Development of an Inducible Transcriptional Control System in *Plasmodium falciparum* with Applications to Targeted Genome Editing

by

Jeffrey C. Wagner

A.B. Genetics, Cell and Developmental Biology modified with Engineering Sciences (2006) Dartmouth College, Hanover, NH

Submitted to the Department of Biological Engineering in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Biological Engineering at the Massachusetts Institute of Technology

June 2014

© 2014 Massachusetts Institute of Technology. All rights reserved.

Author

Signature redacted

Jeffrey C. Wagner
Department of Biological Engineering, MIT

Certified by

Signature redacted

Jacquin C. Niles
Associate Professor, Department of Biological Engineering, MIT
Thesis Advisor

Approved by

Signature redacted

Forest M. White
Associate Professor, Department of Biological Engineering, MIT
Co-Chair, Course XX Graduate Program Committee
This doctoral thesis has been examined by a committee of the Department of Biological Engineering as follows:

Signature redacted

Certified by

..................................................
Edward F. Delong
Professor, Departments of Biological Engineering and Civil and Environmental Engineering, MIT
Thesis Committee Chair

Signature redacted

Certified by

..................................................
Jacquin C. Niles
Associate Professor, Department of Biological Engineering, MIT
Thesis Advisor

Signature redacted

Certified by

..................................................
K. Dane Wittrup
Professor, Departments of Biological Engineering and Chemical Engineering and Materials Science and Engineering, MIT
Thesis Committee Member
Development of an Inducible Transcriptional Control System in *Plasmodium falciparum* with Applications to Targeted Genome Engineering

by

Jeffrey C. Wagner

Submitted to the Department of Biological Engineering on March 28, 2014 in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Biological Engineering

ABSTRACT

Malaria accounts for over 500,000 deaths each year. While malaria is caused by multiple distinct parasites of the genus *Plasmodium*, *P. falciparum* is responsible for the majority of morbidity and mortality due to the disease. Despite this fact, molecular tools for genetic experimentation in the parasite remain underdeveloped. In particular, the ability to inducibly control gene expression and edit the genome in a site-specific manner present significant challenges. In addition, the building of genetic constructs poses challenges due to the required vector size and the high A+T richness of the *P. falciparum* genetic regulatory elements. This work begins by presenting the first vector family for use in the parasite made up of modular parts and encompassing all selectable markers and replication technologies in current use. It also discusses the development of a 2A like viral peptide tag for use in expression of multiple differentially localizing proteins from a single expression cassette. Based on this work, we were then able to construct vectors to reconstitute the T7 RNA polymerase expression system in *P. falciparum*, functionally creating the first system for directed expression of non-coding RNA in the parasite. We were then able to use this expression system to adapt a clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9 system in the parasite and use it to achieve genome editing at high efficiency at multiple loci. The data imply an adaptable system to readily edit the genome of the parasite and holds promise for the ability to create gene knockouts, perform allelic replacements, and add regulatory elements into the parasite significantly faster than has been previously demonstrated. This also represents the first illustration of the functionality of a CRISPR/Cas9 system in any non-bacterial pathogenic organism. In addition, we were able to introduce the lac repression system in order to regulate the T7 RNA polymerase dependent production of RNA and have created the first inducible expression system for RNA in any apicomplexan parasite. This work provides several new molecular tools and frameworks to aid in the study of, and fight against, malaria.

Advisor: Jacquin C. Niles
Title: Associate Professor of Biological Engineering
ACKNOWLEDGEMENTS

This work represents approximately six-and-a-half years of work and during that period I have been fortunate to receive many different forms of support from MIT and my friends and family.

I would first like to thank the Department of Biological Engineering for offering me a spot in the doctoral program so that I could develop as a scientist. Specifically, I would like to thank the Department Chair Doug Lauffenburger for setting up this unique environment in which to learn and grow. I would like to thank the members of the faculty who have given me both useful feedback and resources with which to perform my experiments. Specifically, I would like to thank John Essigmann, Leona Samson, and the staff of the bio-micro center for allowing me the use of their facilities.

I would also like to say thank you to my earlier scientific mentors at the Dartmouth Medical School and the Max Planck Institute for Neurobiology without whose help I would never have reached this point. My undergraduate advisor George O’Toole has been both a scientific and personal inspiration for me from the point that he gave me my first real job as a researcher as a sophomore in college. I would also like to acknowledge the group of researchers with whom I worked closely during this period of my life and learned how to go about the business of scientific research. I’d specifically like to thank Michael Zegans, Christine Toutain, Russell Monds, Dan MacEachran, Stefan Weinges, and Amparo Acker-Palmer for teaching me in these early stages of my career.

Apart from my work in the lab at MIT, I would like to thank Forest White and Ernest Fraenkel who were my mentors for my teaching assistantship during my second year of graduate school. Their dedication to both science and teaching were a strong positive example for me.

I would like to acknowledge my thesis committee members Ed Delong and Dane Wittrup and to thank them for their help and suggestions over the years. I would, of course, also like to thank my advisor Jacquin Niles for funding my work, but more importantly for all of his guidance during graduate school.

I am also grateful to the NIGMS Biotechnology Training Program which gave me not only three years of funding, but allowed me to perform an internship in the private sector. With that in mind, I would like to thank Jennifer Leeds of the Novartis Institutes for Biomedical Research for supervising my internship and allowing me to learn about conducting science in a non-academic setting.

I would like to thank the other members of the Niles Lab for their help and friendship during this journey. Modern biological science cannot be performed alone and every member of the lab has some hand in this work, whether suggesting useful ideas, helping me with a protocol, or simply allowing me to bounce ideas off of them.

I can’t express my gratitude to all my friends who have been there for me over the years. I would never have made it this far without the support network I have built up around me and I share this accomplishment with all of them.

I would also like to thank my classmates who entered the graduate program at the same time as me in 2007. They are truly a smart, funny and down to Earth group of extremely talented scientists. From help with work in the lab to socialization outside of it, this is truly a special group that I count myself lucky to be a part of.
I would like to thank the members of my family who have been a constant source of love, support, and laughter. My parents Ken and Marianne have always been supportive of my goals and been there to offer advice or counsel when needed, or just a sympathetic ear when called for. My brother Greg, and sisters Mary and Lizzie have also been an enormous positive presence in my life and I thank them for that.

Finally, I would like to dedicate this thesis to two people who have had a big influence on my life. First, to my grandfather Charles Wagner who passed away recently but was a major factor in instilling in me a love of learning and of life. Second, to my niece Abigail. Graduate school has had many ups and downs but even after a bad day, she could always make me smile.

-Jeff Wagner
Cambridge, Massachusetts
March 2014
ATTRIBUTIONS

As typical to any scientific endeavor, this work was done in collaboration with several other people and chapters that have been published or have been submitted for publication have multiple authors. All published or submitted articles are noted at the beginning of each chapter with the relevant citation.

Chapter 2 was co-written by Stephen J. Goldfless, Suresh M. Ganesan, and Jacquin C. Niles, and me, all of the Massachusetts Institute of Technology. Additional authors who aided in this work technically and intellectually were Marcus C.S. Lee and David A. Fiddock, both of Columbia University. All DNA constructs were built by Stephen J. Goldfless, Suresh M. Ganesan, and me. Microscopy was carried out by Suresh M. Ganesan and Marcus C.S. Lee. I would also like to acknowledge Robert M.Q. Shanks of the University of Pittsburgh for his helpful conversations regarding yeast gap repair cloning, and Akhil Vaidya of Drexal University University for providing the DSM-1 reagent used.

Chapter 3 was co-written by Jacquin C. Niles and me. Additional authors who aided in this work technically and intellectually were Randy Platt, Stephen J. Goldfless, and Feng Zhang, all of the Massachusetts Institute of Technology. All constructs and experiments in *P. falciparum* were constructed and carried out by me. Stephen J. Goldfless performed the scanning electron microscopy. Randy Platt performed the *in vitro* cleavage assays. I would like to acknowledge John Essigmann and Nidhi Shrivastav for guiding me in carrying out the *in vitro* T7 RNAP activity assays and for letting me use their lab space to carry them out. I would also like to acknowledge Diana Taylor of the University of Hawaii for providing the KAHRP antibody.

Chapter 4 was all work carried out by me in consultation with Jacquin Niles. I would also like to acknowledge Aleja Falla and Sebastian Nassamu who also have experimented with the regulation of T7 RNAP in *P. falciparum* using the T7 RNAP system for their helpful impact.
TABLE OF CONTENTS

ABSTRACT .......................................................................................................................................................... 3

ATTRIBUTIONS ..................................................................................................................................................... 6

TABLE OF CONTENTS ..................................................................................................................................... 7

CHAPTER 1: INTRODUCTION .......................................................................................................................... 12

1.1 THE DISEASE BURDEN OF MALARIA ........................................................................................................... 12

1.2 THE LIFE CYCLE AND GENOME OF P. FALCIPARUM ......................................................................................... 12

Figure 1.1: The lifecycle of P. falciparum ........................................................................................................... 14

1.3 CURRENT MOLECULAR TOOLS FOR THE MANIPULATION OF P. FALCIPARUM ........................................... 15

1.3.1 Transfection of P. falciparum ....................................................................................................................... 15

1.3.2 Vector design and Constitutive RNA/protein expression in P. falciparum .......................................................... 15

Table 1.1: Selectable markers used in P. falciparum genetic manipulation ............................................................. 17

1.3.3 Conditional RNA and protein expression in P. falciparum ........................................................................... 18

1.3.4 Genome editing technologies in P. falciparum .............................................................................................. 19

1.4 GENOME EDITING WITH CRISPR TECHNOLOGY .......................................................................................... 20

1.4.1 Functionality of CRISPR in bacterial systems .............................................................................................. 20

Figure 1.2: Mechanism of CRISPR action in bacterial systems ............................................................................. 21

1.4.2 Use of CRISPR for genome editing in mammalian cells .................................................................................. 22

1.5 T7 RNA POLYMERASE ................................................................................................................................... 23

1.5.1 Biology and use of the T7 RNAP in prokaryotes ............................................................................................. 23

Figure 1.3: Lacl dependent T7 RNAP regulation ................................................................................................... 25

1.5.2 Uses in eukaryotic systems .......................................................................................................................... 25

1.6 CONCLUSION ................................................................................................................................................... 26

CHAPTER 2: AN INTEGRATED STRATEGY FOR EFFICIENT VECTOR CONSTRUCTION AND MULTI-GENE EXPRESSION IN PLASMODIUM FALCIPARUM ......................................................................................................................... 28

2.1 NOTE ............................................................................................................................................................... 28

2.2 ABBREVIATIONS USED IN THIS WORK ........................................................................................................ 28
2.3 ABSTRACT..............................................................................................................................................28
2.4 INTRODUCTION ...................................................................................................................................... 29
2.5 RESULTS................................................................................................................................................. 32
2.5.1 Vector family design and features .........................................................................................................32
   Figure 2.1: Schematic of different vector assembly strategies ....................................................................34
   Figure 2.2: Schematic summary of the new family of plasmid vectors ......................................................37
   Table 2.2: List of vectors used in this study .............................................................................................39
2.5.2 Vector construction using various cloning methods..............................................................................40
   Figure 2.3: Assembly of P. falciparum vector using three different cloning strategies ....................... 41
2.5.3 Plasmids in this vector family can be maintained as stable episomes and chromosomally integrated in P. falciparum ........................................................................................................................................42
   Figure 2.4: Behavior and copy number of the pfYC vectors after transfection ......................................43
2.5.4 Expanded transgene expression from a single plasmid that is compatible with proper subcellular trafficking........................................................................................................................................44
   Figure 2.5: Use of the T2A peptide for multi-protein expression from a single expression cassette in P. falciparum 46
   Figure 2.6: 2A mediated subcellular targeting in Plasmodium falciparum .............................................49
2.6 DISCUSSION ........................................................................................................................................... 50
2.7 MATERIALS AND METHODS .................................................................................................................. 51
2.7.1 Molecular Biology................................................................................................................................51
2.7.2 Vector Construction...............................................................................................................................52
   Table 2.3: Primers used in these studies ..................................................................................................52
2.7.3 Yeast Homologous Recombination .....................................................................................................54
2.7.4 Gibson Assembly..................................................................................................................................55
2.7.5 Restriction/ligation cloning..................................................................................................................55
2.7.6 Construction of attP-containing (pfYC3 series) vectors .....................................................................56
2.7.7 Construction of the pfcen5-1.5 mini-centromere containing (pfYC4 series) plasmids .....................56
2.7.8 Cloning Plasmodium falciparum amal and trxR genes into pfYC120 ...............................................56
2.7.9 Multi-cistronic constructs using the T2A for evaluating subcellular trafficking rules .....................57
2.7.10 Parasite culture and transfection ................................................................. 57
2.7.11 Monitoring transfection progress by luciferase expression ....................... 57
2.7.12 Parasite DNA extraction and qPCR analysis ............................................... 58

Table 2.4: Quantitative PCR primers used in these studies .................................... 59
2.7.13 Western Blot .................................................................................................. 59
2.7.14 Northern blot ............................................................................................... 60
2.7.15 Fluorescence microscopy .............................................................................. 61

CHAPTER 3: UTILIZATION OF THE CRISPR/SPCAS9 SYSTEM IN CONJUNCTION WITH
T7 RNAP FOR TARGETED GENOME EDITING IN P. FALCIPARUM ......................... 62

3.1 NOTE ............................................................................................................... 62
3.2 ABSTRACT ........................................................................................................ 62
3.3 INTRODUCTION ............................................................................................... 63

Figure 3.1: Schematic of SpCas9, gRNA, and target DNA interaction .................. 64
3.4 RESULTS .......................................................................................................... 65

3.4.1 T7 RNAP is capable of producing non-coding RNA in situ ......................... 65

Figure 3.2: Vector design to test functionality of T7 RNAP in P. falciparum .............. 66
Figure 3.3: T7 RNAP western blot ......................................................................... 67
Figure 3.4: In vitro transcription assay with crude parasite lysate ......................... 67
Figure 3.5: Analysis of T7 dependent RNA production in P. falciparum .................. 68
Figure 3.6: Expansion rate of P. falciparum expressing T7 RNAP ......................... 69
Figure 3.7: T7 dependent RNA expression across the P. falciparum IDC ............... 70
Figure 3.8: Dual luciferase assay to test for translation of T7 RNAP generate RNA ...... 71
3.4.2 T7 RNAP can be used to produce functional gRNAs .................................. 71

Figure 3.9: gRNA design and in vitro cleavage assay results .................................. 73
3.4.3 CRISPR mediated knockout of kahrp ......................................................... 74

Figure 3.10: Vector design for kahrp knockout ...................................................... 75
Figure 3.11: Luciferase levels of kahrp knockout strain at the population level ......... 76
Figure 3.12: PCR confirmation of kahrp knockout at the population level ............... 77
Figure 3.13: Southern blot confirmation of kahrp knockout at the population level .... 78
Figure 3.14: Western blot confirmation of kahrp knockout at the population level.......................... 79
Figure 3.15: Confirmation of kahrp knockout in WT 3D7 strain at the population level.................. 80
Figure 3.16: SEM confirmation of kahrp knockout........................................................................ 81
Figure 3.17: PCR and western blot confirmation of kahrp knockout clones................................. 82
3.4.4 CRISPR mediated knockout of eba-175 ............................................................................. 82
Figure 3.18: Schematic of eba-175 CRISPR mediated knockout vectors........................................ 83
Figure 3.19: Population level data for the CRISPR mediated knockout of eba-175.................... 84
3.4.5 Assessing off target effects of CRISPR/Cas9 in P. falciparum ........................................ 84
Figure 3.20: Assessing the fate of kahrp-gRNA<sup>T</sup>-induced cleavage of the kahrp locus in the absence of a homologous donor plasmid........................................................................................................... 87
3.5 DISCUSSION .............................................................................................................................. 88
3.6 MATERIAL AND METHODS ..................................................................................................... 88
  3.6.1 Molecular Biology.................................................................................................................. 88
  3.6.2 Parasite culture and transfection .......................................................................................... 89
  3.6.3 Assessing T7 RNAP expression in P. falciparum................................................................. 89
  3.6.4 Assessing T7 RNAP activity in P. falciparum...................................................................... 90
  3.6.5 In vitro SpCas9 cleavage assays ............................................................................................ 91
  3.6.6 CRISPR/Cas9 genome editing in P. falciparum................................................................. 91
  3.6.7 Analysis of CRISPR-edited parasite lines.......................................................................... 92
Table 3.5: Primers used in this study ........................................................................................... 93

CHAPTER 4: INDUCIBLE TRANSCRIPTION OF NON-CODING RNA IN P. FALCIPARUM MEDIATED BY THE T7 RNA POLYMERASE AND THE LAC REPRESSOR SYSTEM ..........95

4.1 ABSTRACT................................................................................................................................. 95
4.2 INTRODUCTION ....................................................................................................................... 95
4.3 RESULTS................................................................................................................................... 97
  4.3.1 IPTG Toxicity in P. falciparum............................................................................................. 97
    Figure 4.1: IPTG toxicity in P. falciparum ............................................................................... 98
  4.3.2 Vector design and transfection ............................................................................................ 98
    Figure 4.2: Vector Design for inducible expression of RLae RNA........................................ 99
4.3.3 RNA induction by the addition of IPTG .................................................................................................................................................. 99

Figure 4.3: Normalized levels of Rluc RNA +/- IPTG .................................................................................................................................................. 100

4.3.4 Sensitivity of the Lac repression system to IPTG ....................................................................................................................................... 100

Figure 4.4: Sensitivity of RNA production to IPTG concentration ...................................................................................................................................... 101

4.3.5 Time scale of RNA production upon induction in P. falciparum ....................................................................................................................................... 101

Figure 4.5: RNA induction time course after the addition of IPTG ...................................................................................................................................... 102

4.4 DISCUSSION ........................................................................................................................................................................................................ 102

Figure 4.6: mFold algorithm RNA folding of LacO .................................................................................................................................................. 103

4.5 MATERIAL AND METHODS .......................................................................................................................................................................................... 104

4.5.1 Molecular Biology ............................................................................................................................................................................................ 104

4.5.2 Strain Maintenance and Transfection .................................................................................................................................................... 104

4.5.3 RNA extraction from P. falciparum ............................................................................................................................................................ 105

4.5.4 Quantitative Real-time PCR ................................................................................................................................................................. 105

4.5.5 P. falciparum Expansion Assays ............................................................................................................................................................ 105

4.5.6 In silico folding of RNA .................................................................................................................................................................................. 106

CHAPTER 5: CONCLUSIONS AND FUTURE DIRECTIONS ........................................................................................................................................ 107

5.1 SUMMARY OF THIS WORK ...................................................................................................................................................................................... 107

5.2 FUTURE DIRECTIONS IN THE CONSTRUCTION OF GENETIC ELEMENTS FOR USE IN P. FALCIPARUM ...................................................................................................................................................................................... 107

5.3 FUTURE APPLICATION SPACE FOR THE P. FALCIPARUM CRISPR/SfCAS9 GENOME EDITING TECHNOLOGY ...................................................................................................................................................................................... 109

5.4 FUTURE APPLICATION SPACE FOR NON-CODING RNA REGULATION IN P. FALCIPARUM ...................................................................................................................................................................................... 110

5.5 CONCLUSION .......................................................................................................................................................................................................... 111

BIBLIOGRAPHY ........................................................................................................................................................................................................ 112
Chapter 1: Introduction

1.1 The disease burden of malaria

Malaria remains a major public health concern to this day with an estimated ~600,000 deaths per year, mostly children under the age of five\(^1\). The disease in humans is caused by five species of the genus *Plasmodium* (*P. ovale, P. Vivax, P. Malariae, P. knowlesi*, and *P. falciparum*). Of these species, *P. falciparum* is responsible for the majority of morbidity and mortality\(^1\). Aside from the mortality associated with the disease, the economic burden of malaria has been estimated to be well into the billions of dollars per year\(^2\). There are several antimalarial compounds on the market but many are becoming less effective due to resistant strains appearing in the population\(^3\). There is currently no effective vaccine approved for use.

Apart from pharmacological approaches to malaria control, there has been significant effort to control the spread of malaria by controlling the mosquito vector. Malaria is spread by female mosquitoes of the genus *Anopheles*\(^4\). Attempts at insecticide control, bed netting, and mosquito proofing houses have all met with some success\(^5,6\). However, as mosquito resistance to common insecticides builds up, combined with the aforementioned increased prevalence of drug resistant malaria, the need for new treatment options becomes apparent\(^7\).

1.2 The Life cycle and Genome of *P. falciparum*

The genome of *P. falciparum* was sequenced in 2002\(^8\). The most striking feature of the genome was the observed A+T richness of ~80% with 90% A+T richness in the
non-coding regions. This makes *P. falciparum* the most A+T rich organism sequenced to date\(^8\). The A+T richness presents some technical challenges in terms of polymerase chain reaction (PCR), molecular cloning, and molecular sequencing.

In addition to an atypical genome, the parasite *P. falciparum* undergoes multiple stages of development both in the human and mosquito host (Figure 1.1A)\(^9\). First a mosquito bites an infected person, in so taking up several parasites that have differentiated into male and female gametocytes. The male and female gametocytes then fuse in the mosquito midgut, and traverse the gut wall, eventually migrating to the salivary glands of the mosquito. When the mosquito feeds again, the sporozoites are injected into the human host and travel to the liver where they invade hepatocytes. After a gestation period, the infected hepatocytes rupture releasing the *P. falciparum* merozoites into the bloods stream where they invade erythrocytes and begin the intraerythrocytic development cycle (IDC), which is responsible for the majority of symptoms associated with malaria. This cycle (Figure 1.1B) takes approximately 48 hours after the initial invasion of the host cell to develop new *P. falciparum* merozoites that are able to burst out of the infected red blood cell and invade new blood cells to restart the cycle. During the IDC, parasites are only out side of the blood cell on the order of 1 minute before reinvasion\(^10\). Some parasites also differentiate into male and female gametocytes irreversibly and are taken up by the mosquito to begin the mosquito part of the lifecycle again.
Micro array studies have revealed an interesting pattern of genetic regulation. It appears that stage specific genes only are active right before they are needed throughout the parasite life cycle, forming a wave function of gene regulation that is most likely generated by some unknown master regulators. In addition to the biology carried out by proteins in the parasite, there is increasing interest in the processes influenced by non-coding (ncRNA). While largely not understood, ncRNA has been implicated in such diverse processes as chromosome maintenance, genetic regulation, and surface displayed antigen differentiation.

Despite the clinical relevance of this disease, little is known about the genetic regulation schemes or biological processes that the parasite undergoes in order to complete its lifecycle. In order to elucidate essential genes and pathways that could
inform new drug and vaccination strategies, molecular tools capable of manipulating the parasite during the IDC are needed.

1.3 Current Molecular Tools for the manipulation of *P. falciparum*

1.3.1 Transfection of *P. falciparum*

Transfection of *P. falciparum* is generally accomplished by one of two electroporation methods. The first involves the direct electroporation of early “ring stage” parasites. While functional, the DNA must pass through 4 different membranes in order to enter the nucleus (the red blood cell membrane, the parasitophorous vacuole membrane, the membrane of *P. falciparum* itself, and the nuclear envelope.

The second method, termed “spontaneous uptake”, involves electroporating uninfected red blood cells to preload them with the target DNA. These preloaded cells are then exposed to late stage “schizont” parasites that reinvade and take up the DNA, generally at a higher efficiency than direct electroporation of the parasite. In both cases, the underlying mechanisms that allow the target DNA to reach the nucleus have not been elucidated.

It is also worth noting that, unlike the often studied rodent malarial parasite *Plasmodium berghei, P. falciparum* is not tolerant of the uptake of linear DNA and RNA. This may be due to increase nuclease activity present in *P. falciparum* but remains unknown.

1.3.2 Vector design and Constitutive RNA/protein expression in *P. falciparum*
While gene prediction analyses have identified ~5500 genes in the _P. falciparum_ genome, the regulatory elements governing gene expression remain poorly understood. There have been some efforts to identify the minimal 5’ and 3’ untranslated regions (UTRs) but the vectors for constitutive RNA and protein expression contain large amounts of highly A+T rich DNA. These large regions lead to frequent genetic deletions and rearrangements when propagated in an _E. coli_ host.

Using these expression cassettes, four orthogonal selectable markers have been developed: neomycin phosphotransferase II (nptII) which conveys resistance to neomycin analogs, blasticidin S-deaminase (BSD) which conveys resistance to blasticidin, human dihydrofolate reductase (hDHFR) which conveys resistance to the compound 1,6-Dihydro-6,6-dimethyl-1-[3-(2,4,5-trichlorophenoxy)propoxy]-1,3,5-triazine-2,4-diamine, more simply known as WR99210, and most recently yeast dihydroorotate dehydrogenase (yDHODH) which conveys resistance to the compound 5-methyl[1,2,4]triazolo[1,5-a]pyrimidin-7-yl)naphthalen-2-ylamine which is more simply known as DSM-1. These selectable markers are summarized in Table 1.1.

Most plasmid designs contain two _P. falciparum_ expression cassettes, one to express the selectable marker, and one to express a single protein of interest. Expression of multiple proteins from a single cassette in _P. falciparum_ was shown to be possible by using the viral 2A like tag from the _Thosea asigna_ virus (T2A) to prevent peptide bond formation between two linked open reading frames.
<table>
<thead>
<tr>
<th>Selectable Marker Short Name</th>
<th>Neomycin phosphotransferase II</th>
<th>Blasticidin S deaminase</th>
<th>Human dihydrofolate reductase</th>
<th>Yeast dihydroorotate dehydrogenase</th>
</tr>
</thead>
<tbody>
<tr>
<td>Selectable marker Full Name</td>
<td>Neomycin</td>
<td>Blasticidin</td>
<td>Human</td>
<td>Yeast</td>
</tr>
<tr>
<td>Chemical selection agent short name</td>
<td>G418</td>
<td>blasticidin</td>
<td>WR99210</td>
<td>DSM-1</td>
</tr>
<tr>
<td>Chemical full name</td>
<td>$O$-$2$-$Amino$-$2$,7$-$dideoxy$-$D$-$glycero$-$\alpha$-$D$-$gluco$-$heptopyranosyl$-$($1\rightarrow4$)$O$-$3$-$desoxy$-$4$-$C$-$methyl$-$3$-$ (methylamino)$-$\beta$-$D$-$arabinopyranosyl$-$($1\rightarrow6$)$D$-$streptamin</td>
<td>$4$-$amino$-$1$-$[(4$-$((3S)$-$3$-$amino$-$5$-$<a href="methyl">(aminomino)methyl</a> amino)pentanoyl]amino)$-$2$,3$,4$-$trideoxy$-$\beta$-$D$-$erythro$-$hex$-$2$-$enopyranuronosyl]pyrimidin$-$2$-(1H)$-$one</td>
<td>$1$,6$-$Dihydro$-$6$,6$-$dimethyl$-$1$-$[3$-$$(2$,4$,5$-$trichlorophenox y)propoxy]$-$1$,3$,5$-$triazine$-$2$,4$-$diamine</td>
<td>5$-$methyl$[$1$,2$,4$]$triazolo[1$,5$$\alpha$]$pyrimidin$-$7$-$yl$]naphthalen$-$2$-$ylamine</td>
</tr>
<tr>
<td>Chemical Target</td>
<td>P. falciparum ribosome</td>
<td>P. falciparum ribosome</td>
<td>P. falciparum dihydrofolate reductase</td>
<td>P. falciparum dihydroorotate dehydrogenase</td>
</tr>
<tr>
<td>Chemical Structure</td>
<td><img src="image1.png" alt="Structure" /></td>
<td><img src="image2.png" alt="Structure" /></td>
<td><img src="image3.png" alt="Structure" /></td>
<td><img src="image4.png" alt="Structure" /></td>
</tr>
<tr>
<td>Selection concentration used in this work</td>
<td>500μM</td>
<td>6μM</td>
<td>2.5nM</td>
<td>1.5μM</td>
</tr>
<tr>
<td>Reference</td>
<td>Mamoun et al.25</td>
<td>Mamoun et al.25</td>
<td>Fidock et al.26</td>
<td>Ganesan et al.27</td>
</tr>
</tbody>
</table>
The processes by which non-coding RNA is made in the cell are poorly understood. Despite finding evidence of many small nuclear RNAs and tRNAs, there is little work dealing with the initiation of transcription of ncRNAs and, therefore, no readily available vectors for the expression of ncRNA.

1.3.3 Conditional RNA and protein expression in *P. falciparum*

Conditional expression of RNA and proteins has long presented problems in *P. falciparum*. The lack of any known endogenous inducible systems and the dissimilarity to other organisms at the genetic level suggests the use of orthogonal parts may be beneficial in order to achieve regulated control of gene expression. There are several systems that have attempted to achieve this with mixed results.

One such system which conditionally controls transcription uses the Tet operator fused to an activator domain from either *Toxoplasma gondii* or *P. falciparum* ApiAP2 transcriptional regulator protein in order to achieve decrease activation by introducing tetracycline to the system. While shown to have some utility, it does not appear to be functional in regulating a wide variety of proteins.

The current state of the art is the use of an inducible stabilization domain that is fused to and degrades the protein in the absence of a small molecule ligand. While useful in some situations, the FKBP domain is large and thus may interfere with protein folding and function. Likewise, as it must be fused at the N- or C-terminus of the protein, it may interfere with protein localization. Finally, as it depends upon cytoplasmic machinery for degradation of the protein, it is not functional when directed to a compartment that cannot be accessed by the degradation machinery.
Recent work in our lab has demonstrated the use of RNA aptamers that bind to the tetracycline repressor and allow for post-transcriptional regulation of a produced RNA\textsuperscript{34}. This system can be successfully reconstituted in \textit{P. falciparum} and appears to be useful in regulation a wide array of genes at the translational level\textsuperscript{35}.

There is no widely usable system for the regulated production of RNA (both coding and non-coding) in \textit{P. falciparum}.

1.3.4 \textit{Genome editing technologies in P. falciparum}

As additional techniques become available for regulated expression of RNA and protein in \textit{P. falciparum} robust and flexible methods for the directed engineering of the genome become essential. As plasmid copy number can be highly variable in the parasite\textsuperscript{25,36}, integrating cargo of interest onto the genome allows for a more uniform gene dosage. Integrated strains can also be designed to preserve the native regulatory machinery and therefore, the natural gene dosage and temporal regulation of the protein of interest.

General methods of transfection and drug cycling to generate chromosomal integrants are effective but time consuming\textsuperscript{37}. Generation of integrants by this technique can take on the order of 6 months to a year and seems highly dependent upon site selection. This timing can be improved by the addition of negative selection markers to the plasmid\textsuperscript{38} but still remains a highly variable process.

A commonly used system for adding cargo onto the parasite genome in a site specific manner is the Bxb1 integrase system which expresses a site specific recombinase and allows insertion at a complementary site on the genome\textsuperscript{39}. The advantage of this
system is that it is simple to use and highly efficient. The disadvantage is that the site
(which does not occur naturally in *P. falciparum*) must be added onto the genome by
another method in order to generate a strain capable of accepting a donor plasmid. It
therefore relies heavily on pre-made strains and suffers in terms of flexibility of insertion
sites in the genome.

More recently, it has been shown that zinc-finger endonucleases (ZFNs) can be
engineered to create double strand breaks at a specific point in the genome\(^29\). While this
solves the issue of flexibility in terms of loci in the genome to be edited, creation of the
ZFNs is both time consuming and expensive. In addition, the predictive models of ZFN
action still require *in situ* verification to be used with any confidence.

Recently, a new genome editing system based on the bacterial Clustered
Regularly Short Interspaced Palindromic Repeats (CRISPR) system has been shown to
have utility for simple, directed genome editing in a variety of eukaryotic organisms\(^{40-42}\).

1.4 Genome editing with CRISPR Technology

1.4.1 Functionality of CRISPR in bacterial systems

The Clustered Regularly Short Interspaced Palindromic Repeats (CRISPR)
system acts as a primitive bacterial immune system to defend against infection by
bacteriophage. Initial computational studies identified arrays of short palindromic repeats
of 20-50 bases separated by “spacers” which were later identified to have high sequence
similarity to genes known to be prevalent in bacteriophages (Figure 1.2a)\(^{43,44}\). Additional
studies revealed several genes associated with these CRISPR arrays which were termed
CRISPR associated genes or *cas* genes\(^{45}\). Several additional studies have since elucidated some of the mechanism of action.

![Diagram of a typical CRISPR locus in the bacteriophage.](image)

**Figure 1.2: Mechanism of CRISPR action in bacterial systems**

This figure is adapted from Marraffini\(^{46}\)

**Part A:** Diagram of a typical CRISPR locus in the bacteriophage. The *cas* genes are indicated by arrows, repeats by grey boxes, and spacers by different colored lines. The leader contains the promoter region to produce an RNA containing several repeats and spacers before processing.

**Part B:** Adaptation phase of CRISPR action where new spacers are acquired.

**Part C:** crRNA biogenesis where short RNAs are processed by Cas enzymes.

**Part D:** Targeting on an invading bacteriophage with homology to the acquired spacers.

The basic principle of the CRISPR system is summarized in Figure 1.2b-d. The CRISPR response can be divided into three steps. In the first, invading DNA is recognized by the CRISPR associated genes that incorporate them into the CRISPR array at the DNA level. Next, The DNA is transcribed and processed to generate targeting RNAs. Finally, the targeting RNAs interact with more *cas* genes to target the invading
DNA from the same or similar bacteriophage and cleave it by creating a double strand break (DSB)\textsuperscript{47}.

There are considered to be three types of CRISPR systems, each with their own mechanisms of action. Types I and III involve several Cas proteins with specific functions in order to carry out the actions necessary to complete the CRISPR process, while type II systems rely primarily on a large, multifunctional protein known as Cas9, with additional modular RNAs to carry out the processes\textsuperscript{47}. The most well studied member of the Cas9 gene family is the Cas9\textsuperscript{9} from the bacterium \textit{Streptococcus pyogenes} (SpCas9). As type II systems involve fewer proteins and more modular parts, they were identified for potential use in genome editing in higher organisms.

\textbf{1.4.2 Use of CRISPR for genome editing in mammalian cells}

Once the mechanism for CRISPR based action in bacteria was established, interest turned towards adapting the system for use in editing the genomes of higher organisms. Initial studies indicated three potential necessary components for targeting genomic DNA with a type II CRISPR system in mammalian cells: The nuclease (Cas9), the cleaved CRISPR RNA containing two repeats flanking a spacer (crRNA), and an adapter RNA capable of binding a crRNA and the Cas9 (tracerRNA)\textsuperscript{40}. Later studies revealed that the crRNA and the tracerRNA could be fused into a single guide RNA (gRNA) capable of directing Cas9 to the locus of interest and allowing it to create a DSB\textsuperscript{40,42,48}. In mammalian cells, this resulted in activation of the non-homologous end joining pathway (NHEJ) which generated insertions or deletions (indels) at the cut site.
In a system where it is desirable for the RNA to be produced endogenously, the organism of interest’s RNA pol III U6 promoter typically generates the gRNA. In situations where the RNA can simply be injected, the gRNA is usually produced with a T7 RNAP based in vitro reaction. In order for this system to work in *P. falciparum*, a new method of endogenous gRNA production is necessary, as the parasite RNA pol III promoters remain undefined.

At this point, the basics CRISPR system producing a Cas9 protein and a gRNA has been shown to be functional in mammalian cells, zebrafish, mice, and rats. Use of this technology for genome editing in *P. falciparum* is demonstrated in Chapter 3.

1.5 T7 RNA Polymerase

1.5.1 Biology and use of the T7 RNAP in prokaryotes

T7 bacteriophage was first identified in 1945 as a lytic bacteriophage of *E. coli*. The T7 RNA polymerase (T7 RNAP) was identified in 1970 and it was realized that it had a high specificity and high efficiency for its naturally occurring promoter, found at several points on the bacteriophage. In 1986, the polymerase was used as a system to direct the expression of cloned genes. This demonstrated the ability to produce a protein with the simple two-component system of the T7 RNAP and the T7 promoter. The mechanism of action is simply a binding event where the T7 RNAP recognizes the T7 promoter and begins transcription with high efficiency. Important residues for the action have been mapped and the entire structure has been crystallized. Additional studies into the T7 promoter have revealed essential residues and established the ability to modulate the expression levels of RNA from the promoter through mutation.
T7 RNAP has since become the most commonly used tool for \textit{in vitro} transcription and \textit{E. coli} based protein expression owing to its high specificity to its promoter, its requirement of no host factors in order to create transcripts, and its high efficiency.

In addition to the uses of the polymerase for constitutive expression, the T7 RNAP lends itself well to regulated expression of RNA and proteins in bacteria. Initial studies utilized the P\textsubscript{L} promoter from the lambda bacteriophage to produce the polymerase in a temperature dependent manner\textsuperscript{69}. This technique was supplanted by the now most common method of regulation, which relies on the well-studied lac repression system. By this system, the lac operator (lacO) is placed directly upstream of the T7 promoter allow the lac repressor (LacI) to bind to the DNA in the absence of the inducer molecule Isopropyl β-D-1-thiogalactopyranoside (IPTG)\textsuperscript{61}. Upon addition of IPTG, LacI binds the molecule and can no longer bind to lacO, allowing T7 RNAP to initiate transcription. LacI dependent regulation is summarized in Figure 1.3.
Figure 1.3: LacI dependent T7 RNAP regulation

T7 RNAP (blue) is unable to access the T7 promoter when LacI (red) is bound to it in the absence of IPTG. Upon the addition of IPTG, LacI binds and undergoes a conformational shift which renders it unable to bind the lac operator, allowing T7 RNAP to access the promoter and transcribe RNA.

1.5.2 Uses in eukaryotic systems

The ease of use in bacterial systems has generated interest in using T7 RNAP for RNA and protein production in eukaryotic organisms. There are two principal challenges associated with adapting T7 RNAP to use in eukaryotic systems. The first is nuclear localization as prokaryotes have no nucleus and can access the DNA without the specific localization generally required in eukaryotes. The second is the inability to provide the 5'-7-methylguanosine cap which confers stability to the RNA, as well as allowing the assembly of the translational machinery on the RNA in order to produce a protein\(^ {62,63}\). The addition of the cap is generally coupled to the initiation of transcription by the eukaryotic RNA polymerase II, making it's addition by T7 RNAP complicated\(^ {64}\).  

25
With these challenges in mind, the T7 RNAP has been adapted for both RNA and protein production in multiple eukaryotes. In *Saccharomyces cerevisiae*, studies demonstrated that the SV40 nuclear localization signal was sufficient to target the T7 RNAP to the nucleus and to increase the amount of T7 directed transcription\(^6\). Further studies demonstrated the RNA produced from T7 RNAP could be polyadenylated, spliced, and exported from the nucleus, but was unable to produce a protein\(^{66}\). Additional work demonstrated that T7 RNAP could be used to produce RNA in mammalian cells, and also protein through the addition of an internal ribosome entry site (IRES) which is an RNA element typically used in viruses to bypass the need for the 5'-7'-methylguanosine cap\(^6\). An additional example of a eukaryotic system that is able to employ T7 RNAP to produce protein comes from the organism *Trypanosoma brucei*\(^{68}\). Unlike most eukaryotes, the capping mechanism of *T. brucei* is not coupled to the initiation of transcription and so the T7 RNAP generated transcripts become capped and are thus recognized by the usual translation machinery\(^{69}\). The capping machinery in *P. falciparum* is known to be similar to that in *S. cerevisiae* making it unlikely that transcription-independent capping occurs\(^{70}\).

### 1.6 Conclusion

The current state of molecular tools in *P. falciparum* remains in need of development. This work aims to address the lack of tools for the controlled production of ncRNA in the parasite. First I present the construction of the first complete vector family in *P. falciparum* which can be used with multiple modern cloning strategies and is modular to allow for easy adaptation to new purposes. In addition I explore the use of
viral 2A like peptides to allow for production and localization of multiple proteins from a single parasite expression cassette.

Next I present the first demonstration of the use T7 RNAP for the production of ncRNA in the parasite and then use this to give the first demonstration of using a CRISPR system for site directed mutagenesis of the parasite genome. To my knowledge, this represents the first use of CRISPR for genome editing in any eukaryotic pathogen.

Finally, I demonstrate the use of the well-defined lac repression system to modulate the production of RNA in a T7 RNAP dependent manner in *P. falciparum*, opening the door to future studies of ncRNA in the parasite.

Overall, this work establishes several novel molecular tools for use in *P. falciparum* and should act as the foundation for future research in several areas.
Chapter 2: An Integrated Strategy for Efficient Vector Construction and Multi-Gene Expression in Plasmodium falciparum

2.1 Note

This chapter is comprised of text and figures now published in Malaria Journal\textsuperscript{36}

2.2 Abbreviations used in this work

nptII: Neomycin phosphotransferase II; bsd: Blasticidin S deaminase; hdhfr: Human dihydrofolate reductase; ydhodh: Yeast dihydroorotate dehydrogenase

2.3 Abstract

Construction of plasmid vectors for transgene expression in the malaria parasite \textit{Plasmodium falciparum} presents major technical hurdles. Traditional molecular cloning by restriction and ligation often yields deletions and re-arrangements when assembling low-complexity, A+T rich parasite DNA. Furthermore, the use of large 5’- and 3’-untranslated regions of DNA sequence (UTRs) to drive transgene transcription limits the number of expression cassettes that can be incorporated into plasmid vectors.

To address these challenges, two high fidelity cloning strategies, namely yeast homologous recombination and the Gibson assembly method, were evaluated for constructing \textit{P. falciparum} vectors. Additionally, some general rules for reliably using the viral 2A-like peptide to express multiple proteins from a single expression cassette while preserving their proper trafficking to various subcellular compartments were assessed.
Yeast homologous recombination and Gibson assembly were found to be effective strategies for successfully constructing *P. falciparum* plasmid vectors. Using these cloning methods, a validated family of expression vectors that provide a flexible starting point for user-specific applications was created. These are also compatible with traditional cloning by restriction and ligation, and contain useful combinations of all the commonly used utility features for enhancing plasmid segregation and site-specific integration in *P. falciparum*. Additionally, application of a 2A-like peptide for the synthesis of multiple proteins from a single expression cassette, together with some rules for combinatorially directing proteins to discrete subcellular compartments were established.

A set of freely available, sequence-verified and functionally validated parts that offer greater flexibility for constructing *P. falciparum* vectors having expanded expression capacity is provided.

2.4 Introduction

Malaria continues to be a leading cause of morbidity and mortality worldwide. Nearly 50% of the global population is at risk, and in 2010 there were an estimated 219 million cases and 660,000 deaths\(^1\). *Plasmodium falciparum* is the parasite pathogen responsible for the most virulent disease. No vaccine is clinically approved to prevent malaria. Treatment relies heavily on the use of a limited number of anti-malarial drugs to which resistance is increasingly widespread\(^7\), which makes it critical to identify new and
effective drugs. Using genetic approaches to validate potential drug targets in *P. falciparum* is pivotal to this effort. However, the process of constructing the plasmid vectors needed for these studies is time-consuming and inefficient, and imposes a significant barrier to genetically manipulating the parasite.

Several aspects of parasite biology interact to create this challenge. First, the parasite’s genome is extremely A+T rich (80-90%)\(^8\), and extended regions of low complexity sequence are common\(^{71,72}\). Second, regulatory regions upstream (5’untranslated regions of DNA sequence (UTR)) and downstream (3’UTR) of coding sequences are poorly defined in *P. falciparum*, and large regions of putative regulatory DNA are needed to facilitate robust transgene expression\(^{23}\). Very few 5’ and 3’ UTRs have been precisely mapped. As a result, 1-2 kb 5’ and 3’ UTRs are frequently selected on the assumption that these comprise the information necessary to support efficient transcription\(^{21,22,25}\). These long UTRs are close to 90% in A+T composition. Third, the mean coding sequence (CDS) length in *P. falciparum* (excluding introns) is 2.3 kb, nearly twice that of many model organisms\(^8\).

Gene complementation is a powerful strategy, used extensively in forward genetics studies in other organisms, but this approach is under-utilized in *P. falciparum* due in large part to the challenges associated with efficiently assembling the necessary complementing constructs\(^{73}\). The ability to more routinely construct expression vectors for complementation studies is highly synergistic with the increasing rate at which genome-wide insertional mutagenesis studies are identifying candidate genes associated with growth, cell cycle and other phenotypic defects in *P. falciparum*\(^{74,75}\). In constructing
over-expression, complementation and gene targeting vectors in *P. falciparum*, long A+T rich regions must be cloned into final plasmids that can exceed 10 kb. It is recognized that the traditional and commonly used restriction/ligation-based cloning method is inefficient for assembling these vectors, and often yields plasmids with regions that are deleted and/or re-arranged\textsuperscript{24}. Consequently, time-consuming screening of large numbers of bacterial clones is needed to increase the probability of recovering the intact target vector, if it is at all present.

In addition to the vector assembly challenges, typical over-expression vectors are limited in the number of transgenes that can be simultaneously expressed. In the most common format, two expression cassettes are available, and one of these is dedicated to expressing a selectable marker\textsuperscript{25}. Increasing the expression capacity of a single plasmid can be accomplished by introducing additional 5'UTR-CDS-3'UTR cassettes, but this further complicates vector construction for reasons described above. This problem has been circumvented in several eukaryotes through the use of a viral 2A-like peptide that prevents peptide bond formation between two specific and adjacent amino acids during translation and results in the production of two separate proteins from a single expression cassette\textsuperscript{28}. Recently, the a 2A-like peptide has been shown to be functional in *P. falciparum*\textsuperscript{29}, but its broader utility with respect to proteins that are trafficked to different subcellular parasite compartments has not been examined.

Here, an inexpensive and straightforward strategy for more robustly and flexibly assembling *P. falciparum* vectors is introduced, while simultaneously maximizing the amount of transgenic information expressible from a single plasmid without using
additional 5'UTR-CDS-3'UTR expression cassettes. This has been achieved by developing a family of vectors that integrate use of high fidelity and robust DNA assembly by yeast homologous recombination and \textit{in vitro} assembly by the isothermal chew-back-anneal Gibson method with traditional restriction/ligation-based cloning.

Additionally, several desirable utility features have been consolidated in this vector family, including: site-specific integration mediated by the \textit{Bxb1} integrase; improved plasmid segregation mediated by either Rep20 elements or a \textit{P. falciparum} mini-centromere (pfcen5-1.5); and all of the currently used \textit{P. falciparum} selection markers.

Lastly, the broader utility of using a viral 2A-like peptide to achieve expression from a single cassette of multiple genes targeted to distinct parasite subcellular compartments has been demonstrated. This resource is freely available through the Malaria Research and Reference Reagent Resource Center (MR4: www.mr4.org).

2.5 Results

2.5.1 Vector family design and features

In creating this plasmid vector resource several useful design criteria have been incorporated, namely: (1) access to multiple, orthogonal and high-fidelity strategies for cloning a target fragment into the identical context; (2) pre-installed utility features including access to all commonly used \textit{P. falciparum} selection markers (bsd1, hdhfr, ydhodh and nptII), plasmid integration sequences (attP sites), and plasmid segregation/maintenance features such as Rep20 and the mini-centromere, pfcen5-1.5; (3) sufficient modularity to permit straightforward tailoring for user-specific needs; and,
(4) ease of manipulation using reagents that are readily prepared in-house or commercially available at low cost.

This vector family framework includes access to yeast homologous recombination (HR)\textsuperscript{76}, Gibson assembly\textsuperscript{77} and restriction/ligation as central cloning strategies (Figure 2.1). The challenges associated with the traditionally used restriction/ligation method when cloning P. falciparum sequences have been described\textsuperscript{24}. It is thought that the observed genomic deletions and re-arrangements are related to the long A+T rich regions in combination with the restriction and ligation process and the instability of these constructs in E. coli. Though inefficient overall, this strategy is used successfully, and so it is preserved as an option that interfaces directly with the more efficient yeast HR and Gibson strategies.
Figure 2.1: Schematic of different vector assembly strategies
Homology-based (yeast HR and Gibson assembly) and traditional restriction/ligation cloning strategies selected are shown as part of an integrated framework for the orthogonal assembly of Plasmodium falciparum constructs. Beginning with a common primer set, PCR products and the desired vector backbone (see Figure 2.2 for details), the identical target plasmid can be assembled using any of these approaches individually or in parallel.

A major advantage of both Gibson and yeast HR strategies over traditional restriction ligation based cloning is that they do not require enzymatic digestion of the inserted fragment, which can impose constraints on cloning target DNA that contains these restriction sites internally. Rather, as they depend on homologous ends overlapping with a digested vector, the insert does not need to be digested. This allows greater flexibility by permitting a larger set of restriction sites on the vector to be used. Yeast HR
requires more overall time compared to Gibson and restriction/ligation cloning, as S.
cerevisiae grows more slowly than E. coli. However, this strategy efficiently yields target
constructs and all the key components can be inexpensively generated in-house\textsuperscript{80}. The
Gateway\textsuperscript{®} strategy (Life Technologies) has also been used to construct P. falciparum
vectors\textsuperscript{81}. This approach has not been included in the current study, as it is significantly
more expensive than the methods described here. However, when needed, the features
required for enabling Gateway\textsuperscript{®} cloning should be straightforward to introduce into the
framework described below.

The overall architecture of this new vector family and the built-in utility features
are summarized in Figure 2.2, and is derived from the pfGNr plasmid previously
deposited as MRA-462 in MR4. This plasmid contains bacterial (pMB1) and yeast
(CEN6/ARS4) origins of replication, and the kanMX4 gene under the control of a hybrid
bacterial/yeast promoter to facilitate selection of bacterial or yeast colonies on kanamycin
or G-418, respectively. This plasmid contains two P. falciparum gene expression
cassettes consisting of the commonly used 5'/3'UTR pairs PfCaM/Pfhsp86 and
PcDT/PfHRPII arranged head-to-head to improve transcriptional efficiency\textsuperscript{82}. In P.
falciparum, plasmid selection using G-418 is enabled by a gfp-nptII gene fusion
expressed from the PcDT/PfHRPII cassette, and a 2×Rep2O element to enhance plasmid
segregation during replication is also present\textsuperscript{78}.

From this vector, a library of eight base plasmids was first created in which each
of the four frequently used \textit{P. falciparum} selection markers was cloned into one of the
two \textit{P. falciparum} expression cassettes. For ease of reference, a nomenclature to describe
the various vector family members was defined. Plasmids are designated as pfYCxA\textsubscript{B},

35
where \( x \) is a series number indicating the presence of a specific set of utility features (1 = Rep20/yCEN, 2 = Rep20/yCEN/2\( \times \)att\( P \), 3 = 2\( \times \)att\( P \) and 4 = pfcen5-1.5) and \( A \) and \( B \) denote the resistance marker expressed from the PfCaM/PfHsp86 (cassette A) and PcDT/PfHRPII (cassette B) UTR pairs, respectively (0 = no marker; 1 = nptII; 2 = bsd; 3 = hdhfr; and 4 = ydhodh). Introducing the 2\( \times \)att\( P \) site, which facilitates site-specific integration mediated by the Bxb1 integrase into compatible att\( B \) strains\(^{39} \), at the SalI site yields the pfYC2 plasmid series. Two representative members, namely pfYC220:FL and pfYC240:FL (Table 2.1), were generated in this study and provide a standardized approach for easily generating the entire set. Both the pfYC1 and pfYC2 plasmids facilitate manipulation through yeast homologous recombination, Gibson assembly and traditional restriction/ligation cloning to provide the greatest flexibility in assembling a specific construct.
Figure 2.2: Schematic summary of the new family of plasmid vectors
Plasmids are designated by the pfYC prefix, a series number (1-4) and a number (0-4) defining the resistance marker present in expression cassette A (5′PfCaM/3′PfHsp86 UTRs) or B (5′-PcDT/3′PfHRPII UTRs). The series number is defined by specific utility features included in the plasmid as follows: 1 = yeast CEN/ARS origin to enable plasmid maintenance in *S. cerevisiae* during yeast HR and a 2 × Rep20 element to improve plasmid segregation in *P. falciparum* [19]; 2 = same as in 1, but with a 2 × attP element added to enable site-specific chromosomal integration into existing attB+ strains [18]; 3 = 2 × attP element is present, but the yeast CEN/ARS origin and 2 × Rep20 elements have been eliminated; and 4 = the pfcen5-1.5 mini-centromere element is included to facilitate plasmid segregation and maintenance at single copy in *P. falciparum* [20], while the yeast origin, 2 × Rep20 and 2 × attP elements have been eliminated. *P. falciparum* resistance markers are designated as: 0 = none; 1 = nptII (G-418 resistance); 2 = bsd (Blasticidin S resistance); 3 = hdhfr (WR99210 resistance); and 4 = ydhodh (DSM-1 resistance). A non-resistance gene cloned into the available expression cassette is indicated by a colon followed by the gene name (e.g., pfYC110:FL indicates that the nptII and firefly luciferase genes are present in expression cassettes A and B, respectively). Three HindIII sites present on the base plasmid are noted, as they are useful for topologically mapping these vectors and derivatives to screen for potential rearrangements and large insertions or deletions.
A limited set of pfYC3 (pfYC320:FL and pfYC340:FL) plasmids have also been generated, and these retain the attP site but not the Rep20 and CEN6/ARS4 elements from the pfYC2 plasmid series. Elimination of the Rep20 and CEN6/ARS4 elements from plasmids intended for integration into the P. falciparum genome may be desirable, as the Rep20 element has the potential to induce transcriptional silencing in a subtelomeric chromosomal context. Likewise, the S. cerevisiae-derived CEN6/ARS4 element could possibly behave aberrantly when integrated into a P. falciparum chromosome. A limited set of pfYC4 plasmids (pfYC402:FL and pfYC404:FL) has been made in which the Rep20 and CEN6/ARS4 elements in the pfYC1 series have been replaced by the mini-centromere pfcen5-1.5. The option to use yeast homologous recombination in the pfYC3 and pfYC4 series is eliminated. However, Gibson assembly and/or traditional restriction/ligation can be used to generate final constructs that are immediately ready for integration. Validated procedures for generating the complete set as dictated by user needs have also been provided.
<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Selection</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>pfYC101</td>
<td>G-418</td>
<td>Neo resistance under pCDT</td>
</tr>
<tr>
<td>pfYC102</td>
<td>Blasticidin S</td>
<td>BSD resistance under pCDT</td>
</tr>
<tr>
<td>pfYC103</td>
<td>WR99210</td>
<td>hDHFR resistance under pCDT</td>
</tr>
<tr>
<td>pfYC104</td>
<td>DSM-1</td>
<td>yDHODH resistance under pCDT</td>
</tr>
<tr>
<td>pfYC110</td>
<td>G-418</td>
<td>Neo resistance under pfCAM</td>
</tr>
<tr>
<td>pfYC120</td>
<td>Blasticidin S</td>
<td>BSD resistance under pfCAM</td>
</tr>
<tr>
<td>pfYC130</td>
<td>WR99210</td>
<td>hDHFR resistance under pfCAM</td>
</tr>
<tr>
<td>pfYC140</td>
<td>DSM-1</td>
<td>yDHODH resistance under pfCAM</td>
</tr>
<tr>
<td>pfYC101:Fluc</td>
<td>G-418</td>
<td>Neo resistance under pCDT, Fluc</td>
</tr>
<tr>
<td>pfYC102:Fluc</td>
<td>Blasticidin S</td>
<td>BSD resistance under pCDT, Fluc</td>
</tr>
<tr>
<td>pfYC103:Fluc</td>
<td>WR99210</td>
<td>hDHFR resistance under pCDT, Fluc</td>
</tr>
<tr>
<td>pfYC104:Fluc</td>
<td>DSM-1</td>
<td>yDHODH resistance under pCDT, Fluc</td>
</tr>
<tr>
<td>pfYC110:Fluc</td>
<td>G-418</td>
<td>Neo resistance under pfCAM, Fluc</td>
</tr>
<tr>
<td>pfYC120:Fluc</td>
<td>Blasticidin S</td>
<td>BSD resistance under pfCAM, Fluc</td>
</tr>
<tr>
<td>pfYC130:Fluc</td>
<td>WR99210</td>
<td>hDHFR resistance under pfCAM, Fluc</td>
</tr>
<tr>
<td>pfYC140:Fluc</td>
<td>DSM-1</td>
<td>yDHODH resistance under pfCAM, Fluc</td>
</tr>
<tr>
<td>pfYC120:Fluc</td>
<td>Blasticidin S</td>
<td>BSD resistance under pfCAM, Fluc</td>
</tr>
<tr>
<td>pfYC120:Fluc</td>
<td>WR99210</td>
<td>hDHFR resistance under pfCAM, Fluc</td>
</tr>
<tr>
<td>pfYC140:Fluc</td>
<td>DSM-1</td>
<td>yDHODH resistance under pfCAM, Fluc</td>
</tr>
<tr>
<td>pfYC120:Fluc</td>
<td>Blasticidin S</td>
<td>BSD resistance under pfCAM, Fluc</td>
</tr>
<tr>
<td>pfYC120:Fluc</td>
<td>WR99210</td>
<td>hDHFR resistance under pfCAM, Fluc</td>
</tr>
<tr>
<td>pfYC140:Fluc</td>
<td>DSM-1</td>
<td>yDHODH resistance under pfCAM, Fluc</td>
</tr>
<tr>
<td>pfYC120:Fluc</td>
<td>Blasticidin S</td>
<td>BSD resistance under pfCAM, Fluc</td>
</tr>
<tr>
<td>pfYC120:Fluc</td>
<td>WR99210</td>
<td>hDHFR resistance under pfCAM, Fluc</td>
</tr>
<tr>
<td>pfYC140:Fluc</td>
<td>DSM-1</td>
<td>yDHODH resistance under pfCAM, Fluc</td>
</tr>
<tr>
<td>pfYC120:Fluc</td>
<td>Blasticidin S</td>
<td>BSD resistance under pfCAM, Fluc</td>
</tr>
<tr>
<td>pfYC120:Fluc</td>
<td>WR99210</td>
<td>hDHFR resistance under pfCAM, Fluc</td>
</tr>
<tr>
<td>pfYC140:Fluc</td>
<td>DSM-1</td>
<td>yDHODH resistance under pfCAM, Fluc</td>
</tr>
<tr>
<td>pfYC120:Fluc</td>
<td>Blasticidin S</td>
<td>BSD resistance under pfCAM, Fluc</td>
</tr>
<tr>
<td>pfYC120:Fluc</td>
<td>WR99210</td>
<td>hDHFR resistance under pfCAM, Fluc</td>
</tr>
<tr>
<td>pfYC140:Fluc</td>
<td>DSM-1</td>
<td>yDHODH resistance under pfCAM, Fluc</td>
</tr>
</tbody>
</table>
2.5.2 *Vector construction using various cloning methods*

Several vectors were constructed to illustrate the ability to successfully clone firefly and *Renilla* luciferase reporter genes, and two native *P. falciparum* genes (*amal* and *trxR* both ~1.85 kb and ~70% in A+T content) into this vector family using all three cloning strategies. Using yeast HR or Gibson assembly, firefly or *Renilla* luciferase was cloned into the available expression cassette of the entire pfYC series (Table 2.1). All vectors were sequenced and topologically mapped by HindIII restriction digestion. As shown in Figure 2.3A, final plasmids with the expected topology can be assembled using these methods. Similarly, the candidate *P. falciparum* genes *amal* and *trxR* were inserted into pfYC120 using the three vector assembly methods in parallel. Cloning reactions were carried out using the same insert and vector preparations to minimize differences between the materials used in each reaction. Five colonies derived from each cloning method were screened for each gene target and mapped by HindIII digestion to establish proper assembly of the target vector (Figure 2.3B). Gibson assembly yielded topologically correct plasmids for both gene targets. However, under the conditions tested, yeast HR and restriction/ligation yielded the expected plasmid for *trxR* only. Overall, these data show that all three methods can be used to successfully clone native *P. falciparum* genes into this new vector family. Importantly, these independent cloning strategies allow use of the same plasmid backbone and insert combinations to assemble the identical final construct, thus improving the flexibility and overall ease with which *P. falciparum* vectors are made.
Figure 2.3: Assembly of *P. falciparum* vector using three different cloning strategies

(A) The firefly luciferase gene (*FL* = 1.65 kb) was cloned into the pfYC1 and pfYC3 series (Additional file 3) using either yeast HR or Gibson assembly. Topological mapping with HindIII digestion yields three fragments, as *FL* and the selection markers do not contain HindIII sites. A 3.9 kb fragment is released from the pfYC1 series whether *FL* is present or not (Figure 2). The fragments containing cassettes A and B from pfYC10x:*FL plasmids are (1.5 kb + *FL*) = 3.2 kb and (1.7 + selection marker size) kb, respectively. Similarly, the fragments containing cassettes A and B from pfYC1x0:*FL plasmids are (1.5 + selection marker size) kb and (1.7 + *FL* size) = 3.4 kb, respectively. The sizes of the different selection markers are: nptII (0.8 kb); hdhfr (0.6 kb); bsd (0.4 kb) and ydhodh (0.95 kb). This analysis confirms correct insertion of *FL* without gross plasmid rearrangements or insertions/deletions. (B) Two native *P. falciparum* genes, *ama1* (apical membrane antigen 1; PF3D7_1133400; 1.87 kb) and *trxR* (thioredoxin reductase; PF3D7_0923800.1; 1.85 kb) were cloned in parallel using restriction/ligation, Gibson assembly and yeast HR, and the same PCR products and digested pfYC120 vector. Successful gene insertion is expected to yield three HindIII digestion products that include: a backbone fragment (denoted as C); cassette B with the *ama1* or *trxR* gene inserted (denoted as B when no insert is present and B' when containing the proper insert); and cassette A containing the bsd gene (denoted as A) upon. As a reference, the parent pfYC120 plasmid yields products denoted as A, C and B upon HindIII digestion. Altogether, the three strategies yielded the desired final plasmid, with Gibson assembly successfully yielding both *ama1* and *trxR* constructs. The asterisk in the yeast HR *trxR* panel denotes sample degradation that occurred during storage prior to analysis by gel electrophoresis.
2.5.3 *Plasmids in this vector family can be maintained as stable episomes and chromosomally integrated in* *P. falciparum*

Toward establishing this vector family as a verified resource and a framework for routine use in *P. falciparum* transgenic experiments, their ability to yield stable episomal and integrated *P. falciparum* lines was evaluated. The entire pfYC1AB:FL vector set was transfected either singly or in a paired combination (pfYC110/pfYC120) into *P. falciparum* strain 3D7. Transfected parasites were selected using the appropriate drug(s), and growth was monitored by following luciferase activity. As shown in Figure 2.4A, parasites transfected with these plasmids were successfully selected with typical kinetics\(^{84,85}\). Interestingly, under the conditions tested, the pfYC104:FL- and pfYC140:FL- transfected parasites selected with DSM-1 emerged more rapidly than parasites selected with Blasticidin S, WR99210 or G-418. Dual plasmid transfected parasites emerged at rates similar to those observed in single plasmid transfections (Figure 2.4B). Copy numbers for the various pfYC1 plasmids were also determined by quantitative PCR using the single-copy chromosomal \(\beta\)-actin gene as a reference. These data indicate that plasmids selected with Blasticidin S, DSM-1 and G-418 are maintained at an average \(\leq\) five copies, and for WR99210 at \(\sim\) ten copies per parasite genome (Figure 2.4C). This is consistent with results using other *P. falciparum* vectors\(^{25,79}\), indicating that the pfYC vector family behaves similar to currently used plasmids and is suitable for use in transgenic experiments.
Figure 2.4: Behavior and copy number of the pfYC vectors after transfection
(A and B) The entire pfYC1xx:FL plasmid series was either transfected individually (A)
or as a single pair (pfYC110:FL + pfYC120:RL) (B) under the appropriate drug selection
initiated on day 4 post-transfection (arrow). Firefly and Renilla luciferase levels were
monitored to assess parasite population growth kinetics until a parasitaemia ≥1% was
attained. (C). The copy number of each plasmid per parasite genome was determined for
both the single and double transfections. (D) PCR confirmation of chromosomal
integration of pfYC320 and pfYC340 at the cg6 locus in the P. falciparum 3D7-attB
strain. The β-actin gene was PCR amplified as a positive control.

Frequently, the ability to site-specifically integrate constructs is preferred to ensure
stable, homogeneous transgene expression at single copy. The pfYC3 plasmid series is
designed to accomplish this by combining cloning strategy flexibility and a site-specific
integration *attP* utility feature\(^{39}\), while eliminating plasmid elements that are potentially deleterious when chromosomally integrated (Rep20 and CEN6/ARS4). As validation of this desired behaviour, 3D7-*attB* parasites were transfected with pfYC320:FL and pfYC340:FL. Stable parasite lines expressing FL were selected under Blasticidin S or DSM-1 pressure, respectively, and site-specific integration at the cg6 locus was detected by PCR both at the population level and in isolated clones (Figure 2.4D). Overall, these data collectively show that the pfYC vector family provides a robust and complementary set of high efficiency and timesaving cloning strategies for enabling routine assembly of DNA constructs that can be successfully used in *P. falciparum* transgenic experiments. Of note, while we have assembled two representative pfYC4 series plasmids containing the *pfcen5*-1.5 centromere element as a useful starting point for future use, we have not evaluated these in transfections.

2.5.4 *Expanded transgene expression from a single plasmid that is compatible with proper subcellular trafficking*

The ability to simultaneously express multiple proteins from a single plasmid irrespective of their subcellular localization can be highly useful, as it reduces the need for doing sequential transfections and potentially exhausting the limited set of available selection markers. The virus-derived 2A-like peptide sequences (2A tags), which have been used successfully in mammalian, yeast, plant and protozoan contexts to enable polycistronic expression from a single eukaryotic mRNA\(^{28,29}\) were used to accomplish this. These 2A tags mediate peptide bond “skipping” between conserved glycine and proline residues, yielding one protein with a short C-terminal extension encoded by the
tag, and the other with an N-terminal proline. The small size (eight conserved amino acid positions) and broad cross-species functionality of the 2A tag makes it an attractive candidate for application to *P. falciparum*, an organism in which this technology has not been extensively explored. As an entire expression cassette is usually committed exclusively to expressing a selection marker, an initial experiment was designed to address whether the *Thosea asigna* virus 2A-like sequence (T2A) could be used to expand the number of genes expressed from this cassette without compromising the ability to select transfected parasites. T2A with a short, N-terminal linker region\(^{28}\) was inserted between the FL and *nptII* genes in cassette A to generate pfYC101:FL-2A-nptII.

A control construct containing a non-functional tag (T2Am), in which two conserved residues are mutated to alanine\(^{86}\) was also generated (Figure 2.5A). These plasmids were transfected into *P. falciparum* 3D7 under G-418 selection pressure and obtained resistant parasites with FL activity (Figure 2.5B), demonstrating the production of functional nptII and FL proteins in both cases. The ability of T2A to produce distinct FL and nptII proteins from a single mRNA was confirmed by Western and Northern blot (Figure 2.5C and 2.5D, respectively). As expected, mutating T2A to T2Am eliminates the formation of discrete proteins, but does not alter the size of the *FL-nptII* mRNA. This initial characterization, in addition to demonstrating T2A functionality in *P. falciparum*, highlights the potential for using T2A to recover valuable expression capacity by encoding additional information into existing selection marker cassettes while eliminating the unpredictability of how a protein fusion will function.
Figure 2.5: Use of the T2A peptide for multi-protein expression from a single expression cassette in *P. falciparum*

(A) Schematic of *FL-nptII* and control constructs. (B) Both T2A- and T2Am-containing constructs produce active FL. (C) Western blot detection of FL- and nptII-containing proteins. (D) Northern blot analysis of FL-containing transcripts in transfected parasites. 3D7 + FL indicates the inclusion of a synthetic FL mRNA produced by *in vitro* transcription.

Next, the flexibility with which T2A can be used to produce dicistronic messages encoding proteins destined for distinct subcellular compartments within the parasite and its RBC host was examined. Several dicistronic constructs encoding an N-terminal Venus yellow fluorescent protein (vYFP) and a C-terminal tdTomato protein (tdTom) separated by T2A were built in the pfYC120 vector. Previously validated apicoplast, mitochondrial and RBC export targeting sequences derived, respectively, from: acyl carrier protein (PF13_0208500; aa 1-60 = ATS)\(^7\), HSP60 (PF13_1015600; aa 1-68 = MTS)\(^8\), and knob-associated histidine-rich protein (PF13_0202000; aa 1-69 = PEX)\(^9\) were used.
Seven contexts were created in which a different protein targeting signal (or none at all) was placed immediately upstream of vYFP and/or tdTom as follows: (a) vYFP-2A-tdTom; (b) vYFP-2A-ATS-tdTom; (c) vYFP-2A-MTS-tdTom; (d) vYFP-2A-PEX-tdTom; (e) MTS-vYFP-2A-MTS-tdTom; (f) PEX-vYFP-2A-PEX-tdTom; and, (g) ATS-vYFP-2A-tdTom. vYFP and tdTom trafficking were evaluated using fluorescence imaging microscopy, and distinguished production of vYFP versus a possible fusion to tdTom by Western blot (Figure 2.6).

Overall, when no targeting sequence is upstream of vYFP, the downstream tdTom is faithfully trafficked to the subcellular compartment expected based on the associated targeting sequence. Similarly, when vYFP and tdTom are associated with the same targeting sequence (parasite cytosol, mitochondrion and RBC cytosol tested), both are trafficked as separate proteins to the same subcellular compartment, as expected. For the ATS-vYFP-2A-tdTom construct, vYFP is trafficked to the apicoplast as expected. Interestingly, a substantial fraction of the tdTom is mislocalized to the apicoplast with some signal distributed in the parasite’s cytoplasm. By Western blot, vYFP is detected as both the isolated protein and the tdTom fusion (~100 kDa). Presumably, the fusion product accounts for the majority of the mislocalized tdTom, while the cytosolic fraction arises due to the expected T2A behaviour. These data suggest that ‘ribosome skipping’ might be less efficient and/or the downstream protein is more often misdirected when the upstream protein is apicoplast-targeted, at least within the context tested by the present constructs.

This outcome is reminiscent of “slipstreaming” observed when using T2A for multi-cistronic expression of secreted proteins in mammalian cells, though this
phenomenon is thought to be primarily influenced by the C-terminal portion of the upstream protein\textsuperscript{90}, which does not vary across the constructs used here. However, since the vYFP-2A-ATS-tdTom construct exhibits the expected subcellular targeting patterns, and given the other combinations in which proper subcellular trafficking was observed, define some rules for using T2A to successfully achieve multi-cistronic protein expression with proper subcellular targeting.
Figure 2.6: 2A mediated subcellular targeting in *Plasmodium falciparum*

Various targeting sequences were N-terminally fused to an upstream vYFP and a downstream tdTom reporter separated by T2A. The vYFP and tdTom proteins were localized using direct fluorescence microscopy imaging. Mitochondria were stained with MitoTracker (MT), and nuclei with Hoechst 33342. Legend: ATS = apicoplast targeting sequence; MTS = mitochondrial targeting sequence and PEX = protein export element.
2.6 Discussion

A validated set of broadly useful plasmid vectors has been developed that enables versatile assembly of *P. falciparum* constructs, a frequently time-consuming process given the A+T richness and large size of final vectors. This has been achieved by integrating simultaneous access to the high efficiency and inexpensively available yeast homologous recombination, Gibson assembly and conventional restriction/ligation cloning strategies. All three strategies are technically straightforward and use the same restriction enzyme-digested vector and PCR products generated with the same primer set. In principle, all three strategies can be executed in parallel or used interchangeably without the need for new genetic reagents and can be used to successfully clone both reporter and native *P. falciparum* genes. Additionally, several widely used utility features for enhancing plasmid segregation and site-specific chromosomal integration have been pre-installed, and validated operations to enable user-tailored modifications to this vector family are provided.

In addition to improving the ease of constructing new *P. falciparum* expression vectors, a strategy for increasing the amount of expressible information that can be encoded on a single plasmid using a minimal set of 5'3'-UTR cassettes was also developed. From a technical standpoint, this is especially useful as it simplifies the vector construction process by reducing overall plasmid size and instability during propagation in *E. coli*. Practically, this provides more efficient avenues for addressing questions in parasite biology requiring the co-expression of multiple genes. For example, several antigenically variant, multi-gene families, such as PfEMP1, STEVORs and RIFINs, are combinatorially expressed by the parasite to modulate host-parasite interactions, such as
immune recognition and evasion\textsuperscript{12,91}. Therefore, the ability to achieve pre-determined expression of specific combinations of these proteins could prove useful in understanding their combined contributions to these outcomes. Furthermore, subsets of proteins involved in multi-gene pathways must often be trafficked to distinct subcellular compartments. The tricarboxylic acid cycle\textsuperscript{92,93}, lipid and isoprenoid biosynthesis\textsuperscript{94} and haem biosynthesis\textsuperscript{95} involve multiple proteins distributed between the cytosol, mitochondrion and apicoplast, or exclusively targeted to one of these organelles. Also, a substantial fraction of the parasite-encoded proteome is trafficked to the host RBC, including the antigenically variant STEVOR and RIFIN families, and a significant number of these trafficked proteins play essential but poorly understood roles\textsuperscript{96}. Therefore, strategies for achieving multi-cistronic expression while simultaneously preserving faithful protein trafficking will be broadly useful for studying parasite biology. Altogether, these openly available tools and validated methods should provide a convenient option for routinely generating \textit{P. falciparum} plasmid vectors.

2.7 Materials and Methods

2.7.1 Molecular Biology

Unless otherwise indicated, enzymes were from New England Biolabs (Ipswich, MA, USA) and chemicals were from Sigma-Aldrich (St. Lois, MO, USA) or Research Products International (Mt. Prospect, IL, USA). High fidelity (HF) restriction enzymes were used when available. PCR was routinely performed with Phusion DNA polymerase in HF Buffer, or with a 15:1 (v:v) mixture of Hemo KlenTaq:Pfu Turbo (Agilent, Santa Clara, CA, USA) in Hemo KlenTaq Buffer. The latter conditions permit PCR
amplification directly from parasite culture samples, usually included at 5% of the total reaction volume. Plasmids were prepared for transfection with maxi columns (Epoch Life Science, Missouri City, TX, USA) or the Xtra Midi Kit (Clontech, Mountain View, CA, USA).

2.7.2 Vector Construction

The primers used in these studies can be found in table 2.2.

<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Primer Sequence (5'-&gt;3')</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>SG311</td>
<td>GAAATATATCAGACGTCTCCCCCGGGACC ATGGAAGAGCAGCAAAAAACATAAAGAAAG GCC</td>
<td>FL probe (forward)</td>
</tr>
<tr>
<td>SG313</td>
<td>GACCCCATTTGTGAGTACATAAAATATTATATTAT ATAACTCGAGTTACAAACTCGGACTTTCCGC</td>
<td>FL probe (reverse)</td>
</tr>
<tr>
<td>SG763</td>
<td>AGCATGTGCATGGCCATCCCTTT</td>
<td>Amplifying uniquely identifying regions of vYFPtdTom cassette (forward)</td>
</tr>
<tr>
<td>SG764</td>
<td>TGACCTCCTCGCCCTTGCTCA</td>
<td>Amplifying uniquely identifying regions of vYFPtdTom cassette (reverse)</td>
</tr>
<tr>
<td>SG502</td>
<td>AGTAGCATCACCTTCACCTTCACC</td>
<td>Sequencing from vYFP in dicistronic vYFP-tdTom constructs (reverse)</td>
</tr>
<tr>
<td>SG646</td>
<td>CTGCCTTATCCCAAAGATCCAAACG</td>
<td>Sequencing from vYFP in dicistronic vYFP-tdTom constructs (forward)</td>
</tr>
<tr>
<td>SG702</td>
<td>TAGGGTGACACTATAGAATACTCAAGCTTG GCGGCCGCCCAGCTCGAATTCCGGGTTT GT</td>
<td>Tandem <code>attP</code> (forward)</td>
</tr>
<tr>
<td>-------</td>
<td>---------------------------------------------------------------</td>
<td>----------------------</td>
</tr>
<tr>
<td>SG703</td>
<td>AGTTAATTCATCAAATAGCATGCTTGAGTGTCGACGCCAGGGTTTCCCAGTACGA</td>
<td>Tandem <code>attP</code> (reverse)</td>
</tr>
<tr>
<td>SG814</td>
<td>GGACATTTGTTTTAACGAGCAGG ACGCGT TGAATTGTCCCCACGCACCCGGCGCCC</td>
<td>Recircularizion at PmlI/MluI sites (forward)</td>
</tr>
<tr>
<td>SG815</td>
<td>CCTGCTCGTTTTAAACAATGTCC CACGTG ATGAAAAAGGACCCAGGTGGCA</td>
<td>Recircularizion at PmlI/MluI sites (reverse)</td>
</tr>
<tr>
<td>SG864</td>
<td>GCAGGTCGACGCCAGGGTTT</td>
<td>Confirming pfYC integration at cg6-<code>attB</code> (forward)</td>
</tr>
<tr>
<td>SG865</td>
<td>GACGCGCGCGCAAGAGCAACT</td>
<td>Confirming pfYC integration at cg6-<code>attB</code> (reverse)</td>
</tr>
<tr>
<td>SG928</td>
<td>GGGCGCGGCACGGGACAATTCACCGGCT TAATTTAAATATATATATATTTAGACTTA</td>
<td><code>pfcen5-1.5 PCR amplification</code> (forward)</td>
</tr>
<tr>
<td>SG894</td>
<td>TGCCACCTGGGTCCTTTTCATCACGTG TATGTATAATTAATTAATTAATTTACA CAC</td>
<td><code>pfcen5-1.5 PCR amplification</code> (reverse)</td>
</tr>
<tr>
<td>SG369</td>
<td>CCAATGCTTAATCAGTGAGGC</td>
<td>Sequencing primer: <code>yCEN/or pfcen5-1.5 region</code></td>
</tr>
<tr>
<td>Trx pcDT F</td>
<td>AAATATATACACACACCTAAAACCTTACAA AGTATCCTAGGAAAAATGAACATGTAATT</td>
<td><code>trxR PCR amplification</code> (forward)</td>
</tr>
<tr>
<td>Trx pcDT R</td>
<td>TTTAATCTATTATTTAAATTTAATAGGG GTACCCGCAGTTATCCACATTTCCACCC</td>
<td><code>trxR PCR amplification</code> (reverse)</td>
</tr>
<tr>
<td>AMA1 pcDT F</td>
<td>AAATATATACACACACCTAAAACCTTACAA AGTATCCTAGGAAAAATGAGAAAATTA</td>
<td><code>ama1 PCR amplification</code> (forward)</td>
</tr>
<tr>
<td>AMA1 pcDT R</td>
<td>TTTAATCTATTATTTAAATTTAATAGGG GTACCCGCAGTTATAGGTTTTCCTCA</td>
<td><code>ama1 PCR amplification</code> (reverse)</td>
</tr>
</tbody>
</table>
2.7.3 Yeast Homologous Recombination

Yeast homologous recombination (HR) vector construction was carried out by standard methods\textsuperscript{76,97}. Variable amounts of vector backbone (typically 0.1-2 μg) were digested using standard methods to generate linearized vector. PCR was carried out using standard techniques to generate fragments for insertion bearing 20-40 bp homology to the desired flanking regions on the vector. Competent \textit{Saccharomyces cerevisiae} W303-1B was prepared as described\textsuperscript{80} and frozen at -80°C. Either unpurified or column-purified PCR product was co-transformed with either unpurified or column-purified linearized vector. A wide range of concentrations of both linearized vector and PCR product were observed to be efficacious. Transformed yeast were plated on YPD agar (10 g/L yeast extract, 20 g/L peptone, 20 g/L dextrose, 20 g/L agar) supplemented with 400 mg/L G-418 disulphate and allowed to grow for 48-72 hours at 30°C. Typical yields were 10-100 colonies for a negative control transformation lacking PCR insert, and 50-1,000 colonies for the complete HR reaction.

Colonies were then either harvested by plate scraping or picked and grown overnight in YPD + 400 μg/mL G-418. Cells were then treated with 2 U zymolyase (Zymo Research, Irvine, CA, USA) in 250 μL buffer ZB (10 mM sodium citrate pH 6.5, 1 M sorbitol, 25 mM EDTA and 40 mM dithiothreitol) for 1 hour at 37°C to generate spheroplasts. Yeast spheroplasts were lysed with the addition of 250 μL buffer MX2 (0.2 M NaOH and 10 g/L sodium dodecyl sulphate). Plasmid DNA was then purified either by spin column (Epoch Life Science, Missouri City, TX, USA) or alcohol precipitation. The recovered DNA was transformed directly into \textit{Escherichia coli} DH5α or EPI300 (Epicentre Biotechnologies, Madison, WI, USA) prepared with a Z-Competent
Transformation Kit (Zymo Research) or transformed by electroporation. Occasionally, the DNA mixture was drop dialyzed against water for 20 min prior to transformation to increase electroporation. Plasmid DNA was isolated from colonies and assayed for correct vector assembly by restriction digest, diagnostic PCR and/or DNA sequencing.

2.7.4 **Gibson Assembly**

Isothermal chew-back-anneal assembly, commonly known as Gibson assembly, was carried out as described\(^77\). Briefly, vector and PCR product were prepared in the same way as for yeast HR assembly. Fragments were combined with either a home-made or commercially available Gibson Assembly Master Mix. The home-made Master Mix was prepared by combining 699 μL water, 320 μL of 5x isothermal reaction buffer (500 mM Tris-Cl, pH 7.5, 250 mg/mL PEG-8000, 50 mM MgCl\(_2\), 50 mM dithiothreitol, 1 mM each of four dNTPs, 5 mM NAD), 0.64 μL T5 Exonuclease (Epicentre, 10 U/μL), 20 μL Phusion DNA polymerase (2 U/μL) and 160 μL Taq DNA ligase (40 U/μL). This solution was divided into 20 μL aliquots and stored at \( -20^\circ C \). Generally, >100 ng of linearized vector was added to the mixture with an equal volume of PCR insert, generating a variable vector: insert ratio. The mixture was incubated at 50°C for 1 hour and 0.5 μL was transformed into *E. coli* as described above.

2.7.5 **Restriction/ligation cloning**

Restriction/ligation cloning was carried out by standard techniques. Ligations were incubated overnight at 16°C and heat inactivated prior to transformation.
2.7.6 Construction of attP-containing (pfYC3 series) vectors

The 2×attP fragment was PCR amplified from pLN-ENR-GFP with primers SG702/703. pfYC120:FL and pfYC140:FL were digested with SalI and combined with gel-purified PCR product in a Gibson assembly reaction to obtain pfYC220:FL and pfYC240:FL, respectively. These vectors were then digested with MluI and PmlI to release the fragment containing the Rep20 and CEN/ARS elements. Approximately 100 ng of digested vector was then combined in a Gibson assembly reaction containing 50 nM each SG814 and SG815 to recircularize the vector while adding a unique PmeI site between the MluI and PmlI sites. This yielded pfYC320:FL and pfYC340:FL. For integration at the cg6 locus, these two vectors were co-transfected with pINT39 (~50 μg each) into P. falciparum 3D7-attB (MRA-845 from MR4: www.mr4.org). Clonal populations were obtained by limiting dilution and integration verified by PCR using the SG864/865 primer pair.

2.7.7 Construction of the pfcen5-1.5 mini-centromere containing (pfYC4 series) plasmids

The pfcen5-1.5 element was amplified from P. falciparum 3D7 genomic DNA with primers SG894/SG928. pfYC102:FL and pfYC104:FL were digested with MluI and PmlI and the gel-purified backbone lacking the 2×Rep20 element was attached to the SG894/SG928 PCR product by Gibson assembly to yield pfYC402:FL and pfYC404:FL, respectively. Clones were verified by restriction digest and by sequencing with SG369.

2.7.8 Cloning Plasmodium falciparum ama1 and trxR genes into pfYC120
The *ama1* and *trxR* genes were amplified from *P. falciparum* 3D7 genomic DNA using the AMA1 pcDT F/R and Trx pcDT F/R primer pairs, respectively. Restriction/ligation, Gibson assembly and yeast HR were performed as described above.

2.7.9 Multi-cistronic constructs using the T2A for evaluating subcellular trafficking rules

After Western blot and microscopic imaging analysis, the identity of each strain was re-verified by PCR amplifying and sequencing a uniquely identifying fragment of the transfected construct using the SG763/764 and SG502/646 primer pairs, respectively.

2.7.10 Parasite culture and transfection

*Plasmodium falciparum* strain 3D7 parasites were cultured under 5% O₂ and 5% CO₂ in RPMI-1640 media supplemented with 5 g/L Albumax II (Life Technologies), 2 g/L NaHCO₃, 25 mM HEPES-K pH 7.4, 1 mM hypoxanthine and 50 mg/L gentamicin. Transfections used ~50 µg of each plasmid and were performed by the spontaneous DNA uptake method¹⁹ or by direct electroporation of ring-stage cultures²⁶. Transgenic parasites were selected with 2.5 mg/L Blasticidin S, 1.5 µM DSM-1, 5 nM WR9920 (Jacobus Pharmaceuticals) and/or 250 mg/L G-418 beginning 2-4 days after transfection.

2.7.11 Monitoring transfection progress by luciferase expression

Firefly and *Renilla* luciferase levels were measured every fourth day after transfection using the Dual-Luciferase Reporter Assay System (Promega). Samples for measurement were prepared by centrifugation of 1.25 mL of parasite culture at 2°
haematocrit to generate an ~25 μL parasite-infected RBC pellet. Pellets were either used immediately or stored at -80°C until needed for luciferase measurements.

2.7.12 Parasite DNA extraction and qPCR analysis

DNA was harvested from schizont stage parasites at ≥ 5% parasitaemia. Infected red blood cells (RBCs) were treated with 0.1 mg/mL saponin in PBS to release parasites, which were either immediately used or stored in liquid nitrogen for later analysis. Parasites were lysed for 1 hour at 50°C in 200 μL of 50 mM Tris-Cl pH 8.0, 50 mM EDTA, 0.5 mg/mL SDS and 10 μL of protease solution (Qiagen). After adding RNase A (28 U; Qiagen), reactions were incubated at 37°C for 5 min. After adding 20 μL of 6 M sodium perchlorate, DNA was isolated by phenol/chloroform extraction and ethanol precipitation.

qPCR reactions (20 μL each) contained 1× Thermopol Buffer, 0.2 mM dNTPs, 200 nM relevant primer pair (Table 2.3), 0.5× SYBR Green I (Life Technologies), 0.1 μL Taq DNA polymerase and 5 μL of a DNA dilution. Thermocycling was performed on a Roche LightCycler 480 II for 40 cycles according to the following programme: 95°C for 20 sec; denature: 95°C for 3 sec; anneal/extend: 60°C for 30 sec; fluorescence measurement. Vector-borne amplicons (drug resistance marker genes) and a native chromosomal locus (β-actin) were quantified by comparison with plasmid or PCR-amplified DNA standards, respectively.
### Table 2.4: Quantitative PCR primers used in these studies

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
<th>Target Gene</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neo (forward)</td>
<td>CATCCTGATCGACAAAGACCG</td>
<td>nptII</td>
<td>This work</td>
</tr>
<tr>
<td>Neo (reverse)</td>
<td>CCTGCAGAGAAAGATTCATCATC</td>
<td>nptII</td>
<td>This work</td>
</tr>
<tr>
<td>BSD (forward)</td>
<td>GCTGTCCATACGTGCTTTC</td>
<td>Blasticidin S deaminase</td>
<td>This work</td>
</tr>
<tr>
<td>BSD (reverse)</td>
<td>TGGCAACCTGACTTGTATCG</td>
<td>Blasticidin S deaminase</td>
<td>This work</td>
</tr>
<tr>
<td>hDHFR (forward)</td>
<td>GAGGTTTGGTGTCATTCTCTGG</td>
<td>Human dihydrofolate reductase</td>
<td>This work</td>
</tr>
<tr>
<td>hDHFR (reverse)</td>
<td>AGAACATGGGCATCGGC</td>
<td>Human dihydrofolate reductase</td>
<td>This work</td>
</tr>
<tr>
<td>yDHODH (forward)</td>
<td>TCCACCTGTACCCACTTTGG</td>
<td>Yeast dihydroorotate dehydrogenase</td>
<td>This work</td>
</tr>
<tr>
<td>yDHODH (reverse)</td>
<td>GATGTTGAGAAGAGAGTGTAAG</td>
<td>Yeast dihydroorotate dehydrogenase</td>
<td>This work</td>
</tr>
<tr>
<td>Pf-β-actin (forward)</td>
<td>AAAGAAGCAAGCAAGAATCCG</td>
<td><em>P. falciparum</em> β-actin (PF3D7_1246200)</td>
<td>Augagneur et al. 98</td>
</tr>
<tr>
<td>Pf-β-actin (reverse)</td>
<td>TGATGTTGCAAGGGTTTGAAGA</td>
<td><em>P. falciparum</em> β-actin (PF3D7_1246200)</td>
<td>Augagneur et al. 98</td>
</tr>
</tbody>
</table>

#### 2.7.13 Western Blot

Approximately $10^6$ late-stage parasites were harvested by lysis of infected RBCs with 0.5 g/L saponin and then lysed by heating in urea sample buffer (40 mM dithiothreitol, 6.4 M urea, 80 mM glycylglycine, 16 g/L SDS, 40 mM Tris-Cl, pH 6.8) at 95°C for 10 min. After separation by SDS-PAGE, proteins were transferred to a PVDF membrane and probed with an antibody against firefly luciferase (FL) (Promega G7451), neomycin...
phosphotransferase II (Millipore 06-747) or green fluorescent protein (Abcam ab1218).

Blots were then imaged using a horseradish peroxidase-coupled secondary antibody and SuperSignal West Femto substrate (Thermo Scientific).

2.7.14 Northern blot

Total RNA was purified from infected RBCs with a combination of Tri Reagent RT Blood (Molecular Research Center) and an RNeasy Mini Kit (Qiagen). One mL of parasite culture at 20% haematocrit and ~10% late-stage parasitaemia was frozen on liquid nitrogen and thawed with the addition of 3 mL Tri Reagent RT Blood. After phase separation with 0.2 mL BAN (Molecular Research Center), 2 mL of the upper aqueous phase was mixed with 2 mL ethanol and applied to an RNeasy Mini column for purification according to the manufacturer’s instructions. Total RNA (6.5 mg) for each sample (with or without the addition of 6 pg FL RNA generated by in vitro transcription) was mixed with an equal volume of denaturing sample buffer (95% formamide, 0.25 g/L SDS, 0.25 g/L bromophenol blue, 0.25 g/L xylene cyanol, 2.5 g/L ethidium bromide, 50 mM EDTA) and heated at 75°C for 10 min before loading on a 1% TAE agarose gel. Electrophoresis was performed at 80 V for 75 min and RNA was transferred to a Nylon membrane (Pall Biodyne Plus) by downward capillary transfer in 50 mM NaOH for 90 min. After UV fixation (Stratagene Stratalinker, 1.25 mJ), the membrane was probed and imaged with the North2South Chemiluminescent Detection Kit (Thermo). The biotinylated FL probe was prepared from a DNA template generated by PCR with primers SG311 and SG313.
2.7.15 Fluorescence microscopy

For live cell imaging, parasite cultures were incubated for 20 min with 30 nM MitoTracker Deep Red FM (Life Technologies). Infected RBCs were then washed with phosphate-buffered saline and applied to poly-L-lysine-coated glass-bottom culture dishes (MatTek, Ashland, MA, USA). Attached cells were overlaid with RPMI media (free of phenol red and Albumax II) containing 2 μg/mL Hoechst 33342 (Sigma), and imaged immediately at room temperature using a Nikon Ti-E inverted microscope with a 100x objective and a Photometrics CoolSNAP HQ2 CCD camera. Images were collected with the Nikon NIS Elements software and processed using ImageJ99.
Chapter 3: Utilization of the CRISPR/SpCas9 System in Conjunction with T7 RNAP for Targeted Genome Editing in *P. falciparum*

3.1 Note

This chapter has been adapted from a manuscript currently under review by *Nature methods*.

3.2 Abstract

Malaria caused by the protozoan parasite *Plasmodium falciparum* remains a global problem. The ability to perform genetic manipulations, such as selectively knocking out genes, is central to studying the parasite’s biology. Standard approaches are inefficient and present a substantial bottleneck in functional genetics studies. Here we show that the *Streptococcus pyogenes* Cas9 DNA endonuclease (SpCas9) and guide RNAs (gRNAs) produced in situ using T7 RNA polymerase (T7 RNAP) can be used to efficiently edit the *P. falciparum* genome. We show that gRNAs specific to two native chromosomal genes, namely the knob associated histidine rich protein (*kahrp = PF3D7_0202000*) and erythrocyte binding antigen 175 (*eba-175 = PF3D7_0731500*), successfully targeted these genes for SpCas9-mediated disruption. In both instances, the expected modifications occurred within the time required to generate easily manipulated transgenic parasite populations (4-6 weeks). Furthermore, we observed ~100% and ~50% conversion of the *kahrp* and *eba-175* loci, respectively, at the population level. Our results demonstrate a straightforward strategy for rapidly and efficiently achieving site-specific genome editing in *P. falciparum*. We anticipate these findings will broadly
facilitate basic biological studies, as well as enable drug and vaccine target validation efforts by improving the ease and efficiency with which functional genetics in *P. falciparum* can be accomplished.

3.3 Introduction

Malaria is a major cause of global morbidity and mortality, and new strategies for treating and preventing this disease are needed. The protozoan parasite, *Plasmodium falciparum*, is a major cause of the estimated 219 million cases and 655,000 deaths from malaria per year. New molecular tools are needed to enable and increase the speed of research into malaria parasites. The most commonly used approach for modifying chromosomal loci in *P. falciparum* relies on spontaneous single- or double-crossover recombination using plasmids containing sequence homologous to the target region. This is extremely inefficient, and requires many months to achieve the desired outcome.

Circular plasmids are used to transform *P. falciparum*, and these are preferentially maintained as episomes. Isolating rare parasites with the desired chromosomal integration event from a high episomally-transformed background requires protracted on/off selection drug cycling and/or negative selection procedures, none of which increase the initial fraction of the population genetically modified as desired. Zinc finger nucleases (ZFNs) have recently been used to efficiently achieve targeted knockouts and allele replacements in *P. falciparum*. Despite the effectiveness of this technology in some situations, each time a different genomic region is targeted an intensive effort is required to develop and validate new sequence-specific nucleases that are not guaranteed to function effectively. Recently, efficient site-specific genome editing using
SpCas9 has been validated in several organisms. Here, a guide RNA (gRNA) is used to direct SpCas9 to a target site defined by a 5'-NGG-3' protospacer adjacent motif (PAM) and an upstream 20-nucleotide sequence selected based on Watson-Crick base pairing with the target site (Figure 3.1). These straightforward implementation rules make this a potentially versatile option for broadly accessible use in *P. falciparum*.

**Figure 3.1: Schematic of SpCas9, gRNA, and target DNA interaction**
The gRNA base pairs with a 20 nucleotide target DNA region (blue solid line) defined by an NGG protospacer adjacent motif (PAM - red solid line). SpCas9 induces double strand DNA cleavage (red arrow head) three nucleotides in the 5' direction from the PAM site.

Despite interest in non-coding RNAs (ncRNAs) in the malaria parasite, to our knowledge no methods exist for the production of well-defined non-coding RNAs in *P. falciparum*. We therefore reconstituted the T7 RNA Polymerase (T7 RNAP) RNA production system in *P. falciparum* and established its ability to produce high levels of non-coding RNA from the T7 promoter during the Intraerythrocytic Development Cycle (IDC) of the parasite. We then utilized this system to conduct targeted knockouts of genes using the CRISPR system *in situ*. By transfecting and selecting for parasites which
contained the T7 RNAP, the Cas9 nuclease from *Streptococcus pyogenes*, and a T7 RNAP driven gRNA, we discovered that we could stably knockout genes below the detectable limit in a population without cloning by limiting dilution. Our results demonstrate the first use of T7 RNAP in any apicomplexan parasite to date, as well a highly efficient knockout system using easy to produce components *in situ*.

3.4 Results

3.4.1 *T7 RNAP is capable of producing non-coding RNA in situ*

Implementing *SpCas9* RNA-guided nucleases in *P. falciparum* requires a validated strategy for producing gRNAs *in situ*. While the strong U6 promoter transcribed by the endogenous Pol III has been used to produce functional gRNAs in other organisms\(^40,42,103\), a functional U6 promoter has not been defined in *P. falciparum*. Several *P. falciparum* Pol II promoters have been described, however target transcripts produced from these promoters have long, heterogeneously sized 5’ and 3’ flanking regions\(^105,106\). Given this, we chose T7 RNAP for producing defined gRNAs in *P. falciparum*, as it uses well-characterized promoter and terminator sequences to make transcripts of defined size and in high yield in several organisms\(^60,66–68\).

We first validated that functional T7 RNAP can be expressed in *P. falciparum*, as this had not previously been reported. We constructed two plasmid vectors: one for producing T7 RNAP (pT7 RNAP); and the other from which the full-length *Renilla* luciferase (RLuc) coding sequence could be transcribed as a reporter of *in situ* T7 RNAP activity (pT7-RLuc\(^T\); Figure 3.2).
Figure 3.2: **Vector design to test functionality of T7 RNAP in *P. falciparum***

Plasmids used to test T7 RNAP expression and activity. In **pT7 RNAP**, T7 RNAP is expressed using the \( \text{PcDT} 5'\text{-}Pfhrp2 3'\text{-UTR} \) pair, and selection with G418 is facilitated by \( \text{gfp-nptII} \) expressed from the \( \text{PfCam} 5'\text{-}Pfhsp86 3'\text{-UTR} \) pair. In the reporter plasmid, **pT7 RL1**, the Renilla luciferase (RL) gene is expressed from a T7 promoter/T7 terminator cassette. **pT7 RL2** and **pT7** are control plasmids in which the T7 promoter and the RL expression cassette, respectively, have been deleted from **pT7 RL1**, **pT7 RL1**, **pT7 RL2** and **pT7** encode: (i) a Blasticidin S deaminase (bsd) selection marker controlled by the \( \text{PfCam} 5'\text{-}Pfhsp86 3'\text{-UTR} \) pair; and (ii) a firefly luciferase gene controlled by the \( \text{PcDT} 5'\text{-}Pfhrp2 3'\text{-UTR} \) pair. The latter is used to monitor the progress of transfections.

We transfected *P. falciparum* 3D7 with either **pT7 RNAP** alone or together with **pT7-RLuc \(^T\)**, and selected for parasites in which these plasmids were episomally maintained. Western blot analysis confirmed that intact T7 RNAP is produced (Figure 3.3), and *in vitro* transcription using parasite lysates showed that the enzyme is active (Figure 3.4).
Western blot analysis indicates that T7 RNAP protein is produced by the parasite. Blots were probed with anti-T7 RNAP and anti-NPTII antibodies (loading control).

In vitro transcription was carried out with radio labeled nucleotide as described in material and methods. Lane 1=Recombinant T7 RNAP, Lane 2= Recombinant T7 RNAP mixed with WT 3D7 lysate, Lane 3= No lysate, Lane 4=WT 3D7 lysate, Lane 5= Lysate from parasites transfected with the PfGNr plasmid (MRA-462; www.mr4.org) that confers G418 resistance but does not encode T7 RNAP; Lane 6 = Lysate from *P. falciparum* transfected with *pT7 RNAP*.

Based on quantitative RT-PCR and Northern blot analyses, we confirmed the *in situ* activity of T7 RNAP and its ability to produce RNA of the expected size (Figure 3.5).
Figure 3.5:  Analysis of T7 dependent RNA production in *P. falciparum*

A. Northern blot analysis shows that RL transcripts of the expected size are synthesized only when the *pT7 RL1* and *pT7 RNAP* plasmids are co-transfected. A probe to Blasticidin S deaminase was used as a loading control. B. Quantitative RT-PCR analysis of normalized RLuc transcript levels produced after the indicated transfections. Significant RLuc transcript synthesis is observed only when *pT7 RL1* (reporter with T7 promoter) and *pT7 RNAP* (T7 RNAP expressed) are co-transfected.

Importantly, T7 RNAP expression both in the absence and presence of an expression cassette driven by the T7 promoter is well tolerated. No gross effects on parasite viability in the short term (Figure 3.6) or after prolonged periods of continuous culture (data not shown) were observed.
Figure 3.6: Expansion rate of *P. falciparum* expressing T7 RNAP
Expansion was measured by flow cytometry after fixing and staining with SYBR gold dye. Expansion rate was normalized to the -T7 RNAP, - T7 RLuc Cassette strain.

We next sought to determine if the T7 RNAP was capable of making RNA across the entire intraerythrocytic development cycle (IDC). Time course data revealed this to be the case (Figure 3.7).
Figure 3.7: T7 dependent RNA expression across the *P. falciparum* IDC

RNA levels were measured by quantitative PCR at 5 different time points during the parasite IDC. Each measurement was normalized to blasticidin S deaminase (BSD) levels as a loading control (left axis). The raw BSD levels for the strain containing the T7 RNAP are graphed on the right axis and demonstrate the expected pattern of expression during the IDC\(^{11}\). Error bars represent the standard of deviation for three technical replicates.

Finally, we demonstrated that the RLuc RNA is not translated into protein. Dual luciferase measurements showed no difference in RLuc levels in the presence or absence of T7 RNAP (Figure 3.8). This is likely due to the T7 RNAP not providing the 5'-7'-methylguanosine cap which generally recruits the eukaryotic translation machinery to the RNA\(^{62,63}\).
3.4.2 T7 RNAP can be used to produce functional gRNAs

Guide RNAs produced in vitro using T7 RNAP and linearized plasmid DNA templates have been used to successfully induce SpCas9-mediated editing in zebrafish and Drosophila melanogaster. However, since P. falciparum maintains circular plasmids, a T7 terminator is required to specify transcription termination. However, the T7 terminator is mostly transcribed, and thus, gRNAs produced from such templates will have the T7 terminator-derived stem-loop region at their 3’-termini (gRNA\textsuperscript{T}, Figure 3.9). As the impact of this additional sequence on gRNA function was unknown, we first tested whether a gRNA\textsuperscript{T} can direct specific cleavage of a DNA template in vitro. We PCR amplified target DNA fragments from both the parasite kahrp and eba-175 loci (nucleotides 619-2267 and 141-1363, respectively) containing predicted gRNA target
sites. We *in vitro* transcribed test *kahrp*- and *ebal75*- gRNA$^T$s, and a control pUC19-gRNA$^T$ from circular plasmid templates. We then co-incubated *Sp*Cas9-containing cell lysates, DNA templates and either test (*kahrp*- and *ebal75*- gRNA$^T$) or control (pUC19-gRNA$^T$) to test cleavage activity. As shown in Figure 3.9, both the *kahrp*- and *ebal75*- gRNA$^T$s specifically and efficiently cleaved their respective target DNA, while the pUC19-gRNA$^T$ did not cleave the *kahrp* or *ebal75* templates. Thus, the additional sequence introduced when using the T7 terminator does not interfere with *Sp*Cas9/gRNA$^T$-mediated target cleavage.
**Figure 3.9: gRNA design and in vitro cleavage assay results**

In vitro assay to assess cleavage activity of selected gRNA\(^\text{T}\) sequences. Predicted secondary structures for kahrp- and ebal75- gRNA\(^\text{T}\)s with the additional T7 terminator-derived sequence are shown. These in vitro transcribed gRNA\(^\text{T}\)s induced the expected cleavage of their respective target DNA in lysates obtained from SpCas9 expressing HEK293 cells. The control pUC19-gRNA\(^\text{T}\) did not induce cleavage of either kahrp or ebal75 target DNA.
3.4.3 CRISPR mediated knockout of kahrp

Next, we determined whether gRNA\textsuperscript{T} produced \textit{in situ} can be used to successfully disrupt chromosomal loci in \textit{P. falciparum}. We selected the \textit{kahrp} locus on \textit{P. falciparum} chromosome 2 as an initial target. This gene is involved in the formation of 'knobby' projections on the surface of \textit{P. falciparum}-infected red blood cells, and its disruption produces infected red blood cells that have a 'smooth' surface phenotype\textsuperscript{37}. For experimental ease and generalizability with which new loci can be targeted, we constructed two base plasmids (Figure 3.10). The first (pCas9-gRNA\textsuperscript{T}) delivers \textit{SpCas9} and the gRNA\textsuperscript{T} targeting a specified locus. The second (pT7 RNAP-HR) delivers T7 RNAP and encodes a homologous region to repair the \textit{SpCas9/gRNA}\textsuperscript{T}-induced DNA double strand break. We designed the homologous region such that successful chromosomal editing results in an in-frame transcriptional fusion of a 2A-like peptide from the \textit{Thosea asigna} virus (T2A) to the \textit{Renilla} luciferase coding sequence\textsuperscript{36} with the upstream target gene fragment. A stop codon is included at the end of the \textit{Renilla} gene to terminate translation. Thus, successful editing is expected to result in \textit{Renilla} luciferase expression if the targeted gene's promoter is transcriptionally active.
Figure 3.10: Vector design for kahrp knockout

Generalized schematic of the plasmids used to achieve genome editing. \( p\text{Cas}9\)-gRNA\(^T\) encodes \( \text{SpCas}9\) and the \( \text{bsd} \) selection marker under the control of the \( \text{PcDT} \) 5'--\( \text{Pfh}2\) 3'-- and \( \text{PfCam} \) 5'--\( \text{Pfhs}86 \) 3'--UTR pairs, respectively. The gRNA\(^T\) is expressed from a T7 promoter-driven cassette. For homology directed repair of the induced double strand break, pT7 RNAP is modified to include homologous regions flanking a T2A-RLuc gene. Successful repair of the induced strand break eliminates the original gRNA target site, and creates a translational fusion between the upstream fragment of the disrupted target and the T2A-RLuc genes. Consequently, RLuc is placed under the control of the targeted gene’s promoter.

We initially created three versions of \( p\text{Cas}9\)-gRNA\(^T\) from which either no gRNA (\( p\text{Cas}9\)-No gRNA\(^T\)), \( p\text{UC}19\)-gRNA\(^T\) (\( p\text{Cas}9\)-pUC19 gRNA\(^T\)) or \( \text{kahrp} \)-gRNA\(^T\) (\( p\text{Cas}9\)-kahrp gRNA\(^T\)) is expressed. We separately co-transfected these plasmids with pT7 RNAP-HR\(^{kahrp}\) and continuously selected for episomal retention of both over the typical period.
(~ 4-6 weeks) required for obtaining easily manipulated cultures (≥ 1% parasitemia). We used either a 3D7™ parasite line with a reference firefly luciferase gene site-specifically integrated at the cg6 locus or a parental 3D7 line (knob positive) in our experiments.

We monitored Renilla and firefly luciferase levels periodically during the course of transfection as a simple readout of successful editing within the parasite population. On post-transfection day 33, we detected a substantial increase in relative Renilla to firefly luciferase signal for parasites transfected with kahrp-gRNA, but not the control pUC19-gRNA (Figure 3.11).

![Figure 3.11: Luciferase levels of kahrp knockout strain at the population level](image)

Luciferase assays were performed with the Promega Dual luciferase assay kit with RLuc levels normalized to FLuc levels. RLuc expression is observed when the kahrp locus is targeted for cleavage and repair by a kahrp-gRNA and a suitable donor plasmid. A similar increase in RLuc signal is not observed when a non-targeting pUC19-gRNA control is used.

We used PCR to analyze isolated genomic DNA to determine whether the kahrp gene had been disrupted by insertion of the T2A-RLuc cassette in the parasite population expressing kahrp-gRNA, but not the pUC19-gRNA. Primer pairs, p1/p2 and p3/p4,
designed to detect repair upstream and downstream of the cut site, respectively, yielded diagnostic PCR products only in the case for kahrp-gRNA\textsuperscript{T}, but not pUC19-gRNA\textsuperscript{T} transfections (Figure 3.12).

![Diagram](image)

**Figure 3.12:** PCR confirmation of kahrp knockout at the population level

PCR primers to specifically detect homology-directed repair at a target cut site in the kahrp locus amplify products of the expected size for kahrp-gRNA\textsuperscript{T} but not pUC19-gRNA\textsuperscript{T} transfected parasite population.

Next, we performed Southern blot analysis to confirm that the targeted locus had been modified as expected, and to estimate the frequency of editing within the parasite.
population. These data indicated that virtually the entire parasite population transfected with *kahrp*-gRNA\textsuperscript{T}/pT7 RNAP-HR\textsuperscript{*kahrp} had been successfully edited, while in *pUC19*-gRNA\textsuperscript{T}/pT7 RNAP-HR\textsuperscript{*kahrp}-transfected parasites, the native *kahrp* locus remained intact (Figure 3.13).

![Southern blot confirmation of *kahrp* knockout at the population level](image)

**Figure 3.13:** Southern blot confirmation of *kahrp* knockout at the population level

Southern blot analysis of a *kahrp*-gRNA\textsuperscript{T} transfected population indicates that the native locus is undetectable, and only a band corresponding to the edited locus is present. Only the intact native locus is detectable in the *pUC19*-gRNA\textsuperscript{T}-transfected population.

Consistent with these findings, we were unable to detect KAHRP protein expression by Western blot analysis of the *kahrp*-gRNA\textsuperscript{T} transfected parasite population. In contrast, we readily detected KAHRP in the *pUC19*-gRNA\textsuperscript{T} control parasites (Figure 3.14).
Figure 3.14: Western blot confirmation of *kahrp* knockout at the population level

Western blot analysis confirms that KAHRP protein is undetectable in the *kahrp*-gRNA$^T$ transfected population, but present in the *pUC19*-gRNA$^T$ transfected population.

In a parallel and independent experiment, we examined editing induced by the *kahrp*-gRNA$^T$ compared to a no gRNA$^T$ control in a parental 3D7 strain. Beginning on day 20, RLuc levels steadily increased above background, and became elevated in parasites transfected with *kahrp*-gRNA$^T$ but not the no gRNA$^T$ control (Figure 3.15a). As before, PCR analysis revealed editing at the *kahrp* locus only in the *kahrp*-gRNA$^T$ transfection, and Western blot confirmed that virtually no KAHRP protein was expressed at the population level (Figure 3.15b).
Figure 3.15: Confirmation of kahrp knockout in WT 3D7 strain at the population level

In an independent co-transfection of 3D7 parasites with pT7 RNAP-HR\textsuperscript{kahrp} and pCas9-kahrp gRNA\textsuperscript{T}, editing of the kahrp locus is observed. (a) Renilla expression levels gradually increase over time for the kahrp-gRNA\textsuperscript{T} transfection, but not a control transfection in which no gRNA\textsuperscript{T} is produced (pT7 RNAP-HR\textsuperscript{kahrp} and pCas9-No gRNA\textsuperscript{T} plasmids). (b) PCR and Western blot analyses confirm the expected editing event that leads to disruption of the kahrp gene in the kahrp-gRNA\textsuperscript{T} transfection, but not the no gRNA\textsuperscript{T} control.

Additionally, using scanning electron microscopy (SEM), we confirmed that parasites transfected with the no gRNA control plasmid retained the ‘knobby’ phenotype associated with an intact kahrp gene. However, the kahrp-gRNA\textsuperscript{T}-transfected parasites appeared ‘smooth’, as expected for kahrp-null parasites (Figure 3.16).
Finally, we analyzed four clones obtained by limiting dilution from the kahrp-gRNA\textsuperscript{T} transfected parasite pool by PCR and Western blot. The data confirmed that these all had their kahrp gene disrupted, and did not express KAHRP protein (Figure 3.17).
Figure 3.17: PCR and western blot confirmation of kahrp knockout clones
Confirmation PCR (A) and Western blot (B) analyses as in Figure 3.12 and Figure 3.14, respectively, of cloned parasites derived from the kahrp-gRNA\(^T\) edited pool. The unedited pUC9-gRNA\(^T\) pool with an intact native kahrp locus is used as a control.

3.4.4 CRISPR mediated knockout of eba-175

Next, we examined SpCas9-mediated editing at a second, unrelated genomic locus towards establishing the broader utility of this approach in the parasite. We selected the eba-175 gene on Chromosome 7, which encodes a parasite ligand used during invasion of red blood cells\(^{107}\). We constructed ebal75-gRNA\(^T\) expression (pCas9-eba175 gRNA\(^T\))
and donor (pT7 RNAP-HR\textsuperscript{eba175}) plasmids (Figure 3.18) and transfected these as described before.

Figure 3.18: Schematic of \textit{eba-175} CRISPR mediated knockout vectors
Schematic of the plasmids used for genome editing. \textit{pCas9-eba175 gRNA\textsuperscript{T}} and \textit{pCas9-pUC19 gRNA\textsuperscript{T}} encode \textit{SpCas9} and the \textit{bsd} selection marker as described in Figure 3.10, and produce the \textit{eba175 (test)} and \textit{pUC19 (control) gRNA\textsuperscript{T}} respectively, from a T7 promoter-driven cassette.

To evaluate editing, we isolated genomic DNA from parasites transfected in parallel with \textit{pCas9-eba175 gRNA\textsuperscript{T}/pT7 RNAP-HR\textsuperscript{eba175} (test)} and \textit{pCas9-pUC19 gRNA\textsuperscript{T}/pT7 RNAP-HR\textsuperscript{eba175} (control)}. Using PCR and sequencing of the resulting products, we determined that the expected editing events upstream and downstream of the induced cut site had both occurred in the \textit{eba175-gRNA\textsuperscript{T}}, but not the control \textit{pUC19-gRNA\textsuperscript{T}}, expressing parasite population (Figure 3.19). Through Southern blot analysis, we again confirmed that the targeted locus was modified as expected (Figure 3.19). Furthermore, these data
suggested that ~50% of parasites within the population had been successfully edited, further supporting the high efficiency of the CRISPR/Cas9 system in *P. falciparum.

Figure 3.19: Population level data for the CRISPR mediated knockout of *eba-175*

The *eba175* locus and gRNA target site are illustrated, together with the expected editing outcome and location of diagnostic PCR primers to assess locus modification (not drawn to scale). PCR analysis reveals the presence of the expected editing outcome in the parasite population transfected with *eba175*-gRNA, but not *pUC19*-gRNA. Southern blot analysis of the parasite population transfected with *eba175*-gRNA shows that ~50% parasites have an edited locus. No evidence of editing is observed in the *pUC19*-gRNA transfected population.

3.4.5 Assessing off target effects of CRISPR/Cas9 in *P. falciparum*
The potential for unintended gene disruptions due to induced off-target strand breaks and repair by the error-prone non-homologous end joining (NHEJ) mechanism has been described in human cells\textsuperscript{108-111}. This has led to exploration of strategies to mitigate these undesirable outcomes\textsuperscript{112,113}. Therefore, we sought preliminary insight into how significant such off-target events could be in \textit{P. falciparum}. Notably, bioinformatics analyses revealed that \textit{P. falciparum} lacks canonical NHEJ components\textsuperscript{8,114}. Additionally, a recent study examining chromosomal double strand breaks induced by the meganuclease I-Sce1 showed that these are very efficiently and exclusively repaired by homologous donor sequence when present. In the absence of a suitable donor, however, an NHEJ-like repair process of unknown mechanism that resulted in elimination of the I-Sce1 target site was observed\textsuperscript{115}. To understand how an off-target SpCas9/gRNA\textsuperscript{T}-induced chromosomal double strand break might be processed in \textit{P. falciparum}, we simulated such an event by expressing the \textit{kahrp}-gRNA\textsuperscript{T} used earlier to mediate efficient editing of the \textit{kahrp} locus, but in the absence of a suitable homologous donor plasmid [pCas9-\textit{kahrp} \textit{gRNA}\textsuperscript{T}/pT7 RNAP]. Control parasites expressing \textit{pUC19}-gRNA\textsuperscript{T} [pCas9-\textit{pUC19} \textit{gRNA}\textsuperscript{T}/pT7 RNAP] were transfected in parallel (Figure 3.20a). In two independent experiments, we observed no gross defects in relative growth between the two parasite lines generated, as both reached working parasitemia levels at similar times post-transfection, and qualitatively expanded comparably thereafter. To determine whether NHEJ-like events had occurred within the \textit{kahrp} region targeted for cleavage, we isolated total genomic DNA from both pCas9-\textit{kahrp} \textit{gRNA}\textsuperscript{T}/pT7 RNAP lines, and sequenced twenty independent clones from each. In all cases, the \textit{gRNA}\textsuperscript{T} target site remained intact (Figure 3.20b), in contrast to disruption of the I-Sce1 site that is seen
after meganuclease-mediated cleavage and repair\textsuperscript{115}. These data suggest that NHEJ-like events occurring at \textit{SpCas9/kahrp-gRNA}\textsuperscript{T}-induced cleavage sites are likely to be infrequent. Presently, we do not understand why cleavage induced by \textit{SpCas9/gRNA} versus I-Sce1 meganuclease in the absence of a suitable donor sequence produces different repair outcomes. However, discrepancies in how double strand breaks induced by various nuclease platforms, including zinc finger nucleases, TALENs and I-Sce1, are repaired have been described\textsuperscript{116}. Understanding the basis for these differences in repair is an active research area given the implications for improving genome-editing efficiency.

With respect to applications in \textit{P. falciparum}, our findings taken together with previous genome editing studies using zinc finger nucleases\textsuperscript{29} and I-Sce1\textsuperscript{115} suggest that homologous repair events rather than potentially deleterious off-target NHEJ-like outcomes should be strongly favored during CRISPR/Cas9-mediated genome editing.
**Figure 3.20:** Assessing the fate of kahrp-gRNA$^T$-induced cleavage of the kahrp locus in the absence of a homologous donor plasmid

**A.** Schematic illustration of the plasmids used in this experiment. pT7 RNAP and the pCas9- kahrp gRNA$^T$ plasmids are exactly those previously described in Figure 3.2 and Figure 3.10. **B.** The region of kahrp targeted for cleavage by kahrp-gRNA$^T$ is identical in sequence to the native locus in twenty bacterial colonies analyzed. The black arrow indicates the predicted cut site.
3.5 Discussion

Altogether, we demonstrate that CRISPR/Cas9 components combined with gRNAs produced in situ by T7 RNA polymerase can be used to substantially increase the efficiency and technical ease with which genes can be deleted in *P. falciparum*. Indeed, with donor plasmids of a design routinely used for double crossover-mediated knockouts, we found no PCR-detectable integration events in the absence of gRNAs directed to two distinct loci. However, within an identical timeframe, between ~50-100% of parasites expressing gRNAs targeting these loci were edited as expected. We show that continuously selecting for the plasmids used to express the CRISPR/Cas9 and T7 RNA polymerase components induces editing in a sufficiently large proportion of the parasite population to facilitate directly proceeding to isolating edited clones by limiting dilution. These findings promise to substantially accelerate the process for achieving targeted gene disruptions in *P. falciparum*, as it eliminates the very protracted and cumbersome selection drug cycling protocols needed to enrich the extremely rare recombination events achieved with commonly used unassisted single- and double- crossover approaches. Therefore, the ease with which the CRISPR/Cas9-T7 RNA polymerase system can be implemented in *P. falciparum* and programmed to target virtually any gene of interest makes it a highly attractive strategy for rapidly and efficiently conducting functional genetics studies in *P. falciparum*.

3.6 Material and Methods

3.6.1 Molecular Biology
All plasmids used in this study were assembled using previously described methods. Restriction enzymes and Gibson assembly master mix were purchased from New England Biolabs.

3.6.2 Parasite culture and transfection

*P. falciparum* strain 3D7 parasites were grown under 5% O₂ and 5% CO₂ in RPMI-1640 media supplemented with 5 g/L Albumax II (Life Technologies), 2 g/L sodium bicarbonate, 25 mM HEPES pH 7.4 (pH adjusted with potassium hydroxide), 1 mM hypoxanthine and 50 mg/L gentamicin. For strains containing plasmids, appropriate selection drug combinations based on the markers used were added to media as follows: 2.5 mg/L blasticidin S and/or 250 mg/L G418 (Research Products International). Single and double vector transfections were carried out by the spontaneous uptake method using ~50 µg of maxi-prepped plasmid DNA, and 8 square wave electroporation pulses of 365 V for 1 ms each, separated by 0.1 seconds. Drug selection was initiated 4 days post transfection.

3.6.3 Assessing T7 RNAP expression in *P. falciparum*

To establish production of T7 RNAP protein, schizont stage parasite infected red blood cells transfected with *pT7 RNAP* were saponin-lysed, and the lysate components separated by SDS-PAGE. Western blot analysis was carried out using a mouse anti-T7 RNAP monoclonal antibody (Novagen) at a 1:5,000 dilution in Tris-buffered saline, 0.1% Tween 20 and 5% milk. Secondary antibody treatment was performed with HRP.
conjugated goat anti-mouse (Novagen) at a dilution of 1:5000. Development was carried out with the Supersignal west femto development kit (Pierce).

3.6.4 Assessing T7 RNAP activity in P. falciparum

To assess T7 RNAP enzymatic activity, 3D7 parasites were transfected with two vectors simultaneously according to the protocol listed above, using ~50ug for each vector. The base plasmid pfGNr (MRA-462; www.mr4.org) was used for the no T7 RNAP control cases. Total RNA for quantitative RT-PCR and Northern blot analysis was isolated from saponin-lysed schizont stage parasites. Pellets were stored in liquid nitrogen pending RNA extraction. RNA was extracted using Trizol (Life Technologies) and further purified using RNeasy purification kits (Qiagen).

For quantitative real time PCR, extracted RNA was treated with Turbo DNAse kit (Ambion#AM1907) and reverse transcribed to cDNA using the SuperScript III First Strand Synthesis System (Invitrogen #18080-051). cDNA was quantified using gene specific primers (Table 3.1 and SYBR Green on a LightCycler 480 Instrument II (Roche Applied Science). The thermocycling program used was as follows: Initial denaturation at 95°C for 5 min; followed by 45 cycles of 95°C for 30 s, 60°C for 30 s, and 72°C for 30s; followed by a melting curve analysis. cDNA levels were quantified relative to standard curves generated using authentic plasmid templates.

Northern blot analysis was performed using the TurboBlotter kit (Whatman) for transfer, and the North2South kit (Thermo Scientific) for development. Authentic RNA standards were synthesized by in vitro transcription using the T7 Megascript kit (Ambion).
3.6.5 In vitro SpCas9 cleavage assays

*In vitro* cleavage assays were performed as previously described. Briefly, HEK 293FT cells were transfected with the SpCas9 expression plasmid pX165 in 6-well plates using Lipofectamine 2000 (Life Technologies) according to the manufacturer’s protocol. Forty eight hours post-transfection, cells were lysed with 250 μl Lysis Buffer (20 mM HEPES pH 7.5, 100 mM potassium chloride, 5 mM magnesium chloride, 1 mM dithiothreitol, 5% glycerol, 0.1% Triton X-100) supplemented with a protease inhibitor cocktail (Roche). Lysates were sonicated for 10 minutes and cell debris pelleted by centrifugation for 20 min at 5,000 g. Lysates containing SpCas9 protein were aliquoted and stored at -80 °C.

gRNAs were *in vitro* transcribed using the T7 Megascript kit (Ambion) using a circular plasmid template containing a cassette for T7 RNAP dependent expression of the specific gRNA as a template. The target DNA was amplified from *P. falciparum* genomic DNA with primers specific to the gene being tested (Table 3.1).

3.6.6 CRISPR/Cas9 genome editing in *P. falciparum*

Experiments were performed either in WT 3D7, 3D7:AttB:FLuc, or NF54:AttB. Either a no gRNA control or a control with a guide sequence designed to cut the pUC plasmid in the open reading frame of the ampicillin resistance gene but not present in the *P. falciparum* genome was used as a negative control. ~50ug of each plasmid were transfected into the desired target strain simultaneously with drug pressure applied four days post transfection. Luciferase levels were measured periodically to determine if genomic insertion had occurred in the case targeting the *kahrp* gene. Parasites could be
visualized in the expected time course for a standard transfection (~4-6 weeks) with most analyses being performed over the next several weeks.

3.6.7 Analysis of CRISPR-edited parasite lines

Firefly and Renilla luciferase levels were measured using the Dual-Luciferase Reporter Assay System (Promega). Infected red blood cells were lysed using passive lysis buffer supplied with the kit, and measurements made according to the manufacturer’s instructions on a 20/20" luminometer (Turner Biosystems).

PCR analyses to assess modification of the targeted loci were carried out using a 15:1 (v:v) mixture of Hemo KlenTaq:Pfu Turbo (Agilent) in Hemo KlenTaq Buffer on genomic DNA purified from cultures using the QIAamp DNA blood mini kit (Qiagen) after saponin lysis. Primers used for confirmation can be found in Table 3.1.

Samples for KAHRP western blot analysis were obtained by repeated hypotonic lysis of infected red blood cells in water, followed by high-speed centrifugation (21000 g for 1 min). The membrane fraction was recovered, solubilized in 1x SDS loading buffer, and separated by SDS-PAGE on For KAHRP detection, a mouse monoclonal anti-KAHRP antibody (mAb 89) provided by Diane Taylor was used at a 1:1,000 dilution in combination with a secondary antibody treatment of HRP conjugated goat anti-mouse (Novagen) at a dilution of 1:5000. Development was carried out with the Supersignal west femto development kit (Pierce).

Southern blots were carried out on DNA purified from infected red blood cells using the QIAamp DNA blood mini kit (Qiagen) after saponin lysis. Samples were
restriction enzyme digested overnight with HindII/BamHI/PstI and HindIII/XbaI for analysis of the \textit{kahrp} and \textit{eba175} loci, respectively. Blots were processed using the TurboBlotter kit (Whatman) for transfer, and the North2South kit (Thermo Scientific) for development.

Scanning electron microscopy (SEM) was performed as described by Rug et al.\textsuperscript{117} on a JEOL 5600LV SEM instrument operated at 10 kV.

<table>
<thead>
<tr>
<th>Table 3.5: Primers used in this study</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>KAHRP confirmation (5'-&gt;3')</strong></td>
</tr>
<tr>
<td>P1</td>
</tr>
<tr>
<td>P2</td>
</tr>
<tr>
<td>P3</td>
</tr>
<tr>
<td>P4</td>
</tr>
<tr>
<td><strong>EBA175 confirmation (5'-&gt;3')</strong></td>
</tr>
<tr>
<td>P1</td>
</tr>
<tr>
<td>P2</td>
</tr>
<tr>
<td>P3</td>
</tr>
<tr>
<td>P4</td>
</tr>
<tr>
<td><strong>TrxR control</strong></td>
</tr>
<tr>
<td>F</td>
</tr>
<tr>
<td>R</td>
</tr>
<tr>
<td><strong>KAHRP amplification for in vitro assay</strong></td>
</tr>
<tr>
<td>F</td>
</tr>
<tr>
<td>R</td>
</tr>
<tr>
<td><strong>EBA175 amplification for in vitro assay</strong></td>
</tr>
<tr>
<td>F</td>
</tr>
<tr>
<td>R</td>
</tr>
<tr>
<td><strong>BSD qPCR primers</strong></td>
</tr>
<tr>
<td>BSD qPCR</td>
</tr>
<tr>
<td>BSD qPCR F</td>
</tr>
<tr>
<td>BSD qPCR R</td>
</tr>
<tr>
<td>RLuc qPCR primers</td>
</tr>
<tr>
<td>------------------</td>
</tr>
<tr>
<td>RLuc qPCR F</td>
</tr>
<tr>
<td>RLuc qPCR R</td>
</tr>
</tbody>
</table>
Chapter 4: Inducible Transcription of Non-Coding RNA in *P. falciparum* Mediated by the T7 RNA Polymerase and the Lac Repressor system

4.1 Abstract

There is increasing interest in the role of non-coding RNA (ncRNA) in the human malaria parasite *Plasmodium falciparum*. In order to elucidate the role of ncRNA in the parasite, a system for inducible expression of ncRNA is necessary. At this point, no native genetic switches in the parasite have been identified that would be amenable to such an inducible system. Therefore, using an orthogonal system for the production of ncRNA in the parasite is attractive.

Previous work in our lab has shown that the well studied T7 RNA Polymerase (T7 RNAP) is suitable for the production of ncRNA in *P. falciparum*. T7 RNAP has been used in conjunction with several regulatory systems for the inducible production of RNA and proteins in a variety of organisms, with the use of the lac repressor system being the most prevalent. Here we present the adaptation of this system to inducible ncRNA production in *P. falciparum*.

4.2 Introduction

Despite accounting for over 500,000 deaths each year, much about that parasites that cause malaria remains a mystery. Challenges faced by those who wish to study malaria parasites include difficult (or non existent) culturing conditions, inefficient methods of DNA delivery, and the lack of molecular tools for the study of DNA, RNA, and proteins.
Of the parasites responsible for human malaria, *P. falciparum* is responsible for the majority of morbidity and mortality. The development of *in situ* culture methods allows the maintenance of *P. falciparum* in a laboratory setting\textsuperscript{18} and electroporation methods have allowed for the uptake of foreign DNA\textsuperscript{18,19}. While past work has allowed for the expression of recombinant proteins in the parasite\textsuperscript{23}, to date there exists no system for the production of non-coding RNA within the parasite.

Non-coding RNA (ncRNA) is of increasing interest in *P. falciparum*. Recent studies have demonstrated the presence of long non-coding RNAs (lncRNAs) that are thought to play a role in chromosomal maintenance as well as potentially gene regulation\textsuperscript{13,14,17}. In addition ncRNA has been implicated in the process of antigenic selection and switching by which *P. falciparum* will heritably produce and display only one out of \(~60\) different antigens on the surface of the infected red blood cell\textsuperscript{12,15,16,119}. Finally, it has been suggested that antisense based RNA mechanisms can be used for genetic regulation in the parasite, although this remains controversial\textsuperscript{112,120}. Elucidation of these processes may lead to new drug targets and vaccination strategies.

Recent work in our lab has established the utility of the T7 RNA polymerase (T7 RNAP) for the production of ncRNA in *P. falciparum* (see Chapter 3). While T7 RNAP originates in a bacteriophage of *Escherichia coli* and has been established as efficacious for protein production in this host\textsuperscript{60}, it has also been established to be functional in several eukaryotic systems\textsuperscript{65,67,68}. Notably, T7 RNAP lacks the capping machinery generally required by most eukaryotes in order to mediate ribosomal binding to a target mRNA, which absent intervention of another system results in the production of non-coding RNA which can interact with most RNA processing machinery\textsuperscript{66}.
In addition, several regulatory technologies exist for controlling the T7 RNAP mediated production of RNA\textsuperscript{61,121}. Of these systems, the most ubiquitous is the use of the well-studied lac repression system in \textit{E. coli} which utilizes the lac repressor, LacI\textsuperscript{61}. This system depends on the galactose analog Isopropyl \(\beta\)-D-1-thiogalactopyranoside (IPTG) to induce transcription by binding to LacI and preventing it's interaction with the canonical lac operator (lacO)\textsuperscript{122}. A diagram of this interaction is given in Figure 1.3. This work has been expanded to show functionality of the lac repressor system in eukaryotic organisms\textsuperscript{123}. The closely related T3 RNA polymerase has also been used in conjunction with the lac repressor system to regulate RNA production in mammalian cells\textsuperscript{124}.

Here, we present the first demonstration of the functionality of the T7 RNAP and lac repressor system for the controlled production of RNA in \textit{P. falciparum}. RNA is produced from a canonical T7 promoter/lacO combination and is able to be regulated 10-20 fold upon induction with IPTG. This work lays the foundation for further studies into non-coding RNA in the parasite.

4.3 Results

4.3.1 IPTG Toxicity in \textit{P. falciparum}

In order for IPTG to act as a suitable inducer molecule for the LacI repressed T7 system in \textit{P. falciparum}, toxicity effects on the parasite needed to be established. To test this, a simple expansion assay was performed at a range of IPTG concentrations (Figure 4.1). The expansion assay showed no growth defects up to a concentration of 10mM, the highest concentration of IPTG tested. This indicates that IPTG may be a suitable inducer molecule for use in \textit{P. falciparum}.
Figure 4.1: IPTG toxicity in *P. falciparum*

*P. falciparum* expansion assay at a range of IPTG concentrations. Parasitemia was normalized to an expansion of *P. falciparum* containing no IPTG. Error bars represent a standard of deviation for three biological replicates.

### 4.3.2 Vector design and transfection

In order to test the ability of LacI to modulate the production of RNA from a T7 promoter, a dual vector system was set up (Figure 4.2). The first vector contains the T7 RNAP while the second vector contains LacI and the T7 expression cassette. Both vectors were transfected into the parasite simultaneously with parasite appearing on the time scale of a normal transfection (3-4 weeks).
4.3.3 RNA induction by the addition of IPTG

As it had been previously demonstrated that T7 RNAP was capable of driving RNA production in situ (See Chapter 3) we sought to determine if we could regulate the production of RNA using the lac repression system.

Cultures were then synchronized and split into +/- 1mM IPTG flasks and maintained for two invasion cycles (approximately 96 hours). RNA was then harvested and the levels of RLuc RNA was measured by quantitative real-time PCR (qPCR) and normalized to the levels of BSD RNA (Figure 4.3).

The results clearly showed a >10 fold increase in the RNA levels in the presence of IPTG, indicating that the Lac repressor is able to regulate RNA production in situ. To further probe the mechanism, we deleted the lac operator and showed that IPTG no longer had an effect on the levels of RNA produced, further suggesting that the system was functioning as it has been shown to function in other organisms (Figure 4.3).
Figure 4.3: Normalized levels of Rluc RNA +/- IPTG
Quantitative PCR results for regulated production of RNA from a T7 promoter in the presence or absence of 1mM IPTG. Presence and absence of various components of the system are noted by + or - below the graph. White bars indicate the -IPTG condition while black bars indicate the +IPTG condition. Error bars represent the standard deviation of three technical replicates. This graph represents typical results.

Interestingly, while IPTG had no effect on the levels of RNA in the ΔLacO case, the total levels of RNA were approximately 10 fold less than the levels of RNA observed in the case with the lac operator and the presence of IPTG. This suggests that the lacO is essential for the regulated production of RNA from the T7 promoter but also enhances the overall RNA levels in the cell.

4.3.4 Sensitivity of the Lac repression system to IPTG

We next sought to determine if the 1mM concentration that we had chosen represented the maximally induced state for RNA production. To test this, various
concentrations of IPTG up to 5mM were incubated for two life cycles with the parasite. RNA was harvested and quantified as before (Figure 4.4). The results indicate that 1mM is the maximally induced state and that lower concentrations of IPTG are also capable of achieving the same level of induction.

![Figure 4.4: Sensitivity of RNA production to IPTG concentration](image)

RNA was assayed at varying concentrations of IPTG. Fold induction was calculated as the normalized RNA level at a specific concentration divided by the normalized RNA concentration with no IPTG added. Levels of BSD were used as the normalizing factor in each case as before. Measurements were also performed in a strain lacking the T7 RNAP to confirm the absence of RNA induction at the maximal concentration of IPTG as well as the no IPTG condition. Error bars represent standard of deviation of three technical replicates.

**4.3.5 Time scale of RNA production upon induction in P. falciparum**

Finally, we sought to determine the rate at which induction occurs in the parasite upon the addition of IPTG to the media. For this, a time course was carried out with RNA harvested at various points after the addition of 1mM IPTG to late trophozoites/early schizonts. We saw that the induction began to occur as short as 20 minutes after addition of the media, with maximal levels of induction achieved (based on the 48 hour induction experiments in Figure 4.3) by 8 hours post induction (Figure 4.5).
Figure 4.5: RNA Induction time course after the addition of IPTG
IPTG was added to synchronized cultures with periodic harvesting over the next 8 hours. Error bars represent standard of deviation of three technical replicates. Fold induction was calculated as the normalized RNA level at a specific concentration divided by the normalized RNA concentration with no IPTG added at the initial time point. Levels of BSD were used as the normalizing factor in each case as before. –T7 RNAP indicates a culture without T7 RNAP to confirm the lack of RNA production over time.

4.4 Discussion

The work presented here establishes that the lac repressor system can be used to regulate the production of RNA from a T7 promoter. The induction is both sensitive and rapid upon the addition of IPTG. Previously, we established that the RNA is not translated (Figure 3.8), making this an ideal system for the study of non-coding RNA.

Interestingly, the lac operator seemed to enhance the overall levels of T7 RNAP generated RNA present in the parasite. While this could be due to differences in the efficiency of initiation of the T7 RNAP, the ΔLacO construct preserves the GGG initiation triplet that has been established as being sufficient for efficient T7 RNAP initiation\(^5^8\). Instead, we believe it possible that the lac operator forms a stem loop that
offers the RNA some protection from 5' degradation by native *P. falciparum* RNAses.

Computational prediction of RNA structure shows a strong stem loop ($\Delta G = -15.7$ kcal/mol) formed by the lac operator in the context of the RLuc RNA (Figure 4.4).

![Figure 4.6: mFold algorithm RNA folding of LacO](image)

Results of the mFold algorithm on the folding of the RNA formed by the combination of the T7 promoter and LacO. The initial GG required for efficient initiation of T7 RNAP is highlighted in blue. Yellow indicates the bases which make up LacO.

It is possible that this stem loops confers stability similar to the stability given by a 5'-7'-methylguanosine cap$^{62,63}$. It is worth noting that the system in the repressed state still does produce a significant quantity of RNA. This could be modulated by either adding an additional lacO site which could provide additional regulation, or by mutating the T7 promoter to
modulate the amount of RNA produced to the desired range\textsuperscript{58,58}. These modifications would likely allow the regulated production of RNA across a range of concentrations, thus making this technology broadly suitable for use in studies of ncRNA in \textit{P. falciparum}.

4.5 Material and Methods

4.5.1 Molecular Biology

Vectors were constructed using either yeast homologous recombination\textsuperscript{76,97} or Gibson assembly\textsuperscript{77} utilizing the commercially available master mix (NEB #E2611S). The base vector used for construction of additional vectors was the pfGNr vector (MR4 #MRA-462, www.mr4.org).

4.5.2 Strain Maintenance and Transfection

\textit{P. falciparum} strain 3D7 was maintained under 5% O\textsubscript{2} and 5% CO\textsubscript{2} in RPMI-1640 media supplemented with 5 g/L Albumax II (Life Technologies), 2 g/L NaHCO\textsubscript{3}, 25 mM HEPES-K pH 7.4, 1 mM hypoxanthine and 50 mg/L gentamicin. For strains containing vector(s) appropriate drug combinations based on selectable markers were also added to the media (2.5 mg/L blasticidin S, 250 mg/L G418 (RPI)). The media was supplemented with IPTG when necessary as indicated.

Single vector transfections were carried out by the spontaneous uptake method\textsuperscript{19} utilizing ~50ug of maxi prepped plasmid DNA and 8 square wave electroporation pulses of 365V for 1ms each, separated by 0.1 seconds. Double vector transfections were carried
out identically with ~50ug or each plasmid in the transfection mixture. Drug selection was applied 4 days post transfection.

4.5.3 RNA extraction from *P. falciparum*

Parasites were saponin lysed in the schizont stage unless otherwise noted and stored in Liquid nitrogen until RNA extraction. RNA extraction was carried out using the Trizol system for parasite lysis (Life Technologies #15596-026) in conjunction with the RNeasy extraction kit for RNA purification (Qiagen #74104).

4.5.4 Quantitative Real-time PCR

Extracted RNA was treated with Turbo DNase kit (Ambion#AM1907) and reverse transcribed to cDNA using the SuperScript III First Strand Synthesis System (Invitrogen #18080-051). cDNA was quantified using gene specific primers and SYBR green dye on the Roche 48- LightCycler instrument. Primer sets were compared to a standard curve generated for each set simultaneous to cDNA measurement to correct for amplification efficiency of different primer sets.

4.5.5 *P. falciparum* Expansion Assays

Synchronized cultures were fixed >3 hours in 1% formaldehyde Acid Citrate Dextrose media. Fractions of cultures were then stained with SYBR Green and measured on an Accurri C6 flow cytometer set to measure 100000 events. Samples were gated for blood cells based on scatter and parasitemia was calculated by percentage of the population positive for SYBR green fluorescence.
4.5.6  *In silico folding of RNA*

*In silico* RNA folding was performed with the mfold algorithm online server (http://mfold.rna.albany.edu/) using default settings\textsuperscript{125}.
Chapter 5: Conclusions and Future Directions

5.1 Summary of this work

This work has introduced several new molecular tools for use in the highly clinically relevant parasite Plasmodium falciparum. First, I present the creation of the first vector family for use in P. falciparum containing all currently used selectable markers and replication elements, and show that they can be used not only with standard cloning, but with yeast homologous recombination cloning and Gibson assembly. In addition I probe the use of the Thosea asigna virus 2A (T2A)-like viral peptide for use in localization of multiple proteins expressed from the same expression cassette.

Next I present the first demonstration of using the CRISPR/SpCas9 system for genomic editing in any non-bacterial infectious disease. In the process of demonstrating this, I also introduce the first example of using T7 RNAP for directed transcription of non-coding RNA (ncRNA) in any apicomplexan parasite.

Finally, I show that the lac repression system can be used to regulate T7 RNAP directed RNA production in P. falciparum, generating the first inducible transcriptional system for the production of non-coding RNA in the parasite. Overall, this work develops several new molecular tools that will greatly aid in the studies of P. falciparum and allow for the uncovering of new drug targets and potential vaccination strategies.

5.2 Future directions in the construction of genetic elements for use in P. falciparum

As was presented here, there is now a modular family of vectors for use in P. falciparum augmenting the other vectors currently in use in the parasite. While the
modularity and the use of Gibson assembly and yeast gap repair to build these vectors has allowed the process to be streamlined significantly, assembly is still prone to rearrangements and deletions as part of the assembly. This is likely due to the *E. coli* host having issue replicating such A+T rich regions and selecting for plasmids that are able to eliminate the *P. falciparum* expression cassettes. It appears to be somewhat strain specific\textsuperscript{24}, but the mechanism of this action is unknown and discovery of the true reason behind these rearrangements would further simplify the process of vector construction for use in the parasite.

Incorporating the use of other assembly technologies would also be a useful direction to pursue. This work demonstrates the lack of a single technology for use in all vector assembly situations and by no means represents a survey of all existing assembly technologies. For example, modifying the vectors for use with the Gateway® cloning system\textsuperscript{126}, which allows for highly efficient site-specific recombination to be used for vector assembly, could prove highly useful. As new technologies for cloning are introduced, it will be important to test them out for *P. falciparum* vector assembly to further develop what is usable and what isn’t.

The other major issue to be addressed is size limitations in the plasmids. As noted, the expression cassettes used in *P. falciparum* vectors are poorly defined and tend to encompass large, A+T rich regions. As interest grows in expressing multiple genes of *P. falciparum* simultaneously, these constructs begin to approach the traditional size limitations of replicating plasmids in bacteria\textsuperscript{127}. Ongoing work is needed to allow for the assembly of larger constructs for use in the parasite in bacteria. Two approaches to this are currently being developed in our lab. The first involves the use of bacterial artificial
chromosomes to allow for larger constructs and thus, more *P. falciparum* expression cassettes. The second involves the use of linear vectors that are shown to be more amenable to cloning of large fragments in bacteria\textsuperscript{128}. These constructs can then be circularized for transfection into *P. falciparum*\textsuperscript{129}. In both cases, expanding the theoretical limit to vector size will allow for more complex experiments.

**5.3 Future application space for the *P. falciparum* CRISPR/SpCas9 genome editing technology**

While this work establishes the functionality of the CRISPR/Cas9 system in *P. falciparum* there is still a great deal of unexplored application space. This work establishes the utility of the system for directed gene knockouts, but implies several other potential applications as well as modifications that would render the tool more useful.

From a technical standpoint, there are several components that remain constant in order to carry out CRISPR based genome editing, making it possible to generate a strain with all of the common components integrated, greatly simplifying the transfection process. A single vector carrying both the *T7 RNAP* and the *S. pyogenes cas9* genes could be integrated at the CG6 locus using the *bxb1* site-specific integrase system\textsuperscript{39}, to create a strain that constitutively expresses both the *T7 RNAP* and *SpCas9* proteins. Site-specific genome editing would then simply require a plasmid with a site-specific gRNA, and a region for HR mediated repair.

In addition, the question remains about the requirements for the design of the gRNA for site specific targeting. As discussed previously, *T7 RNAP* initiates most efficiently with the pattern GG. Therefore, the targeting site with the adjacent protospacer
adjacent motif (PAM) is constrained to be 5'-GGN_{18}NGG-3'. It remains to be seen in *P. falciparum* whether or not the SpCas9 protein will tolerate a GG overhang at the 5' end of the gRNA, making the required pattern for a gRNA target with PAM 5'-N_{20}NGG-3' and effectively reducing the constrains on selecting a targeting site.

From an application-space perspective, while utilization of the CRISPR/SpCas9 system for gene knockouts is useful, additional applications in allelic replacement and the insertion of gene regulatory machinery to control essential *P. falciparum* genes could prove more useful. For allelic replacement, the HR regions would simply need to be modified to contain the desired changes to the allele, and a suitable cut site found within the genome. Likewise for the addition of regulatory elements, one would need to find a suitable cut site near the 5' or 3' end of the gene, and then build an HR construct designed to recombine and introduce a regulatory element such as the FKBP destabilization domain\(^{33}\), or the TetR aptamer\(^{14}\).

5.4 Future Application space for non-coding RNA regulation in *P. falciparum*

While this work establishes the utility of the lac repression system to regulate the production of non-coding RNA (ncRNA) in the parasite, it does not address any biological applications. The functionality of ncRNA in the parasite remains very poorly understood at this juncture, due in part to the lack of inducible molecular tools for its study. This work provides a framework with which the questions about the various type of ncRNA found in the cell can be addressed. Potentially, the most interesting line of questioning would be to investigate the *var* gene epigenetic variation. There are
approximately 60 members of the var gene family which produces the surface antigen exported to the red blood cell surface. Of the 60 members, only one is ever expressed in a single cell. The method of this action is unknown, but it is known that an ncRNA is produced from the specific var gene intron which, when eliminated, leads to a breakdown in monoallelic expression\textsuperscript{16}. The ability to inducibly produce ncRNA has the potential to lead to further studies of this mechanism.

5.5 Conclusion

No scientific endeavor is ever truly complete. Here I present several new molecular tools for use in \textit{P. falciparum}. This creates a framework for those continuing this work to build and probe new designs and applications. Some of these designs are discussed above but as time goes on, it is my hope that those who follow will come up with ideas that I haven’t even touched upon and to use this work as a foundation for new ways of understanding and combating malaria.
5. Lindsay, S. W., Emerson, P. M. & Charlwood, J. D. Reducing malaria by mosquito-proofing houses. TRENDS Parasitol. 18, 510–514 (2002).


74. Balu, B. et al. piggyBac is an effective tool for functional analysis of the Plasmodium falciparum genome. BMC Microbiol. 9, 83 (2009).


