gametocytes. As such, it would be interesting to measure the level of PF10_0083 in the stressed parasites. We observed that PF10_0083 levels remained high in the late stage gametocyte, and suggested that PF10_0083 might play other roles after gametocyte maturation. Therefore, the roles of PF10_0083 after gametocyte development, specifically in the mosquito host, should be determined.

6.6 Conclusion

In this thesis, we characterized a putative TF, PF10_0083, a CCCH-type ZnF. Using over-expression, endogenous-tagged and knock-out cell lines, we demonstrated that PF10_0083 was localized to the cytoplasm during the ring and trophozoite stages of the IDC. In the schizont stage, the protein was localized predominantly at the nuclear periphery. We showed that PF10_0083 was not expressed in all the schizonts. We

found that only a sub-population of schizonts, around 3.5% schizonts in the over-expression cell line, and 1.72% schizonts in the endogenous-tagged cell line, expresses PF10_0083. We also observed that the over-expression of PF10_0083 resulted in a lower parasite invasion rate. Under stress conditions, we demonstrated that the over-expression of PF10_0083 resulted in enhanced gametocyte production, while the knock-out of PF10_0083 caused decreased gametocyte production. In addition, we showed that PF10_0083 is capable of associating with DNA. We demonstrated that PF10_0083 preferentially binds to the IGRs on the genome, and that this binding positively correlated with the up-regulation of gametocytes genes. As such, we conclude that PF10_0083 promotes the differentiation of asexual parasites to sexual forms under stress conditions. Further studies are required to study how PF10_0083 gets transported to the nucleus during schizont stage. Furthermore, the binding of PF10_0083 to its DNA targets requires validation. The role of PF10_0083 in the gametocytes remains to be determined.

Chapter 7 Appendix

Supplementary figures

CHA-1	1	2	3	4	5	6	7	8	9	10	11	12
A												
B		1	1	7	7	13	13	19	19	diluent A	diluent A	
C		2	2	8	8	14	14	20	20	diluent A	diluent A	
D		3	3	9	9	15	15	21	21	diluent A	diluent A	
E		4	4	10	10	16	16	22	22			
F		5	5	11	11	17	17	23	23			
G		6	6	12	12	**18**	**18**	24	24			
H												

CHA-2	1	2	3	4	5	6	7	8	9	10	11	12
A												
B		25	25	31	31	37	37	43	43	diluent B	diluent B	
C		26	26	32	32	38	38	44	44	diluent B	diluent B	
D		27	27	33	33	39	39	45	45	diluent B	diluent B	
E		28	28	34	34	40	40	46	46			
F		29	29	35	35	41	41	47	47			
G		**30**	**30**	36	36	42	42	48	48			
H												
		Day 10	Day 12									

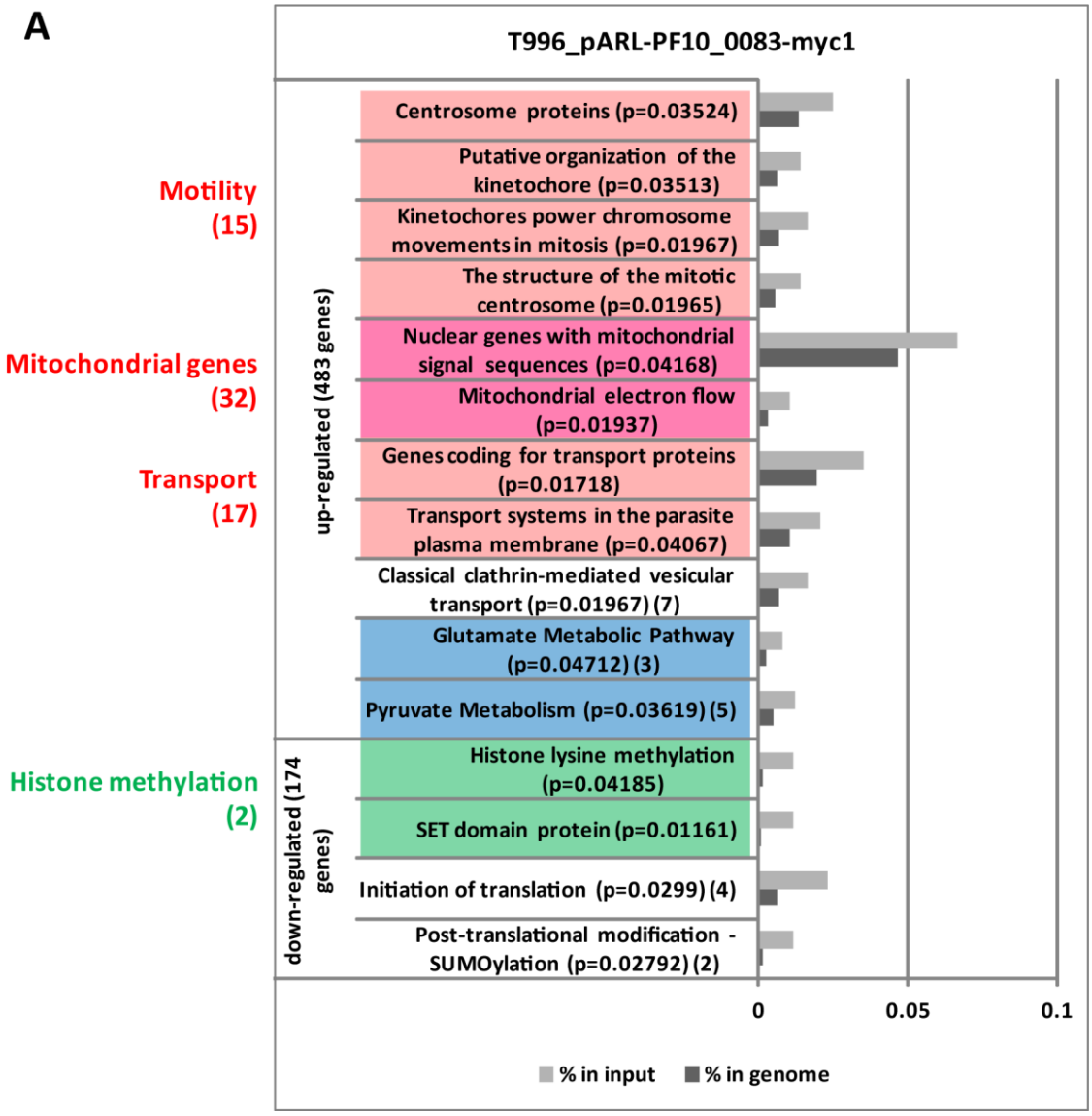
KO-1	1	2	3	4	5	6	7	8	9	10	11	12
A												
B		1	1	7	7	13	13	19	19	diluent A	diluent A	
C		2	2	8	8	14	14	20	20	diluent A	diluent A	
D		3	3	9	9	15	15	21	21	diluent A	diluent A	
E		4	4	10	10	16	16	22	22			
F		5	5	11	11	17	17	23	23			
G		6	6	12	12	18	18	**24**	**24**			
H												

KO-2	1	2	3	4	5	6	7	8	9	10	11	12
A												
B		25	25	31	31	37	37	43	43	diluent B	diluent B	
C		26	26	32	32	38	38	44	44	diluent B	diluent B	
D		27	27	**33**	**33**	39	39	45	45	diluent B	diluent B	
E		28	28	34	34	40	40	46	46			
F		29	29	35	35	41	41	47	47			
G		30	30	36	36	42	42	48	48			
H												
		Day 10	Day 12									

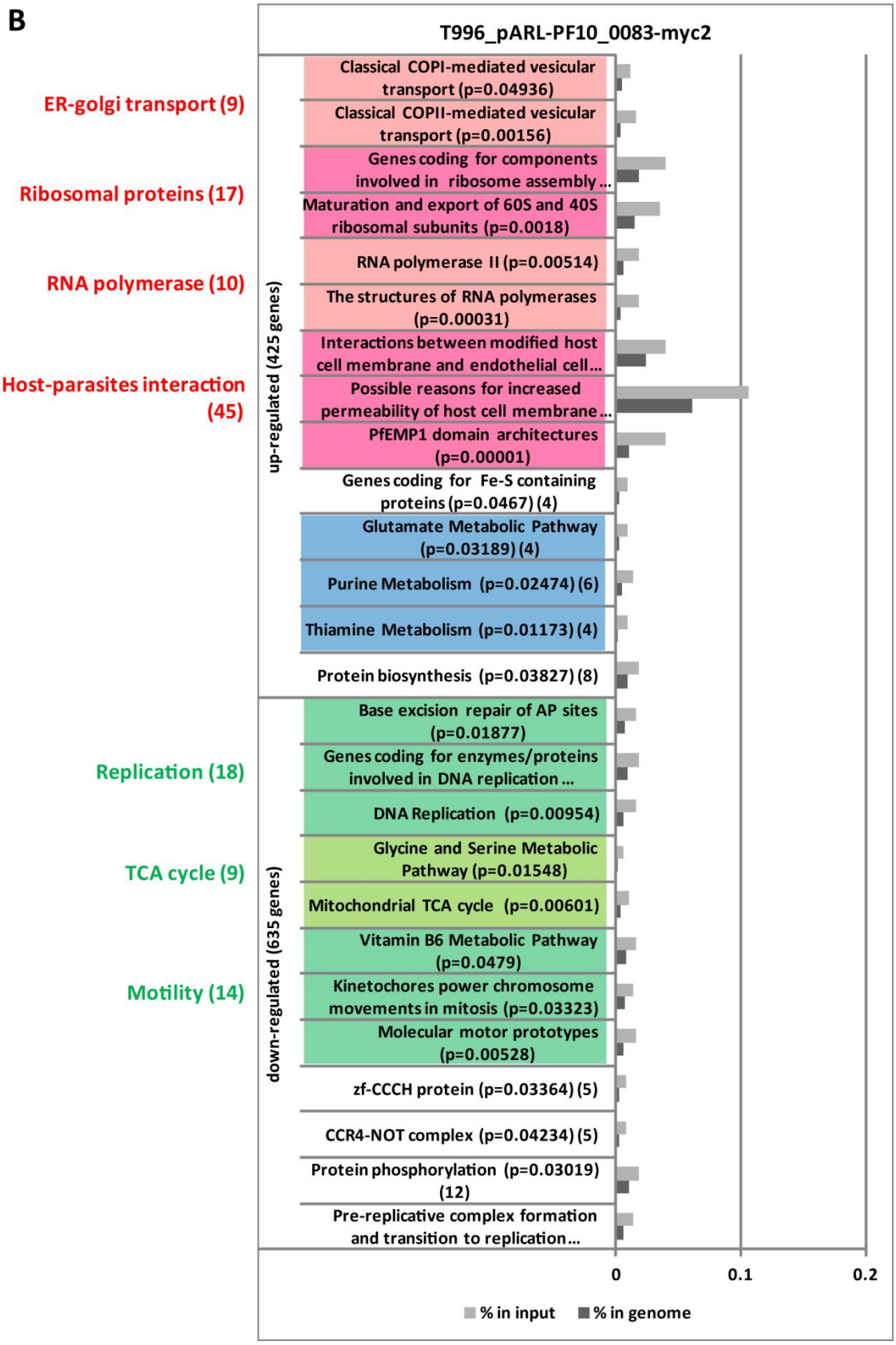
Supplementary Figure 1 Cloning by limiting dilution.

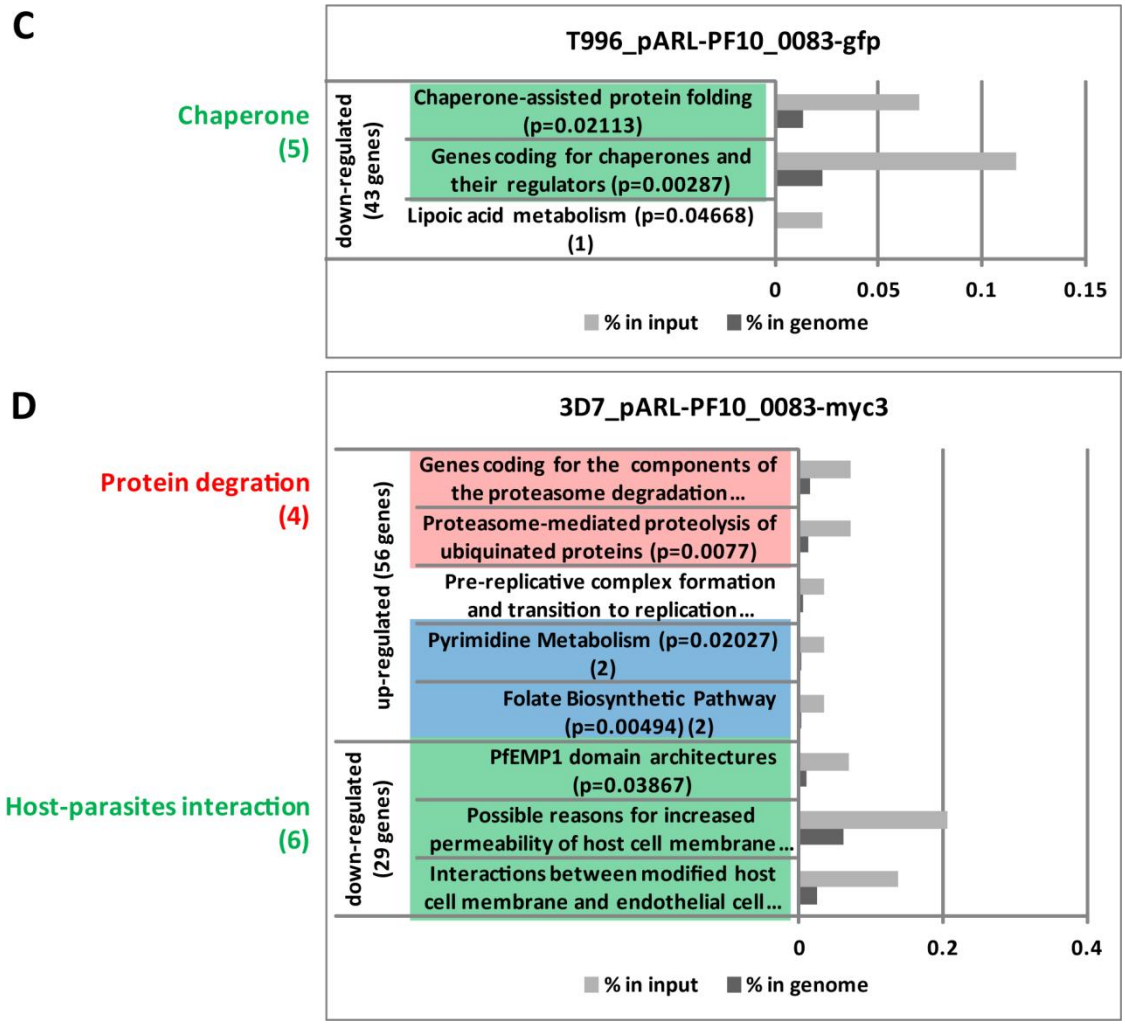
A total of 48 clones of PF10_003-CHA and PF10_003-KO parasites were prepared in 2 96-well plates. Purple and blue boxes are clones where parasites were first observed in

day 10 and day 12 after cloning by limiting dilution. In PF10_003-CHA, 21 out of 48 clones were positive. PF10_0083-CHA18 and CHA30 were randomly selected for further studies (highlighted in bold). In PF10_003-KO, 15 out of 48 clones were positive. PF10_0083-KO24 and KO33 were randomly selected for further studies (highlighted in bold). White boxes are clones that parasites did not develop. Diluent A and diluent B are positive controls of the experiment.



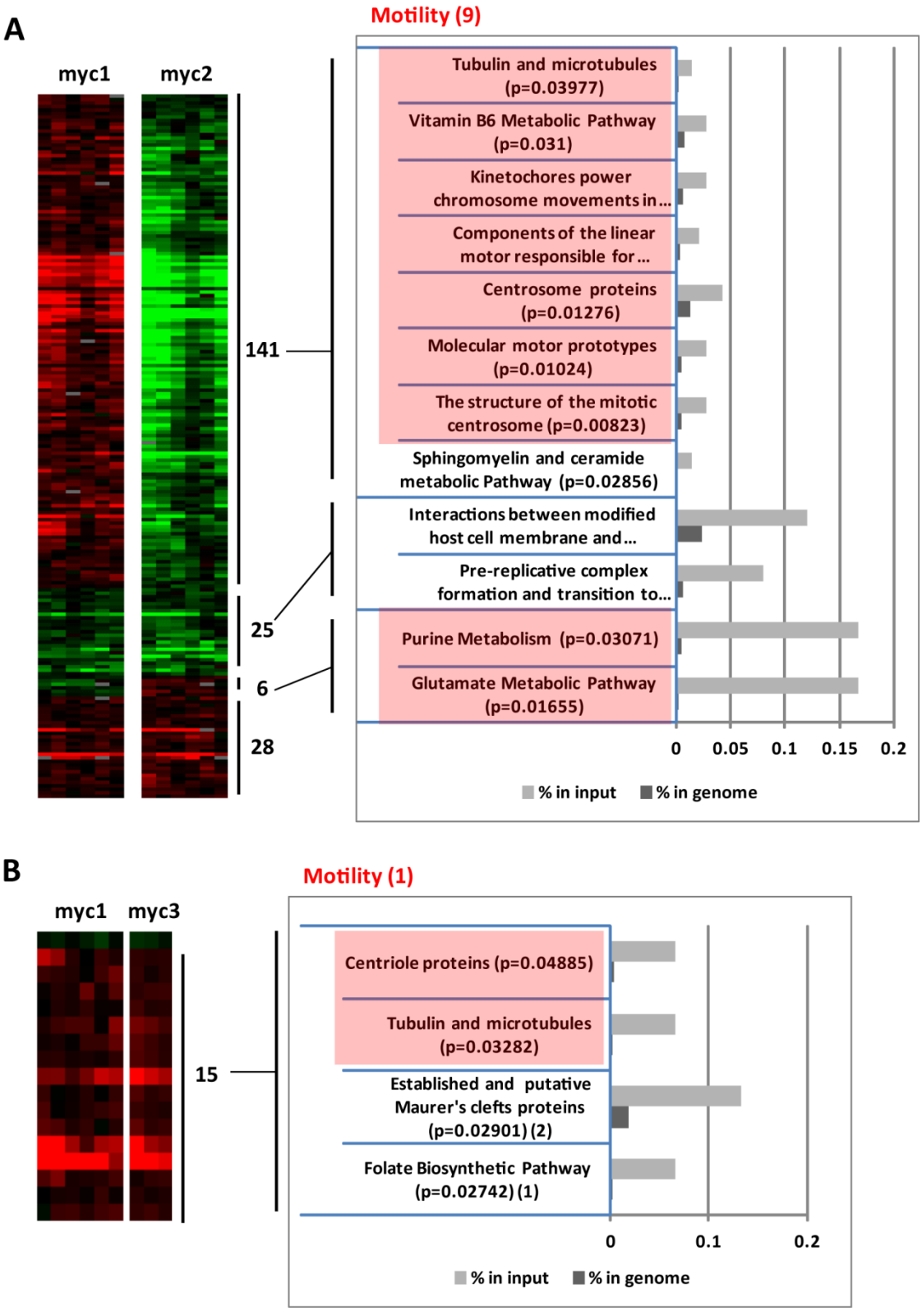
B


T996_pARL-PF10_0083-myc2
ER-golgi transport (9)
Ribosomal proteins (17)
RNA polymerase (10)
Host-parasites interaction (45)
Replication (18)
TCA cycle (9)
Motility (14)
up-regulated (425 genes)
down-regulated (635 genes)
Classical COPI-mediated vesicular transport (p=0.04936)
Classical COPII-mediated vesicular transport (p=0.00156)
Genes coding for components involved in ribosome assembly ...
Maturation and export of 60S and 40S ribosomal subunits (p=0.0018)
RNA polymerase II (p=0.00514)
The structures of RNA polymerases (p=0.00031)
Interactions between modified host cell membrane and endothelial cell...
Possible reasons for increased permeability of host cell membrane ...
PfEMP1 domain architectures (p=0.00001)
Genes coding for Fe-S containing proteins (p=0.0467) (4)
Glutamate Metabolic Pathway (p=0.03189) (4)
Purine Metabolism (p=0.02474) (6)
Thiamine Metabolism (p=0.01173) (4)
Protein biosynthesis (p=0.03827) (8)
Base excision repair of AP sites (p=0.01877)
Genes coding for enzymes/proteins involved in DNA replication ...
DNA Replication (p=0.00954)
Glycine and Serine Metabolic Pathway (p=0.01548)
Mitochondrial TCA cycle (p=0.00601)
Vitamin B6 Metabolic Pathway (p=0.0479)
Kinetochores power chromosome movements in mitosis (p=0.03323)
Molecular motor prototypes (p=0.00528)
zf-CCCH protein (p=0.03364) (5)
CCR4-NOT complex (p=0.04234) (5)
Protein phosphorylation (p=0.03019) (12)
Pre-replicative complex formation and transition to replication...
0
0.1
0.2
% in input
% in genome



Supplementary Figure 2 Functional enrichment of genes affected by PF10_0083 over-expression.

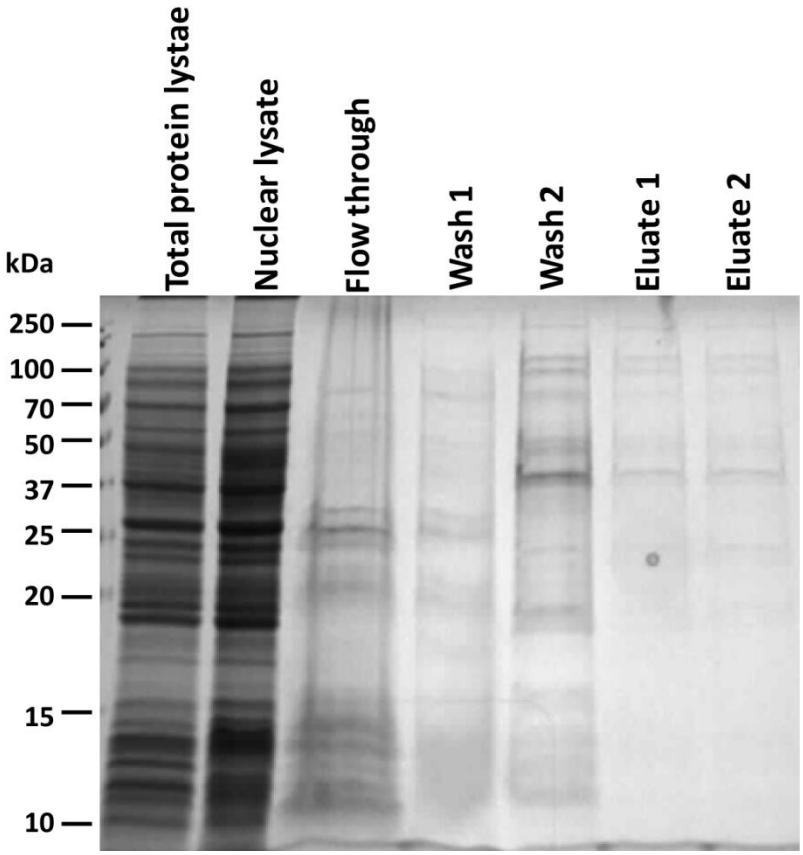
Functional enrichment (MPM: Malaria Parasites Metabolic Pathways) of up-regulated genes and down-regulated genes that were affected in (A) T996-pARL-PF10_0083-myc1, (B) T996-pARL-PF10_0083-myc2, (C) T996-pARL-PF10_0083-gfp, and (D) 3D7-pARL-PF10_0083-myc3. The light grey bars represent the pathway related genes in the over-expression cell lines, and the dark grey bars represent the pathway related genes in the genome. The number of genes in each pathway is shown in brackets. Blue boxes represent functional groups enriched in the up-regulated gene group in all the 4 cell lines. All pathways have a binomial p-value less than 0.05.



Supplementary Figure 3 Functional enrichment of the differentially expressed genes affected in more than 2 over-expression cell lines.

Hierarchical clustered genes that were transcriptionally altered in (A) T996-pARL-PF10_0083-myc1 and T996-pARL-PF10_0083-myc2, and (B) T996-pARL-PF10_0083-myc1 and 3D7-pARL-PF10_0083-myc3. Red and green represents genes that had higher

and lower mRNA levels in the transfected parasites compared to control parasites. Bar graphs depict the functional enrichment (MPM: Malaria Parasites Metabolic Pathways) of genes that were transcriptionally altered. The light grey bars represent the pathway related genes in the over-expression cell lines, and the dark grey bars represent the pathway related genes in the genome. The number of genes in each pathway is shown in brackets. All pathways have a binomial p-value less than 0.05



Supplementary Figure 4 Protein co-immunoprecipitation.

Protein co-immunoprecipitation was performed on T996-PF10_0083-myc1 over-expression cell line using the anti-myc antibody. The total protein lysate, nuclear lysate, wash 1and 2, and eluate 1 and 2 were separated on a 12% SDS-PAGE system. The proteins were visualized by silver stain. The expected size of PF10_0083 is ~47 kDa. In eluate 1 and 2, PF10_0083 was not detected.

Chapter 8 References

1. Cox, F.E., *History of human parasitology.* Clin Microbiol Rev, 2002. **15**(4): p. 595-612.
2. Cox, F.E., *History of the discovery of the malaria parasites and their vectors.* Parasit Vectors, 2010. **3**(1): p. 5.
3. Hempelmann, E. and K. Krafts, *Bad air, amulets and mosquitoes: 2,000 years of changing perspectives on malaria.* Malar J, 2013. **12**(1): p. 232.
4. Poinar, G., Jr., *Plasmodium dominicana n. sp. (Plasmodiidae: Haemospororida) from Tertiary Dominican amber.* Syst Parasitol, 2005. **61**(1): p. 47-52.
5. W.H.O, *World Malaria Report 2011.* World Health Organization, 2011.
6. W.H.O, *World Malaria Report 2012.* World Health Organization, 2012.
7. White, N.J., *Plasmodium knowlesi: the fifth human malaria parasite.* Clin Infect Dis, 2008. **46**(2): p. 172-3.
8. Yakob, L., M.B. Bonsall, and G. Yan, *Modelling knowlesi malaria transmission in humans: vector preference and host competence.* Malar J, 2010. **9**: p. 329.
9. Miller, L.H., et al., *The pathogenic basis of malaria.* Nature, 2002. **415**(6872): p. 673-9.
10. Silvie, O., et al., *Interactions of the malaria parasite and its mammalian host.* Curr Opin Microbiol, 2008. **11**(4): p. 352-9.
11. Vaughan, A.M., A.S. Aly, and S.H. Kappe, *Malaria parasite pre-erythrocytic stage infection: gliding and hiding.* Cell Host Microbe, 2008. **4**(3): p. 209-18.
12. Stanway, R.R., J. Schmuckli-Maurer, and V.T. Heussler, *Analysis of liver stage development in and merozoite release from hepatocytes.* Methods Mol Biol, 2013. **923**: p. 411-27.
13. Cowman, A.F. and B.S. Crabb, *Invasion of red blood cells by malaria parasites.* Cell, 2006. **124**(4): p. 755-66.
14. Bozdech, Z., et al., *The transcriptome of the intraerythrocytic developmental cycle of Plasmodium falciparum.* PLoS Biol, 2003. **1**(1): p. E5.
15. Matuschewski, K., *Getting infectious: formation and maturation of Plasmodium sporozoites in the Anopheles vector.* Cell Microbiol, 2006. **8**(10): p. 1547-56.
16. Butcher, G.A., *Antimalarial drugs and the mosquito transmission of Plasmodium.* Int J Parasitol, 1997. **27**(9): p. 975-87.
17. Carter, R. and L.H. Miller, *Evidence for environmental modulation of gametocytogenesis in Plasmodium falciparum in continuous culture.* Bull World Health Organ, 1979. **57 Suppl 1**: p. 37-52.
18. Kwiatkowski, D., et al., *TNF concentration in fatal cerebral, non-fatal cerebral, and uncomplicated Plasmodium falciparum malaria.* Lancet, 1990. **336**(8725): p. 1201-4.
19. Smalley, M.E. and J. Brown, *Plasmodium falciparum gametocytogenesis stimulated by lymphocytes and serum from infected Gambian children.* Trans R Soc Trop Med Hyg, 1981. **75**(2): p. 316-7.
20. Barkakaty, B.N., G.K. Sharma, and N.K. Chakravorty, *Studies on efficacy of treatment with sulfamethoxazole + trimethoprim and sulfalene +*

pyrimethamine combinations in Plasmodium falciparum malaria of known and unknown resistant status. J Commun Dis, 1988. **20**(3): p. 165-74.
21. Buckling, A.G., et al., *Adaptive changes in Plasmodium transmission strategies following chloroquine chemotherapy.* Proc Biol Sci, 1997. **264**(1381): p. 553-9.
22. Price, R.N., et al., *Effects of artemisinin derivatives on malaria transmissibility.* Lancet, 1996. **347**(9016): p. 1654-8.
23. Jeffery, G.M. and D.E. Eyles, *Infectivity to mosquitoes of Plasmodium falciparum as related to gametocyte density and duration of infection.* Am J Trop Med Hyg, 1955. **4**(5): p. 781-9.
24. McKenzie, F.E., G.M. Jeffery, and W.E. Collins, *Gametocytemia and fever in human malaria infections.* J Parasitol, 2007. **93**(3): p. 627-33.
25. Inselburg, J., *Stage-specific inhibitory effect of cyclic AMP on asexual maturation and gametocyte formation of Plasmodium falciparum.* J Parasitol, 1983. **69**(3): p. 592-7.
26. Kongkasuriyachai, D. and N. Kumar, *Functional characterisation of sexual stage specific proteins in Plasmodium falciparum.* Int J Parasitol, 2002. **32**(13): p. 1559-66.
27. Talman, A.M., et al., *Gametocytogenesis: the puberty of Plasmodium falciparum.* Malar J, 2004. **3**: p. 24.
28. Dyer, M. and K.P. Day, *Commitment to gametocytogenesis in Plasmodium falciparum.* Parasitol Today, 2000. **16**(3): p. 102-7.
29. Silvestrini, F., P. Alano, and J.L. Williams, *Commitment to the production of male and female gametocytes in the human malaria parasite Plasmodium falciparum.* Parasitology, 2000. **121 Pt 5**: p. 465-71.
30. Smith, T.G., et al., *Commitment to sexual differentiation in the human malaria parasite, Plasmodium falciparum.* Parasitology, 2000. **121 (Pt 2)**: p. 127-33.
31. Schneider, P., et al., *Quantification of Plasmodium falciparum gametocytes in differential stages of development by quantitative nucleic acid sequence-based amplification.* Mol Biochem Parasitol, 2004. **137**(1): p. 35-41.
32. Florens, L., et al., *A proteomic view of the Plasmodium falciparum life cycle.* Nature, 2002. **419**(6906): p. 520-6.
33. Lobo, C.A., R.N. Konings, and N. Kumar, *Expression of early gametocyte-stage antigens Pfg27 and Pfs16 in synchronized gametocytes and non-gametocyte producing clones of Plasmodium falciparum.* Mol Biochem Parasitol, 1994. **68**(1): p. 151-4.
34. Rawlings, D.J., et al., *Alpha-tubulin II is a male-specific protein in Plasmodium falciparum.* Mol Biochem Parasitol, 1992. **56**(2): p. 239-50.
35. Severini, C., et al., *The production of the osmiophilic body protein Pfg377 is associated with stage of maturation and sex in Plasmodium falciparum gametocytes.* Mol Biochem Parasitol, 1999. **100**(2): p. 247-52.
36. Pologe, L.G., *Aberrant transcription and the failure of Plasmodium falciparum to differentiate into gametocytes.* Mol Biochem Parasitol, 1994. **68**(1): p. 35-43.
37. Lobo, C.A., et al., *Disruption of the Pfg27 locus by homologous recombination leads to loss of the sexual phenotype in P. falciparum.* Mol Cell, 1999. **3**(6): p. 793-8.
38. Alano, P., et al., *Plasmodium falciparum: parasites defective in early stages of gametocytogenesis.* Exp Parasitol, 1995. **81**(2): p. 227-35.

39. Gardiner, D.L., et al., *Implication of a Plasmodium falciparum gene in the switch between asexual reproduction and gametocytogenesis.* Mol Biochem Parasitol, 2005. **140**(2): p. 153-60.
40. Eksi, S., et al., *Plasmodium falciparum gametocyte development 1 (Pfgdv1) and gametocytogenesis early gene identification and commitment to sexual development.* PLoS Pathog, 2012. **8**(10): p. e1002964.
41. Pace, T., et al., *Set regulation in asexual and sexual Plasmodium parasites reveals a novel mechanism of stage-specific expression.* Mol Microbiol, 2006. **60**(4): p. 870-82.
42. Dobson, S., et al., *Characterization of a unique aspartate-rich protein of the SET/TAF-family in the human malaria parasite, Plasmodium falciparum, which inhibits protein phosphatase 2A.* Mol Biochem Parasitol, 2003. **126**(2): p. 239-50.
43. Kafsack, B.F., et al., *A transcriptional switch underlies commitment to sexual development in malaria parasites.* Nature, 2014. **507**(7491): p. 248-52.
44. Sinha, A., et al., *A cascade of DNA-binding proteins for sexual commitment and development in Plasmodium.* Nature, 2014. **507**(7491): p. 253-7.
45. Le Roch, K.G., et al., *Global analysis of transcript and protein levels across the Plasmodium falciparum life cycle.* Genome Res, 2004. **14**(11): p. 2308-18.
46. Hall, N., et al., *A comprehensive survey of the Plasmodium life cycle by genomic, transcriptomic, and proteomic analyses.* Science, 2005. **307**(5706): p. 82-6.
47. Miao, J., et al., *Puf mediates translation repression of transmission-blocking vaccine candidates in malaria parasites.* PLoS Pathog, 2013. **9**(4): p. e1003268.
48. Greenwood, D., *Conflicts of interest: the genesis of synthetic antimalarial agents in peace and war.* J Antimicrob Chemother, 1995. **36**(5): p. 857-72.
49. Achan, J., et al., *Quinine, an old anti-malarial drug in a modern world: role in the treatment of malaria.* Malar J, 2011. **10**: p. 144.
50. Ridley, R.G., *Medical need, scientific opportunity and the drive for antimalarial drugs.* Nature, 2002. **415**(6872): p. 686-93.
51. Rehwagen, C., *WHO ultimatum on artemisinin monotherapy is showing results.* BMJ, 2006. **332**(7551): p. 1176.
52. Nosten, F. and N.J. White, *Artemisinin-based combination treatment of falciparum malaria.* Am J Trop Med Hyg, 2007. **77**(6 Suppl): p. 181-92.
53. Lengeler, C., *Insecticide-treated bednets and curtains for preventing malaria.* Cochrane Database Syst Rev, 2000(2): p. CD000363.
54. Rehwagen, C., *WHO recommends DDT to control malaria.* BMJ, 2006. **333**(7569): p. 622.
55. W.H.O, *WHO position statement on integrated vector management.* Wkly Epidemiol Rec, 2008. **83**(20): p. 177-81.
56. Worrall, E. and U. Fillinger, *Large-scale use of mosquito larval source management for malaria control in Africa: a cost analysis.* Malar J, 2011. **10**: p. 338.
57. Alonso, P.L., et al., *Efficacy of the RTS,S/AS02A vaccine against Plasmodium falciparum infection and disease in young African children: randomised controlled trial.* Lancet, 2004. **364**(9443): p. 1411-20.
58. Alonso, P.L., et al., *Duration of protection with RTS,S/AS02A malaria vaccine in prevention of Plasmodium falciparum disease in Mozambican children:*

single-blind extended follow-up of a randomised controlled trial. Lancet, 2005. **366**(9502): p. 2012-8.
59. Agnandji, S.T., et al., *First results of phase 3 trial of RTS,S/AS01 malaria vaccine in African children.* N Engl J Med, 2011. **365**(20): p. 1863-75.
60. Mueller, A.K., et al., *Plasmodium liver stage developmental arrest by depletion of a protein at the parasite-host interface.* Proc Natl Acad Sci U S A, 2005. **102**(8): p. 3022-7.
61. Mueller, A.K., et al., *Genetically modified Plasmodium parasites as a protective experimental malaria vaccine.* Nature, 2005. **433**(7022): p. 164-7.
62. Gardner, M.J., et al., *Genome sequence of the human malaria parasite Plasmodium falciparum.* Nature, 2002. **419**(6906): p. 498-511.
63. Le Roch, K.G., et al., *Discovery of gene function by expression profiling of the malaria parasite life cycle.* Science, 2003. **301**(5639): p. 1503-8.
64. Foth, B.J., et al., *Quantitative time-course profiling of parasite and host cell proteins in the human malaria parasite Plasmodium falciparum.* Mol Cell Proteomics, 2011. **10**(8): p. M110 006411.
65. Dzikowski, R., M. Frank, and K. Deitsch, *Mutually exclusive expression of virulence genes by malaria parasites is regulated independently of antigen production.* PLoS Pathog, 2006. **2**(3): p. e22.
66. Chookajorn, T., P. Ponsuwanna, and L. Cui, *Mutually exclusive var gene expression in the malaria parasite: multiple layers of regulation.* Trends Parasitol, 2008. **24**(10): p. 455-61.
67. Scherf, A., J.J. Lopez-Rubio, and L. Riviere, *Antigenic variation in Plasmodium falciparum.* Annu Rev Microbiol, 2008. **62**: p. 445-70.
68. Coulson, R.M., N. Hall, and C.A. Ouzounis, *Comparative genomics of transcriptional control in the human malaria parasite Plasmodium falciparum.* Genome Res, 2004. **14**(8): p. 1548-54.
69. Lanzer, M., D. de Bruin, and J.V. Ravetch, *Transcription mapping of a 100 kb locus of Plasmodium falciparum identifies an intergenic region in which transcription terminates and reinitiates.* EMBO J, 1992. **11**(5): p. 1949-55.
70. Lanzer, M., D. de Bruin, and J.V. Ravetch, *A sequence element associated with the Plasmodium falciparum KAHRP gene is the site of developmentally regulated protein-DNA interactions.* Nucleic Acids Res, 1992. **20**(12): p. 3051-6.
71. Templeton, T.J., et al., *Comparative analysis of apicomplexa and genomic diversity in eukaryotes.* Genome Res, 2004. **14**(9): p. 1686-95.
72. Callebaut, I., et al., *Prediction of the general transcription factors associated with RNA polymerase II in Plasmodium falciparum: conserved features and differences relative to other eukaryotes.* BMC Genomics, 2005. **6**: p. 100.
73. McAndrew, M.B., et al., *Characterisation of the gene encoding an unusually divergent TATA-binding protein (TBP) from the extremely A+T-rich human malaria parasite Plasmodium falciparum.* Gene, 1993. **124**(2): p. 165-71.
74. Ruvalcaba-Salazar, O.K., et al., *Recombinant and native Plasmodium falciparum TATA-binding-protein binds to a specific TATA box element in promoter regions.* Mol Biochem Parasitol, 2005. **140**(2): p. 183-96.
75. Smale, S.T. and J.T. Kadonaga, *The RNA polymerase II core promoter.* Annu Rev Biochem, 2003. **72**: p. 449-79.
76. Kuehner, J.N. and D.A. Brow, *Quantitative analysis of in vivo initiator selection by yeast RNA polymerase II supports a scanning model.* J Biol Chem, 2006. **281**(20): p. 14119-28.

77. Watanabe, J., et al., *Analysis of transcriptomes of human malaria parasite Plasmodium falciparum using full-length enriched library: identification of novel genes and diverse transcription start sites of messenger RNAs.* Gene, 2002. **291**(1-2): p. 105-13.
78. Plessy, C., et al., *Promoter architecture of mouse olfactory receptor genes.* Genome Res, 2012. **22**(3): p. 486-97.
79. Schug, J., et al., *Promoter features related to tissue specificity as measured by Shannon entropy.* Genome Biol, 2005. **6**(4): p. R33.
80. Kim, T.H., et al., *A high-resolution map of active promoters in the human genome.* Nature, 2005. **436**(7052): p. 876-80.
81. Carninci, P., et al., *Genome-wide analysis of mammalian promoter architecture and evolution.* Nat Genet, 2006. **38**(6): p. 626-35.
82. Ohler, U., et al., *Computational analysis of core promoters in the Drosophila genome.* Genome Biol, 2002. **3**(12): p. RESEARCH0087.
83. Molina, C. and E. Grotewold, *Genome wide analysis of Arabidopsis core promoters.* BMC Genomics, 2005. **6**: p. 25.
84. Antequera, F. and A. Bird, *Number of CpG islands and genes in human and mouse.* Proc Natl Acad Sci U S A, 1993. **90**(24): p. 11995-9.
85. Saxonov, S., P. Berg, and D.L. Brutlag, *A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters.* Proc Natl Acad Sci U S A, 2006. **103**(5): p. 1412-7.
86. Yamashita, R., et al., *Genome-wide analysis reveals strong correlation between CpG islands with nearby transcription start sites of genes and their tissue specificity.* Gene, 2005. **350**(2): p. 129-36.
87. Akalin, A., et al., *Transcriptional features of genomic regulatory blocks.* Genome Biol, 2009. **10**(4): p. R38.
88. Ohler, U., *Identification of core promoter modules in Drosophila and their application in accurate transcription start site prediction.* Nucleic Acids Res, 2006. **34**(20): p. 5943-50.
89. Engstrom, P.G., et al., *Genomic regulatory blocks underlie extensive microsynteny conservation in insects.* Genome Res, 2007. **17**(12): p. 1898-908.
90. Ernst, J. and M. Kellis, *Discovery and characterization of chromatin states for systematic annotation of the human genome.* Nat Biotechnol, 2010. **28**(8): p. 817-25.
91. Bernstein, B.E., et al., *A bivalent chromatin structure marks key developmental genes in embryonic stem cells.* Cell, 2006. **125**(2): p. 315-26.
92. Nozaki, T., et al., *Tight associations between transcription promoter type and epigenetic variation in histone positioning and modification.* BMC Genomics, 2011. **12**: p. 416.
93. Balu, B., P.L. Blair, and J.H. Adams, *Identification of the transcription initiation site reveals a novel transcript structure for Plasmodium falciparum maebl.* Exp Parasitol, 2009. **121**(1): p. 110-4.
94. Barik, S., R.E. Taylor, and D. Chakrabarti, *Identification, cloning, and mutational analysis of the casein kinase 1 cDNA of the malaria parasite, Plasmodium falciparum. Stage-specific expression of the gene.* J Biol Chem, 1997. **272**(42): p. 26132-8.
95. Russell, K., et al., *Analysis of the spatial and temporal arrangement of transcripts over intergenic regions in the human malarial parasite Plasmodium falciparum.* BMC Genomics, 2013. **14**: p. 267.

96. Crabb, B.S. and A.F. Cowman, *Characterization of promoters and stable transfection by homologous and nonhomologous recombination in Plasmodium falciparum.* Proc Natl Acad Sci U S A, 1996. **93**(14): p. 7289-94.
97. Horrocks, P. and B.J. Kilbey, *Physical and functional mapping of the transcriptional start sites of Plasmodium falciparum proliferating cell nuclear antigen.* Mol Biochem Parasitol, 1996. **82**(2): p. 207-15.
98. Militello, K.T., et al., *Identification of regulatory elements in the Plasmodium falciparum genome.* Mol Biochem Parasitol, 2004. **134**(1): p. 75-88.
99. Porter, M.E., *Positive and negative effects of deletions and mutations within the 5' flanking sequences of Plasmodium falciparum DNA polymerase delta.* Mol Biochem Parasitol, 2002. **122**(1): p. 9-19.
100. Osta, M., et al., *A 24 bp cis-acting element essential for the transcriptional activity of Plasmodium falciparum CDP-diacylglycerol synthase gene promoter.* Mol Biochem Parasitol, 2002. **121**(1): p. 87-98.
101. Horrocks, P. and M. Lanzer, *Mutational analysis identifies a five base pair cis-acting sequence essential for GBP130 promoter activity in Plasmodium falciparum.* Mol Biochem Parasitol, 1999. **99**(1): p. 77-87.
102. Elemento, O., N. Slonim, and S. Tavazoie, *A universal framework for regulatory element discovery across all genomes and data types.* Mol Cell, 2007. **28**(2): p. 337-50.
103. Young, J.A., et al., *In silico discovery of transcription regulatory elements in Plasmodium falciparum.* BMC Genomics, 2008. **9**: p. 70.
104. Pace, T., et al., *Targeted terminal deletions as a tool for functional genomics studies in Plasmodium.* Genome Res, 2000. **10**(9): p. 1414-20.
105. Horrocks, P., et al., *Stage-specific promoter activity from stably maintained episomes in Plasmodium falciparum.* Int J Parasitol, 2002. **32**(10): p. 1203-6.
106. Kocken, C.H., et al., *Precise timing of expression of a Plasmodium falciparum-derived transgene in Plasmodium berghei is a critical determinant of subsequent subcellular localization.* J Biol Chem, 1998. **273**(24): p. 15119-24.
107. Triglia, T., et al., *Apical membrane antigen 1 plays a central role in erythrocyte invasion by Plasmodium species.* Mol Microbiol, 2000. **38**(4): p. 706-18.
108. Wickham, M.E., J.K. Thompson, and A.F. Cowman, *Characterisation of the merozoite surface protein-2 promoter using stable and transient transfection in Plasmodium falciparum.* Mol Biochem Parasitol, 2003. **129**(2): p. 147-56.
109. Deitsch, K.W., M.S. Calderwood, and T.E. Wellems, *Malaria. Cooperative silencing elements in var genes.* Nature, 2001. **412**(6850): p. 875-6.
110. Dechering, K.J., et al., *Isolation and functional characterization of two distinct sexual-stage-specific promoters of the human malaria parasite Plasmodium falciparum.* Mol Cell Biol, 1999. **19**(2): p. 967-78.
111. Wong, E.H., S. Hasenkamp, and P. Horrocks, *Analysis of the molecular mechanisms governing the stage-specific expression of a prototypical housekeeping gene during intraerythrocytic development of P. falciparum.* J Mol Biol, 2011. **408**(2): p. 205-21.
112. Bourbon, H.M., *Comparative genomics supports a deep evolutionary origin for the large, four-module transcriptional mediator complex.* Nucleic Acids Res, 2008. **36**(12): p. 3993-4008.
113. Cai, G., et al., *Mediator structural conservation and implications for the regulation mechanism.* Structure, 2009. **17**(4): p. 559-67.

114. Asturias, F.J., et al., *Conserved structures of mediator and RNA polymerase II holoenzyme.* Science, 1999. **283**(5404): p. 985-7.
115. Myers, L.C. and R.D. Kornberg, *Mediator of transcriptional regulation.* Annu Rev Biochem, 2000. **69**: p. 729-49.
116. Malik, S. and R.G. Roeder, *The metazoan Mediator co-activator complex as an integrative hub for transcriptional regulation.* Nat Rev Genet, 2010. **11**(11): p. 761-72.
117. Borggrefe, T. and X. Yue, *Interactions between subunits of the Mediator complex with gene-specific transcription factors.* Semin Cell Dev Biol, 2011. **22**(7): p. 759-68.
118. Brzovic, P.S., et al., *The acidic transcription activator Gcn4 binds the mediator subunit Gal11/Med15 using a simple protein interface forming a fuzzy complex.* Mol Cell, 2011. **44**(6): p. 942-53.
119. Esnault, C., et al., *Mediator-dependent recruitment of TFIIH modules in preinitiation complex.* Mol Cell, 2008. **31**(3): p. 337-46.
120. Soutourina, J., et al., *Direct interaction of RNA polymerase II and mediator required for transcription in vivo.* Science, 2011. **331**(6023): p. 1451-4.
121. Kagey, M.H., et al., *Mediator and cohesin connect gene expression and chromatin architecture.* Nature, 2010. **467**(7314): p. 430-5.
122. Bjorklund, S. and C.M. Gustafsson, *The yeast Mediator complex and its regulation.* Trends Biochem Sci, 2005. **30**(5): p. 240-4.
123. Kremer, S.B., et al., *Role of Mediator in regulating Pol II elongation and nucleosome displacement in Saccharomyces cerevisiae.* Genetics, 2012. **191**(1): p. 95-106.
124. Mukundan, B. and A. Ansari, *Novel role for mediator complex subunit Srb5/Med18 in termination of transcription.* J Biol Chem, 2011. **286**(43): p. 37053-7.
125. Khorosjutina, O., et al., *A chromatin-remodeling protein is a component of fission yeast mediator.* J Biol Chem, 2010. **285**(39): p. 29729-37.
126. Zhu, X., et al., *Histone modifications influence mediator interactions with chromatin.* Nucleic Acids Res, 2011. **39**(19): p. 8342-54.
127. Lemieux, K. and L. Gaudreau, *Targeting of Swi/Snf to the yeast GAL1 UAS G requires the Mediator, TAF IIs, and RNA polymerase II.* Embo j, 2004. **23**(20): p. 4040-50.
128. Huang, Y., et al., *Mediator complex regulates alternative mRNA processing via the MED23 subunit.* Mol Cell, 2012. **45**(4): p. 459-69.
129. Thiaville, M.M., et al., *Activated transcription via mammalian amino acid response elements does not require enhanced recruitment of the Mediator complex.* Nucleic Acids Res, 2008. **36**(17): p. 5571-80.
130. Holstege, F.C., et al., *Dissecting the regulatory circuitry of a eukaryotic genome.* Cell, 1998. **95**(5): p. 717-28.
131. Ansari, S.A., Q. He, and R.H. Morse, *Mediator complex association with constitutively transcribed genes in yeast.* Proc Natl Acad Sci U S A, 2009. **106**(39): p. 16734-9.
132. Thompson, C.M. and R.A. Young, *General requirement for RNA polymerase II holoenzymes in vivo.* Proc Natl Acad Sci U S A, 1995. **92**(10): p. 4587-90.
133. Cantin, G.T., J.L. Stevens, and A.J. Berk, *Activation domain-mediator interactions promote transcription preinitiation complex assembly on promoter DNA.* Proc Natl Acad Sci U S A, 2003. **100**(21): p. 12003-8.

134. Rana, R., et al., *Med25 is required for RNA polymerase II recruitment to specific promoters, thus regulating xenobiotic and lipid metabolism in human liver.* Mol Cell Biol, 2011. **31**(3): p. 466-81.
135. Bischoff, E. and C. Vaquero, *In silico and biological survey of transcription-associated proteins implicated in the transcriptional machinery during the erythrocytic development of Plasmodium falciparum.* BMC Genomics, 2010. **11**: p. 34.
136. Iyer, L.M., et al., *Comparative genomics of transcription factors and chromatin proteins in parasitic protists and other eukaryotes.* Int J Parasitol, 2008. **38**(1): p. 1-31.
137. Hampsey, M., *Molecular genetics of the RNA polymerase II general transcriptional machinery.* Microbiol Mol Biol Rev, 1998. **62**(2): p. 465-503.
138. Aravind, L., et al., *Plasmodium biology: genomic gleanings.* Cell, 2003. **115**(7): p. 771-85.
139. Essien, K., S. Hannenhalli, and C.J. Stoeckert, Jr., *Computational analysis of constraints on noncoding regions, coding regions and gene expression in relation to Plasmodium phenotypic diversity.* PLoS One, 2008. **3**(9): p. e3122.
140. Chervitz, S.A., et al., *Comparison of the complete protein sets of worm and yeast: orthology and divergence.* Science, 1998. **282**(5396): p. 2022-8.
141. Riechmann, J.L., et al., *Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes.* Science, 2000. **290**(5499): p. 2105-10.
142. Lespinet, O., et al., *The role of lineage-specific gene family expansion in the evolution of eukaryotes.* Genome Res, 2002. **12**(7): p. 1048-59.
143. Emes, R.D., et al., *Comparison of the genomes of human and mouse lays the foundation of genome zoology.* Hum Mol Genet, 2003. **12**(7): p. 701-9.
144. Balaji, S., et al., *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, 2005. **33**(13): p. 3994-4006.
145. De Silva, E.K., et al., *Specific DNA-binding by apicomplexan AP2 transcription factors.* Proc Natl Acad Sci U S A, 2008. **105**(24): p. 8393-8.
146. Campbell, T.L., et al., *Identification and genome-wide prediction of DNA binding specificities for the ApiAP2 family of regulators from the malaria parasite.* PLoS Pathog, 2010. **6**(10): p. e1001165.
147. Boschet, C., et al., *Characterization of PfMyb1 transcription factor during erythrocytic development of 3D7 and F12 Plasmodium falciparum clones.* Mol Biochem Parasitol, 2004. **138**(1): p. 159-63.
148. Gissot, M., et al., *PfMyb1, a Plasmodium falciparum transcription factor, is required for intra-erythrocytic growth and controls key genes for cell cycle regulation.* J Mol Biol, 2005. **346**(1): p. 29-42.
149. Klug, A., *Zinc finger peptides for the regulation of gene expression.* J Mol Biol, 1999. **293**(2): p. 215-8.
150. Matthews, J.M. and M. Sunde, *Zinc fingers--folds for many occasions.* IUBMB Life, 2002. **54**(6): p. 351-5.
151. Hall, T.M., *Multiple modes of RNA recognition by zinc finger proteins.* Curr Opin Struct Biol, 2005. **15**(3): p. 367-73.
152. Gamsjaeger, R., et al., *Sticky fingers: zinc-fingers as protein-recognition motifs.* Trends Biochem Sci, 2007. **32**(2): p. 63-70.
153. Ravasi, T., et al., *Systematic characterization of the zinc-finger-containing proteins in the mouse transcriptome.* Genome Res, 2003. **13**(6B): p. 1430-42.

154. Miller, J., A.D. McLachlan, and A. Klug, *Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes.* EMBO J, 1985. **4**(6): p. 1609-14.
155. Laitaoja, M., J. Valjakka, and J. Janis, *Zinc Coordination Spheres in Protein Structures.* Inorg Chem, 2013. **52**(19): p. 10983-10991.
156. Krishna, S.S., I. Majumdar, and N.V. Grishin, *Structural classification of zinc fingers: survey and summary.* Nucleic Acids Res, 2003. **31**(2): p. 532-50.
157. Zhong, Z., et al., *Identification of a novel human zinc finger gene, ZNF438, with transcription inhibition activity.* J Biochem Mol Biol, 2007. **40**(4): p. 517-24.
158. Sham, M.H., et al., *The zinc finger gene Krox20 regulates HoxB2 (Hox2.8) during hindbrain segmentation.* Cell, 1993. **72**(2): p. 183-96.
159. Laity, J.H., B.M. Lee, and P.E. Wright, *Zinc finger proteins: new insights into structural and functional diversity.* Curr Opin Struct Biol, 2001. **11**(1): p. 39-46.
160. Lee, B.M., et al., *Dynamical behavior of the HIV-1 nucleocapsid protein.* J Mol Biol, 1998. **279**(3): p. 633-49.
161. Grishin, N.V., *Treble clef finger--a functionally diverse zinc-binding structural motif.* Nucleic Acids Res, 2001. **29**(8): p. 1703-14.
162. Zhu, W., et al., *The N-terminal domain of TFIIB from Pyrococcus furiosus forms a zinc ribbon.* Nat Struct Biol, 1996. **3**(2): p. 122-4.
163. Qian, X., et al., *Novel zinc finger motif in the basal transcriptional machinery: three-dimensional NMR studies of the nucleic acid binding domain of transcriptional elongation factor TFIIS.* Biochemistry, 1993. **32**(38): p. 9944-59.
164. Liang, S.D., et al., *DNA sequence preferences of GAL4 and PPR1: how a subset of Zn2 Cys6 binuclear cluster proteins recognizes DNA.* Mol Cell Biol, 1996. **16**(7): p. 3773-80.
165. Ladomery, M. and G. Dellaire, *Multifunctional zinc finger proteins in development and disease.* Ann Hum Genet, 2002. **66**(Pt 5-6): p. 331-42.
166. Kim, J.S. and C.O. Pabo, *Transcriptional repression by zinc finger peptides. Exploring the potential for applications in gene therapy.* J Biol Chem, 1997. **272**(47): p. 29795-800.
167. Carroll, D., *Progress and prospects: zinc-finger nucleases as gene therapy agents.* Gene Ther, 2008. **15**(22): p. 1463-8.
168. Lai, W.S., et al., *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, 2000. **275**(23): p. 17827-37.
169. Blackshear, P.J., *Tristetraprolin and other CCCH tandem zinc-finger proteins in the regulation of mRNA turnover.* Biochem Soc Trans, 2002. **30**(Pt 6): p. 945-52.
170. Baou, M., A. Jewell, and J.J. Murphy, *TIS11 family proteins and their roles in posttranscriptional gene regulation.* J Biomed Biotechnol, 2009. **2009**: p. 634520.
171. Hurt, J.A., et al., *A conserved CCCH-type zinc finger protein regulates mRNA nuclear adenylation and export.* J Cell Biol, 2009. **185**(2): p. 265-77.
172. Soucek, S., A.H. Corbett, and M.B. Fasken, *The long and the short of it: the role of the zinc finger polyadenosine RNA binding protein, Nab2, in control of poly(A) tail length.* Biochim Biophys Acta, 2012. **1819**(6): p. 546-54.

173. Carballo, E., W.S. Lai, and P.J. Blackshear, *Feedback inhibition of macrophage tumor necrosis factor-alpha production by tristetraprolin.* Science, 1998. **281**(5379): p. 1001-5.
174. Lai, W.S. and P.J. Blackshear, *Interactions of CCCH zinc finger proteins with mRNA: tristetraprolin-mediated AU-rich element-dependent mRNA degradation can occur in the absence of a poly(A) tail.* J Biol Chem, 2001. **276**(25): p. 23144-54.
175. De, J., et al., *Identification of four CCCH zinc finger proteins in Xenopus, including a novel vertebrate protein with four zinc fingers and severely restricted expression.* Gene, 1999. **228**(1-2): p. 133-45.
176. Tabara, H., et al., *pos-1 encodes a cytoplasmic zinc-finger protein essential for germline specification in C. elegans.* Development, 1999. **126**(1): p. 1-11.
177. Tenenhaus, C., et al., *PIE-1 is a bifunctional protein that regulates maternal and zygotic gene expression in the embryonic germ line of Caenorhabditis elegans.* Genes Dev, 2001. **15**(8): p. 1031-40.
178. Huang, N.N., et al., *MEX-3 interacting proteins link cell polarity to asymmetric gene expression in Caenorhabditis elegans.* Development, 2002. **129**(3): p. 747-59.
179. Shimada, M., H. Kawahara, and H. Doi, *Novel family of CCCH-type zinc-finger proteins, MOE-1, -2 and -3, participates in C. elegans oocyte maturation.* Genes Cells, 2002. **7**(9): p. 933-47.
180. Morking, P.A., et al., *TcZFP1: a CCCH zinc finger protein of Trypanosoma cruzi that binds poly-C oligoribonucleotides in vitro.* Biochem Biophys Res Commun, 2004. **319**(1): p. 169-77.
181. Caro, F., et al., *Protein interactions within the TcZFP zinc finger family members of Trypanosoma cruzi: implications for their functions.* Biochem Biophys Res Commun, 2005. **333**(3): p. 1017-25.
182. Benz, C., et al., *The Trypanosoma brucei zinc finger protein ZC3H18 is involved in differentiation.* Mol Biochem Parasitol, 2011. **177**(2): p. 148-51.
183. Ouna, B.A., et al., *The Trypanosoma brucei CCCH zinc finger proteins ZC3H12 and ZC3H13.* Mol Biochem Parasitol, 2012. **183**(2): p. 184-8.
184. Wang, L., et al., *OsLIC, a Novel CCCH-Type Zinc Finger Protein with Transcription Activation, Mediates Rice Architecture via Brassinosteroids Signaling.* PLoS One, 2008. **3**(10): p. e3521.
185. Kong, Z., et al., *A novel nuclear-localized CCCH-type zinc finger protein, OsDOS, is involved in delaying leaf senescence in rice.* Plant Physiol, 2006. **141**(4): p. 1376-88.
186. Li, Z. and T.L. Thomas, *PEI1, an embryo-specific zinc finger protein gene required for heart-stage embryo formation in Arabidopsis.* Plant Cell, 1998. **10**(3): p. 383-98.
187. Kim, D.H., et al., *SOMNUS, a CCCH-type zinc finger protein in Arabidopsis, negatively regulates light-dependent seed germination downstream of PIL5.* Plant Cell, 2008. **20**(5): p. 1260-77.
188. Grabowska, A., et al., *Characterization of CsSEF1 gene encoding putative CCCH-type zinc finger protein expressed during cucumber somatic embryogenesis.* J Plant Physiol, 2009. **166**(3): p. 310-23.
189. Zhu, J.K., *Salt and drought stress signal transduction in plants.* Annu Rev Plant Biol, 2002. **53**: p. 247-73.

190. Deng, H., et al., *A CCCH-type zinc finger nucleic acid-binding protein quantitatively confers resistance against rice bacterial blight disease.* Plant Physiol, 2012. **158**(2): p. 876-89.
191. Guo, Y.H., et al., *GhZFP1, a novel CCCH-type zinc finger protein from cotton, enhances salt stress tolerance and fungal disease resistance in transgenic tobacco by interacting with GZIRD21A and GZIPR5.* New Phytol, 2009. **183**(1): p. 62-75.
192. Sun, J., et al., *The CCCH-type zinc finger proteins AtSZF1 and AtSZF2 regulate salt stress responses in Arabidopsis.* Plant Cell Physiol, 2007. **48**(8): p. 1148-58.
193. Peng, X., et al., *CCCH-type zinc finger family in maize: genome-wide identification, classification and expression profiling under abscisic acid and drought treatments.* PLoS One, 2012. **7**(7): p. e40120.
194. Gritzmacher, C.A. and R.T. Reese, *Protein and nucleic acid synthesis during synchronized growth of Plasmodium falciparum.* J Bacteriol, 1984. **160**(3): p. 1165-7.
195. Shock, J.L., K.F. Fischer, and J.L. DeRisi, *Whole-genome analysis of mRNA decay in Plasmodium falciparum reveals a global lengthening of mRNA half-life during the intra-erythrocytic development cycle.* Genome Biol, 2007. **8**(7): p. R134.
196. Deitsch, K., et al., *Mechanisms of gene regulation in Plasmodium.* Am J Trop Med Hyg, 2007. **77**(2): p. 201-8.
197. Jakymiw, A., et al., *The role of GW/P-bodies in RNA processing and silencing.* J Cell Sci, 2007. **120**(Pt 8): p. 1317-23.
198. Shaw, P.J., et al., *Characterization of human malaria parasite Plasmodium falciparum eIF4E homologue and mRNA 5' cap status.* Mol Biochem Parasitol, 2007. **155**(2): p. 146-55.
199. Sonenberg, N. and A.G. Hinnebusch, *New modes of translational control in development, behavior, and disease.* Mol Cell, 2007. **28**(5): p. 721-9.
200. Parker, R. and U. Sheth, *P bodies and the control of mRNA translation and degradation.* Mol Cell, 2007. **25**(5): p. 635-46.
201. Abaza, I. and F. Gebauer, *Trading translation with RNA-binding proteins.* RNA, 2008. **14**(3): p. 404-9.
202. Mair, G.R., et al., *Regulation of sexual development of Plasmodium by translational repression.* Science, 2006. **313**(5787): p. 667-9.
203. Li, B., M. Carey, and J.L. Workman, *The role of chromatin during transcription.* Cell, 2007. **128**(4): p. 707-19.
204. Kouzarides, T., *Chromatin modifications and their function.* Cell, 2007. **128**(4): p. 693-705.
205. Belotserkovskaya, R. and S.L. Berger, *Interplay between chromatin modifying and remodeling complexes in transcriptional regulation.* Crit Rev Eukaryot Gene Expr, 1999. **9**(3-4): p. 221-30.
206. Euskirchen, G., R.K. Auerbach, and M. Snyder, *SWI/SNF chromatin-remodeling factors: multiscale analyses and diverse functions.* J Biol Chem, 2012. **287**(37): p. 30897-905.
207. Ramirez-Carrozzi, V.R., et al., *A unifying model for the selective regulation of inducible transcription by CpG islands and nucleosome remodeling.* Cell, 2009. **138**(1): p. 114-28.

208. Hirschhorn, J.N., et al., *Evidence that SNF2/SWI2 and SNF5 activate transcription in yeast by altering chromatin structure.* Genes Dev, 1992. **6**(12A): p. 2288-98.
209. Klochendler-Yeivin, A., et al., *The murine SNF5/INI1 chromatin remodeling factor is essential for embryonic development and tumor suppression.* EMBO Rep, 2000. **1**(6): p. 500-6.
210. Reyes, J.C., et al., *Altered control of cellular proliferation in the absence of mammalian brahma (SNF2alpha).* EMBO J, 1998. **17**(23): p. 6979-91.
211. Versteege, I., et al., *Truncating mutations of hSNF5/INI1 in aggressive paediatric cancer.* Nature, 1998. **394**(6689): p. 203-6.
212. Jenuwein, T. and C.D. Allis, *Translating the histone code.* Science, 2001. **293**(5532): p. 1074-80.
213. MacDonald, V.E. and L.J. Howe, *Histone acetylation: where to go and how to get there.* Epigenetics, 2009. **4**(3): p. 139-43.
214. Garcia-Dominguez, M. and J.C. Reyes, *SUMO association with repressor complexes, emerging routes for transcriptional control.* Biochim Biophys Acta, 2009. **1789**(6-8): p. 451-9.
215. Greer, E.L. and Y. Shi, *Histone methylation: a dynamic mark in health, disease and inheritance.* Nat Rev Genet, 2012. **13**(5): p. 343-57.
216. Latham, J.A. and S.Y. Dent, *Cross-regulation of histone modifications.* Nat Struct Mol Biol, 2007. **14**(11): p. 1017-24.
217. Nakayama, J., et al., *Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly.* Science, 2001. **292**(5514): p. 110-3.
218. Morris, S.A., et al., *Identification of histone H3 lysine 36 acetylation as a highly conserved histone modification.* J Biol Chem, 2007. **282**(10): p. 7632-40.
219. Bird, A., *The essentials of DNA methylation.* Cell, 1992. **70**(1): p. 5-8.
220. Chen, R.Z., et al., *DNA hypomethylation leads to elevated mutation rates.* Nature, 1998. **395**(6697): p. 89-93.
221. Panning, B. and R. Jaenisch, *RNA and the epigenetic regulation of X chromosome inactivation.* Cell, 1998. **93**(3): p. 305-8.
222. Li, E., C. Beard, and R. Jaenisch, *Role for DNA methylation in genomic imprinting.* Nature, 1993. **366**(6453): p. 362-5.
223. Reik, W., W. Dean, and J. Walter, *Epigenetic reprogramming in mammalian development.* Science, 2001. **293**(5532): p. 1089-93.
224. Li, E., T.H. Bestor, and R. Jaenisch, *Targeted mutation of the DNA methyltransferase gene results in embryonic lethality.* Cell, 1992. **69**(6): p. 915-26.
225. Jones, P.A., *DNA methylation errors and cancer.* Cancer Res, 1996. **56**(11): p. 2463-7.
226. Robertson, K.D. and A.P. Wolffe, *DNA methylation in health and disease.* Nat Rev Genet, 2000. **1**(1): p. 11-9.
227. Tate, P.H. and A.P. Bird, *Effects of DNA methylation on DNA-binding proteins and gene expression.* Curr Opin Genet Dev, 1993. **3**(2): p. 226-31.
228. Boyes, J. and A. Bird, *Repression of genes by DNA methylation depends on CpG density and promoter strength: evidence for involvement of a methyl-CpG binding protein.* EMBO J, 1992. **11**(1): p. 327-33.
229. Wolffe, A.P., *Packaging principle: how DNA methylation and histone acetylation control the transcriptional activity of chromatin.* J Exp Zool, 1998. **282**(1-2): p. 239-44.

230. Miao, J., et al., *The malaria parasite Plasmodium falciparum histones: organization, expression, and acetylation.* Gene, 2006. **369**: p. 53-65.
231. Cui, L., et al., *PfGCN5-mediated histone H3 acetylation plays a key role in gene expression in Plasmodium falciparum.* Eukaryot Cell, 2007. **6**(7): p. 1219-27.
232. Hakimi, M.A. and K.W. Deitsch, *Epigenetics in Apicomplexa: control of gene expression during cell cycle progression, differentiation and antigenic variation.* Curr Opin Microbiol, 2007. **10**(4): p. 357-62.
233. Salcedo-Amaya, A.M., et al., *Dynamic histone H3 epigenome marking during the intraerythrocytic cycle of Plasmodium falciparum.* Proc Natl Acad Sci U S A, 2009. **106**(24): p. 9655-60.
234. Trelle, M.B., et al., *Global histone analysis by mass spectrometry reveals a high content of acetylated lysine residues in the malaria parasite Plasmodium falciparum.* J Proteome Res, 2009. **8**(7): p. 3439-50.
235. Cui, L. and J. Miao, *Cytotoxic effect of curcumin on malaria parasite Plasmodium falciparum: inhibition of histone acetylation and generation of reactive oxygen species.* Antimicrob Agents Chemother, 2007. **51**(2): p. 488-94.
236. Cui, L., et al., *Histone acetyltransferase inhibitor anacardic acid causes changes in global gene expression during in vitro Plasmodium falciparum development.* Eukaryot Cell, 2008. **7**(7): p. 1200-10.
237. Chaal, B.K., et al., *Histone deacetylases play a major role in the transcriptional regulation of the Plasmodium falciparum life cycle.* PLoS Pathog, 2010. **6**(1): p. e1000737.
238. Gupta, A.P., et al., *Dynamic epigenetic regulation of gene expression during the life cycle of malaria parasite Plasmodium falciparum.* PLoS Pathog, 2013. **9**(2): p. e1003170.
239. Chookajorn, T., et al., *Epigenetic memory at malaria virulence genes.* Proc Natl Acad Sci U S A, 2007. **104**(3): p. 899-902.
240. Lopez-Rubio, J.J., et al., *5' flanking region of var genes nucleate histone modification patterns linked to phenotypic inheritance of virulence traits in malaria parasites.* Mol Microbiol, 2007. **66**(6): p. 1296-305.
241. Cortes, A., et al., *Epigenetic silencing of Plasmodium falciparum genes linked to erythrocyte invasion.* PLoS Pathog, 2007. **3**(8): p. e107.
242. Ponts, N., et al., *Genome-wide mapping of DNA methylation in the human malaria parasite Plasmodium falciparum.* Cell Host Microbe, 2013. **14**(6): p. 696-706.
243. Cremer, T. and C. Cremer, *Chromosome territories, nuclear architecture and gene regulation in mammalian cells.* Nat Rev Genet, 2001. **2**(4): p. 292-301.
244. Babu, M.M., et al., *Eukaryotic gene regulation in three dimensions and its impact on genome evolution.* Curr Opin Genet Dev, 2008. **18**(6): p. 571-82.
245. Hakim, O., M.H. Sung, and G.L. Hager, *3D shortcuts to gene regulation.* Curr Opin Cell Biol, 2010. **22**(3): p. 305-13.
246. Cremer, T., et al., *Role of chromosome territories in the functional compartmentalization of the cell nucleus.* Cold Spring Harb Symp Quant Biol, 1993. **58**: p. 777-92.
247. Cremer, T. and M. Cremer, *Chromosome territories.* Cold Spring Harb Perspect Biol, 2010. **2**(3): p. a003889.
248. Ferrai, C., et al., *Gene positioning.* Cold Spring Harb Perspect Biol, 2010. **2**(6): p. a000588.

249. Guelen, L., et al., *Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions.* Nature, 2008. **453**(7197): p. 948-51.
250. Lanctot, C., et al., *Dynamic genome architecture in the nuclear space: regulation of gene expression in three dimensions.* Nat Rev Genet, 2007. **8**(2): p. 104-15.
251. Brown, K.E., et al., *Dynamic repositioning of genes in the nucleus of lymphocytes preparing for cell division.* Mol Cell, 1999. **3**(2): p. 207-17.
252. Craig, J.M. and W.A. Bickmore, *The distribution of CpG islands in mammalian chromosomes.* Nat Genet, 1994. **7**(3): p. 376-82.
253. Houtsmuller, A.B., et al., *Action of DNA repair endonuclease ERCC1/XPF in living cells.* Science, 1999. **284**(5416): p. 958-61.
254. Carmo-Fonseca, M., L. Mendes-Soares, and I. Campos, *To be or not to be in the nucleolus.* Nat Cell Biol, 2000. **2**(6): p. E107-12.
255. Carmo-Fonseca, M., *The contribution of nuclear compartmentalization to gene regulation.* Cell, 2002. **108**(4): p. 513-21.
256. Misteli, T., J.F. Caceres, and D.L. Spector, *The dynamics of a pre-mRNA splicing factor in living cells.* Nature, 1997. **387**(6632): p. 523-7.
257. Cha, J.H., *Transcriptional dysregulation in Huntington's disease.* Trends Neurosci, 2000. **23**(9): p. 387-92.
258. Tapscott, S.J. and C.A. Thornton, *Biomedicine. Reconstructing myotonic dystrophy.* Science, 2001. **293**(5531): p. 816-7.
259. Ralph, S.A. and A. Scherf, *The epigenetic control of antigenic variation in Plasmodium falciparum.* Curr Opin Microbiol, 2005. **8**(4): p. 434-40.
260. Duraisingh, M.T., et al., *Heterochromatin silencing and locus repositioning linked to regulation of virulence genes in Plasmodium falciparum.* Cell, 2005. **121**(1): p. 13-24.
261. Dzikowski, R., et al., *Mechanisms underlying mutually exclusive expression of virulence genes by malaria parasites.* EMBO Rep, 2007. **8**(10): p. 959-65.
262. Lopez-Rubio, J.J., L. Mancio-Silva, and A. Scherf, *Genome-wide analysis of heterochromatin associates clonally variant gene regulation with perinuclear repressive centers in malaria parasites.* Cell Host Microbe, 2009. **5**(2): p. 179-90.
263. Ralph, S.A., C. Scheidig-Benatar, and A. Scherf, *Antigenic variation in Plasmodium falciparum is associated with movement of var loci between subnuclear locations.* Proc Natl Acad Sci U S A, 2005. **102**(15): p. 5414-9.
264. Voss, T.S., et al., *A var gene promoter controls allelic exclusion of virulence genes in Plasmodium falciparum malaria.* Nature, 2006. **439**(7079): p. 1004-8.
265. Weiner, A., et al., *3D nuclear architecture reveals coupled cell cycle dynamics of chromatin and nuclear pores in the malaria parasite Plasmodium falciparum.* Cell Microbiol, 2011. **13**(7): p. 967-77.
266. Issar, N., et al., *Differential sub-nuclear localisation of repressive and activating histone methyl modifications in P. falciparum.* Microbes Infect, 2009. **11**(3): p. 403-7.
267. Luah, Y.H., et al., *A moonlighting function of Plasmodium falciparum histone 3, mono-methylated at lysine 9?* PLoS One, 2010. **5**(4): p. e10252.
268. Volz, J., et al., *Potential epigenetic regulatory proteins localise to distinct nuclear sub-compartments in Plasmodium falciparum.* Int J Parasitol, 2009. **40**(1): p. 109-21.
269. Rosario, V., *Cloning of naturally occurring mixed infections of malaria parasites.* Science, 1981. **212**(4498): p. 1037-8.

270. Hu, G., et al., *Selection of long oligonucleotides for gene expression microarrays using weighted rank-sum strategy.* BMC Bioinformatics, 2007. **8**: p. 350.
271. Waterkeyn, J.G., B.S. Crabb, and A.F. Cowman, *Transfection of the human malaria parasite Plasmodium falciparum.* Int J Parasitol, 1999. **29**(6): p. 945-55.
272. Flueck, C., et al., *Plasmodium falciparum heterochromatin protein 1 marks genomic loci linked to phenotypic variation of exported virulence factors.* PLoS Pathog, 2009. **5**(9): p. e1000569.
273. Zhao, R., et al., *Gene bookmarking accelerates the kinetics of post-mitotic transcriptional re-activation.* Nat Cell Biol, 2011. **13**(11): p. 1295-304.
274. Villaverde, A. and M.M. Carrio, *Protein aggregation in recombinant bacteria: biological role of inclusion bodies.* Biotechnol Lett, 2003. **25**(17): p. 1385-95.
275. Dale, G.E., et al., *Improving protein solubility through rationally designed amino acid replacements: solubilization of the trimethoprim-resistant type S1 dihydrofolate reductase.* Protein Eng, 1994. **7**(7): p. 933-9.
276. Farinas, E.T., T. Bulter, and F.H. Arnold, *Directed enzyme evolution.* Curr Opin Biotechnol, 2001. **12**(6): p. 545-51.
277. Silvestrini, F., et al., *Protein export marks the early phase of gametocytogenesis of the human malaria parasite Plasmodium falciparum.* Mol Cell Proteomics, 2010. **9**(7): p. 1437-48.
278. Saliba, K.S. and M. Jacobs-Lorena, *Production of Plasmodium falciparum gametocytes in vitro.* Methods Mol Biol, 2013. **923**: p. 17-25.
279. Fivelman, Q.L., et al., *Improved synchronous production of Plasmodium falciparum gametocytes in vitro.* Mol Biochem Parasitol, 2007. **154**(1): p. 119-23.
280. Brockelman, C.R., *Conditions favoring gametocytogenesis in the continuous culture of Plasmodium falciparum.* J Protozool, 1982. **29**(3): p. 454-8.
281. Graves, P.M., R. Carter, and K.M. McNeill, *Gametocyte production in cloned lines of Plasmodium falciparum.* Am J Trop Med Hyg, 1984. **33**(6): p. 1045-50.
282. Chutmongkonkul, M., W.A. Maier, and H.M. Seitz, *A new model for testing gametocytocidal effects of some antimalarial drugs on Plasmodium falciparum in vitro.* Ann Trop Med Parasitol, 1992. **86**(3): p. 207-15.
283. Bruce, M.C., et al., *Commitment of the malaria parasite Plasmodium falciparum to sexual and asexual development.* Parasitology, 1990. **100 Pt 2**: p. 191-200.
284. Besson, A., S.F. Dowdy, and J.M. Roberts, *CDK inhibitors: cell cycle regulators and beyond.* Dev Cell, 2008. **14**(2): p. 159-69.
285. Cong, F., et al., *Requirement for a nuclear function of beta-catenin in Wnt signaling.* Mol Cell Biol, 2003. **23**(23): p. 8462-70.
286. Whiteside, S.T. and S. Goodbourn, *Signal transduction and nuclear targeting: regulation of transcription factor activity by subcellular localisation.* J Cell Sci, 1993. **104 (Pt 4)**: p. 949-55.
287. Moll, T., et al., *The role of phosphorylation and the CDC28 protein kinase in cell cycle-regulated nuclear import of the S. cerevisiae transcription factor SWI5.* Cell, 1991. **66**(4): p. 743-58.
288. Pomeranz, M.C., et al., *The Arabidopsis tandem zinc finger protein AtTZF1 traffics between the nucleus and cytoplasmic foci and binds both DNA and RNA.* Plant Physiol, 2010. **152**(1): p. 151-65.

289. Arnoys, E.J. and J.L. Wang, *Dual localization: proteins in extracellular and intracellular compartments.* Acta Histochem, 2007. **109**(2): p. 89-110.
290. de Lanerolle, P. and L. Serebryannyy, *Nuclear actin and myosins: life without filaments.* Nat Cell Biol, 2011. **13**(11): p. 1282-8.
291. Perkins, N.D., *Integrating cell-signalling pathways with NF-kappaB and IKK function.* Nat Rev Mol Cell Biol, 2007. **8**(1): p. 49-62.
292. Cockell, M. and S.M. Gasser, *Nuclear compartments and gene regulation.* Curr Opin Genet Dev, 1999. **9**(2): p. 199-205.
293. Ishii, K., et al., *Chromatin boundaries in budding yeast: the nuclear pore connection.* Cell, 2002. **109**(5): p. 551-62.
294. Ponts, N., et al., *Nucleosome landscape and control of transcription in the human malaria parasite.* Genome Res, 2010. **20**(2): p. 228-38.
295. Lieb, J.D., et al., *Promoter-specific binding of Rap1 revealed by genome-wide maps of protein-DNA association.* Nat Genet, 2001. **28**(4): p. 327-34.
296. Zeitlinger, J., et al., *Program-specific distribution of a transcription factor dependent on partner transcription factor and MAPK signaling.* Cell, 2003. **113**(3): p. 395-404.
297. Gaudet, J. and S.E. Mango, *Regulation of organogenesis by the Caenorhabditis elegans FoxA protein PHA-4.* Science, 2002. **295**(5556): p. 821-5.
298. Johnson, D.S., et al., *Genome-wide mapping of in vivo protein-DNA interactions.* Science, 2007. **316**(5830): p. 1497-502.
299. Fernandez, P.C., et al., *Genomic targets of the human c-Myc protein.* Genes Dev, 2003. **17**(9): p. 1115-29.
300. Robertson, G., et al., *Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing.* Nat Methods, 2007. **4**(8): p. 651-7.
301. Sandmann, T., et al., *A core transcriptional network for early mesoderm development in Drosophila melanogaster.* Genes Dev, 2007. **21**(4): p. 436-49.
302. Zeitlinger, J., et al., *Whole-genome ChIP-chip analysis of Dorsal, Twist, and Snail suggests integration of diverse patterning processes in the Drosophila embryo.* Genes Dev, 2007. **21**(4): p. 385-90.
303. Li, X.Y., et al., *Transcription factors bind thousands of active and inactive regions in the Drosophila blastoderm.* PLoS Biol, 2008. **6**(2): p. e27.
304. Lin, S. and A.D. Riggs, *The general affinity of lac repressor for E. coli DNA: implications for gene regulation in procaryotes and eucaryotes.* Cell, 1975. **4**(2): p. 107-11.
305. Chen, X., et al., *Integration of external signaling pathways with the core transcriptional network in embryonic stem cells.* Cell, 2008. **133**(6): p. 1106-17.
306. Tanay, A., *Extensive low-affinity transcriptional interactions in the yeast genome.* Genome Res, 2006. **16**(8): p. 962-72.
307. Vakoc, C.R., et al., *Proximity among distant regulatory elements at the beta-globin locus requires GATA-1 and FOG-1.* Mol Cell, 2005. **17**(3): p. 453-62.
308. Yochum, G.S., et al., *A beta-catenin/TCF-coordinated chromatin loop at MYC integrates 5' and 3' Wnt responsive enhancers.* Proc Natl Acad Sci U S A, 2010. **107**(1): p. 145-50.
309. Ganapathi, M., et al., *Extensive role of the general regulatory factors, Abf1 and Rap1, in determining genome-wide chromatin structure in budding yeast.* Nucleic Acids Res, 2011. **39**(6): p. 2032-44.

310. Krungkrai, J., P. Prapunwattana, and S.R. Krungkrai, *Ultrastructure and function of mitochondria in gametocytic stage of Plasmodium falciparum.* Parasite, 2000. **7**(1): p. 19-26.
311. Bongaerts, G., *Do the mitochondria of malaria parasites behave like the phoenix after return in the mosquito? Regeneration of degenerated mitochondria is required for successful Plasmodium infection.* Med Hypotheses, 2005. **64**(3): p. 599-601.
312. Peter, M., et al., *FAR1 links the signal transduction pathway to the cell cycle machinery in yeast.* Cell, 1993. **73**(4): p. 747-60.
313. Peter, M. and I. Herskowitz, *Direct inhibition of the yeast cyclin-dependent kinase Cdc28-Cln by Far1.* Science, 1994. **265**(5176): p. 1228-31.
314. McCord, R., et al., *Rfm1, a novel tethering factor required to recruit the Hst1 histone deacetylase for repression of middle sporulation genes.* Mol Cell Biol, 2003. **23**(6): p. 2009-16.
315. Mantrova, E.Y. and T. Hsu, *Down-regulation of transcription factor CF2 by Drosophila Ras/MAP kinase signaling in oogenesis: cytoplasmic retention and degradation.* Genes Dev, 1998. **12**(8): p. 1166-75.
316. Oh, J.E., J.A. Han, and E.S. Hwang, *Downregulation of transcription factor, Sp1, during cellular senescence.* Biochem Biophys Res Commun, 2007. **353**(1): p. 86-91.
317. Wen, X., et al., *Downregulation of the transcription factor KLF4 is required for the lineage commitment of T cells.* Cell Res, 2011. **21**(12): p. 1701-10.
318. Hoetelmans, R.W., et al., *Effects of acetone, methanol, or paraformaldehyde on cellular structure, visualized by reflection contrast microscopy and transmission and scanning electron microscopy.* Appl Immunohistochem Mol Morphol, 2001. **9**(4): p. 346-51.
319. Melan, M.A. and G. Sluder, *Redistribution and differential extraction of soluble proteins in permeabilized cultured cells. Implications for immunofluorescence microscopy.* J Cell Sci, 1992. **101 (Pt 4)**: p. 731-43.
320. Denard, B., et al., *The membrane-bound transcription factor CREB3L1 is activated in response to virus infection to inhibit proliferation of virus-infected cells.* Cell Host Microbe, 2011. **10**(1): p. 65-74.
321. Kim, Y.S., et al., *A membrane-bound NAC transcription factor regulates cell division in Arabidopsis.* Plant Cell, 2006. **18**(11): p. 3132-44.
322. Kopan, R., *Notch: a membrane-bound transcription factor.* J Cell Sci, 2002. **115**(Pt 6): p. 1095-7.
323. Bujalka, H., et al., *MYRF is a membrane-associated transcription factor that autoproteolytically cleaves to directly activate myelin genes.* PLoS Biol, 2013. **11**(8): p. e1001625.
324. Pradel, G., *Proteins of the malaria parasite sexual stages: expression, function and potential for transmission blocking strategies.* Parasitology, 2007. **134**(Pt.14): p. 1911-29.