UC Riverside
UC Riverside Electronic Theses and Dissertations

Title
Discovering the Multi-Faceted Roles of Ubiquitylation within the Human Malaria Parasite, Plasmodium falciparum

Permalink
https://escholarship.org/uc/item/3sd8z8fr

Author
Chung, Duk-Won Doug

Publication Date
2012

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA
RIVERSIDE

Discovering the Multi-Faceted Roles of Ubiquitylation within the Human Malaria Parasite, *Plasmodium falciparum*

Doctor of Philosophy

in

Cell, Molecular and Developmental Biology

by

Duk-Won Doug Chung

March 2012

Dissertation Committee:
Dr. Karine G. Le Roch Ph.D., Chairperson
Dr. Jeffrey Bachant Ph.D.
Dr. Constance Nugent Ph.D.

The Dissertation of Duk-Won Doug Chung is approved:

____

____

____

Committee Chairperson

University of California, Riverside

ABSTRACT OF THE DISSERTATION

Discovering the Multi-Faceted Roles of Ubiquitylation within the Human Malaria Parasite, *Plasmodium falciparum*

by

Duk-Won Doug Chung

Doctor of Philosophy, Graduate Program in Cell, Molecular and Developmental Biology
University of California, Riverside. March 2012
Dr. Karine G. Le Roch, Chairperson

Malaria is one of the deadliest infectious diseases, infecting 300-500 million and killing up to one million people globally every year. It is the world's leading cause of death in children and prevalent in sub-Saharan Africa, South Asia, and parts of South America, adversely impacting economic development in many impoverished countries. Though treatments for malaria do exist, the rising drug-resistances gained by the elusive malaria parasites counteract the efficacy of current treatments and beckon the need for new anti-malarials and new drug targets.

One limiting factor for the development of new antimalarials is our poor understanding of the basic biology of the *P. falciparum*, the most lethal of the parasite species that cause human malaria. To better understand the parasite biology, we investigate the importance of post-translational modification, specifically ubiquitylation, in regulating *Plasmodium* biology. On a genome-wide

scale, we identify the *Plasmodium* "ubiquitome", which is the population of proteins that are likely to be modified by ubiquitin. We have found that the *Plasmodium* ubiquitome takes up a large portion of the entire *Plasmodium* protein population, which suggests that ubiquitylation is an important aspect of the *Plasmodium* biology. Furthermore, our findings show that the roles of ubiquitin in the *Plasmodium* has both conserved and also parasite-specific functions, such as invasion, hemoglobin metabolism, and liver stage-specific purposes.

To fuller expand our understanding of the ubiquitylation pathway in the malaria parasite, we investigate the *Plasmodium* endoplasmic reticulum-associated degradation (ERAD) pathway, as well as a newly identified duplicated ERAD-like system. In eukaryotic cells, the ERAD pathway serves to recognize misfolded proteins within the ER lumen and label them with ubiquitin for proteasome degradation. Here, we characterize the *Plasmodium* ERAD system with a combination of localization studies, *in vitro* biochemical assays, and knockout experiments. Altogether, our findings indicate that the putative ubiquitylating components of the *Plasmodium* ERAD system localizes to the their expected regions (the ER and cytosol), has ubiquitylating properties and is likely essential to the parasite, leaving the possibility for exploiting the *Plasmodium* ERAD system for antimalarial targeting.

Lastly, we characterize the duplicated ERAD system and validated that it targets the *Plasmodium*-specific organelle called the apicoplast. The apicoplast is a four membrane-bound organelle that is essential for the parasite's survival. The functions of the apicoplast (i.e. isoprenoid biosynthesis and fatty acid synthesis) are mostly fulfilled by ~500 nucleus-encoded proteins that are transported to the outermost apicoplast compartment *via* the secretory pathway. However, how these proteins import into the apicoplast through its multiple membranes remain unclear. Here, by using bioinformatics analysis, cloning techniques, recombinant proteins, biochemical assays and fluorescent microscopy, we have biologically characterized the ERAD-like core ubiquitylating components, showing that they are capable of ubiquitylation, and localize them to the apicoplast. Our data indicates that the ERAD-like system likely has similar translocative properties as the traditional ERAD system of the ER and functions to transport apicoplast-targeted proteins across apicoplast membranes. Because of both the essentiality and specificity of the apicoplast to the malaria parasite, we propose that this apicoplast ERAD-like system would make a solid candidate for anti-malaria drug targeting.

Overall, our investigation validates the importance of ubiquitylation within the *Plasmodium* biology and also highlights a few potential antimalarial drug targets that are specific to the deadly parasite.

TABLE OF CONTENTS

Introduction: An overview of malaria: the global impact of the disease and the biological approaches to understanding the regulatory aspects of the malaria parasite 1

Malaria: The disease 2

The historical and current global impact of malaria 2

The life cycle of the malaria parasite 11

Strategies for malaria and vector control 15

The Apicoplast 17

Post-translational regulation of the malaria parasite life cycle 20

Ubiquitylation 22

The ubiquitin proteasome system 24

The ERAD system and duplicated ERAD-like system in the apicoplast 29

References 35

Chapter 1: Unraveling the ubiquitome of the human malaria parasite 42

Chapter Preface 43

Introduction 45

Results 48

Discussion 62

Chapter Concluding remarks 66

Material and Methods 67

References 75

Chapter 2:

Elucidating the ubiquitin-dependent ERAD system in the human malaria parasite 80

Chapter Preface 81

Introduction 82

Results 86

Discussion 99

Chapter Concluding remarks 103

Material and Methods 104

References 110

Chapter 3:

Ubiquitin-dependent protein import into the *Plasmodium* apicoplast *via* a duplicated ERAD-like system 113

Chapter Preface 114

Introduction 115

Results 118

Discussion 142

Chapter Concluding remarks 150

Material and Methods 151

References 163

Conclusion: Some afterthoughts and perspectives 168

References 175

Appendix: 177

A1: Post-translational modifications in *Plasmodium*: more than you think. 178
Mol. Biochem. Parasitol. 168: 123–134 (2009)

A2: Targeting the *Plasmodium* ubiquitin/proteasome system with anti-malarial compounds: promises for the future. 231
Infect Disord Drug Targets 10: 158–164 (2010)

A3: Deciphering the Ubiquitin-Mediated Pathway in Apicomplexan Parasites: A Potential Strategy to Interfere with Parasite Virulence. 258
PLoS ONE 3: e2386 (2008)

A4: Genomics and integrated systems biology in *Plasmodium falciparum*: a path to malaria control and eradication. 328
Parasite Immunology 34: 50–60

A5: Genome-Wide Analysis of Gene Expression. 364
Encyclopedia of Biological Chemistry, 2nd Edition. (2012)

A6: Plasmid construct maps and additional scientific papers 386

A7: List of published papers 401

LIST OF FIGURES

INTRODUCTION

Figure I.1	A global map of boundaries of malaria transmission in 1990	5
Figure I.2	A global map of the boundaries of malaria transmission in 2011	9
Figure I.3	Malaria transmission areas and reported drug resistances	10
Figure I.4	A graphical overview of the *Plasmodium falciparum* life cycle	13
Figure I.5	Light microscope images of *Plasmodium falciparum* at various stages	14
Figure I.6	The apicoplast	18
Figure I.7	Ultrastructures of the apicoplast of the *Plasmodium falciparum*	19
Figure I.8	General flow chart of ubiquitylation	24
Figure I.9	An overview of the possible drug targets within the *Plasmodium* ubiquitylating and proteasome degradation systems	28
Figure I.10	A simplified graphical model of the ERAD system	30
Figure I.11	A graphical model of apicoplast import using previously reported finding	34

CHAPTER 1

Figure 1.1	Ubiquitylation in *P. falciparum* at ring, trophozoite and schizont stages	49
Figure 1.2	Comparison of proteomes for *P. falciparum*, *S. cerevisiae* and *A. thaliana*	52

Figure 1.3	Selective immunoprecipitation and MudPIT identification of ubiquitin-conjugates	56

CHAPTER 2

Figure 2.1	Model of the ERAD system and proteasome degradation	84
Figure 2.2	*In silico* domain architecture of *Plasmodium* ERAD ubiquitylating enzymes	87
Figure 2.3	Recombinant PF14_0215 is able to confer *in vitro* ubiquitylation	89
Figure 2.4	PF14_0215 localizes to the ER membranes	92
Figure 2.5	Knock-out strategy for PF14_0215	93
Figure 2.6	*In vitro* ubiquitylation of PFL1245w and PFL0190w	95
Figure 2.7	PFL1245w and PFL0190w mainly localize to the cytosol	97
Figure 2.8	PFL0190w mainly localize to the cytosol	98

CHAPTER 3

Figure 3.1	*In silico* domain architecture of putative *Plasmodium* ERAD-like ubiquitylating proteins that target the apicoplast	122
Figure 3.2	Recombinant RING domains of PFC0740c and PFC0510w have *in vitro* E3 ubiquitin ligase activity with human UbcH5a and UbcH13	125
Figure 3.3	HA tagging strategy and eventual death of HA-tagged PFC0740c strains	127
Figure 3.4	The N-terminus of PFC0740c is sufficient to target GFP to the apicoplast	129
Figure 3.5	GFP tagging strategy and apicoplast localization of PFC0510w	131

Figure 3.6 Gene disruption of ERAD-like ubiquitylating components 134

Figure 3.7 PF13_0344 and MAL13P1.227 have *in vitro* ubiquitin E1 activating and E2 conjugating activity, respectively 137

Figure 3.8 Putative deubiquitylating enzymes PF10_0308 and Mal8P1.126 mediate DUB activity and localize to the apicoplast, respectively 140

Figure 3.9 Overview model of the *Plasmodium* ERAD and ERAD-like systems 145

APPENDIX A1

Figure A1.1 General depiction of the roles of PTMs in *Plasmodium* 229

APPENDIX A2

Figure A2.1 Graphical depiction of the possible drug targets within the *Plasmodium* UPS 251

APPENDIX A3

Figure A3.1 Representation of the ubiquitin-mediated pathways 305

Figure A3.2 Color matrix representation of by-domain diversity for the 13 proteomes 307

Figure A3.3 Dendrogram tree of ubiquitin and ubiquitin-like modifiers in *Plasmodium spp*., *cryptosporidium spp.* and *T. gondii* 308

Figure A3.4 Dendrogram tree of ubiquitin and ubiquitin-like activating enzymes in *Plasmodium spp*., *cryptosporidium spp.* and *T. gondii* 309

Figure A3.5	Dendrogram tree of ubiquitin and ubiquitin-like conjugating enzymes in *Plasmodium spp*., *cryptosporidium spp.* and *T. gondii*	310
Figure A3.6	Dendrogram tree of ubiquitin and ubiquitin-like conjugating enzymes in *Plasmodium spp*., *cryptosporidium spp.* and *T. gondii*	311
Figure A3.7	Domain architecture of RING and RING-like E3 ligases in *P. falciparum* compared to *S. cerevisiae*	314
Figure A3.8	Proposed representation of *Plasmodium falciparum* erythrocytic cell cycle and two predicted regulatory complexes, a Skp1- Cullin-F-box (SCF) complex and an anaphase-promoting complex-related (APC/C-related)	316

APPENDIX A4

Figure A4.1	Methods for next-generation sequencing (NGS) of the human malaria parasite's genome	353
Figure A4.2	A system biology approach to understand the parasite biology for the design of new drug and vaccines strategies	354

APPENDIX A5

Figure A5.1	Serial Analysis of Gene Expression (SAGE)	382
Figure A5.2	Microarray	383
Figure A5.3	RNA-Seq with second generation and third generation sequencing technology	385

APPENDIX A6

Figure A6.1	Plasmid used for the recombinant cloning of PF14_0215 (PfHRD1) with GST and 6xHIS	386
Figure A6.2	Plasmid used to knockout the PF14_0215 (PfHRD1) gene	387
Figure A6.3	Plasmid used for the recombinant cloning of PFL0190w (PfUBC) with 6xHIS	388
Figure A6.4	Plasmid used for the recombinant cloning of PFL1245w (PfUBA1) with 6xHIS	389
Figure A6.5	Plasmid used for the recombinant cloning of PFC0510w (PfHRD1p) with GST and 6xHIS	390
Figure A6.6	Plasmid used for the recombinant cloning of PFC0740c (PfHRD3p) with GST and 6xHIS	391
Figure A6.7	Plasmid used for the recombinant cloning of Mal13P1.227 (PfUBCp) with 6xHIS	392
Figure A6.8	Plasmid used for the recombinant cloning of PF13_0344 (PfUBA2p) with 6xHIS	393
Figure A6.9	Plasmid used for the recombinant cloning of PF10_0308 (PfOTUp) with 6xHIS	394
Figure A6.10	Plasmid used to fuse GFP to the endogenous PFC0510w gene	395
Figure A6.11	Plasmid used to fuse a 3xHA tag to the endogenous PFC0740c gene	396
Figure A6.12	Plasmid used to disrupt the PFC0740c gene	397
Figure A6.13	Plasmid used to disrupt the PFC0510w gene	398
Figure A6.14	Plasmid used to disrupt the PF13_0182 gene	399
Figure A6.15	Vector that was given to us by the Przyborski lab for episomal expression of genes tagged with GFP	400

LIST OF TABLES

Table 1.1	Ubiquitin target predictions by *UbPred*	51
Table 1.2	Ubiquitin marks on ubiquitin itself found in *P. falciparum*	58
Table 1.3	Ubiquitin-conjugates specifically found in trophozoite stage	59
Table 1.4	Components of the ubiquitin/proteasome system identified by immunoprecipitation coupled to MudPIT	61
Table 3.1	Summary of *in silico* predicted apicoplast ubiquitylating proteins in *P. falciparum* and their homologues in other Apicomplexa	119
Table A1.1	Commonly studies post-translational modifications in *Plasmodium* and their putative biological functions	230
Table A3.1	Predicted number of UPS components in the 13 analyzed genomes	306
Table A3.2	Annotated list of E3 ubiquitin and ubiquitin-like ligases in *P. falciparum*, with their homologs in *T. gondii*, *C. parvum* and *S. cerevisiae*	312

INTRODUCTION

An overview of malaria: the global impact of the disease and the biological approaches to understanding the regulatory aspects of the malaria parasite.

Malaria: The disease

The causative agents of human malaria are parasitic protozoan, which belong to one of five *Plasmodium* species: *P. vivax, P. ovale, P. malarie, P. falciparum, and P. knowlesi*. The time between initial malarial infection and the appearance of symptoms ranges from 9 to 40 days, depending on the malarial species. *P. falciparum* has the shortest incubation period of 9–14 days and *P. malariae* with the longest. The common early symptoms of malaria are fever, chills, headache, sweats, fatigue, nausea and vomiting, which may fluctuate in intensity and duration as the disease progresses. Other common symptoms include dry cough, muscle pains and an enlarged spleen. In severe cases, malaria can lead to anemia, renal failure and the impairment of the respiratory and central nervous systems, leading to seizures and loss of consciousness. (Sadanand, 2010). The intensity and recovery rate from malaria vary depending on the general health of the individual and also the malarial species one is infected with. In particularly serious cases, infection with *P. falciparum* may be lethal.

The historical and current global impact of malaria

The origins and date of emergence of malaria is still debated. Yet, most can agree that malaria predates human history and likely arose in our primate predecessors from the Ethiopian regions of Africa (Schlagenhauf, 2004). With humans, the malaria parasites coevolved and spread with human travel, beginning

from the valley of the Nile, to the Mediterranean, then to Asia, Europe and finally to the Americas (Schlagenhauf, 2004).

Malaria comes from the Italian words “mal’ aria”, which means “bad air” or “evil air”. The early Greeks, as far as 850 BC, observed that people with malarial symptoms usually lived around marshy areas and swamps. And thus it was believed for many centuries that it was the haze rising from these wetlands that caused malaria (Cox, 2010). The first mention of these words, mal’aria, to describe a disease that killed so many people was recorded in Marco Cornaro’s book “Scritture della laguna” published in Venice in 1440 (Neghina *et al*, 2010). However, the first descriptions of malaria’s signature clinical symptoms (anemia, splenomegaly, fever and cold sweat) were recorded much earlier in documents that date to 2700 BC in ancient China. Additional early accounts were found on clay tablets in Mesopotamia from 2000 BC, on papyrus in Egypt from 1570 BC and in Hindu texts that date as far back as 600 BC (Cox, 2010). Since the beginning of human history, malaria had a part in determining the rise and fall of kingdoms, the outcomes of wars and the economic status of nations.

Before the 17th century, most of the treatments for malaria consisted of bleeding and purging, though numerous other unorthodox methods were also employed, such as eating garlic in sour wine and wearing the largest tooth of a fish as an amulet (Schlagenhauf, 2004). Though the origins of its discovery is still

debated, it wasn't until the 1600's that the use of chichona bark was discovered to be the first effective treatment against malaria. A couple of centuries later around 1820, two French researchers were able to extract the active anti-malarial compound, quinine, from the cinchona bark. Thereafter, quinine was widely implemented for malaria treatment and the cultivation of cinchona trees were highly sought after, especially during World War I. In 1880, Charles Louis Alphonse Laveran, a French army doctor working in Algeria, was the first to observe the malarial parasite within human blood cells. Shortly after the discovery of malaria's protozoan causative agent, a British doctor named Ronald Ross was the first to suggest that the mosquito was responsible for transmitting malaria. In 1899, while working in Sierra Leone, Ross demonstrated that the female anopheles mosquito was responsible for the transmission of malaria between humans (Cox, 2010) and was later award the Nobel Prize in Medicine in 1902. Around the same time, a group of Italian scientists also showed evidence for the anopheles mosquito being the vector for human malaria.

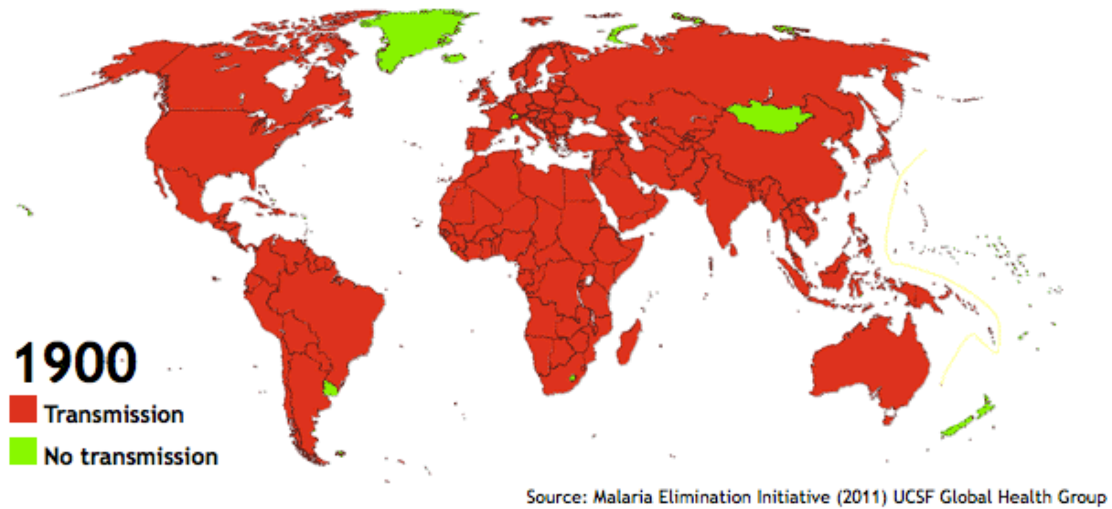


Figure I.1 - A global map of the boundaries of malaria transmission in 1900.
(Source: Malaria Elimination Initiative (2011) UCSF Global Health Group.)

By the start of the 1900s, the transmission of malaria practically reached the entire world (Figure I.1) and had become a global problem, literally stifling national economies and even infrastructures. Take for instance, the Panama Canal, which is an 82-kilometer canal that cuts through the Isthmus of Panama and connects the Atlantic and Pacific oceans. The Panama Canal is considered one of the biggest factors for the increase of global trade and integration of world economies during the early 1900s. However, the construction of the Panama Canal was significantly delayed by malaria and yellow fever, causing sickness and death among the workers. In 1906, 21,000 out of 26,000 employees who were working on the canal contracted malaria at one point during their work stint (www.cdc.gov). Only after concerted efforts to reduce malarial transmission in the regions were undertaken,

construction was able to progress to a sustainable rate, which eventually led to its completion by 1914.

Elsewhere in the world, other efforts to combat malaria were being undertaken. During the same year the Panama Canal was finished, United States Public Health service requested funds from the U.S. congress to investigate, combat and control malaria in the United States. Another notable achievement was the synthesis of the antimalarial chloroquine by a German scientist named Hans Andersag at Bayer I.G. Farbenindustrie A.G. laboratories in 1934. Originally named resochin by Andersag, chloroquine is a synthetic derivative of quinine. The advantages of chloroquine over quinine were that it was easier to produce and had a significantly longer half-life of 60 days, rather than 8-10 hours, respectively (Petersen *et al*, 2011). Though chloroquine had significant anti-malarial properties, it was initially considered too toxic for human use. However, after more than a decade of clinical research, chloroquine was established to have therapeutic value and was introduced into clinical practice in 1947. Furthermore, Dichloro-diphenyl-trichloroethane (DDT), which was first synthesized in 1874 by another German scientist Othmer Zeidler, was later discovered to have potent insecticidal properties in 1939 and was widely used to prevent the spread of malaria shortly after World War II. In 1947, the United States sponsored a national malaria eradication campaign spearheaded by the Communicable Disease Center (currently known as Centers for Disease Control and Prevention) (CDC). This eradication campaign

primarily consisted of administering chloroquine and spraying DDT in rural households and in regions that had a high risk of malaria transmission. By 1951, malaria was considered eliminated in the United States.

In 1955, spurred on by the successful elimination of malaria in the US and the urgent belief that time and money were of the essence, the World Health Organization (WHO) ambitiously proposed a global malaria eradication campaign. Using a combination of DDT spraying, surveillance, and anti-malarial treatment, the elimination of malaria in some nations with temperate climates and seasonal malaria transmission were successful (Nájera *et al*, 2011). However, several nations (i.e. Indonesia, Afghanistan, Haiti, and Nicaragua) had negligible success, while other nations (most of sub-Saharan Africa) were completely excluded from the eradication campaign (Petersen *et al*, 2011). One major obstacle to the campaign was the rise of chloroquine-resistance, which was first observed along the Thai-Cambodia border and Columbia in the late 1950s and later to South-East Asia, Africa, Papua New Guinea and the Phillipines by the late 1970s (Mita *et al*, 2009). The only other available drug alternative at the time was sulfadoxine–pyrimethamine, though it also faced drug-resistant parasites about a year after implementation. The malaria eradication campaign faced many other obstacles: the emergence of insecticide resistance, wars between nations, massive population movements, the lack of sustained funding from donor countries, and the absence of community participation (Petersen *et al*, 2011). These obstacles made the long-term

maintenance of the malaria eradication campaign untenable and thus the effort for its completion was eventually abandoned by 1978.

Over a century after the discovery quinone, several more antimalarial compounds were either synthesized or discovered. One notable antimalarial is artemisinin. In 1971, Chinese scientists discovered artemisinin, which is derived from the leaves of the *Artemisia annua* (also known as qing hao or Chinese sweetworm), which has been used in China as an herbal remedy for centuries. Though it is complex and costly to extract, artemisinin was found to be an effective anti-malarial and introduced to the world in 1979. Semi-synthetic versions of artemisinin, such as Dihydroartemisinin (DHA), were eventually produced and yielded greater anti-malarial efficacy, long half-lives and lower toxicity in humans. Artemisinin is useful in severe cases because of its broad range of action, killing malaria parasites in both the early and mature stages. However, like its predecessors, artemisinin encountered resistant parasites, first in Cambodia and eventually to other parts of South-East Asia (Porter-Kelley *et al*, 2010). Because of the rise of resistance, the WHO recommended the use of Artemisinin Combination Therapies (ACTs), which are permutations of artemisinin derivatives blended with other antimalarials such as mefloquine or sulfadoxine. The reasoning behind this strategy was two-fold: increase the efficacy of treatment and also to lower the risk of producing more artemisinin-resistant parasites (Porter-Kelley *et al*, 2010).

However, the downside to ACTs is that they are more costly than any other mono-therapy.

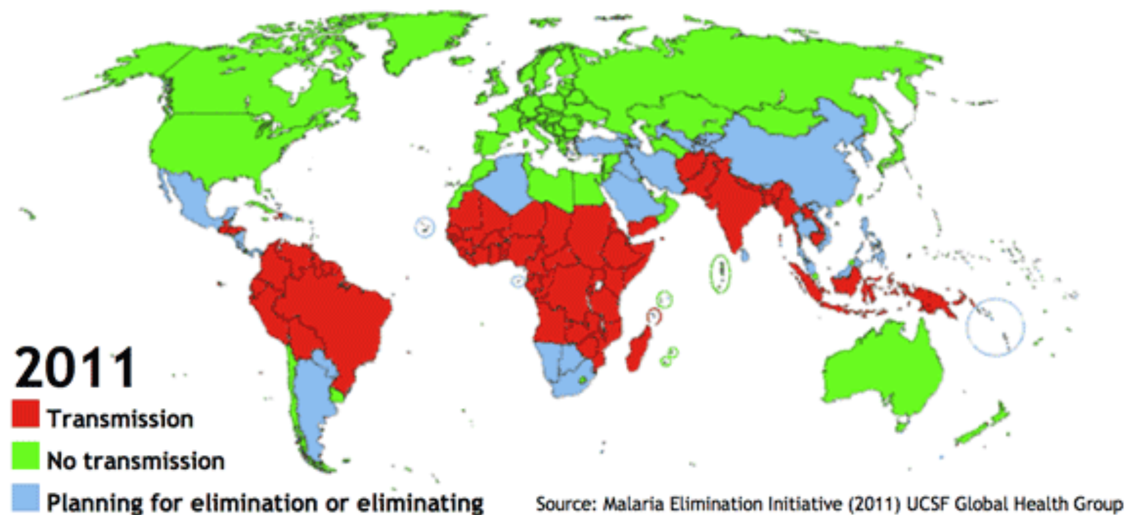


Figure I.2 - A global map of the boundaries of malaria transmission in 2011.
(Source: Malaria Elimination Initiative (2011) UCSF Global Health Group

Today, because of the milestone discoveries of the malaria parasite and vector, the success of the DDT as an insecticide, the advent of less toxic and more effective synthetic antimalarials, and the implementation of bed-nets, the spread of the deadly malaria parasite has greatly been reduced since the turn of the 1900th century and have led to the successful malaria eradication in many parts of the world (Figure I.2) (Raghavendra *et al*, 2011). However, despite the great advancements in our understanding and fight against the malarial parasite, malaria is still one of the world's deadliest infectious diseases, infecting 300-500 million and killing up to one million people globally every year (Garcia, 2010). It is the leading

cause of death in children and prevalent in sub-Saharan Africa, South Asia, and parts of South America (Figure I.2), adversely impacting economic development in many impoverished countries (Sadanand, 2010). The reasons for the continued prevalence of malaria in these regions is compounded with many factors that include limited access to quality diagnostic facilities, the inability of those infected to pay for preventative measures (i.e. bed-nets) and treatments, social and political barriers, and also the rise of resistance against the commonly used antimalarials (i.e. quinone, chloroquine, mefloquine, primaquine, artemisinin) in the past several decades (O'Brien *et al*, 2011) (Figure I.3).

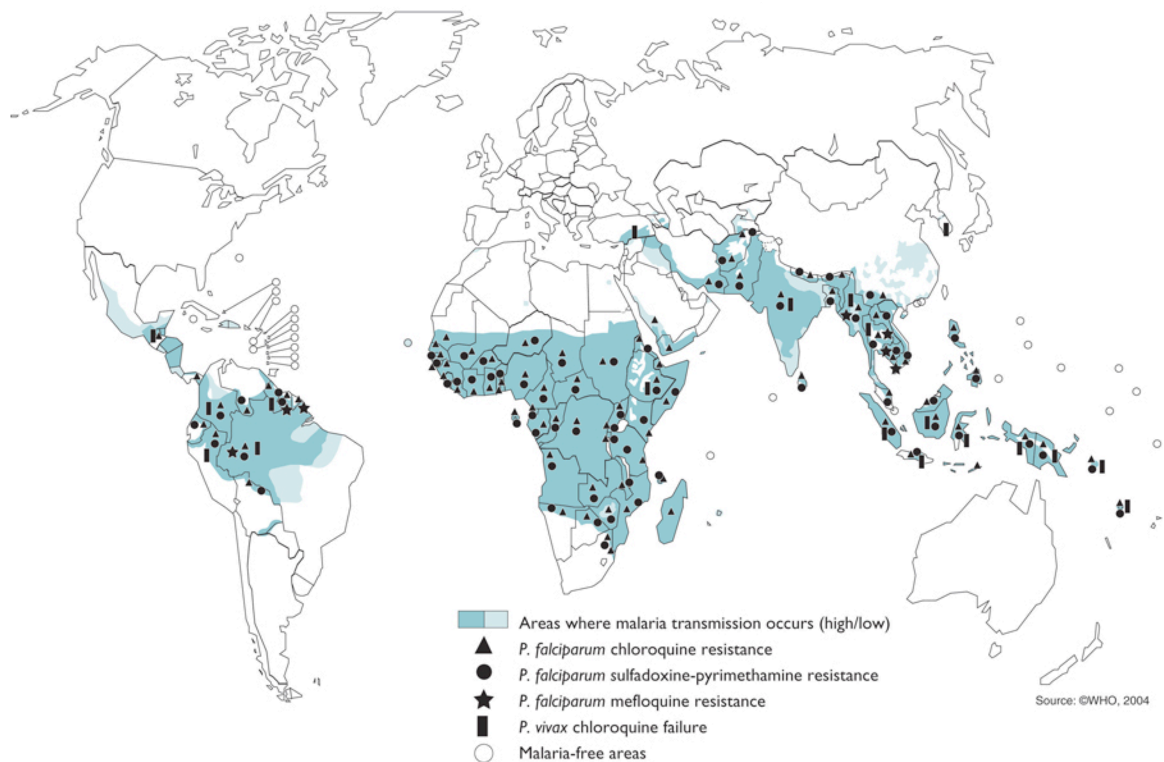


Figure I.3 - Malaria transmission areas and reported drug resistance from 2004
(Source: World Health Organization, 2004)

There are several lessons that we could learn from the previously failed malaria eradication campaign. One of the most important lessons is that it is fair to reason that in any country that malaria is still endemic today, the problem is not so simple and cannot be solved with a single uniform strategy. It is essential to identify the biological, physical, social and cultural barriers that have proven to block the success of malaria control in the past and make all the necessary efforts to circumvent them in the future. In addition, having a direct link between research from various fields (i.e. ecology, sociology, economy, biology, chemistry, political science), government agencies, and organizations working first-hand in the field, would be highly beneficial. The fight against malaria can only be won by the concerted involvement of multi-disciplinary programs with both vertical and horizontal approaches.

The life cycle of the malaria parasite

Malaria can be caused in humans by five species of *Plasmodium* (*P. vivax, P. ovale, P. malarie, P. falciparum, and P. knowlesi*) with *Plasmodium falciparum* being the most lethal and being the most drug-resistant (Parija & Praharaj, 2011). The vector in which the malaria parasites are spread is the female *Anopheles* mosquito. The female mosquito can become infected with the malaria parasite when it takes a blood-meal from a malaria-infected human and draws in mature *Plasmodium* gametocytes (Figure I.4). The male and female gametes fuse in the gut of the mosquito and generate a zygote, which undergoes rapid meiosis to transform into a

motile ookinete and then into an oocyte within the mid-gut wall. From the oocyte, sporozoites are formed within 10-22 days, which then bud and migrate to the salivary glands of the mosquito (Baton & Ranford-Cartwright, 2005). From then on, the next time the infected female mosquito bites another human to take an additional blood-meal, the sporozoites are injected into the new human host and travel to the liver. There, they replicate and differentiate into merozoites, which are eventually released into the blood stream and invade red blood cells (RBC).

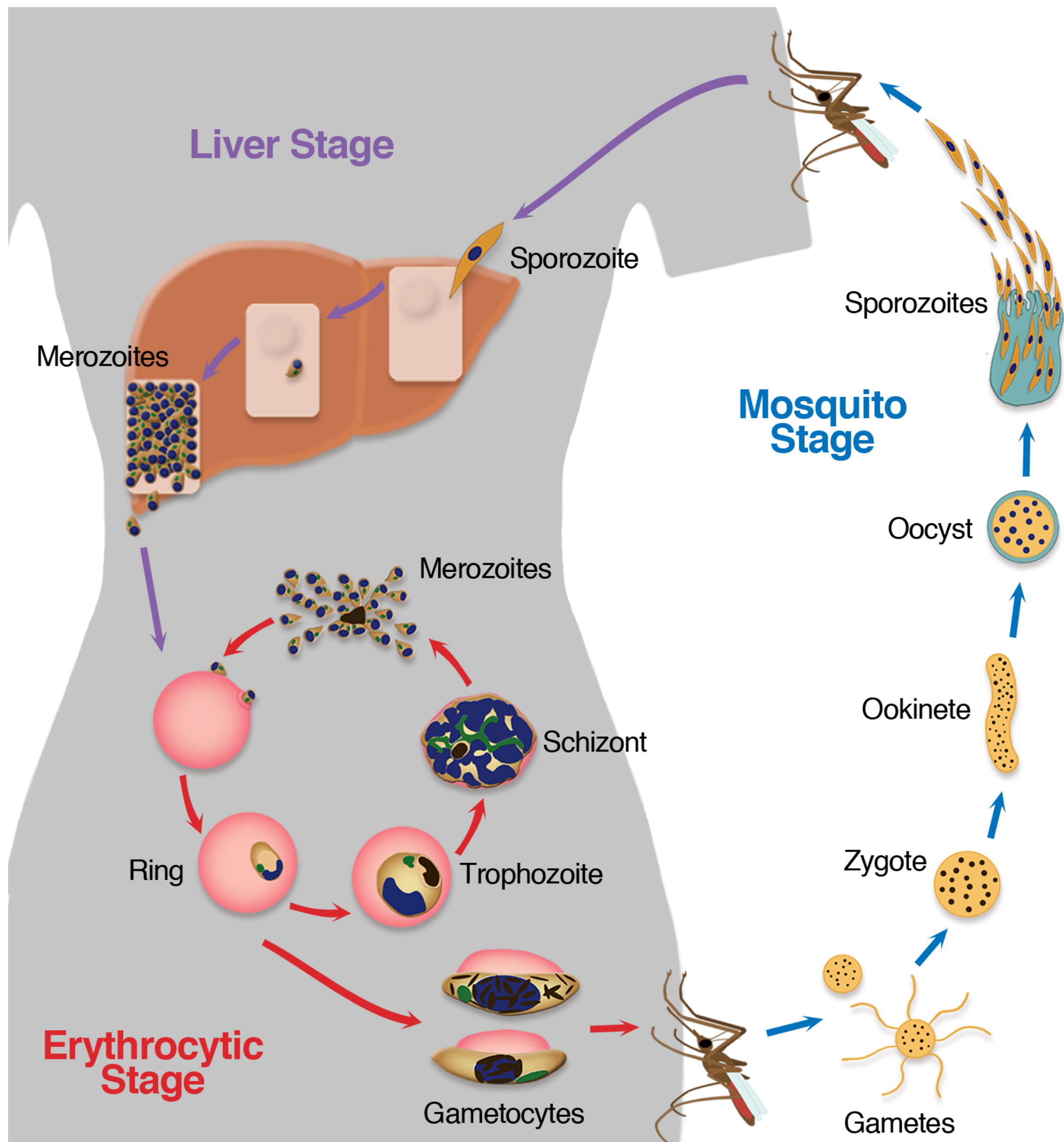


Figure I.4 – A graphical overview of the *Plasmodium falciparum* life cycle
(Source: S. Cervantes, Le Roch lab)

The *P. falciparum* inhabits either the human host or the mosquito vector. The parasite is introduced to the human by mosquitos during blood feeding in the form of a sporozoite, which then invades the liver (liver stage) and replicates into merozoites. Merozoites subsequently invade red blood cells, where the parasite enters into a repetitive replicative cycle (erythrocytic stage). Parasites can also transform into gametocytes, where they are sucked up by a new mosquito vector and mature into sporozoites, thus completing the life cycle

In the RBC, the parasites will feed, grow and passage through the sequential three major blood stages: ring, trophozoite and schizont. Four rounds of nuclear division and the multiplication of organelles occurs in the schizont stage producing up to 32 merozoites, which will egress and reinvade a new uninfected RBC, thus continuing cycle. This asexual cycle is known as the intra-erythrocytic cycle with a typical length of time of around 48 hours for *P. falciparum* (Tilley *et al*, 2011) (Figure I.5). After a few weeks of propagation, some of the parasites transform into either male or female gametocytes (Figure I.5), which can be taken up by a mosquito vector and thus furthering the spread of the parasite (Dixon *et al*, 2008).

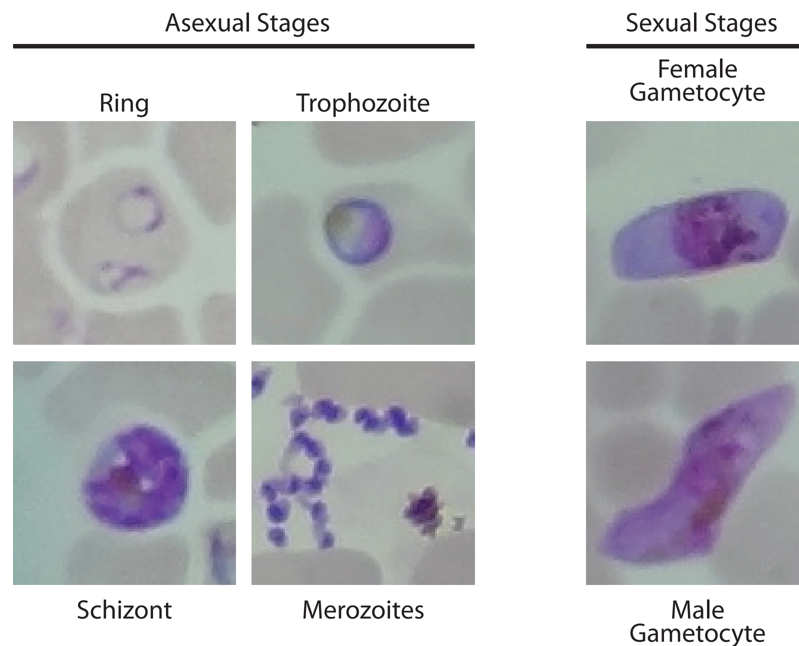


Figure I.5 – Light microscope images of *Plasmodium falciparum* at various stages.
(Source: D. Chung, Le Roch Lab)

Prepared with Giemsa staining, blood smears with infected RBCs show the *P. falciparum* within the asexual erythrocytic stages in the left panels and the sexual stages in the right two panels.

Strategies for malaria and vector control

Various strategies are currently being explored in order to prevent and eliminate malaria. One major approach is through the development of a malaria vaccine. Vaccines have been and continue to be effective against global diseases (i.e. smallpox, polio, measles, rubella). Promising is that it has been documented that individuals in endemic countries have acquired malarial immunity and that numerous potential malarial vaccines have reached human clinical trials. A major obstacle to designing a malarial vaccine is that the malaria parasite predominantly resides within RBCs and are shielded from detection by antibodies. That is why much of the vaccine research today are targeting stages of the parasite life cycle, such as the sporozoite and merozoite stages, where the parasite is momentarily exposed before invading hepatic cells or RBCs (Targett & Greenwood, 2008). Most recently, a vaccine called RTS,S has reached phase III clinical trials with reports of up to 50% reduction in the incidence of malaria among young children (The RTS,S Clinical Trials Partnership, 2011). However, using a more realistic approach, the efficacy of RTS,S against malaria is calculated to be closer to 34-36% (Duncan & Hill, 2011). Though there has been much progress, a universally effective vaccine has yet to be achieved.

Vector control is another huge antimalarial endeavor. There has been continual effort to educate and to provide tools to limit exposure to mosquitos and to reduce the mosquito population in endemic regions. Notably, a couple of research

teams were able to create transgenic malaria mosquitoes that are *Plasmodium*-resistant (Ito *et al*, 2002; Corby-Harris *et al*, 2010). However, successful replacement of current mosquito populations with the *Plasmodium*-resistant transgenic lines would rely upon a drive mechanism, posing many difficulties and possible unforeseen consequences.

Furthermore, due to the rise of antimalarial drug-resistance by the parasites, there has been a resurgence of effort to find novel drugs and drug targets. One such method is a discovery-based approach where high-throughput screenings of thousands of natural and synthetic compounds are being run in order to potentially uncover the next major antimalarial (McNamara & Winzeler, 2011; Prudhomme *et al*, 2008). With the tremendous amount of high-throughput operations, an unprecedented identification of over 10,000 new compounds, with at least some antimalarial efficacy, were reported. Since most high-throughput systems implemented cell-based assays, these novel compounds would inherently lack defined modes of action. Though a minority of compounds may already have recognizable targets, a major challenge is to assign these novel molecules to their modes of action.

On the flip side, other groups have devoted their efforts to characterizing the different aspects of *Plasmodium* biology in order to possibly expose the parasite's vital components and essential processes with the intention of targeting them for

drug development. A few examples include identifying major regulatory mechanisms that control parasite invasion, host-immune evasion, egress, transcription, translation, gametocytogenesis, growth and metabolism.

The Apicoplast

Plasmodium parasites belong to the phylum Apicomplexa, which, with the exception of *Cryptosporia* and *Gregarines* (Valigurová *et al*, 2007), are distinguished from other eukaryotic cells because they contain a non-photosynthetic plant-like organelle (or plastid) called the apicoplast (See Figure I.6 (Waller *et al*, 2000)). The apicoplast is a result of a secondary endosymbiosis when an ancestor of Apicomplexa engulfed a red alga and created a four membrane-bound plastid (Oborník *et al*, 2009). It is involved in heme synthesis, fatty acid metabolism, and isoprenoid biosynthesis (see (Fleige *et al*, 2010) for a review).

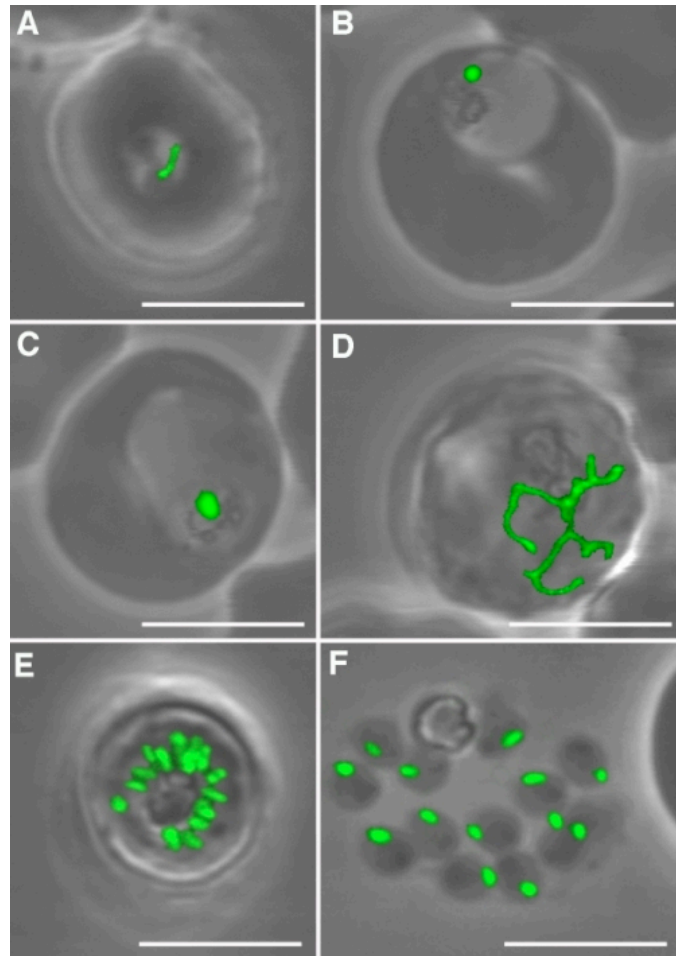


Figure I.6 - The apicoplast.
(Source: Waller *et al*, 2000)

The single apicoplast in **(A)** enlarges and elongates to form the branched forms shown in **(D)**. They divide **(E)** and segregate into the free merozoites in **(F)**. GFP has been fused to the acyl carrier protein, an apicoplast membrane protein.

The apicoplast contains a ~35kb long circular genome encoding ~30 proteins (Wilson *et al*, 1996), but most of its functions, such as isoprenoid biosynthesis, are fulfilled by proteins encoded within the parasite nuclear DNA. The trafficking pathways involved in the import of ~500 nucleus-encoded proteins (Foth *et al*, 2003) through the four apicoplast membranes remain unclear (Figure I.7,

(Maréchal & Cesbron-Delauw, 2001)). Apicoplast targeting is thought to involve the secretory pathway, through the endoplasmic reticulum (ER), and a bipartite leader peptide located at the n-terminus of transported proteins (Tonkin *et al*, 2008). The bipartite domain consists of a signal peptide (cleaved off after passage into the ER) followed by a transit peptide that avoids secretion and is sufficient for apicoplast stromal targeting of nucleus-encoded proteins (Waller *et al*, 1998). In contradiction, the presence of the bipartite leader peptide is however not required in some cases of apicoplast membrane targeting of membrane-bound proteins (Lim *et al*, 2009). Moreover, there are increasing cases of apicoplast targeting without the presence of a canonical bi-partite signal, including multi-compartment targeting (Ponpuak *et al*, 2007). Evidently, the rules controlling trafficking to the apicoplast remain unclear and research is still ongoing.

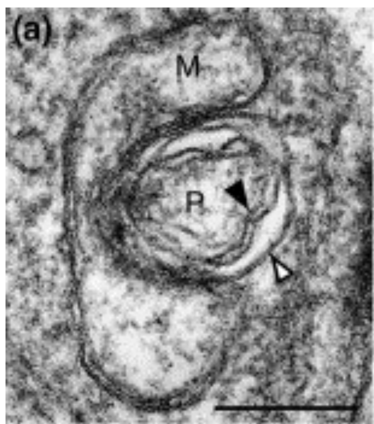


Figure I.7 – Ultrastructures of the apicoplast of the *Plasmodium falciparum*.
(Source: Marechal *et al*, 2001).

Black arrow heads indicate envelope membranes, white arrowheads indicate outermost membranes.

Labels: **M**, mitochondria; **P**, plastid

Shortly after its discovery, the apicoplast has been the focus of great interest within the parasite community because it was shown to be essential and its

disruption causes a delayed-death phenomenon (Sullivan *et al*, 2000). In theory, the apicoplast reveals itself as a great candidate for anti-malarial drug targeting because it is both parasite-specific and necessary for parasite survival.

Post-translational regulation of the malaria parasite life cycle

There has been significant progress in elucidating how the malaria parasite regulates its life cycle. The sequencing of the *Plasmodium* genomes, along with subsequent comparative bioinformatics approaches, transcriptome and proteome analyses, have created a vast amount of information regarding protein prediction and their hypothetical functions in both the human host and mosquito vector (Bozdech *et al*, 2003; Le Roch *et al*, 2004, 2003; Gardner *et al*, 2002). While microarray analyses have demonstrated a remarkable change in steady-state mRNA levels during parasite development, only a relatively few regulatory motifs and transcription regulators have been uncovered so far (Coulson *et al*, 2004). In addition, unlike other organisms, there seems to be few transcriptional changes in *Plasmodium* following exposure to external stimuli (Le Roch *et al*, 2008; Guerra *et al*, 2008). These findings imply that parasite transcription could be hard-wired (Ganesan *et al*, 2008). This rigidity in transcription suggests that post-transcriptional and post-translational regulations are likely to play major roles in regulating the parasite life cycle.

By definition, post-translational modifications (PTMs) must add or subtract a specific mass difference and not be particular to any one protein (Hoffman *et al*, 2008). In addition to changing protein mass, PTMs can also alter the protein charge and conformation. Such changes modify the protein's enzyme activity, binding affinity and hydrophobicity (Clark *et al*, 2005). The spectrum of the PTMs within a cell is immense and varies with respect to specificity and abundance. For example, some types of PTMs such as phosphorylation and ubiquitylation are universally employed to regulate a broad host of functions, have relatively high abundance, and have a wide range of target substrates. On the other hand, some PTMs such as acetylation are highly specific in their roles, have a relatively low abundance, and may target only a few proteins or even one target substrate at a time. Furthermore, PTMs can either modify a protein at one specific amino acid residue, or be associated with numerous different residues. The potential diversity of protein modifications seem limitless due to the fact that an individual protein can undergo a single modification or multiple types of modifications at several sites, possibly producing multiple protein isoforms, each with a unique biological activity. One can easily see why PTMs are responsible for a major increase in complexity from genome to proteome. For example, the human genome contains approximately 30,000 open reading frames, but is predicted to give rise to roughly 1.8 million different protein variants (Jensen, 2004). In addition, PTMs are reversible and thus provide the flexibility and adaptability that are essential for mediating rapid cellular responses to the cell's constantly changing conditions. While transcriptional

regulation can control cell progression in a temporal manner, post-transcriptional regulation of proteins can provide faster mechanisms of positive or negative regulation in most signaling pathways. Thus, post-translational modifications are vitally essential for the survival of all kinds of cells, with *Plasmodium* not being an exception.

Ubiquitylation

One major post-transcriptional regulation mechanism in signaling and cell progression in eukaryotic cells is ubiquitylation, a process where a small 76 amino acid peptide called ubiquitin is covalently attached to protein substrates (Hershko & Ciechanover, 1998). Although ubiquitylation is more familiarly known for its association with proteasome degradation, it is now well documented that mono- and poly-ubiquitylation of targeted substrates can mediate a wide array of cellular processes (*e.g.* cell proliferation, cell stress response, transcription, cell death, DNA repair, intracellular trafficking, endocytosis and signal transduction) in a proteasome-independent manner (Chen & Sun, 2009). Ubiquitylation involves the covalent attachment of a ubiquitin moiety to lysine residues of protein substrates *via* the hierarchical intervention of a ubiquitin-activating enzyme (E1), a ubiquitin-conjugating enzyme (E2), and a ubiquitin ligase (E3) that is usually involved in specific substrate recognition (Figure I.8) (Pickart, 2004; Laney & Hochstrasser, 1999). Using ATP, E1 enzymes adenylate ubiquitin at its c-terminus, creating a high energy mixed anhydride bond. The sulfhydryl group of the E1 active-site cysteine

then attacks the anhydride bond, forming a high-energy thioester bond between the E1 enzyme and ubiquitin, thus expelling AMP. Ubiquitin is then passed to the active-site cysteine of the E2 enzyme. Lastly, with the aid of an E3 ligase, ubiquitin is covalently attached to a target protein substrate at the ε-amino group of a lysine. Ubiquitin can also be specifically cleaved off its targets (or matured in the case of ubiquitin-fusion proteins) by deubiquitylating enzymes (DUB).

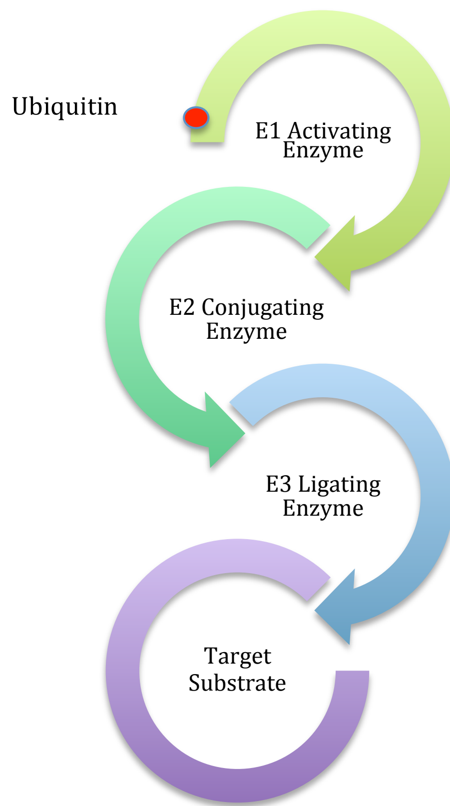


Figure I.8 – General flow chart of ubiquitylation.

Ubiquitin, an 8 kDa protein, is covalently attached to target substrates by a series of ubiquitylating enzymes. Briefly, using ATP, E1 activating enzymes primes the c-terminus of ubiquitin by adenylation and subsequently is attached to the active-site cysteine of the E1 enzyme. Ubiquitin is then transferred to the active-site cysteine of an E2 conjugating enzyme. Finally, ubiquitin is attached to target subtrates at lysine residues with the aid of E3 ligases.

The ubiquitin proteasome system

As mentioned previously, ubiquitylation is best recognized for its role within the ubiquitin proteasome system (UPS), where the degradation of proteins by the UPS has many known roles in cell cycle regulation, transcription and

transmembrane signaling (Jung *et al*, 2009; Weissman *et al*, 2011). Over the past decades, our increased understanding of the importance of the UPS in many different human and infectious diseases has elicited a wide interest in deciphering the components of this pathway in many living organisms. While we have only begun to investigate and understand the hundreds of genes involved in this ubiquitous pathway, it is already becoming clear that many different components of this system can be targets for inhibition of various human diseases, with pathogen-caused diseases not being an exception.

The first success in targeting the UPS against a human disease has been demonstrated by the inhibition of the proteasome for cancer treatment. Bortezomib (Velcade; Millennium), a dipeptidyl boronic acid inhibitor of the proteasome has been the first drug approved by the FDA for the treatment of relapsed or refractory multiple myeloma (Adams *et al*, 1999; Roccaro *et al*, 2006). While one can argue that the proteasome is ubiquitous to all human cell lines, the efficacy and limited toxicity observed so far by this proteasome inhibitor in cancer patients results from the fact that rapidly dividing cells are more sensitive to the inhibitor than non-dividing cells.

As with all other eukaryotic cells, regulated protein degradation is an essential aspect of *Plasmodium* biology and may also provide possible drug targets for effective antimalarials. As a proof of concept, about a dozen known proteasome inhibitors have been tested against *Plasmodium* and have shown significant anti-

malaria activity. For example, bortezomib (the proteasome inhibitor mentioned above) along with its boronate analogs MLN-273 and ZL_3B were found to arrest the *P. falciparum* erythrocytic cycle with 50% inhibitory concentrations (IC_{50}) within the low nanomolar range in both wild-type and drug-resistant strains (Lindenthal *et al*, 2005; Reynolds *et al*). In addition, other proteasome inhibitors (ie. Salinosporamide A, epoxomicin, lactacystin, gliotoxin) exhibited promising preliminary results as well (Dick *et al*, 1996; Kreidenweiss *et al*, 2008; Prudhomme *et al*, 2008). Despite the high conservation between human, yeast and *Plasmodium* proteasomes, proteasome inhibitors have effectively exhibited antimalarial properties with limited host toxicity. This observation may be linked to the fact that malaria parasites, like cancer cells, have a high rate of replication and thus are more sensitive to proteasome inhibitors. Furthermore, an amino acid (Y168G) divergence, found in the catalytic domain of the proteasome ß subunit, may possibly be exploited to increase drug specificity against the parasite proteasome (Prudhomme *et al*, 2008). The *Plasmodium* also contains a prokaryotic proteasome, which may also be exploited for drug targeting to due to its parasite specificity. Though the effectiveness of proteasome inhibitors against malaria in human clinical trials remains to be evidenced, targeting the *Plasmodium* degradation pathway has so far shown much promise.

From previous reports mentioned above, we are confident that the *Plasmodium* proteasome systems are important to the parasite and provide

excellent drug target candidates. In addition, we believe the characterization of several more specific components in the UPS opens the door to new therapeutic interventions. For example, E3 ligases are the most numerous and diverse among the ubiquitylating enzymes, while E1 enzymes are the most conserved (Figure I.9) (Ponts *et al*, 2008; Chung & Le Roch, 2010). Because of the significant differences between the *Plasmodium* E3 ligases and their human homologues, E3 ligases present themselves as excellent parasite-specific antimalarial target candidates.

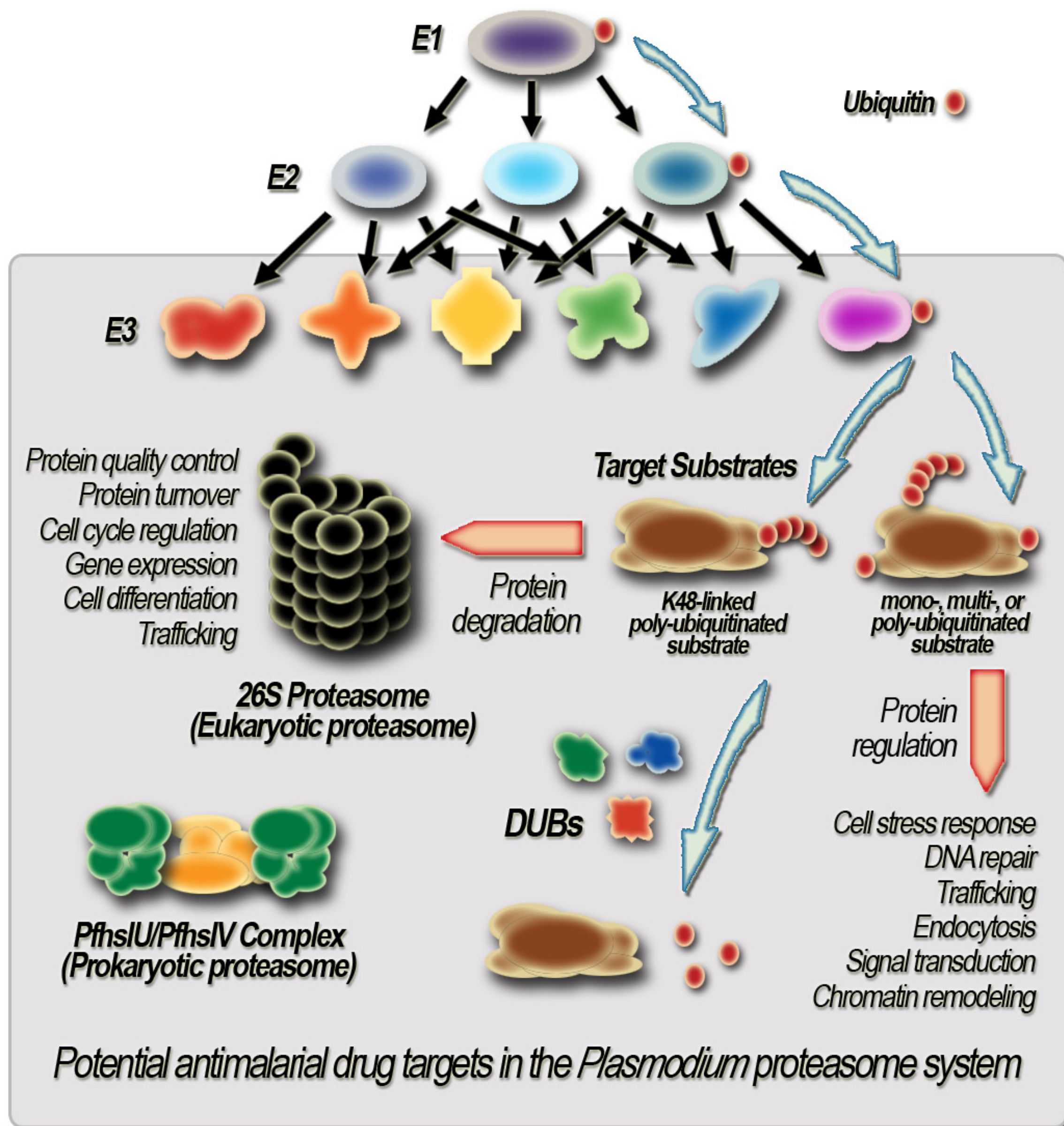


Figure I.9 – An overview of the possible drug targets within the *Plasmodium* ubiquitylating and proteasome degradation systems.
(Source: Chung *et al*, 2010, *Infect Disord Drug Targets* **10**:158-164)

Among the ubiquitylating enzymes, studies show that E1 enzymes are the most conserved while E3 ligases are the most numerous and diverse, with *Plasmodium* not being an exception. Since ubiquitylation is an important aspect of cell biology, *Plasmodium* E3 ligases present themselves as great drug targets because of their diversity and biological importance. Furthermore, *Plasmodium* proteasome systems may also present themselves to be potential candidates for antimalarial targeting, particularly with the several promising proteasome inhibitor studies that have shown significant antimalarial properties.

The ERAD system and duplicated ERAD-like system in the apicoplast

One of the many important aspects of ubiquitin-dependent proteasome degradation is the endoplasmic reticulum-associated degradation (ERAD) system, where misfolded proteins are recognized within the endoplasmic reticulum (ER) lumen and translocated across the ER membrane for proteasome degradation (Bagola *et al*, 2011). The ERAD system maintains protein quality and prevents the accumulation of misfolded proteins within the cells (Figure I.10).

A brief description of the classical eukaryotic ERAD model is that aberrant proteins are recognized by ER luminal chaperone binding proteins (BiP) and protein disulfide isomerases (PDI) to help discriminate properly folded proteins from misfolded proteins (Xie & Ng, 2010). Misfolded proteins are then shuttled to the DER1 translocon complex, which form a hydrophobic pore to allow the retro-translocation of proteins through the ER membrane. Within this translocon complex, an E3 ubiquitin ligase called HRD1 is implicated to interact with membrane-bound proteins needed for retrotranslocation and also help form the hydrophobic pore complex (Sato *et al*, 2009). Besides the HRD1 E3 ligase, the ERAD system consists of other ubiquitylating enzymes, and are subsequently ubiquitylated and retro-translocated to the cytosol where they are poly-ubiquitylated for destruction by the 26S proteasome (Bagola *et al*, 2011; Carvalho *et al*, 2006; Hershko & Ciechanover, 1998).

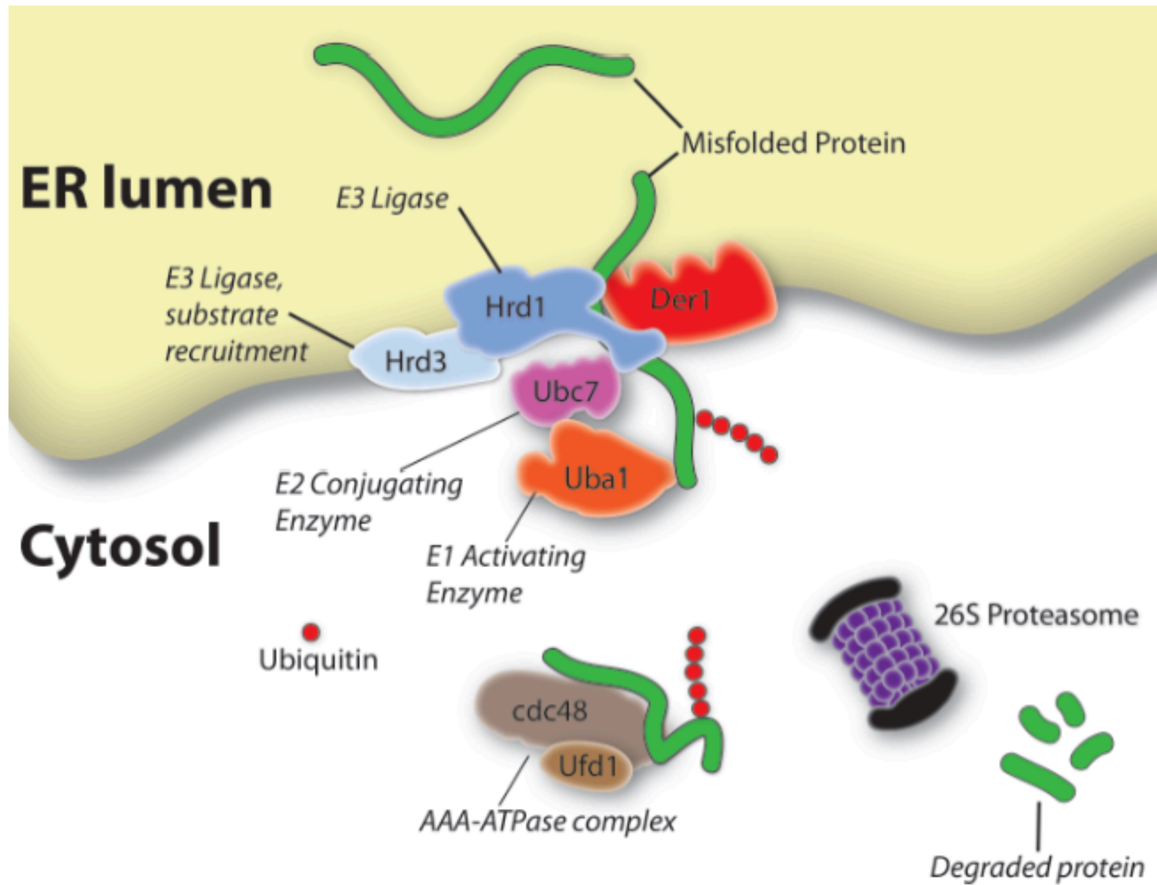


Figure I.10 – A simplified graphical model of the ERAD system
(Source: D. Chung, Le Roch Lab)

The ERAD system recognizes misfolded proteins in the ER lumen and translocates them across the ER membrane. During translocation, the misfolded proteins are ubiquitylating and then shuttled by the cdc48 complex to the 26S proteasome for degradation.

Recent studies have identified proteins that are duplicated homologues of components of the ERAD system, and are believed to target the apicoplast (Hempel *et al*, 2010; Sommer *et al*, 2007; Spork *et al*, 2009; Ponts *et al*, 2008). These proteins are suspected to form an ERAD-like system in Apicomplexa that could be similar to the symbiont-specific ERAD-like machinery (SELMA) found in Chromalveolates, a supergroup of plastid-containing unicellular eukaryotes related to red algae

(Hempel *et al*, 2010; Sommer *et al*, 2007). Similarly to the classical ERAD machinery, the SELMA is implicated in the pre-protein import to the periplastidal compartment (PPC) (Agrawal *et al*, 2009; Hempel *et al*, 2010). In *Plasmodium* and the apicomplexan *Toxoplasma gondii*, the ERAD-like system is strongly suspected to be involved in translocating proteins across the apicoplast membranes (Agrawal *et al*, 2009; Hempel *et al*, 2009; Sommer *et al*, 2007; Spork *et al*, 2009) in a fashion similar to the way the classical ERAD system retro-translocate aberrant proteins from the ER lumen to the cytosol during protein quality control.

There is still debate on how nuclear-encoded apicoplast targeted proteins actually reach the apicoplast. Some believe that apicoplast proteins utilize vesicles to reach the apicoplast while others believe that the apicoplast form transient bridges with the ER, allowing for the apicoplast proteins to float inside (Figure I.11). Perhaps, both scenarios are true. In either case, the apicoplast proteins are able to reach the third most outer membrane, the periplastid membrane. However, how apicoplast proteins are able to translocate across the three inner membranes is still a mystery. Kalonan *et al*. and van Dooren *et al.* reported an existence of a TIC-like protein, in both *Plasmodium* and *Toxoplasma*, at the apicoplast innermembranes and suggest that the TIC-like protein serves to translocate proteins across the innermost membrane into the apicoplast stroma. For import across the periplastid membrane, it is suggested by other groups that the duplicated ERAD-like components are likely playing an important role.

Similar to the ERAD system, preliminary studies indicated that the ERAD-like system may also be ubiquitin-dependent. In the Chromalveolate *Phaeodactylum tricornutum*, ERAD homologues of an E3 ligase and a de-ubiquitylating enzyme with verified *in vitro* ubiquitylating/deubiquitylating activity were localized in the plastid (Hempel *et al*, 2010). In *P. falciparum* and five other *Plasmodia*, multiple components of a putative apicoplast ERAD-like system have been identified (Ponts *et al*, 2008; Sommer *et al*, 2007; Spork *et al*, 2009). Among them were candidates for an apicoplast-targeted ubiquitin (PfApiUB, PF08_0067), two E1 ubiquitin-activating (UBA) enzymes (PF13_0182 and PF13_0344), and one PfApiUBC4 E2 ubiquitin-conjugating (UBC) enzyme (Mal13P1.227). Apicoplast localization was experimentally validated in *P. falciparum* for PfApiUBA1 (PF13_0182), PfApiUB (PF08_0067), and PfsDER1 (PF14_0498) (Kalanon *et al*, 2009; Spork *et al*, 2009). It is suggested that PfsDer1, having multiple transmembrane domains, creates a hydrophobic pore within the apicoplast periplastid membrane (PPM) and allows for the apicoplast-targeted preproteins from the ER to translocate inward, where they would be subsequently ubiquitylated by the identified E1 UBA enzymes and E2 UBC enzyme for further apicoplast import. The identification of SELMA components has brought new perspectives on the mechanisms of how nuclear-encoded proteins are transported to the plastids of Chromalveolates. However, unlike the SELMA system within the Chromalveolate, possible candidates for an E3 ligase and DUB in the *P. falciparum* ERAD-like system have yet to be identified. Since the primary roles of E3 ligases are both the recognition and catalyzation of target substrate for

ubiquitylation, it would be difficult to imagine an ERAD-like system without an E3 ligase (Figure I.11). In addition, it is also hard to picture an ubiquitylating system without a DUB in order to reverse the attachment of ubiquitin and thus replenish the pool of available free ubiquitin.

My thesis project starts with a broad investigation of the post-translational modification ubiquitylation within the *Plasmodium falciparum*. I participate in a genome-wide *in silico* search to identify the *Plasmodium* "ubiquitome", which consists of all the possible proteins that are, at least at one point, ubiquitylated. Then my investigation focuses on characterizing the core ubiquitylating components of the *Plasmodium* ERAD systems. Though the ERAD system has been relatively well characterized in other model eukaryotic systems, there has been little work done on the *Plasmodium* ERAD system, which potentially could offer new venues for drug targets highlighted by the several promising proteasome inhibitor studies. And lastly, my investigation concentrates on a small group of parasite-specific ubiquitylating proteins that appeared to be a duplicated set of ERAD-like proteins and contained features for apicoplast targeting. Through localization studies and functional analyses, it appears that this novel ERAD-like system may work to import proteins into the apicoplast and could offer a new line of drug targets against malaria.

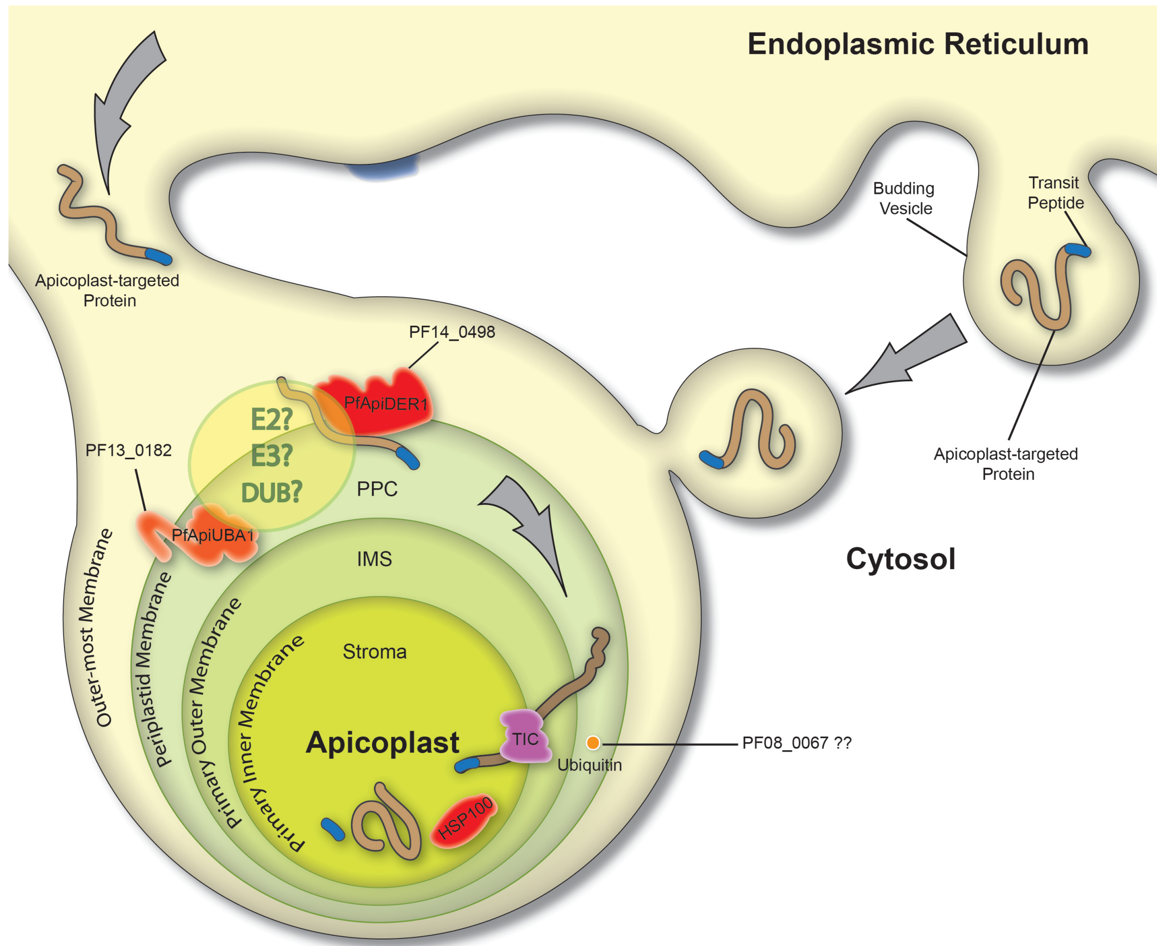


Figure I.11 – A graphical model of apicoplast import using previously reported findings.
(Source: D. Chung, Le Roch Lab)

Apicoplast import of nuclear-encoded proteins is still debated. Some believe that apicoplast-targetd proteins utilize vesicle trafficking, while others believe that the apicoplast form temporary membrane fusion bridges that allow proteins to reach the apicoplast. In either case, imported proteins are able to reach the periplastid membrane. However, the method for translocation across the three most-inner membranes is still mystery. A TIC protein was identified to exist within the innermorst membranes, possibly serving similar roles to the chloroplast innermembrane translocons. Also, another group has localized an UBA1(E1) and an ubiquitin-like protein to the apicoplast, possibly working to ubiquitylate and translocate proteins across the periplastid membrane. However, there are still several ubiquitylating proteins that have not been biologically characterized and are missing that would be necessary for a fully functioning ubiquitylating machinery.

REFERENCES

Adams J, Palombella VJ, Sausville EA, Johnson J, Destree A, Lazarus DD, Maas J, Pien CS, Prakash S & Elliott PJ (1999) Proteasome inhibitors: a novel class of potent and effective antitumor agents. *Cancer Res.* **59:** 2615–2622

Agrawal S, van Dooren GG, Beatty WL & Striepen B (2009) Genetic Evidence that an Endosymbiont-derived Endoplasmic Reticulum-associated Protein Degradation (ERAD) System Functions in Import of Apicoplast Proteins. *Journal of Biological Chemistry* **284:** 33683–33691

Bagola K, Mehnert M, Jarosch E & Sommer T (2011) Protein dislocation from the ER. *Biochim. Biophys. Acta* **1808:** 925–936

Baton LA & Ranford-Cartwright LC (2005) Spreading the seeds of million-murdering death: metamorphoses of malaria in the mosquito. *Trends Parasitol.* **21:** 573–580

Bozdech Z, Llinás M, Pulliam BL, Wong ED, Zhu J & DeRisi JL (2003) The Transcriptome of the Intraerythrocytic Developmental Cycle of Plasmodium falciparum. *PLoS Biol* **1:** e5

Carvalho P, Goder V & Rapoport TA (2006) Distinct Ubiquitin-Ligase Complexes Define Convergent Pathways for the Degradation of ER Proteins. *Cell* **126:** 361–373

Chen ZJ & Sun LJ (2009) Nonproteolytic functions of ubiquitin in cell signaling. *Mol. Cell* **33:** 275–286

Chung D-WD & Le Roch KG (2010) Targeting the Plasmodium ubiquitin/proteasome system with anti-malarial compounds: promises for the future. *Infect Disord Drug Targets* **10:** 158–164

Clark RSB, Bayir H & Jenkins LW (2005) Posttranslational protein modifications. *Crit. Care Med.* **33:** S407–409

Corby-Harris V, Drexler A, Watkins de Jong L, Antonova Y, Pakpour N, Ziegler R, Ramberg F, Lewis EE, Brown JM, Luckhart S & Riehle MA (2010) Activation of Akt Signaling Reduces the Prevalence and Intensity of Malaria Parasite Infection and Lifespan in Anopheles stephensi Mosquitoes. *PLoS Pathog* **6:**

Coulson RMR, Hall N & Ouzounis CA (2004) Comparative genomics of transcriptional control in the human malaria parasite Plasmodium falciparum. *Genome Res.* **14:** 1548–1554

Cox FE (2010) History of the discovery of the malaria parasites and their vectors. *Parasit Vectors* **3:** 5

Dick LR, Cruikshank AA, Grenier L, Melandri FD, Nunes SL & Stein RL (1996) Mechanistic studies on the inactivation of the proteasome by lactacystin: a central role for clasto-lactacystin beta-lactone. *J. Biol. Chem.* **271:** 7273–7276

Dixon MWA, Thompson J, Gardiner DL & Trenholme KR (2008) Sex in Plasmodium: a sign of commitment. *Trends Parasitol.* **24:** 168–175

Duncan CJA & Hill AVS (2011) What is the efficacy of the RTS,S malaria vaccine? *BMJ* **343:** d7728–d7728

Fleige T, Limenitakis J & Soldati-Favre D (2010) Apicoplast: keep it or leave it. *Microbes and Infection* **12:** 253–262

Foth BJ, Ralph SA, Tonkin CJ, Struck NS, Fraunholz M, Roos DS, Cowman AF & McFadden GI (2003) Dissecting Apicoplast Targeting in the Malaria Parasite Plasmodium falciparum. *Science* **299:** 705 –708

Ganesan K, Ponmee N, Jiang L, Fowble JW, White J, Kamchonwongpaisan S, Yuthavong Y, Wilairat P & Rathod PK (2008) A genetically hard-wired metabolic transcriptome in Plasmodium falciparum fails to mount protective responses to lethal antifolates. *PLoS Pathog.* **4:** e1000214

Garcia LS (2010) Malaria. *Clinics in Laboratory Medicine* **30:** 93–129

Gardner MJ, Hall N, Fung E, White O, Berriman M, Hyman RW, Carlton JM, Pain A, Nelson KE, Bowman S, Paulsen IT, James K, Eisen JA, Rutherford K, Salzberg SL, Craig A, Kyes S, Chan M-S, Nene V, Shallom SJ, *et al* (2002) Genome sequence of the human malaria parasite Plasmodium falciparum. *Nature* **419:** 498–511

Guerra CA, Gikandi PW, Tatem AJ, Noor AM, Smith DL, Hay SI & Snow RW (2008) The limits and intensity of Plasmodium falciparum transmission: implications for malaria control and elimination worldwide. *PLoS Med.* **5:** e38

Hempel F, Bullmann L, Lau J, Zauner S & Maier UG (2009) ERAD-Derived Preprotein Transport across the Second Outermost Plastid Membrane of Diatoms. *Mol Biol Evol* **26:** 1781–1790

Hempel F, Felsner G & Maier UG (2010) New mechanistic insights into pre-protein transport across the second outermost plastid membrane of diatoms. *Molecular Microbiology* **76:** 793–801

Hershko A & Ciechanover A (1998) THE UBIQUITIN SYSTEM. *Annu. Rev. Biochem.* **67:** 425–479

Hoffman MD, Sniatynski MJ & Kast J (2008) Current approaches for global post-translational modification discovery and mass spectrometric analysis. *Anal. Chim. Acta* **627:** 50–61

Ito J, Ghosh A, Moreira LA, Wimmer EA & Jacobs-Lorena M (2002) Transgenic anopheline mosquitoes impaired in transmission of a malaria parasite. *Nature* **417:** 452–455

Jensen ON (2004) Modification-specific proteomics: characterization of post-translational modifications by mass spectrometry. *Curr Opin Chem Biol* **8:** 33–41

Jung T, Catalgol B & Grune T (2009) The proteasomal system. *Mol. Aspects Med.* **30:** 191–296

Kalanon M, Tonkin CJ & McFadden GI (2009) Characterization of Two Putative Protein Translocation Components in the Apicoplast of Plasmodium falciparum. *Eukaryotic Cell* **8:** 1146–1154

Kreidenweiss A, Kremsner PG & Mordmüller B (2008) Comprehensive study of proteasome inhibitors against Plasmodium falciparum laboratory strains and field isolates from Gabon. *Malar. J.* **7:** 187

Laney JD & Hochstrasser M (1999) Substrate Targeting in the Ubiquitin System. *Cell* **97:** 427–430

Lim L, Kalanon M & McFadden GI (2009) New proteins in the apicoplast membranes: time to rethink apicoplast protein targeting. *Trends in Parasitology* **25:** 197–200

Lindenthal C, Weich N, Chia YS, Heussler V & Klinkert MQ (2005) The proteasome inhibitor MLN-273 blocks exoerythrocytic and erythrocytic development of Plasmodium parasites. *Parasitology* **131:** 37–44

Maréchal E & Cesbron-Delauw MF (2001) The apicoplast: a new member of the plastid family. *Trends Plant Sci.* **6:** 200–205

McNamara C & Winzeler EA (2011) Target identification and validation of novel antimalarials. *Future Microbiol* **6:** 693–704

Mita T, Tanabe K & Kita K (2009) Spread and evolution of Plasmodium falciparum drug resistance. *Parasitology International* **58:** 201–209

Nájera JA, González-Silva M & Alonso PL (2011) Some Lessons for the Future from the Global Malaria Eradication Programme (1955–1969). *PLoS Med* **8:**

Neghina R, Neghina AM, Marincu I & Iacobiciu I (2010) Malaria, a Journey in Time: In Search of the Lost Myths and Forgotten Stories. *The American Journal of the Medical Sciences* **340:** 492–498

O'Brien C, Henrich PP, Passi N & Fidock DA (2011) Recent clinical and molecular insights into emerging artemisinin resistance in Plasmodium falciparum. *Curr. Opin. Infect. Dis.* **24:** 570–577

Oborník M, Janouskovec J, Chrudimský T & Lukes J (2009) Evolution of the apicoplast and its hosts: From heterotrophy to autotrophy and back again. *International Journal for Parasitology* **39:** 1–12

Parija S & Praharaj I (2011) Drug resistance in malaria. *Indian Journal of Medical Microbiology* **29:** 243

Petersen I, Eastman R & Lanzer M (2011) Drug-resistant malaria: Molecular mechanisms and implications for public health. *FEBS Letters* **585:** 1551–1562

Pickart CM (2004) Back to the Future with Ubiquitin. *Cell* **116:** 181–190

Ponpuak M, Klemba M, Park M, Gluzman IY, Lamppa GK & Goldberg DE (2007) A role for falcilysin in transit peptide degradation in the Plasmodium falciparum apicoplast. *Mol. Microbiol* **63:** 314–334

Ponts N, Yang J, Chung D-WD, Prudhomme J, Girke T, Horrocks P & Le Roch KG (2008) Deciphering the Ubiquitin-Mediated Pathway in Apicomplexan Parasites: A Potential Strategy to Interfere with Parasite Virulence. *PLoS ONE* **3:** e2386

Porter-Kelley JM, Cofie J, Jean S, Brooks ME, Lassiter M & Mayer DG (2010) Acquired resistance of malarial parasites against artemisinin-based drugs: social and economic impacts. *Infect Drug Resist* **3:** 87–94

Prudhomme J, McDaniel E, Ponts N, Bertani S, Fenical W, Jensen P & Le Roch K (2008) Marine Actinomycetes: A New Source of Compounds against the Human Malaria Parasite. *PLoS ONE* **3:**

Raghavendra K, Barik TK, Reddy BPN, Sharma P & Dash AP (2011) Malaria vector control: from past to future. *Parasitology Research* **108:** 757–779

Reynolds JM, El Bissati K, Brandenburg J, Günzl A & Mamoun CB Antimalarial activity of the anticancer and proteasome inhibitor bortezomib and its analog ZL3B. *BMC Clin Pharmacol* **7:** 13–13

Roccaro AM, Hideshima T, Richardson PG, Russo D, Ribatti D, Vacca A, Dammacco F & Anderson KC (2006) Bortezomib as an antitumor agent. *Curr Pharm Biotechnol* **7:** 441–448

Le Roch KG, Johnson JR, Ahiboh H, Chung D-WD, Prudhomme J, Plouffe D, Henson K, Zhou Y, Witola W, Yates JR, Mamoun CB, Winzeler EA & Vial H (2008) A systematic approach to understand the mechanism of action of the bisthiazolium compound T4 on the human malaria parasite, Plasmodium falciparum. *BMC Genomics* **9:** 513

Le Roch KG, Johnson JR, Florens L, Zhou Y, Santrosyan A, Grainger M, Yan SF, Williamson KC, Holder AA, Carucci DJ, Yates JR 3rd & Winzeler EA (2004) Global analysis of transcript and protein levels across the Plasmodium falciparum life cycle. *Genome Res.* **14:** 2308–2318

Le Roch KG, Zhou Y, Blair PL, Grainger M, Moch JK, Haynes JD, De La Vega P, Holder AA, Batalov S, Carucci DJ & Winzeler EA (2003) Discovery of gene function by expression profiling of the malaria parasite life cycle. *Science* **301:** 1503–1508

Sadanand S (2010) Malaria: an evaluation of the current state of research on pathogenesis and antimalarial drugs. *Yale J Biol Med* **83:** 185–191

Sato BK, Schulz D, Do PH & Hampton RY (2009) Misfolded Membrane Proteins Are Specifically Recognized by the Transmembrane Domain of the Hrd1p Ubiquitin Ligase. *Molecular Cell* **34:** 212–222

Schlagenhauf P (2004) Malaria: from prehistory to present. *Infectious Disease Clinics of North America* **18:** 189–205

Sommer MS, Gould SB, Lehmann P, Gruber A, Przyborski JM & Maier U-G (2007) Der1-mediated Preprotein Import into the Periplastid Compartment of Chromalveolates? *Mol Biol Evol* **24:** 918–928

Spork S, Hiss JA, Mandel K, Sommer M, Kooij TWA, Chu T, Schneider G, Maier UG & Przyborski JM (2009) An Unusual ERAD-Like Complex Is Targeted to the Apicoplast of Plasmodium falciparum. *Eukaryotic Cell* **8:** 1134–1145

Sullivan M, Li J, Kumar S, Rogers MJ & McCutchan TF (2000) Effects of interruption of apicoplast function on malaria infection, development, and transmission. *Molecular and Biochemical Parasitology* **109:** 17–23

Targett GA & Greenwood BM (2008) Malaria vaccines and their potential role in the elimination of malaria. *Malar J* **7:** S10

The RTS,S Clinical Trials Partnership (2011) First Results of Phase 3 Trial of RTS,S/AS01 Malaria Vaccine in African Children. *New England Journal of Medicine* **365:** 1863–1875

Tilley L, Dixon MWA & Kirk K (2011) The Plasmodium falciparum-infected red blood cell. *The International Journal of Biochemistry & Cell Biology* **43:** 839–842

Tonkin CJ, Kalanon M & McFadden GI (2008) Protein targeting to the malaria parasite plastid. *Traffic* **9:** 166–175

Valigurová A, Hofmannová L, Koudela B & Vávra J (2007) An Ultrastructural Comparison of the Attachment Sites Between Gregarina steini and Cryptosporidium muris. *Journal of Eukaryotic Microbiology* **54:** 495–510

Waller RF, Keeling PJ, Donald RGK, Striepen B, Handman E, Lang-Unnasch N, Cowman AF, Besra GS, Roos DS & McFadden GI (1998) Nuclear-encoded proteins target to the plastid in Toxoplasma gondii and Plasmodium falciparum. *Proc Natl Acad Sci U S A* **95:** 12352–12357

Waller RF, Reed MB, Cowman AF & McFadden GI (2000) Protein trafficking to the plastid of Plasmodium falciparum is via the secretory pathway. *EMBO J.* **19:** 1794–1802

Weissman AM, Shabek N & Ciechanover A (2011) The predator becomes the prey: regulating the ubiquitin system by ubiquitylation and degradation. *Nat Rev Mol Cell Biol* **12:** 605–620

Wilson (Iain) RJM, Denny PW, Preiser PR, Rangachari K, Roberts K, Roy A, Whyte A, Strath M, Moore DJ, Moore PW & Williamson DH (1996) Complete Gene Map of the Plastid-like DNA of the Malaria ParasitePlasmodium falciparum. *Journal of Molecular Biology* **261:** 155–172

Xie W & Ng DTW (2010) ERAD substrate recognition in budding yeast. *Seminars in Cell & Developmental Biology* **21:** 533–539

CHAPTER 1

Unraveling the ubiquitome of the human malaria parasite.

*Nadia Ponts, Anita Saraf, **Duk-Won Doug Chung**, Alona Harris, Jacques Prudhomme,*

Michael P. Washburn, Laurence Florens, Karine G. Le Roch.

Journal of Biological Chemistry, **286:** 40320–40330, 2011

CHAPTER 1 PREFACE

A comprehensive approach must be undertaken in order to ultimately defeat such a long-battled disease such as malaria. Like in any war, the more knowledge one can acquire about the enemy, the better the chances of victory. The biology of the malaria parasite is complex and still largely mysterious. Yet, through arduous scientific endeavors, mankind has slowly begun to unlock some of malaria's vital secrets. With the sequencing of the *Plasmodium* genomes, we now have the tools to probe and analyze the individual proteins that make up the parasite on a grand scale, first *in silico* with a computer and then followed up with biological validation.

By using immunofluorescent microscopy, biochemical assays, *in silico* prediction, and mass spectrometry analysis using the multidimensional protein identification technology, we investigate the *Plasmodium* ubiquitome, which is the population of proteins that are ubiquitylated at one point within the *Plasmodium* life cycle. Ubiquitylation is a major post-translational modification found in all eukaryotes, yet it is an aspect of *Plasmodium* biology that has been relatively overlooked despite promising proteasome inhibitor studies that have shown significant antimalarial properties. It is our hope that identifying the group of proteins that are ubiquitylated will shed some light on how ubiquitin-dependent proteasome degradation, and ubiquitylation in general, contribute to the regulation of the deadly malarial parasite. Though I am not the primary contributor to this

body of work presented in this chapter (I performed the immunofluorescence experiments and helped with the *in vitro* ubiquitylation assays of parasite lysates), these findings have helped establish the direction for my next two projects and contribute in discovering the various kinds of roles ubiquitylation has within the *P. falciparum*.

INTRODUCTION

Despite significant advances in the research against malaria, many aspects of the parasite's biology remain obscure. Recent studies highlighted the importance of post-translational regulations for the parasite's progression throughout its life cycle (reviewed in (Chung *et al*, 2009)). Among those, protein ubiquitylation is certainly one of the most abundant post-translational modifications of proteins. Ubiquitylation is classically involved in numerous crucial biological processes in eukaryotic cells in a proteasome-dependent or independent manner. The specificity of its components and the wide range of biological processes in which it is involved make the ubiquitylation pathway an important source of suitable targets for drug treatments (see (Petroski, 2008) for a review). Proteasome inhibitors are indeed promising drugs to treat cancer and both autoimmune and infectious disease (see (Berkers & Ovaa, 2010) for a review). In the particular case of *Plasmodium*, the proteasome inhibitors salinosporamide-A (Prudhomme *et al*, 2008), bortezomib (Reynolds *et al*), and MLN-273 (Lindenthal *et al*, 2005) were shown to be efficient antimalarials (salinosporamide-A is currently under drug trial). In addition to their cellular roles, the components of the ubiquitin system are involved in many host-pathogen interactions. Pathogens often utilize the host ubiquitylation system to bypass the infected host immune system (see (Edelmann & Kessler, 2008; Spallek *et al*, 2009) for reviews).

Ubiquitylation consists in the attachment of one (mono-ubiquitylation) or more (multi-ubiquitylation of different lysine residues, or poly-ubiquitylation *i.e.*, the formation of a poly-ubiquitin chain) ubiquitin moieties to a target protein *via* the formation of an iso-peptide bound between the C-terminal di-glycine motif of ubiquitin and a lysine in the target protein. The entire process requires the sequential intervention of three families of enzymes, namely E1 ubiquitin-activating enzymes, E2 ubiquitin-conjugating enzymes, and E3 ubiquitin ligases involved in specific substrate recognition (see (Chung *et al*, 2009; Ponder & Bogyo, 2007; Ponts *et al*, 2008) for reviews). In *P. falciparum*, previous *in silico* studies identified four predicted sources of ubiquitin moeities. The polyubiquitin gene PFL0585w contains five conserved ubiquitin repeats (Ponder & Bogyo, 2007; Ponts *et al*, 2008). The two ubiquitin-fusion proteins $PfUB_{S27a}$ and $PfUB_{L40}$ (PF14_0027 and PF13_0346, respectively) contain an ubiquitin moiety at their N-terminal side (Ponts *et al*, 2008). More recently, one conserved ubiquitin domain was identified in PF08_0067 and is believed to be part of an Endoplasmic Reticulum-Associated protein Degradation (ERAD)-like pathway (Spork *et al*, 2009). In addition to these four genes encoding ubiquitins, more than 100 proteins involved in the ubiquitylation system were identified *in silico* (Spork *et al*, 2009), *i.e.,* about 2% of *P. falciparum*'s total protein-coding genes. By contrast, only a few substrates for ubiquitylation have been identified in *P. falciparum*, including actin (Field *et al*, 1993) and the histone protein H2B (Trelle *et al*, 2009). The quickly reversible character of ubiquitylation (intervention of deubiquitinases) and the rapid degradation of poly-ubiquitylated

proteins (proteasome) render the isolation and analysis of ubiquitylated proteins challenging.

Various methods can be used to identify ubiquitylated proteins both *in silico* (Radivojac *et al*, 2010; Tung & Ho, 2008) and experimentally (Manzano *et al*, 2008; Tan *et al*, 2008; Hjerpe *et al*, 2009; Tomlinson *et al*, 2007; Golebiowski *et al*, 2010), including immunoprecipitation of ubiquitylated proteins coupled to mass spectrometry analysis. In human, specific precipitation of ubiquitin-conjugated proteins identified 345 substrates for ubiquitylation (Hatakeyama *et al*, 2005). Here, we explore *P. falciparum*'s ubiquitome (*i.e.,* all the proteins that are targets for ubiquitylation) using a combination of immunofluorescent localization of ubiquitin conjugates, detection of ubiquitylating activities of various parasite extracts, *in silico* prediction of ubiquitin targets genome-wide using the previously published algorithms *UbPred* (Radivojac *et al*, 2010) and *UbiPred* (Tung & Ho, 2008), and immunoprecipitation of ubiquitin conjugates followed by immunoprecipitates analysis using Multidimensional Protein Identification Technology (MudPIT). The application of these techniques to the analysis of *P. falciparum*'s ubiquitome during its asexual cell cycle permitted the discovery of unknown features. We found that more than half of the parasite's proteome represents possible targets for ubiquitylation. About 200 of these targets were confirmed by MudPIT analysis. The schizont stage in particular was found to contain the largest number of ubiquitylated proteins, whereas the ring stage contained the least. This observation

is consistent with the parasite's reduction in size and activity before maturation into invasive merozoites. The biological implications of these findings are discussed.

RESULTS

Ubiquitylation in Plasmodium falciparum erythrocytic stages

P. falciparum's genome contains more than 100 genes that are potentially encoding the components of the ubiquitylation system, including four sources of ubiquitin (PFL0585w, PF14_0027, PF13_0346, and PF08_0067), eight E1 enzymes (with various specificities for ubiquitin and ubiquitin-like moieties), 14 E2 enzymes, more than 50 E3 ubiquitin ligases, and about 30 deubiquitinases (Ponts *et al*, 2008). To experimentally confirm the presence of an active ubiquitylation system during *P. falciparum*'s asexual cell cycle, fresh (same day) protein extracts were prepared from parasites harvested at ring, trophozoite and schizont stages and tested for the presence of endogenous ubiquitylating activity (Figure 1.1A). Ubiquitylation activity was detected in all three morphological stages of the parasite.

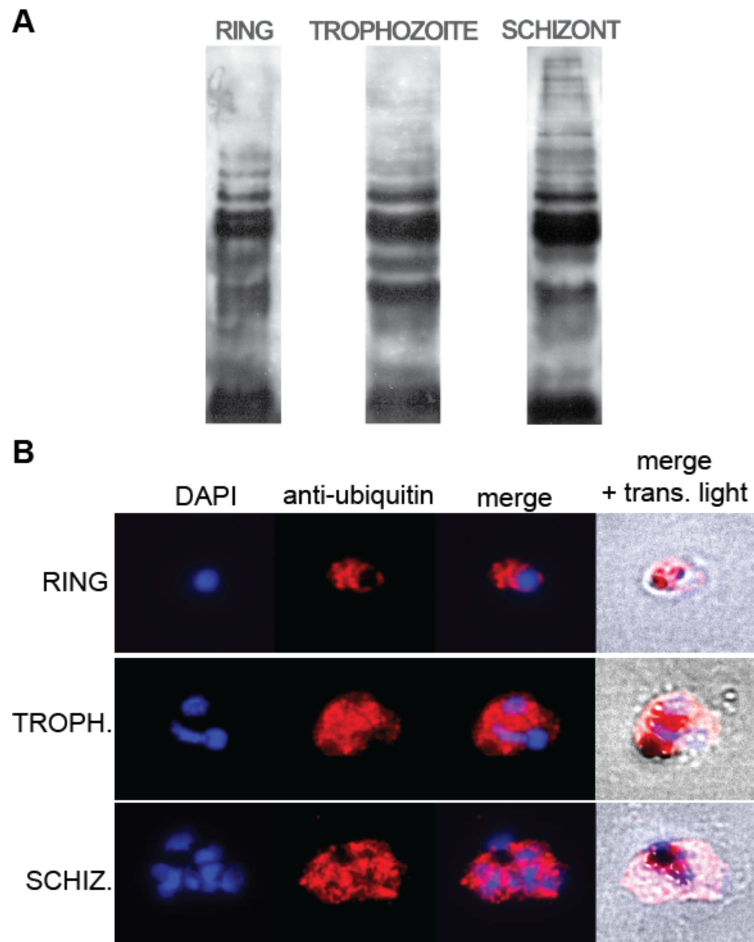


Figure 1.1 - Ubiquitylation in *P. falciparum* at ring, trophozoite and schizont stages.

(A) *In vitro* ubiquitylation activity of parasite's protein extracts prepared at different morphological stages. Ubiquitylation assays were performed in the presence of biotin-ubiquitin and revealed by western-blot streptavidin-HRP with chemilluminescence detection. The presence of multiple bands reveals the presence of various biotin-ubiquitin-conjugates. (**B)** Immunofluorescent *in situ* localization of ubiquitin conjugates in parasites at their ring, trophozoite (TROPH.), or schizont (SCHIZ.) stages. The parasites were visualized under transmitted (trans.) light, their nuclei stained with DAPI (blue), and the ubiquitin-conjugates visualized in the red. Ubiquitin conjugates are present at all investigated stages.

We further investigated the localization of ubiquitin conjugates in intra-erythrocytic parasites by immunofluorescence microscopy using an anti-ubiquitin antibody that does not react with free ubiquitin (which minimizes eventual background due to the intra-cellular pool of free ubiquitin). Results are shown on Figure 1.1B. At all stages, ubiquitin-conjugates are detected in abundance in the entire cell. This pattern is consistent with the fact that ubiquitylation is usually found in both the cytoplasm and the nucleus. In addition, localized intense spots of fluorescence can be observed, consistent with the presence of ubiquitin-conjugates in the parasite's vesicles. These results confirm the presence of a functional ubiquitylation system in all asexual stages of *P. falciparum*.

In silico **identification of putative targets for ubiquitylation**

We explored *in silico* the ubiquitome of *P. falciparum*. We used the algorithm *UbPred* to identify the proteins among the 5446 translated protein-coding sequences that contain favorable ubiquitylation sites (Radivojac *et al*, 2010). Three different cut-offs were used to classify the results: (*i*) all identified proteins (low to high confidence), (*ii*) proteins identified with medium to high confidence, and (*iii*) highly confident results only (see the experimental procedures for more details). Results are summarized in Table 1.1. A total of 5036 proteins were predicted to contain ubiquitylation sites, 3516 of them with a high confidence level (*i.e.,* ~65 % of the input dataset, Figure 1.2A). This predicted abundance of targets for

ubiquitylation is slightly higher than the predictions made for *Saccharomyces cerevisiae* for which we found 52% of its proteome identified by *UbPred* with high confidence (*S. cerevisiae* datasets were used to train and test the accuracy of the algorithm). When the same algorithm was applied to *Arabidopsis thaliana*'s proteome, only 39% of the total proteins were highly confident targets for ubiquitylation. In addition, the average number of ubiquitylation sites per protein is much higher in *P. falciparum* (Table 1.1). Following the same trend, lysine is the second most abundant amino acid in *P. falciparum*'s proteome with a frequency close to 12 %, about double the frequencies observed in *S. cerevisiae* and *A. thaliana*, *i.e.,* 5 % and 6 % respectively (Figure 1.2B).

Table 1.1 - Ubiquitin target predictions by *UbPred* (Radivojac *et al*, 2010)

Organism	Confidence Level	Number of proteins	Percentage of the total proteome	Average number of site per protein
Plasmodium falciparum	Low to high	5036	92.5	18.7
	Medium to high	4446	81.6	14.6
	High	**3516**	**64.6**	**8.7**
Saccharomyces cerevisiae	Low to high	5258	89.3	7.9
	Medium to high	4470	76.0	6.0
	High	3059	52.0	4.1
Arabidopsis thaliana	Low to high	42953	82.9	5.6
	Medium to high	34834	67.2	4.3
	High	20214	39.0	3.1

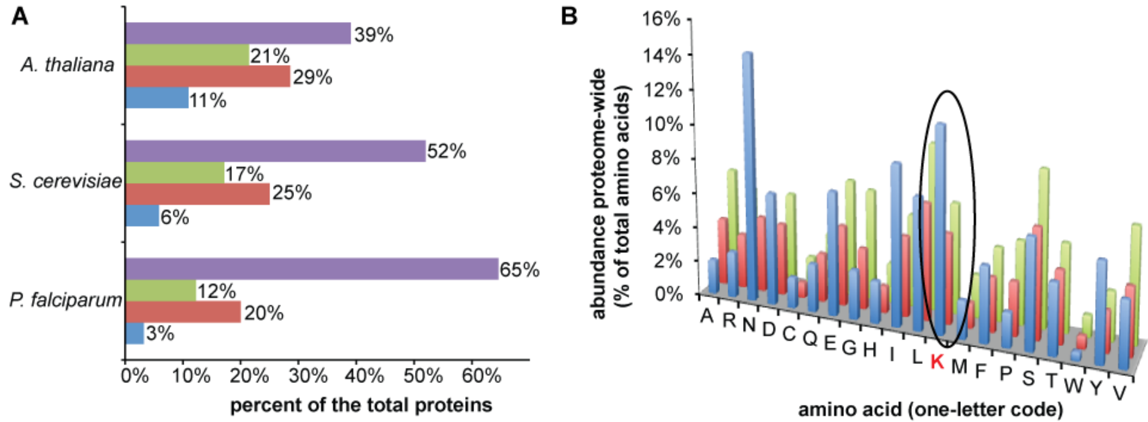


Figure 1.2 - Comparison of proteomes for *P. falciparum*, *S. cerevisiae* and *A. thaliana*.

(A) Portion of the total proteome that, according to the *UbPred* algorithm, is not a target for ubiquitylation (blue bars) or a target for ubiquitylation according to *UbPred* predictions made with low (red bars), medium (green bars), or high confidence (purple bars). The large majority of *P. falciparum*'s proteins are confident putative targets for ubiquitylation. **(B)** Amino acid content in *P. falciparum* (blue bars), *S. cerevisiae* (red bars), and *A. thaliana* (green bars). *P. falciparum*'s proteome contains a higher proportion of lysines than *S. cerevisiae* and *A. thaliana*.

Slightly different results were obtained using *UbiPred*: 2077 proteins were predicted as ubiquitylated with high confidence, *i.e.*, about 38% of the total proteome. This discrepancy might find its origin in the different datasets used to train the predicting algorithms. Both *UbPred* and *UbiPred* were built using ubiquitylated proteins as positive trainers. The negative datasets (non-ubiquitylated sites), however, were chosen differently: verified non-ubiquitylated proteins were used for the training of *UbPred* whereas putative non-ubiquitylated sites were used

to train *UbiPred*. As a consequence, the likelihood of false negatives is higher for results generated with *UbiPred*, consistent with a lower number of predicted targets for ubiquitylation.

The overlap between *UbPred* and *UbiPred* predictions consisted of 1332 proteins. All together, we identified a total of 4261 highly confident putative targets for ubiquitylation using *UbPred* and *UbiPred* in combination (~78% of the parasite's proteome). The high lysine content of the parasite's proteins and the relative abundance of predicted ubiquitin targets may indicate major roles for ubiquitylation in *P. falciparum*'s biology.

We examined in more details the cellular localization of the 4261 highly confident ubiquitin targets identified with *UbPred* and *UbiPred* (used in combination). We found 299 of them contain an apicoplast-targeting signal according to the plasmoAP rules (Foth *et al*, 2003; Tonkin *et al*, 2008), 233 and 154 contain a PEXEL (Marti *et al*, 2004, 2005) or HT (Hiller *et al*, 2004) signal, respectively, and are targeted to the human host erythrocyte membrane, and 1288 have one or more transmembrane domain (Krogh *et al*, 2001) (data queries made on PlasmoDB v6.5). These values do not represent any significant enrichment of ubiquitylated proteins according to the cellular compartment.

MudPIT identification of ubiquitin conjugates in P. falciparum asexual stages

Protein extracts were prepared from *P. falciparum* cultures harvested at ring, trophozoite, or schizont stage. Ubiquitin conjugates were then immunoprecipitated with an anti-ubiquitin antibody that does not recognize free ubiquitin (see Materials and Methods), and visualized by anti-ubiquitin western blot. We tested the immunoprecipitation procedure on an asynchronous population and confirmed that the antibody selectively enriches the sample in ubiquitin conjugates and does not precipitate free ubiquitin (Figure 1.3A). The procedure was performed on stage-specific extract. We found that all asexual intra-erythrocytic stages contain ubiquitylated proteins (Figure 1.3B).

The immunoprecipitated ubiquitin-conjugates were further analyzed by MudPIT (see Materials and Methods). A total of 437 proteins were detected from the anti-ubiquitin immunoprecipitated samples. To increase the confidence of our analysis, we rejected all the proteins that were not significantly enriched in the ubiquitin immunoprecipitations compared to samples precipitated in the absence of specific antibody (see Materials and Methods). This method minimizes the possible number of false positives. The counterpart is nonetheless a higher rate of false negatives since proteins that precipitate non-specifically in the absence of anti-ubiquitin antibody can still be conjugated to ubiquitin. Using this filter, we find 73 *Plasmodium* proteins, ~1.3% of the translated genome, that were specifically

detected in the ubiquitin immunoprecipitations, 24 of them with low spectral counts. For comparison, previous works using similar technique (same antibody and same denaturing conditions) identified 345 ubiquitin conjugates in human cells (Matsumoto *et al*, 2005), ~ 1.7 % of the translated genome, and 200 ubiquitin conjugates in *Arabidopsis thaliana* (Manzano *et al*, 2008), ~ 0.8 % of the translated genome).

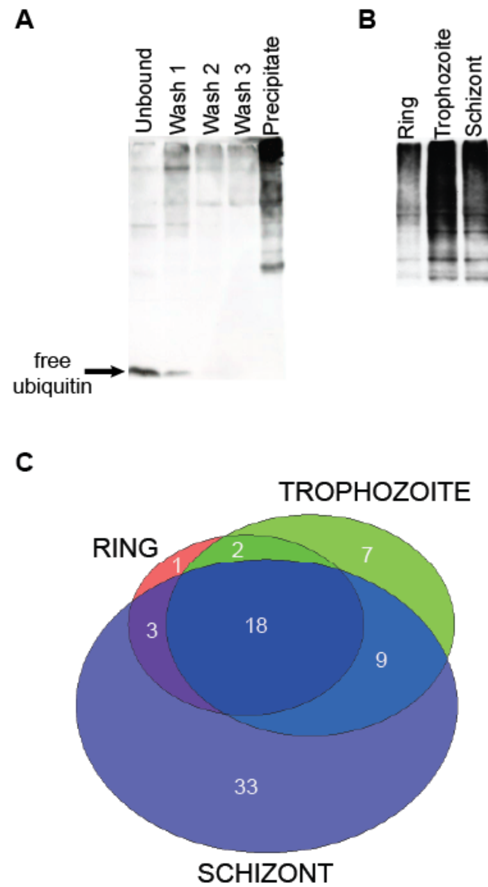


Figure 1.3 - Selective immunoprecipitation and MudPIT identification of ubiquitin-conjugates.

(A) Antibody performances were tested on an asynchronous population of *P. falciparum* during its asexual cycle. The sample was first incubated with anti-ubiquitin antibody and then agarose beads. Proteins that were not captured by the antibody (unbound) were removed by centrifugation. The beads with conjugates antibodies-ubiquitylated proteins were washed three times (wash 1 to wash 3) before the elution of specifically bound ubiquitylated proteins (precipitate). Our results show a significant enrichment in ubiquitin-conjugates in the immunoprecipitate without the presence of free ubiquitin.
(B) Ubiquitin-conjugates content in 10 µL of immunoprecipitated proteins for rings, trophozoites, and schizonts. **(C)** Venn diagram representing the number of different ubiquitin-conjugates identified by MudPIT in each morphological stage.

Ubiquitin-conjugates were the most abundant in schizont stage, with 63 proteins identified (17 of which with low abundance) out of 73 (Figure 1.3C). A total of 33 proteins were specific to this stage (15 of them were found in the low abundance dataset). Among them, 50 are present in the highly confident *in silico*-predicted dataset and 22 were predicted with low to medium confidence with at least one of the algorithms (including ubiquitin itself, gene PFL0585w). The 40S ribosomal protein S15A (PFC0735w) was the only protein identified by neither *UbPred* nor *UbiPred*. Remarkably, ubiquitin (the polyubiquitin gene PFL0585w contains five ubiquitin repeats) was found in the low to medium confidence dataset. Ubiquitin contains seven lysine residues that can all serve for the extension of ubiquitin chains: K6, K11, K27, K29, K33, K48, and K63. *UbPred* identified K63 with low confidence, and *UbiPred* identified K33 (low confidence, score = 0.53) and K48 (medium/high confidence, score = 0.83). The scores for K6, K11, K27, and K29 were below threshold (data not shown). Our MudPIT analysis, nonetheless, found evidences of ubiquitylated K6 and K11 in the samples. Spectra matching such modified peptides from the ubiquitin proteins represented 1.3% and 13.3% of the total spectral count, respectively (Table 2).

Table 1.2. Ubiquitin marks on ubiquitin itself found in *P. falciparum*

Lysine	Number of spectra found in MudPIT†	Score for *UbPred* predictions
K6	8 (1.3%)	0.49
K9	0	0.38
K11	79 (13.3%)	0.54
K27	0	0.38
K29	0	0.56
K48	477 (80.4%)	0.54
K63	29 (4.9%)	0.62

†Out of 593 spectra obtained merging all runs

K48 and K63 were also detected as modified (80.4% and 4.9% of the spectra). K48 is the most abundant linkage usually found in eukaryotic cells and is involved in targeting ubiquitin-tagged substrates to the proteasome 26S for degradation. The roles of K6, K11, K27, K29, and K33 are still not understood (see (Kirkpatrick *et al*, 2005) for a review). These results show that the linkages K6, K11, K48, and K63 (at least) exist in *P. falciparum*.

We found 24 and 36 ubiquitylated proteins in rings and trophozoites, respectively. The 60S ribosomal protein L27a (PFF0885w) was found to be ubiquitylated at ring stage exclusively (5 spectral counts), and seven ubiquitin-conjugates were trophozoite-specific (Table 3). Two of them are members of the ubiquitin/proteasome system in *P. falciparum* (Ponts *et al*, 2008). UFD2 (PFD0265w) is a U-box ubiquitin ligase involved in the endoplasmic reticulum-associated protein degradation pathway (ERAD), and PFD0265w encodes a metallo-

deubiquitinase (mov34) closely related to the PRP8 RNA splicing factor. PF10_0068, PFB0715w, and PF14_0096 are also involved in the metabolism of RNA.

Table 1.3. Ubiquitin-conjugates specifically found in trophozoite stage†

Systematic name	Product description	Function within the ubiquitin/proteasome system††	Spectral count(s)
PF10_0068	RNA binding protein, putative	N/A†††	5
PFB0715w	DNA-directed RNA polymerase II second largest subunit, putative	N/A	2
PFD0265w	pre-mRNA splicing factor, putative	deubiquitinase (mov34)	3
PFF1415c	Heat shock DnaJ protein, putative	N/A	2
MAL8P1.103	conserved *Plasmodium* protein, unknown function	N/A	2
PF08_0020	ubiquitination-mediated degradation component, putative	UFD2 U-box ubiquitin ligase	6
PF14_0096	RNA binding protein Bruno, putative (HoBo)	N/A	2

†Gene names and product descriptions are given according to www.geneDB.org unless specified otherwise.
††Annotation proposed based on protein domain architecture (Ponts *et al*, 2008).
†††Not applicable.

Functional classification of the ubiquitin conjugates identified by MudPIT

We examined in more details the characteristics of the MudPIT dataset. We classified the 73 proteins into 11 functional categories according to their annotation in PlasmoDB v6.5 (Figure 1.4). We find 12% of the dataset consist of chaperonin and other proteins involved in folding, 11% of proteins involved in translation (mostly

ribosomal proteins), RNA metabolism (such as splicing factors), and ubiquitin-dependant metabolic processes. In human cells proteins involved in translation/protein synthesis represent 18 % of the immunoprecipitated ubiquitin-conjugates (Matsumoto *et al*, 2005). This value drops down to 4.5 % in *A. thaliana* (Manzano *et al*, 2008). Proteins involved in transcription represent 4% of the dataset, *i.e.,* twice as much as the proportion found in human cells (Matsumoto *et al*, 2005), but half that of plant (Manzano *et al*, 2008). This group contained a subunit of the RNA polymerase II (PFB0715w) and the putative ApiAP2 transcription factor PFF0200c that is mainly expressed and found ubiquitylated at schizont stage only (spectral count = 4).

Proteins involved in parasite-specific processes (invasion, hemoglobin metabolism, liver stage-specific) represent 8% of the dataset. Remarkably, 11% of the dataset contained proteins of the ubiquitin-proteasome system, an observation consistent with the fact that ubiquitin is covalently and sequentially attached to the various components of the ubiquitin machinery before being transferred on the targeted substrates (Table 1.4). Among them are components of the 26S proteasome (PF08_0109, PF13_0063), the E2 enzymes PFC0255c homolog to MMS2, and two ubiquitin ligases (the “E4” ligase UFD2 PF08_0020, and the zinc “RING” finger, PF10_0046 homolog to the *A. thaliana* CIP8). Altogether, our observations showed that ubiquitin targets were involved in a wide array of biological processes in the parasite including pathogenicity.

Table 1.4 - Components of the ubiquitin/proteasome system identified by immunoprecipitation coupled to MudPIT†

Gene name	Systematic name	Product description	Stage(s)	Spectral count(s)*
Polyubiquitin	PFL0585w	polyubiquitin	Ring, Trophozoite, Schizont	266, 632, 265
Mms2††	PFC0255c	ubiquitin conjugating enzyme E2, putative	Trophozoite, Schizont	2, 8
Ufd2	PF08_0020	ubiquitination-mediated degradation component, putative (U-box ubiquitin ligase)	Trophozoite	6
CIP8	PF10_0046	E3 "RING" ubiquitin ligase, putative	Ring, Trophozoite, schizont	2, 3, 2
N/A	PF11_0142	Ubiquitin domain-containing protein	Trophozoite, Schizont	6, 17
N/A	PF08_0109	Proteasome subunit alpha type 5, putative	Trophozoite, Schizont	2, 6
N/A	PF13_0063	26S proteasome regulatory subunit 7, putative	Trophozoite, schizont	4, 4
Sumo	PFE0285c	Small ubiquitin-related modifier, putative	Ring, Trophozoite, Schizont	2, 3, 4

*Average spectral count among all replicates with a positive count.
†Gene names and product descriptions are given according to www.geneDB.org unless specified otherwise.
††According to BLASTP results performed on the NCBI refseq protein database.
†††Not available

DISCUSSION

The present study proposes more than 4200 putative ubiquitin substrates and validates ~2% of them for ubiquitylation during the intra-erythrocytic cycle of the human malaria parasite *P. falciparum*. The results obtained *in silico* using the two distinct algorithms *UbPred* and *UbiPred* show good overlap. The discrepancies observed between the two algorithms probably originate from their differences of conception. Both *UbPred* and *UbiPred* were built using positive (ubiquitylated) and negative (non-ubiquitylated sites) datasets, with an important difference in the choice of the negative dataset. *UbPred* was trained to recognize non-ubiquitylated sites on *experimentally verified* non-ubiquitylated proteins. The negative dataset for *UbiPred*, on the other hand, consisted of *putative* non-ubiquitylated sites more likely to contain more false negatives that the experimentally verified one. As a consequence, the likelihood of false negatives is higher for results generated with *UbiPred*, consistent with less predicted targets for ubiquitylation. In addition, we found that the two algorithms perform differently depending on the type of linkage that is considered. These differences probably find roots within the nature of training datasets. Less abundant linkages are more difficult to detect. *P. falciparum* seems to contain a relatively high proportion of K11 linkages, 13.3 %. High levels of K11-linkages have been observed in *Saccharomyces cerevisiae* and are suspected to be of major importance in endoplasmic reticulum-associated degradation (ERAD) of proteins (Xu *et al*, 2009). *Plasmodium* possesses a conventional ERAD system. In addition, there are growing evidences of an ERAD-like pathway that is involved in

protein trafficking to the parasite- specific plastid, namely the apicoplast (Spork *et al*, 2009). In humans, K11-linkages are suspected to play a key role in mitotic protein degradation (Matsumoto *et al*, 2010). The potential importance of such linkages in *Plasmodium* (suggested by their abundance) could be major in terms of cell cycle progression, parasite survival, and virulence.

Altogether, we found a high proportion of the proteome being possible targets for ubiquitylation across the various stages of the parasite's life cycle. This observation is consistent with a particularly high lysine content in *P. falciparum*'s proteome that offers many possible anchor points for ubiquitin. In addition, these results represent other pieces of evidence that post-translational modifications in general and ubiquitylation in particular are major regulatory pathways in the malaria parasite.

Our analysis also highlights the limits of the technical approach that was used to identify ubiquitin-conjugates at a large scale. The main challenge is indeed to obtain a reliable negative control to exclude false positives while minimizing the number of false negatives: non-specific interactions do not necessarily invalidate the specific ones. As is, we identified 73 *Plasmodium* ubiquitin conjugates together with 364 potential additional candidates. The rejection of these 364 candidates is directly linked to the baseline drawn by the chosen negative control. The question of how to choose a good control to draw a *blank* baseline receives, indeed, no perfect answer.

Using various experimental strategies, e.g. protein capture *via* ubiquitin-binding domains, could be a piece of solution by permitting a broader discovery of the various ubiquitin conjugates.

The proteins in our dataset of ubiquitin conjugates are involved in a variety of biological processes, such as translation. Many ribosomal proteins are ubiquitylated during the red blood cell cycle. In human cells, the ribosomal subunits S3, S18, L23a, L24, and L28 were previously shown to be ubiquitylated (Matsumoto *et al*, 2005; Spence *et al*, 2000). Our results are consistent with a function for ubiquitylation in the quality control of ribosome biogenesis. Ribosome assembly, indeed, is a highly complex association of numerous and various proteins thus prone to misfolding and other maturation errors. The high proportion of ubiquitylated proteins in our dataset may reflect this propensity to errors and the subsequent targeting to the proteasome after ubiquitin tagging. Another hypothesis is that ubiquitylation of ribosomal proteins plays a role in the regulation of translation, as previously shown in human cells (Spence *et al*, 2000).

Our dataset also contained 11 proteins related to the ubiquitin/proteasome system itself. Since all interactions between ubiquitin and the enzymes catalyzing its transfer to various targets (including ultimately the proteasome) involve the formation of a covalent bond, the identification of a subset of these enzymes is a validation of our dataset. Most of the proteins that we identified are subunits of the

proteasome. In addition, we also found one E2 ubiquitin-conjugating enzymes, two ubiquitin ligases, and the ubiquitin-like PfSUMO (PFE0285c). These results may indicate that ubiquitylation regulates other small modifiers pathways.

Most of the ubiquitin-conjugates that we identified were found during the schizont stage, consistent with a reduction in size and activity of the parasite for maturation into the invasive merozoite. We nonetheless showed that a smaller number of proteins which ubiquitin conjugates were detected in the other intra-erythrocytic stages.

Our results provide the first general view of the *P. falciparum*'s ubiquitome and represent an important step towards the characterization of the ubiquitylation system in the parasite. This study identifies various new targets for ubiquitylation and provides numerous leads for investigating the malaria parasite's biology through post-translational modifications. For example, ubiquitylation can play a role in regulating gene expression at the transcriptional level, as reflected by the identification of a major subunit of the RNA polymerase II and of an ApiAP2 transcription factor in the confident dataset of ubiquitylated proteins. Given the importance of ubiquitylation in the parasite's biology, the ubiquitin targets shown in the present work could represent excellent objects of research for future development of antimalarial drugs.

CONCLUDING REMARKS FOR CHAPTER 1

From our study, ubiquitylation seems to play a prominent role in coordinating the various biological functions with the *Plasmodium*. Not surprisingly, the abundance and specificity of certain ubiquitylated proteins is stage-specific and implicated in a wide range of roles such as RNA metabolism, translation, transcription and, of course, protein degradation. Even more interesting is that we have found a significant portion of the ubiquitome to be involved in parasite-specific roles such as parasite invasion and hemoglobin metabolism. It may be worthwhile to investigate ubiquitylation in parasite-specific roles and possibly exploit them for novel drug target candidates.

MATERIAL AND METHODS

Parasite strain and culture conditions

Sorbitol-synchronized *P. falciparum* parasite strain 3D7 was cultured in human erythrocytes according to previously described protocols (Lambros & Vanderberg, 1979; Le Roch *et al*, 2003; Trager & Jensen, 1976). For each experiment, cultures were harvested 48 hours after the first sorbitol treatment (ring stage), and after 18 and 36 hours (respectively trophozoite and schizont morphological stages monitored by Giemsa staining).

Ubiquitylation assays

Parasite protein extracts were freshly prepared before each ubiquitylation assay. 2-5x10^9 parasites were extracted by 0.15% saponin lysis of red blood cells (15 min of incubation on ice) followed by three washes in ice-cold PBS supplemented with 2 mM PMSF. After 15 min of incubation on ice, the parasite pellet was resuspended in one milliliter of ice-cold cytoplasmic lysis buffer (20 mM HEPES pH 7.9, 10 mM KCl, 1 mM EDTA, 1 mM EGTA, 1 mM DTT, 0.65% Igepal®, 0.5 mM PMSF, Roche Complete Mini EDTA-free Cocktail Protease Inhibitor) and left on ice for five more minutes. After 10 min of centrifugation at 4°C and 1500g the supernatant was collected (cytoplasmic protein extract) and the pelleted nuclei were lysed in 100 µL of nuclei lysis buffer (20 mM HEPES pH 7.9, 0.1 mM NaCl, 0.1 mM EDTA, 0.1 mM EGTA, 1.5 mM $MgCl_2$, 1 mM DTT, 25% glycerol, 1 mM PMSF, Roche Complete Mini EDTA-free

Cocktail Protease Inhibitor) for 20 min at 4˚C under vigorous shaking. The nuclear protein extract was then cleared by centrifugation at 4˚C and 6000g for 10 min and combined to the cytoplasmic extract. Protein concentrations were measured by Bradford assay (Bradford, 1976).

Fresh (*i.e.*, prepared the same day and never frozen) protein extracts were used to perform ubiquitylation assays according to a modified previously published protocol (Qin *et al*, 2008). 16.5 µL of protein extract were mixed on ice with 18.5 µL of ubiquitylation reaction mix (20 mM HEPES pH 7.9, 1.5 mM DTT, 1X Energy Regeneration Solution (BostonBiochem®), five micrograms of biotinylated ubiquitin (BostonBiochem®), and 20 µg of native ubiquitin (BostonBiochem®), and incubated at 30˚C under gentle agitation. After two hours of incubation, the reaction was mixed with Laemmli buffer (Laemmli, 1970), run on a 10% SDS-PAGE, and transferred onto a PVDF membrane. After two hours of incubation with streptavidine proteins coupled to horseradish peroxidase (Jackson ImmunoResearch Laboratories, Inc.), the presence of incorporated biotin-ubiquitin was revealed by electroluminescence (Pierce® ECL Western blotting substrate).

Immunofluorescent localization of ubiquitylated proteins

Immunofluorescence detection of ubiquitylated proteins was performed according to a modified version of a previously described protocol (Tonkin *et al*, 2004). Infected red blood cells were washed once with PBS and an aliquot was

resuspended in 300 to 400 volumes of PBS containing 1% BSA (w/v). 30 μL aliquots of cell suspension were spotted on the surface of a depression slide and completely air-dried. The cells were then fixed for 10 min at room temperature with 4% paraformaldehyde in PBS. After rinsing the slide once with PBS, the cells were blocked with 1% BSA in PBS (w/v) for 30 min at room temperature in a humidified chamber. The liquid was then removed and the slide was incubated with a 1:250 dilution of anti-conjugated ubiquitin antibody (clone FK2, Enzo® Life Sciences) in solution with 1% BSA/PBS (w/v) for two hours at room temperature in a humidified chamber. After liquid removal and three washes with PBS for five minutes, the slide was incubated with 1:250 of secondary antibody (Texas Red®-X goat anti-mouse IgG, Invitrogen™) in solution with 1% BSA/PBS (w/v) for one hour at room temperature in a humidified chamber After liquid removal and three washes with PBS for five minutes, nuclei were stained with DAPI 100 ng/mL for five minutes. The slide was then carefully washed three times five minutes with PBS. The slide was finally mounted with slow-fade mounting medium (Fluoromount-G, Southern Biotech) and images were viewed with a fluorescence microscope within 72 hours of preparing the slide. Images were prepared and mounted using the software ImageJ (http://rsbweb.nih.gov/ij/).

***In silico* identification of putative targets for ubiquitylation**

The complete set of 5446 translated coding sequences for *P. falciparum* was downloaded from PlasmoDB v6.5 (www.plasmodb.org) and analyzed for the

presence of putative ubiquitylation sites using *UbPred* (Radivojac *et al*, 2010) and *UbiPred* (Tung & Ho, 2008). Predictions with *UbPred* were classified by level of confidence, from low (score ranging from 0.62 to 0.69) to medium (score ranging from 0.70 to 0.84) to high (score equal to or higher than 0.85), according to the software instructions. With regard to the predictions made with *UbiPred*, proteins were considered potentially ubiquitylated with high confidence when prediction scores were equal to at least 0.85 (Tung & Ho, 2008).

Immunoprecipitation of ubiquitylated proteins

Protein ubiquitylation leads to the rapid degradation of the target protein by the proteasome, which is a major obstacle to the isolation and analysis of ubiquitylated proteins. To prevent the degradation of ubiquitylated proteins in the sample, parasite cultures were treated with 400 nM of the proteasome inhibitor MG132 (EMD Chemicals Inc.) six hours before harvest. In addition, ubiquitin moieties are rapidly removed *in vivo* by deubiquitylating enzymes (DUB). To ensure maximum detection of targets for ubiquitylation, all reagents were supplemented with 20 mM of N-ethylmaleimide, 2 mM PMSF, and complete mini EDTA-free protease inhibitor cocktail (Roche Diagnostics). Cultures were harvested by five minutes of centrifugation at 800g and 4˚C, followed by three washes in PBS supplemented with NEM, PMSF and anti-protease cocktail. Parasites were extracted with 0.15% w/v saponin (in PBS) for 15 min on ice. Cells were pelleted by centrifugation for 10 min at 3200g and 4˚C, and washed with PBS (supplemented with NEM, PMSF and anti-

protease cocktail) until the supernatant came out clear. Parasite pellets were then flash frozen and stored at -80˚C until immunoprecipitation.

Parasites were thawed in immunoprecipitation buffer (0.5% Triton X-100, 50 mM Tris-HCl pH 7.5, 300 mM NaCl, 20 mM N-ethylmaleimide, 0.5 mM EDTA, 2 mM PMSF, Roche complete mini EDTA-free protease inhibitor cocktail) and lysed by six strokes of sonication (10s each) on a Fisher Dismembrator Model 100 set to power $2^{1/2}$. Lysates were cleared by 15 min of centrifugation at 13000g and 4˚C, and incubated with washed agarose A beads (Invitrogen™) for one hour at 4˚C under constant agitation. After 10 min of centrifugation at 13000g and 4˚C, the supernatant was incubated with anti-conjugated ubiquitin mouse IgG_1 (clone FK2 that does not react with free ubiquitin, Enzo® Life Sciences) overnight at 4˚C. In parallel, negative control samples were left without antibody (beads only). The conjugates and the negative control samples were then incubated with washed agarose A beads for two hours at 4˚C. The beads were collected by 30s of centrifugation at 4000g, washed three times in immunoprecipitation buffer, and eluted with Laemmli buffer. Immunoprecipitates were analyzed by 12% SDS-PAGE, transfer on PVDF membrane, incubation with a rabbit anti-ubiquitin IgG (Millipore) followed by immunodetection with a goat anti-rabbit IgG coupled to horseradish peroxidase (Millipore™) and revelation by electroluminescence (Pierce® ECL Western blotting substrate). Alternatively, immunoprecipitated proteins and

negative controls were eluted three times from the beads with 50 μL of elution buffer (50 mM Tris-HCl pH 6.8, 100 mM DTT, 2% SDS) before MudPIT analysis.

MudPIT analysis of immunoprecipitated proteins

Immunoprecipitated proteins were TCA-precipitated and the pellets were solubilized in 100 mM TRIS-HCl pH 8.5 and 8 M Urea; TCEP (Tris(2-Carboxylethyl)-Phosphine Hydrochloride, Pierce) and CAM (Chloroacetamide, Sigma) were added to a final concentration of 5 mM and 10 mM, respectively. Protein suspensions were digested overnight at 37°C using Endoproteinase Lys-C at 1:50 w/w (Roche). Samples were brought to a final concentration of 2 M urea and 2 mM $CaCl_2$ before performing a second overnight digestion at 37°C using Trypsin (Promega) at 1:100 w/w. Formic acid (5% final) was added to stop the reactions. Samples were loaded on split-triple-phase fused-silica micro-capillary columns (McDonald *et al*, 2002) and placed in-line with an Eksigent NanoLC 2D system. Full MS spectra were recorded on the peptides over a 400 to 1,700 *m/z* range in the Orbitrap at 60K resolution, followed by fragmentation in the ion trap (at 35% collision energy) on the first to fifth most intense ions selected from the full MS spectrum with dynamic exclusion enabled for 90s (Zhang *et al*, 2010). A total of 3, 3, 4 and 7 technical replicates were acquired for the ring, trophozoite, schizont, and negative controls samples, respectively.

MudPIT data analysis

RAW files were extracted into ms2 file format (McDonald *et al*, 2004) using RAW_Xtract v.1.0 (Venable *et al*, 2004). MS/MS spectra were searched using SEQUEST v.27 (rev.9) (Eng *et al*, 1994) with a peptide mass tolerance of 50 ppm, and searched against a protein database combining non-redundant 5439 *Plasmodium falciparum* (PlasmoDB release 5.5) and 30552 human proteins (NCBI 2008-03-04 release), as well as 162 usual contaminants such as human keratins, IgGs, and proteolytic enzymes. To estimate false discovery rates (FDR), each protein sequence was randomized (keeping the same amino acid composition and length) and the resulting "shuffled" sequences were added to the database used for the SEQUEST searches, for a total search space of 72306 amino acid sequences. To account for alkylation by CAM, +57 Da were added statically to cysteine residues for all searches. Peptide/spectrum matches were sorted, selected and compared using DTASelect/CONTRAST (Tabb *et al*, 2002). Peptides had to be full tryptic and at least 7 amino acid long, with a DeltCn cut-off of 0.08, and XCorr minima of 1.8, 2 .0, and 3.0 for singly-, doubly-, and triply-charged spectra. Proteins had to be detected by two such peptides or 1 peptide with 2 spectra in each run. NSAF v7, an in-house developed software, was used to create the final report on all non-redundant proteins detected across the different MudPIT runs, and calculate FDRs, which were on average 0.4% ± 0.3 and 0.7% ± 0.4 at the spectral and protein levels, respectively. Protein abundance was evaluated according to their normalized spectral abundance factors (dNSAF) calculated as the spectral count relative to the length of a given

protein normalized by the sum of the relative spectral counts in the sample (Zhang *et al*, 2010). We used the open source software package "plgem" (written in R and maintained by the BioConductor project) to establish some statistical significance on this dataset (Pavelka *et al*, 2008). Briefly, the runs with the most replicates (negative controls) were used to fit a Power Law Global Error Model on the mean versus standard deviation plots. Then STN ratios against Control as a baseline were calculated using observed mean values and PLGEM-calculated standard deviation values:

$$STN = \frac{ObsMean_{Stage} - ObsMean_{Control}}{CalcSD_{Stage} + CalcSD_{Control}}$$

Finally a series of random re-sampled STN ratios were computed to calculate significance thresholds (p-values) of PLGEM-based STN values. Proteins not detected in negative controls or with p-values ≤ 0.05 in at least one of the three stages examined were considered potential ubiquitylated conjugates. The entire MS/MS dataset was queried for the presence of lysines modified by 114.04 Da ("ubiquitin marks") on the 49 proteins with significant p-values as previously described (Xiang *et al*, 2007).

REFERENCES

Berkers CR & Ovaa H (2010) Drug discovery and assay development in the ubiquitin-proteasome system. *Biochem. Soc. Trans.* **38:** 14–20

Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72:** 248–254

Chung D-WD, Ponts N, Cervantes S & Le Roch KG (2009) Post-translational modifications in Plasmodium: more than you think! *Mol. Biochem. Parasitol.* **168:** 123–134

Edelmann MJ & Kessler BM (2008) Ubiquitin and ubiquitin-like specific proteases targeted by infectious pathogens: Emerging patterns and molecular principles. *Biochim. Biophys. Acta* **1782:** 809–816

Eng JK, McCormack AL & Yates III JR (1994) An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *Journal of the American Society for Mass Spectrometry* **5:** 976–989

Field SJ, Pinder JC, Clough B, Dluzewski AR, Wilson RJ & Gratzer WB (1993) Actin in the merozoite of the malaria parasite, Plasmodium falciparum. *Cell Motil. Cytoskeleton* **25:** 43–48

Foth BJ, Ralph SA, Tonkin CJ, Struck NS, Fraunholz M, Roos DS, Cowman AF & McFadden GI (2003) Dissecting Apicoplast Targeting in the Malaria Parasite Plasmodium falciparum. *Science* **299:** 705 –708

Golebiowski F, Tatham MH, Nakamura A & Hay RT (2010) High-stringency tandem affinity purification of proteins conjugated to ubiquitin-like moieties. *Nat Protoc* **5:** 873–882

Hatakeyama S, Matsumoto M & Nakayama KI (2005) Mapping of ubiquitination sites on target proteins. *Meth. Enzymol.* **399:** 277–286

Hiller NL, Bhattacharjee S, van Ooij C, Liolios K, Harrison T, Lopez-Estraño C & Haldar K (2004) A host-targeting signal in virulence proteins reveals a secretome in malarial infection. *Science* **306:** 1934–1937

Hjerpe R, Aillet F, Lopitz-Otsoa F, Lang V, England P & Rodriguez MS (2009) Efficient protection and isolation of ubiquitylated proteins using tandem ubiquitin-binding entities. *EMBO Rep.* **10:** 1250–1258

Kirkpatrick DS, Denison C & Gygi SP (2005) Weighing in on ubiquitin: the expanding role of mass-spectrometry-based proteomics. *Nat. Cell Biol.* **7:** 750–757

Krogh A, Larsson B, von Heijne G & Sonnhammer EL (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **305:** 567–580

Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227:** 680–685

Lambros C & Vanderberg JP (1979) Synchronization of Plasmodium falciparum erythrocytic stages in culture. *J. Parasitol.* **65:** 418–420

Lindenthal C, Weich N, Chia YS, Heussler V & Klinkert MQ (2005) The proteasome inhibitor MLN-273 blocks exoerythrocytic and erythrocytic development of Plasmodium parasites. *Parasitology* **131:** 37–44

Manzano C, Abraham Z, López-Torrejón G & Del Pozo JC (2008) Identification of ubiquitinated proteins in Arabidopsis. *Plant Mol. Biol.* **68:** 145–158

Marti M, Baum J, Rug M, Tilley L & Cowman AF (2005) Signal-mediated export of proteins from the malaria parasite to the host erythrocyte. *J. Cell Biol.* **171:** 587–592

Marti M, Good RT, Rug M, Knuepfer E & Cowman AF (2004) Targeting malaria virulence and remodeling proteins to the host erythrocyte. *Science* **306:** 1930–1933

Matsumoto M, Hatakeyama S, Oyamada K, Oda Y, Nishimura T & Nakayama KI (2005) Large-scale analysis of the human ubiquitin-related proteome. *Proteomics* **5:** 4145–4151

Matsumoto ML, Wickliffe KE, Dong KC, Yu C, Bosanac I, Bustos D, Phu L, Kirkpatrick DS, Hymowitz SG, Rape M, Kelley RF & Dixit VM (2010) K11-linked polyubiquitination in cell cycle control revealed by a K11 linkage-specific antibody. *Mol. Cell* **39:** 477–484

McDonald WH, Ohi R, Miyamoto DT, Mitchison TJ & Yates III JR (2002) Comparison of three directly coupled HPLC MS/MS strategies for identification of proteins from complex mixtures: single-dimension LC-MS/MS, 2-phase MudPIT, and 3-phase MudPIT. *International Journal of Mass Spectrometry* **219:** 245–251

McDonald WH, Tabb DL, Sadygov RG, MacCoss MJ, Venable J, Graumann J, Johnson JR, Cociorva D & Yates III JR (2004) MS1, MS2, and SQT—three unified,

compact, and easily parsed file formats for the storage of shotgun proteomic spectra and identifications. *Rapid Communications in Mass Spectrometry* **18:** 2162–2168

Pavelka N, Fournier ML, Swanson SK, Pelizzola M, Ricciardi-Castagnoli P, Florens L & Washburn MP (2008) Statistical similarities between transcriptomics and quantitative shotgun proteomics data. *Mol. Cell Proteomics* **7:** 631–644

Petroski MD (2008) The ubiquitin system, disease, and drug discovery. *BMC Biochem.* **9 Suppl 1:** S7

Ponder EL & Bogyo M (2007) Ubiquitin-Like Modifiers and Their Deconjugating Enzymes in Medically Important Parasitic Protozoa. *Eukaryotic Cell* **6:** 1943–1952

Ponts N, Yang J, Chung D-WD, Prudhomme J, Girke T, Horrocks P & Le Roch KG (2008) Deciphering the ubiquitin-mediated pathway in apicomplexan parasites: a potential strategy to interfere with parasite virulence. *PLoS ONE* **3:** e2386

Prudhomme J, McDaniel E, Ponts N, Bertani S, Fenical W, Jensen P & Le Roch K (2008) Marine Actinomycetes: A New Source of Compounds against the Human Malaria Parasite. *PLoS ONE* **3:**

Qin F, Sakuma Y, Tran L-SP, Maruyama K, Kidokoro S, Fujita Y, Fujita M, Umezawa T, Sawano Y, Miyazono K-I, Tanokura M, Shinozaki K & Yamaguchi-Shinozaki K (2008) Arabidopsis DREB2A-interacting proteins function as RING E3 ligases and negatively regulate plant drought stress-responsive gene expression. *Plant Cell* **20:** 1693–1707

Radivojac P, Vacic V, Haynes C, Cocklin RR, Mohan A, Heyen JW, Goebl MG & Iakoucheva LM (2010) Identification, analysis, and prediction of protein ubiquitination sites. *Proteins* **78:** 365–380

Reynolds JM, El Bissati K, Brandenburg J, Günzl A & Mamoun CB Antimalarial activity of the anticancer and proteasome inhibitor bortezomib and its analog ZL3B. *BMC Clin Pharmacol* **7:** 13-13

Le Roch KG, Zhou Y, Blair PL, Grainger M, Moch JK, Haynes JD, De La Vega P, Holder AA, Batalov S, Carucci DJ & Winzeler EA (2003) Discovery of gene function by expression profiling of the malaria parasite life cycle. *Science* **301:** 1503–1508

Spallek T, Robatzek S & Göhre V (2009) How microbes utilize host ubiquitination. *Cell. Microbiol.* **11:** 1425–1434

Spence J, Gali RR, Dittmar G, Sherman F, Karin M & Finley D (2000) Cell cycle-regulated modification of the ribosome by a variant multiubiquitin chain. *Cell* **102:** 67–76

Spork S, Hiss JA, Mandel K, Sommer M, Kooij TWA, Chu T, Schneider G, Maier UG & Przyborski JM (2009) An Unusual ERAD-Like Complex Is Targeted to the Apicoplast of Plasmodium falciparum. *Eukaryotic Cell* **8:** 1134–1145

Tabb DL, McDonald WH & Yates JR 3rd (2002) DTASelect and Contrast: tools for assembling and comparing protein identifications from shotgun proteomics. *J. Proteome Res.* **1:** 21–26

Tan F, Lu L, Cai Y, Wang J, Xie Y, Wang L, Gong Y, Xu B-E, Wu J, Luo Y, Qiang B, Yuan J, Sun X & Peng X (2008) Proteomic analysis of ubiquitinated proteins in normal hepatocyte cell line Chang liver cells. *Proteomics* **8:** 2885–2896

Tomlinson E, Palaniyappan N, Tooth D & Layfield R (2007) Methods for the purification of ubiquitinated proteins. *Proteomics* **7:** 1016–1022

Tonkin CJ, van Dooren GG, Spurck TP, Struck NS, Good RT, Handman E, Cowman AF & McFadden GI (2004) Localization of organellar proteins in Plasmodium falciparum using a novel set of transfection vectors and a new immunofluorescence fixation method. *Mol. Biochem. Parasitol* **137:** 13–21

Tonkin CJ, Foth BJ, Ralph SA, Struck N, Cowman AF & McFadden GI (2008) Evolution of malaria parasite plastid targeting sequences. *Proc. Natl. Acad. Sci. U.S.A.* **105:** 4781–4785

Trager W & Jensen J (1976) Human malaria parasites in continuous culture. *Science* **193:** 673–675

Trelle MB, Salcedo-Amaya AM, Cohen AM, Stunnenberg HG & Jensen ON (2009) Global histone analysis by mass spectrometry reveals a high content of acetylated lysine residues in the malaria parasite Plasmodium falciparum. *J. Proteome Res.* **8:** 3439–3450

Tung C-W & Ho S-Y (2008) Computational identification of ubiquitylation sites from protein sequences. *BMC Bioinformatics* **9:** 310

Venable JD, Dong M-Q, Wohlschlegel J, Dillin A & Yates JR (2004) Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra. *Nat. Methods* **1:** 39–45

Xiang Y, Takeo S, Florens L, Hughes SE, Huo L-J, Gilliland WD, Swanson SK, Teeter K, Schwartz JW, Washburn MP, Jaspersen SL & Hawley RS (2007) The inhibition of polo kinase by matrimony maintains G2 arrest in the meiotic cell cycle. *PLoS Biol.* **5:** e323

Xu P, Duong DM, Seyfried NT, Cheng D, Xie Y, Robert J, Rush J, Hochstrasser M, Finley D & Peng J (2009) Quantitative proteomics reveals the function of unconventional ubiquitin chains in proteasomal degradation. *Cell* **137:** 133–145

Zhang Y, Wen Z, Washburn MP & Florens L (2010) Refinements to label free proteome quantitation: how to deal with peptides shared by multiple proteins. *Anal. Chem.* **82:** 2272–2281

CHAPTER 2

Elucidating the ubiquitin-dependent ERAD system in the human malaria parasite

CHAPTER 2 PREFACE

From our genome-wide investigation of the *Plasmodium* ubiquitome, we have evidence to believe that ubiquitylation is a major regulator of the parasite life cycle. Aside from the more conserved roles of transcription, translation, and RNA metabolism, ubiquitylation also seems to function in many parasite-specific roles such as host cell invasion and hemoglobin degradation. Another major theme that stood out in our previous study is ubiquitin-dependent protein degradation, which is a burgeoning field of interest due to several encouraging reports of proteasome inhibitors that were able to confer antimalarial activity. Despite the growing interest in the *Plasmodium* proteasome system, relatively little investigation has been done to actually characterize the parasite degradation machinery.

In this chapter, we provide an initial biological investigation of the ubiquitylating components of the endoplasmic reticulum-associated degradation (ERAD) system, which is a major pathway in targeting misfolded proteins from the ER to the cytosol for proteasome degradation. Also, in conjunction with the promising proteasome inhibitor studies, we explore the possibility of targeting the *Plasmodium* ERAD system for therapeutic intervention.

INTRODUCTION

Malaria is one of the most deadliest infectious diseases of the world, infecting up to half a billion and killing up to a million people each year (Garcia, 2010). Though there has been progress in combating malaria with multi-drug therapies (van Vugt *et al*, 2011), the high cost treatment and the rise of drug-resistance by the malarial parasites beckon the need for novel and cheap antimalarials. The causative agents of human malaria belong to the *Plasmodium* spp., with *Plasmodium falciparum* as the deadliest species. In hopes of finding possible weak points to exploit, there has been much effort in understanding how the *Plasmodium* regulates its vigorous life cycle.

In recent years, there have been several studies that shown that proteasome inhibitors can have significant antimalarial properties (Lindenthal *et al*, 2005; Reynolds *et al*; Dick *et al*, 1996; Kreidenweiss *et al*; Prudhomme *et al*, 2008), which suggest that protein degradation is an important aspect of *Plasmodium* biology. Though there has been some work in characterizing some of its 26S proteasome (Li *et al*, 2000; Certad *et al*, 1999) and bacteria-like proteasome (Ramasamy *et al*, 2007; Yoo *et al*, 1996; Gardner *et al*, 2002), there has been very little work done on actually characterizing the ubiquitylating machinery of the proteasome degradation system within the *Plasmodium*. Up until now, there has been no functional study that has investigated the ubiquitylating components that make up the ERAD system.

In the classical ERAD model of eukaryotic cells, aberrant proteins are recognized by ER luminal chaperone proteins and protein disulfide isomerases to help discriminate properly folded proteins from misfolded proteins (Xie & Ng, 2010) (Figure 2.1). Misfolded proteins are shuttled to the DER1 translocon complex, which forms a hydrophobic pore to allow the retro-translocation of proteins through the ER membrane. Within this translocon complex, an E3 ubiquitin ligase called HRD1 has multiple functions. It interacts with membrane-bound proteins needed for retro-translocation and helps form the hydrophobic pore complex (Sato *et al*, 2009). It also catalyzes, with the intervention of other ubiquitylating enzymes, the ubiquitylation of the target misfolded protein that is the prerequisite for subsequent retro-translocation to the cytosol and destruction by the 26S proteasome (Bagola *et al*, 2011; Carvalho *et al*, 2006; Hershko & Ciechanover, 1998). Typically, ubiquitylation involves the covalent attachment of a ubiquitin moiety to lysine residues of protein substrates *via* the hierarchical intervention of an E1 ubiquitin-activating enzyme, an E2 ubiquitin-conjugating enzyme, and an E3 ubiquitin ligase that is usually involved in specific substrate recognition (Pickart, 2004; Laney & Hochstrasser, 1999). Ubiquitin can also be specifically cleaved off its targets or matured in the case of ubiquitin-fusion proteins by deubiquitylating enzymes (DUB).

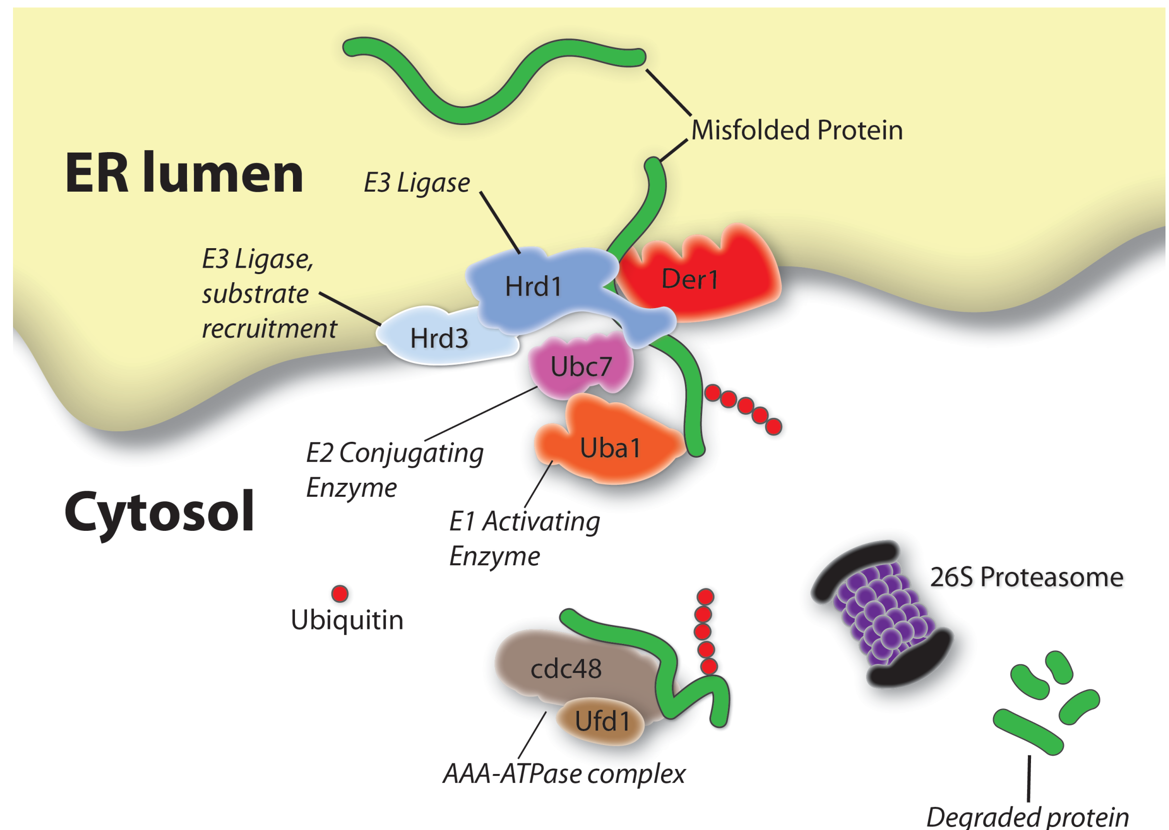


Figure 2.1 – Model of the ERAD system and proteasome degradation
The ERAD system translocates misfolded proteins across the ER membrane, where the aberrant proteins are ubiquitylated by a set of core ubiquitylating enzymes and then shuttled to the 26S proteasome for degradation. Only a few select ERAD components are shown for simplicity.

In order to first detect the components of the ERAD system of the *P. falciparum*, an *in silico* search was performed and a list of probable ERAD candidates was identified. Among the candidates was a group of proteins that consisted of an ubiquitin E1, E2, and E3 enzyme, which makes a full set of ubiquitylating proteins sufficient to promote ubiquitylation. Biological validation of these proteins were performed by constructing recombinant proteins, which were shown to be able to confer *in vitro* ubiquitylation. Immunofluorescence microscopy experiments also

reveals these candidate *Plasmodium* ERAD proteins to localize to their expected respective regions of either the ER membrane or the cytosol. Furthermore, gene disruption and gene replacement experiments suggest that the ubiquitylating components of the *Plasmodium* ERAD system are essential to the parasite survival.

Protein degradation is an essential part of parasite biology as indicated by previous findings that show proteasome inhibitors having significant antimalarial properties (see (Chung & Le Roch, 2010) for a review). Thus, these particular components of the *Plasmodium* ERAD system contribute to our understanding of how protein degradation works within the parasite and represent possible drug target candidates. The analysis presented here is the first functional characterization of the ubiquitylating components of the *P. falciparum* ERAD pathway.

RESULTS

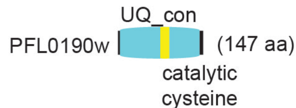
In silico analysis of the Plasmodium ERAD ubiquitylating components

The entire translated *Plasmodium* genome was bioinformatically searched to identify ubiquitylating proteins using a previously published approach (Ponts *et al*, 2008). Based on homology, the most likely ubiquitylating candidates that make up the *Plasmodium* ERAD system were identified: PFL1245w, a putative E1 ubiquitin enzyme; PFL0190w, a putative E2 ubiquitin conjugating enzyme; PF14_0215, a putative E3 ubiquitin ligase.

The putative *Plasmodium* protein that is encoded by the gene PFL1245w has the closest attributes to that of the UBA1, the E1 ubiquitin activating enzyme of the ERAD system. The 1140 amino acid long PFL1245w protein contains an ubiquitin-activating enzyme active site and two ubiquitin-like activating enzymes catalytic domains, with its catalytic cysteine at the N-terminal. PFL0190w was found to have the closest homology to that of the UBC found within the ERAD system. With only 147 amino acids, PFL0190w is a short protein containing a ubiquitin conjugating enzyme domain with its catalytic cysteine found in the middle. (Figure 2.2).

E1 ubiquitin activating enzyme

E2 ubiquitin conjugating enzyme



E3 ring finger ubiquitin ligase

Figure 2.2 – *In silico* domain architecture of *Plasmodium* ERAD ubiquitylating enzymes. Domains were predicted with PFAMscan and pictures were generated with Matlab using an in-house script. Transmembrane (TM) domains were predicted with TMHMM. SP = signal peptide; TM = transmembrane domain; UBQ = ubiquitin domain; UBA/ThiF/UBACT = E1 ubiquitin activating domain; UBQ conj = E2 ubiquitin conjugating domain; zf-C3HC4 = E3 RING finger ubiquitin ligase domain.

Based on homology to the ERAD components of other eukaryotic models, PF14_0215 seems to be closely related to the canonical HRD1 of the ERAD pathway (Sommer *et al*, 2007; Spork *et al*, 2009; Ponts *et al*, 2008). Classically, HRD1 is a multiple transmembrane domain E3 ligase embedded within the ER membrane with its C-terminal end carrying the E3 RING zinc finger (zf-C3HC4) domain facing the cytosol. In addition to having a highly probable signal peptide for ER targeting, PF14_0215 has four predicted N-terminal transmembrane domains and a C-terminal zf-C3HC4 domain, consistent with HRD1 domain architecture (Figure 2.2). Its feature of multiple domains is compatible with an ability to form pores to

participate in the recognition and translocation of misfolded proteins across the ER membrane. Considering these homologies, PF14_0215 is the likely parasite HRD1 involved in the ERAD system.

PF14_0215 is a genuine E3 ubiquitin ligase that localizes in the ER membrane and is likely essential for parasite survival

To determine the function of the HRD1 homologue PF14_0215, its RING domain was cloned in *E. coli*, fused to tandem GST and 6xHIS tags, expressed, and purified (Figure 2.3A). *In vitro* ubiquitylation assays were performed by incubating the purified recombinant PF14_0215 RING domain with commercially available human E1 (UBA1), various human E2 (UBC) enzymes, and commercial purified ubiquitin (Figure 2.3B). Anti-ubiquitin immunoblots show that different patterns of poly-ubiquitin tails were obtained when our recombinant E3 and the commercial E1 were mixed with three of the eight tested E2 conjugating enzymes, UBCH5a, UBCH6, and $UBCH13_{complex}$. When PF14_0215 was removed from the assay, poly-ubiquitin tails were no longer visible (Figure 2.3C), which demonstrates that PF14_0215 catalyzes the formation of poly-ubiquitin tails.

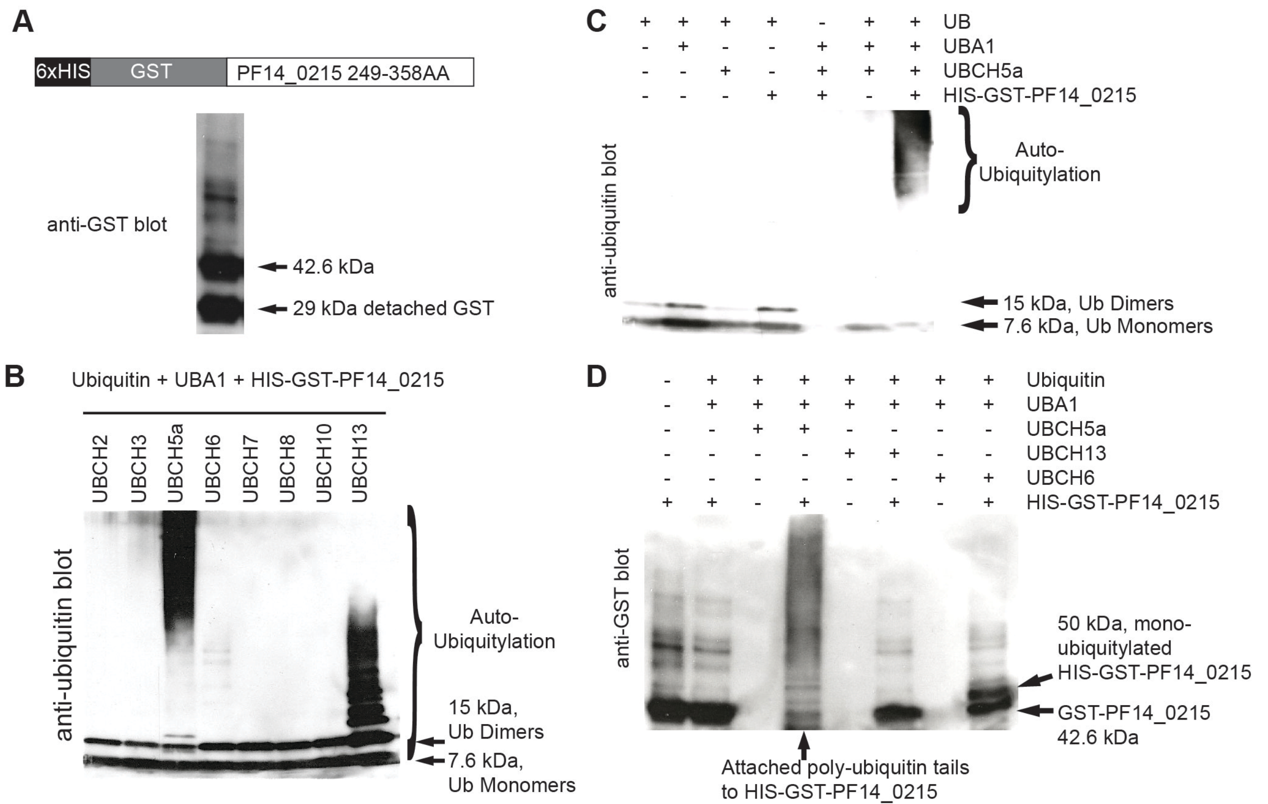


Figure 2.3 – Recombinant PF14_0215 is able to confer *in vitro* ubiquitylation

(**A**) Epitope-tagged recombinant PF14_0215 is depicted and anti-GST blot reveals purification. (**B**) *In vitro* ubiquitylation assays were performed with human UBA1, recombinant PF14_0215 and varying human UBC enzymes. Ubiquitylation is detected with UBCH5a, UBCH6, and UBCH13. (**C**) *In vitro* ubiquitylation assays reveal that ubiquitylation can only be detected when recombinant PF14_215 is added with a complete reaction with UBCH5a (far right lane) but not with partial reactions. (**D**) Indicated by the shifts of molecular weight of GST-fused proteins, GST-tagged recombinant PF14_0215 exhibit different kinds of ubiquitin attachment depending on the UBC used. When incubated with UBCH5a, ubiquitin polymers are attached. With UBCH13, no ubiquitin attachment is observed. With UBCH6, an attachment of a single ubiquitin is mostly detected.

We further characterized the ability of PF14_0215 to mediate auto-ubiquitylation. *In vitro* ubiquitylation assays were followed by anti-GST immunoblots to reveal any shifts in molecular weight from the covalent attachment

of ubiquitin directly onto PF14_0215 itself. Differing ubiquitylation patterns were observed with respect to differing human recombinant E2s. Anti-GST immunoblots (Figure 2.3D) revealed multiple shifts in molecular weight when incubated with the ubiquitin-conjugating enzyme UBCH5a. These results indicate that UBCH5a mediates the attachment of ubiquitin to the RING domains of these genes. On the other hand, our data indicates that the E2 enzyme UBCH13 does not promote auto-ubiquitylation of PF14_0215 (pattern not different from what obtained with E3 alone); only the formation of free poly-ubiquitin tails was observed (Figure 2.3B, last lane). When incubated with UBCH6, a single molecular weight shift of about 8 kDa was observed (Figure 2.3D, last lane). This observation suggests that UBCH6 mediates the attachment of a single ubiquitin to PF14_0215. These *in vitro* assays demonstrate that PF14_0215 is a genuine E3 ubiquitin ligase that catalyzes both the formation of poly-ubiquitin tails and its auto-ubiquitylation.

In order to determine PF14_0215 localization, we performed immunofluorescence assays (IFA) using custom-made antibodies (see Materials and Methods). Using IFA microscopy we located PF14_0215 to reticular structures outside the nuclear regions within the trophozoite and schizont stages of the parasite, similar to the physical attributes of the ER. In the late schizont stages, when the merozoites are being formed and are starting to bud off from the central body, PF14_0215 resides within globular structures, typical of the ER, surrounding each budding merozoite. PF14_0215 was co-immunostained with an antibody

recognizing Plasmepsin V (Klemba & Goldberg, 2005), a *Plasmodium* protein that resides in the ER membrane (Figure 2.4A). Microscopy images exhibited high co-localization between PF14_0215 and Plasmepsin V, further validating that PF14_0215 resides in the ER membrane. For comparative purposes, PF14_0215 was stained in the *P. falciparum* strain, D10 ACP(leader)-GFP, which expresses GFP fused with a signal and transit peptide for the acyl carrier protein (ACP), a typical apicoplast marker (Tonkin *et al*, 2004) (Figure 2.4B). No significant co-localization between PF14_0215 and ACP was seen.

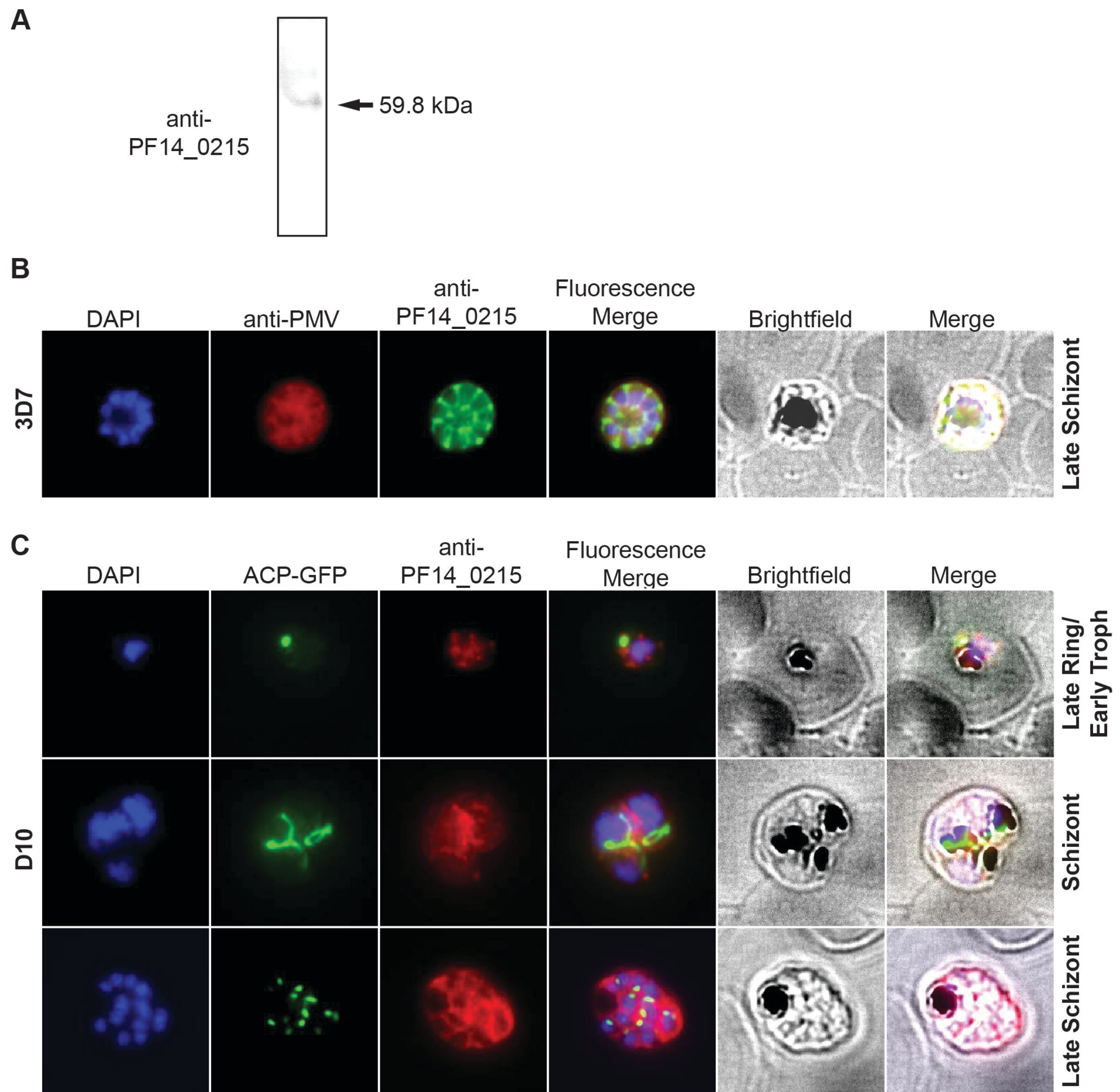


Figure 2.4 – PF14_0215 localizes to the ER membranes

(A) Immunoblot was performed using anti-PF14_0215 custom antibodies on crude parasite extract. A band was detected at the expected size of 59.8 kDa for the PF14_0215 protein. (**B**) IFA experiments show colocalization of the PF14_0215 with PMV, an ER membrane marker. (**C**) Using the D10 strain, which has GFP fused to the leader sequence of ACP, an apicoplast marker, IFA reveals that PF14_0215 does not reside in the apicoplast. Different parasite stages are shown.

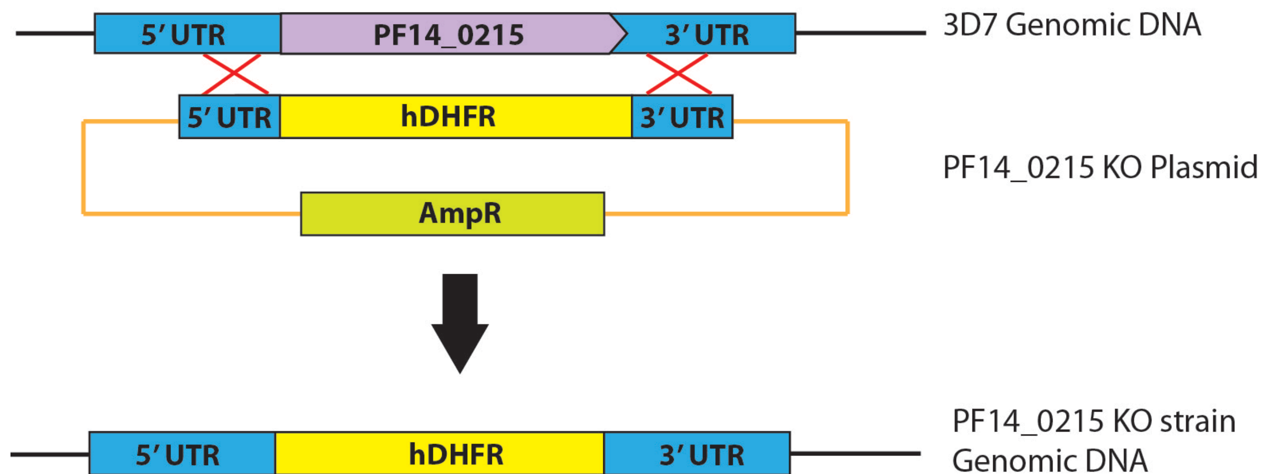


Figure 2.5 – Knock Strategy for PF14_0215
Knockout (KO) strategy of PF14_0215. A PF14_0215 knockout vector was constructed with an human dihydrofolate reductase (hDHFR) selection cassette that is flanked by the 5' UTR and 3' UTR of the PF14_0215. A *Saccharomyces cerevisiae* cytosine deaminase (ScCD) cassette (not shown) was placed outside the PF14_0215 5' UTR and 3' UTR sections to be used for negative selection. Double recombination that would excise the endogenous PF14_0215 gene was never recovered.

To test the essentiality of the PF14_0215 for parasite survival, we constructed a knockout vector that was designed to disrupt the PF14_0215 gene (see Figure 2.5 and Material and Methods for knockout strategy). Despite multiple attempts, gene knockout experiments failed to produce viable parasites indicating that PF14_0215 may be essential for parasite survival. Though the outcome of the knockout experiments are negative results, it evokes the observed essentiality of HRD1 in other eukaryotic model organisms. Evidence showing *in vitro* ubiquitylation activity, ER membrane localization and likely gene essentiality of PF14_215, strongly suggest that it is the E3 ligase HRD1 *Plasmodium* homologue of the classical ERAD system.

PFL1245w and PFL0190w are functional cytosolic ubiquitin activating and conjugating enzymes

PFL1245w and PFL0190w are predicted to be the E1 ubiquitin-activating enzyme and the E2 ubiquitin-conjugating enzyme, respectively. In order to validate the *in silico* predictions, we have recombinantly cloned, expressed and purified PFL1245w and PFL1090w (Figure 2.6A) to test their *in vitro* activities. Since we found that PF14_0215 is likely the *Plasmodium* homologue of HRD1 of the ERAD system, we also tested the interaction of PFL1245w and PFL1090w with PF14_0215. Our results show that recombinant PFL0190w (19.6 kDa) is capable of *in vitro* mono and di-autoubiquitylation activity when incubated with commercially available human UBA1 (Figure 2.6B, second furthest lane from the right). When the recombinant PF14_0215 RING domain was added to the reaction, a large laddering effect (poly-ubiquitin chains) was observed (Figure 2.6B, far right lane) similar to the effect observed using the commercial human UBCH5a (Figure 2.3B, third lane from the left). Anti-GST blots further revealed that these poly-ubiquitin chains were attached to the recombinant GST-tagged RING domain of PF14_0215 (Figure 2.6C). These results show that PFL0190w is a functional E2 ubiquitin-conjugating enzyme that is compatible *in vitro* with recombinant PF14_0215.

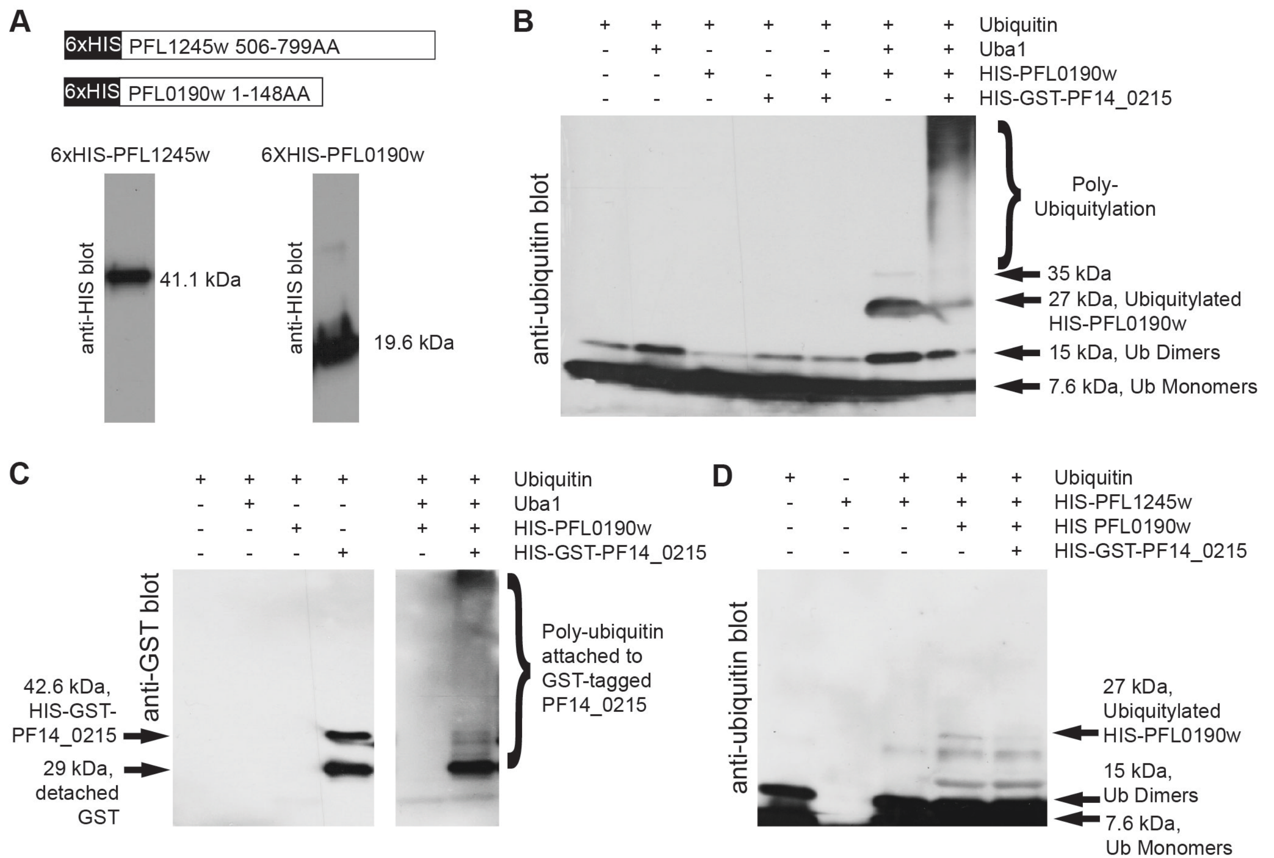


Figure 2.6 – *In vitro* ubiquitylation of PFL1245w and PFL0190w

(**A**) 6xHIS-tagged recombinant PFL1245w (E1) and PFL0190w (E2) are depicted and anti-HIS blots reveal purification. (**B**) *In vitro* ubiquitylation assays were performed with recombinant PFL0190w(E2), which revealed an attachment of a single ubiquitin when incubated with human UBA1 (E1) (second lane from the right) and an increased number of ubiquitylation products when both UBA1 and recombinant PF14_0215 (E3) was added (far right lane). (**C**) When incubated with human UBA1 (E1), recombinant PFL0190w (E2) (along with the other necessary reagents) attaches polymers of ubiquitin to recombinant PF14_0215 (E3) (far right lane). (**D**) Recombinant PFL1245w (E1) is capable of attaching a single ubiquitin to recombinant PFL0190w (E2), which is depicted by the appearance of a 27kDa band (second lane from the right). Extra banding was not detected with the addition of recombinant PF14_0215 (E3) (far right lane).

In vitro ubiquitylation assays were repeated using recombinant parasite E1, PFL1245w together with E2 PFL0190w and E3 PF14_0215 (Figure 2.6D).

PFL1245w was able to ubiquitylate PFL0190w as shown by the presence of a band at 27 kDa. When recombinant *Plasmodium* E3 ligase PF14_0215 was added to the reaction, no increase of ubiquitylating activity could be detected. Considering the mode of action of the classical ERAD-system, it is possible that the full-length of recombinant PF14_0215 as well as additional accessory proteins are required for a full *in vitro* activity. On the whole, these results validate *(i)* PFL1245w as a functional E1 ubiquitin-activating enzyme that can work *in vitro* with the E2 PFL0190w, and *(ii)* PFL0190w as a functional E2 ubiquitin-conjugating enzyme that can work *in vitro* with the E3 ligase PF14_0215.

To further validate PFL1245w and PFL0190w as the likely *Plasmodium* ERAD E1 and E2 enzymes, respectively, we investigated their localization using custom-made antibodies; antibody-specificity was verified by immunoblotting (Figure 2.7A and Figure 2.8A). In other model organisms, the classical ERAD E1 UBA1 and E2 UBC7 proteins are reported to reside in the cytoplasm until they are needed and are recruited to the outer membrane of the ER. Using IFA microscopy, we found that both PFL0190w and PFL1245w are mainly dispersed throughout the cytoplasm with small aggregations scattered throughout the parasite (Figures 2.7B and Figure 2.8B). Immunostaining in D10 ACP(leader)-GFP *Plasmodium* strains eliminated the possibility of an apicoplast localization of these proteins. In addition, when PFL1245w was co-immunostained with PMV, an ER membrane marker, there was

noticeable overlap between the two proteins, which suggests that PFL1245w is likely recruited to the ER membrane as expected (Figure 2.7C).

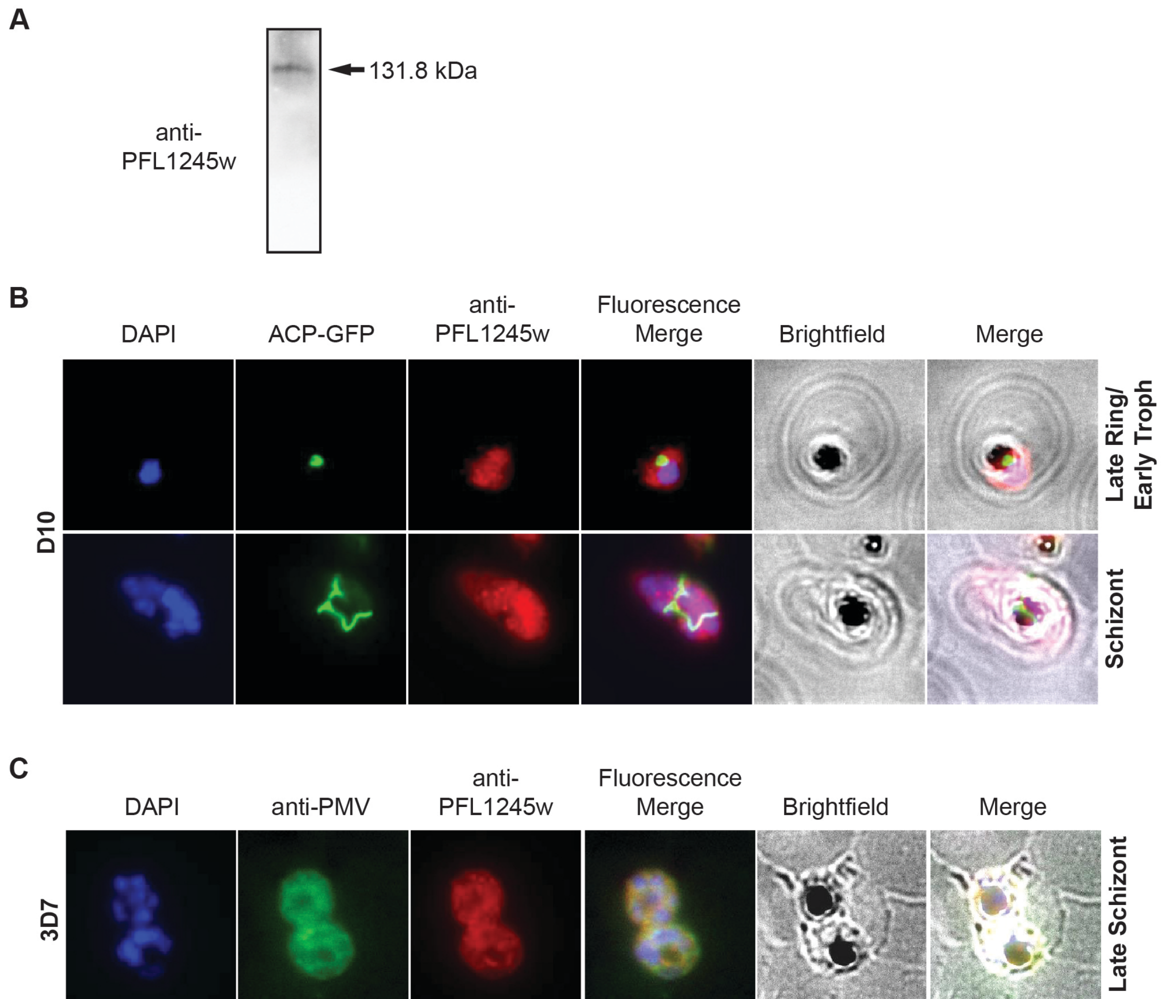


Figure 2.7 – PFL1245w and PFL0190w mainly localize to the cytosol.
(A) Immunoblots were made using anti-PFL1245w custom antibodies on crude parasite protein extracts. A bands at 131.8 kDa was detected, which is the expected size for the PFL1245w protein. (**B**) IFA experiments, using different parasite stages, show that PFL1245w localizes mainly to the cytosol. **(C)** When co-immunostained with PMV, an ER membrane protein marker, there was noticeable overlap between the two proteins, suggesting PFL1245w recruitment to the ER as well.

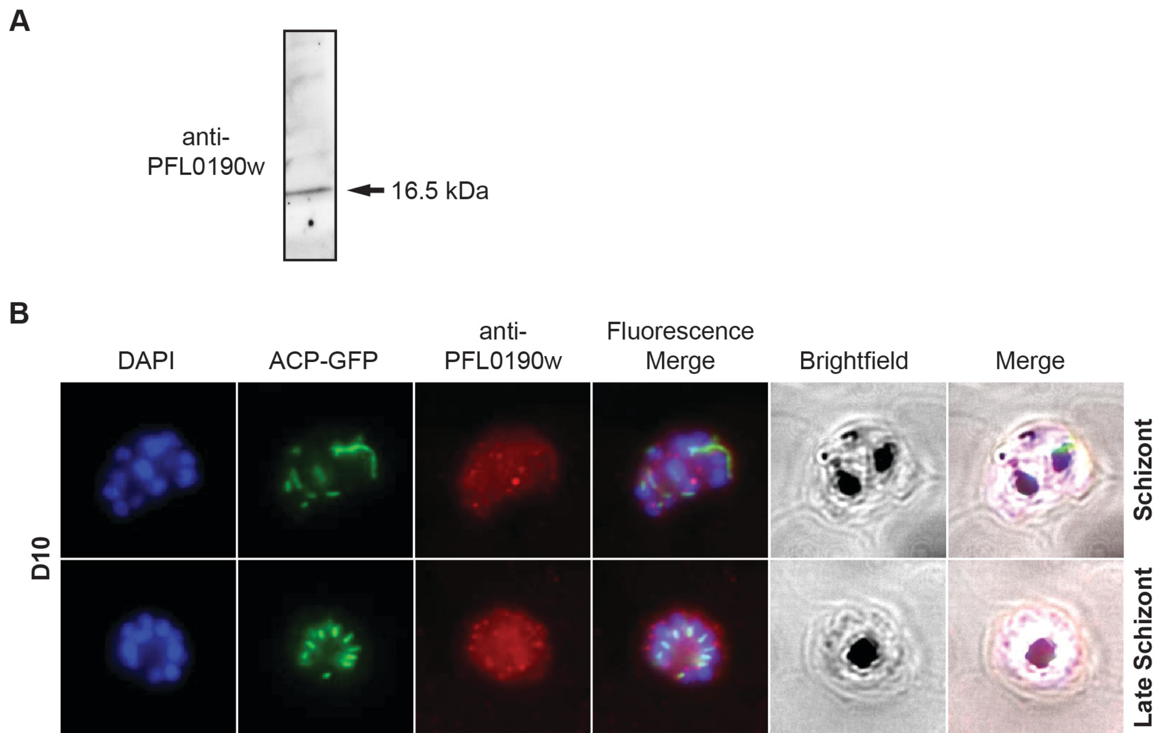


Figure 2.8 – PFL0190w mainly localize to the cytosol.
(A) Immunoblots were made using anti-PFL0190w custom antibodies on crude parasite protein extracts. A band at the 16.5 kDa was detected, which is the expected size for the PFL0190w protein. (**B**) IFA experiments show PFL0190w to localize mainly to the cytosol, even at different parasite stages.

The localization pattern of PFL1245w and PFL0190w indicates cytoplasmic residence with punctual recruitment to the ER that is consistent with a participation in the ERAD pathway. Given their *in vitro* activities in auto-ubiquitylating assays and their pattern of localization, PFL1245w and PFL0190w are likely the *Plasmodium* homologues for E1 UBA1 and E2 UBC7 involved in the parasite's ERAD system.

DISCUSSION

Proteasome inhibitor studies have shown promising antimalarial results and have ignited interest in studying how protein degradation in *Plasmodium* can be targeted for effective drug discovery and synthesis. Though there have been several reports on the efficacy of various proteasome inhibitors on malaria parasites, there have been relatively few studies done on functionally characterizing the malarial protein degradation system. An integral part of protein degradation within eukaryotic cells is the ERAD system, which relies on ubiquitylation in order to shuttle misfolded proteins across the ER membrane and label them for degradation by the 26S proteasome. Here, as an initial study, we show the first characterization of the ubiquitylating components of the *Plasmodium* ERAD system.

According to bioinformatics, *in vitro* ubiquitylation assays, and localization studies, it seems as though the *Plasmodium* ERAD system may function similar to that of other eukaryotic model systems. The domain architecture of these putative ubiquitylating *Plasmodium* ERAD proteins (PFL1245w, PFL0190w, PF14_0215) are homologues to that of their identified counterparts in other eukaryotes (Ponts *et al*, 2008), while localization studies show them localizing to their expected destinations of either the ER membrane or cytosol.

The *in vitro* ubiquitylation assays performed on recombinant versions of PFL1245w, PFL09190w and PF14_0215 revealed that these proteins are capable of

facilitating ubiquitylation and that they could possibly work together. However, further functional studies (*i.e.* pull-down assays followed with mass spectrometery) will need to be performed to show that these core ubiquitylating enzymes do indeed collaborate to conjugate ubiquitin to proteins and work together with other known ERAD proteins (*i.e.* Der1, Cdc48, Ufd1). Future work also consists of experiments that can show that these ubiquitylating proteins are directly involved in labeling misfolded proteins for degradation by the 26S proteasome.

The results from the knock out attempts of PF14_0215 suggest that PF14_0215 is an essential *Plasmodium* gene. As a likely part of the ERAD system, the probable essentiality of PF14_0215 is not surprising since the ERAD system, and protein degradation in general, is known to be an essential part of eukaryotic biology. That is why there is burgeoning interest in proteasome degradation and inhibitors in regards to antimalarial drug discovery. If the protein degradation at the proteasome level can be effectively exploited as a potential antimalarial target, it is reasonable to believe that upstream pathways, such as the ubiquitylating components within the ERAD system, may also serve as viable antimalarial drug targets. Currently, a wide range of ubiquitylating enzymes, from E1 activating enzymes to deubiquitylating enzymes (DUB)s are being screened for inhibitors that may confer anti-cancer properties. In fact, there are already several inhibitors of ubiquitin or ubiquitin-like enzymes that have shown effective results against cancer

(Sun, 2003; Yang *et al*, 2007), with some already in clinical trials (for a review, see Edelmann *et al*, 2011).

Some may argue that targeting a highly conserved system such as the ERAD pathway in the *Plasmodium* is likely to have cross-reactivity with the very similar human host ERAD counterpart and is unlikely to produce a tight parasite-specific drug. However, as mentioned previously, there has already been some success in inhibiting cancer cells without excessively affecting the host cells in preclinical trials. Furthermore, since it has been shown that E3 ligases and DUBs are the most various and divergent within eukaryotes (Ponts *et al*, 2008), higher parasite-specificity can be achieved by investigating essential parasite-specific E3 ligases or DUBs and screening for their inhibitors. In theory, this could provide for an effective strategy to uncover possible drug targets within the *Plasmodium* ERAD degradation pathway that could produce both specific and effective antimalarials.

The biology of the *Plasmodium* ERAD system and overall degradation pathways are worthy themes to investigate. Whether or not protein degradation in protozoan parasites will have any unique functional divergences from the more studied human and yeast models remains to be seen, which can only be uncovered with more vigilant examination. One aspect of parasite biology that we can be certain of is that protein degradation is necessary for parasite survival. Thus, the *Plasmodium* ERAD system offers several excellent antimalarial target candidates,

particularly the E3 ligases and DUB which have been characterized as being more divergent from their human host counterparts. Though we have just begun to explore the machinery that is responsible for protein degradation in the human malaria parasite, we believe that the *Plasmodium* ERAD system may provide great drug target candidates that are both effective and parasite-specific.

CONCLUDING REMARKS FOR CHAPTER 2

Our investigation in this chapter brings a first glimpse into the *Plasmodium* ERAD system and ubiquitin-dependent protein degradation. However, as an initial study, there are still many more mysteries to parasite protein degradation to unravel. From studies in other model organisms and also from our gene-disruption experiments, the *Plasmodium* ERAD system is likely an indispensible pathway by the parasite. And though some may argue that the parasite ERAD system may not be an ideal drug target because it shares many similarities to that of its human host, we offer the more diverse E3 ligases and DUBs of the parasite ERAD pathway as potential parasite-specific drug target candidates. Only with more fervent exploration and careful examination of parasite protein degradation, can we determine whether or not the *Plasmodium* ERAD system is unique from its other eukaryotic counterparts and possibly exploit it for antimalarial intervention.

MATERIALS AND METHODS

Cloning and purification of recombinant proteins

6xHIS tagged recombinant proteins (PFL0190w, PFL1245w, PF14_0215) were constructed by using a modified version of PGS21a (PGS-21aHIS), which has the GST coding region removed. GST removal was done by cutting PGS-21a with ClaI and NcoI and inserting a 6xHIS PCR fragment that was amplified using the primers GATCGAGATCGATCTCGATC and ATCCATGGCCTTACCGCTGCTATGATGATGAT from the PGS-21a plasmid itself. The RING domain of PF14_0215 was amplified with PCR using primers AGGGGATCCCTTAAGCCGCGGCCTTTACATATGACAGCAGAT and AGGAAGCTTTTAACTAGTGCTAGCTTACTTTTGTGTTGTATCATTTTCTG. A segment that contained the E1 activating domain of PFL1245w was amplified with PCR using primers AGGGGATCCCTTAAGCCGCGGGTGGTGAATATTTTTGGGTTGG and AGGAAGCTTTTAACTAGTGCTAGCCCAACCCAAAAATATTCACCAC. The entire gene of PFL0190w was amplified with primers AGGGGATCCCTTAAGCCGCGGATGGCCCTTAAAAGAATAACAAAA and AGGAAGCTTTTAACTAGTGCTAGCTTATTGTGCATATTTTTGTGTCC. The amplified genes were cut with SacII and SpeI and ligated into the PGS-21a that had both the GST and 6xHIS tag or just the 6xHIS tag by itself.

Cloned expression *e.coli* cells were grown to an OD_{600} of >0.5. IPTG was added to a concentration of 1mM for induction and incubated overnight at 12°C.

Cells were spun down and resuspended in lysis buffer (25mM Tris-HCl, pH 7.5, 500mM NaCl, and 1% (v/v) Triton X-100) with protease inhibitor cocktail (Roche) and 1mM AEBSF. Cells were sonicated and spun down.

GST purifications were performed with glutathione agarose (Sigma). Bound proteins were washed three times with GST wash buffer (25mM Tris-HCl, pH 7.5, 300mM NaCl and 1% Triton X-100). GST-tagged proteins were then eluted with GST elution buffer (25mM Tris-HCl, pH 7.5, 150mM NaCl, 15mM reduced glutathione and 0.01% Triton X-100 and 40% (v/v) glycerol). Anti-GST immunoblots were probed with goat anti-GST antibodies (1:5000; GE Healthcare) and donkey anti-goat antibodies conjugated to horseradish peroxidase (HRP) (1:20,000; Jackson Immuno research).

HIS-tagged protein purifications were performed with Ni-NTA beads (Qiagen) and subsequently washed several times with a solution containing 25mM Tris-HCL pH 7.5, 500mM NaCl, 30mM imidazole and 5% glycerol. Purified proteins were eluted with 25mM Tris-HCl pH 7.5, 500mM NaCl, 250mM imidazole and 5% glycerol. Anti-HIS immunoblots were probed with mouse anti-HIS antibodies (1:2500; Millipore) and goat anti-mouse antibodies conjugated to horseradish peroxidase (HRP) (1:10,000; BioRad).

Custom antibodies

The following peptide sequences were used as antigens for the production of affinity-purified antibodies (Fisher Scientific or Genscript) in rabbits: RFKSFQKYRELTKNIETK for PF14_0215, KTDRTKYHQTAKAWTQKYAQ for PFL0190w, CSDQDLVDVLIPSIQFIYK for PFL1245w

Parasite culturing and transfection

The *P. falciparum* 3D7 and D10_ACP (leader) GFP (provided by MR4, MRA568) strains were grown in human O+ red blood cells according to standard protocols (Trager & Jensen, 1976) with the exception that cultures were incubated in gassed flasks. Transfection of parasite cultures were carried out by electroporation of infected human O+ red blood cells as described in (Deitsch *et al*, 2001). Transfectants were selected with WR99210 (Fidock & Wellems, 1997).

Immunofluorescence Assay

Immunofluorescence assay was performed using the protocol described in (Spork *et al*, 2009). Briefly, cells were resuspended in PBS and spun down for a few minutes at low speeds. With the supernatant discarded, the pellet was resuspended in 1 mL of PBS containing 4% paraformaldehyde and 0.0075% glutaraldehyde and incubated at 37°C for 1 hour. After gentle spinning, the fixed pellet was washed (and incubated for 10 minutes) with 1.25M glycine/PBS. After another round of spinning,

the pellet was then incubated in 0.1% TritonX-100/PBS solution (10 minutes) and then spun again and washed with 125mM glycine/PBS for another 10 minutes. After centrifugation, the pellet was blocked in 3% BSA/PBS for at least 1 hour at room temperature. Then the primary antibodies were added and incubated for at least 2 hours at room temperature (or overnight at 4°C). After washing with PBS, secondary antibodies were added and the cells were incubated for about two hours at room temperature and then subjected to a final round of washes. DAPI was added to visualize the nuclei.

Primary antibodies with their respective dilutions are as follows: anti-Mal13P1.227 (1:100), anti-PF14_0215 (1:150), anti-PFL0190w (1:50) and anti-PFL1245w (1:50). Secondary goat anti-rabbit IgG Alexa Fluor® 488 (1:100) or donkey anti-rabbit IgG Alexa Fluor® 568 (1:100) were then used for the respective primary antibodies. DAPI (100ng/mL final concentration) was added and slides were mounted with slow-fade mounting medium (Fluoromount-G; Southern Biotech) and viewed with fluorescence microscopy. Rabbit anti-ACP antibodies (a gift from the Geoffrey I. McFadden lab) were used at a 1:250 dilution (Tonkin *et al*, 2004). Mouse anti-plasmepsin V (PMV) antibodies (Klemba & Goldberg, 2005) were obtained from the Malaria Research and Reference Reagent Resource (MR4) center and used at a 1:20 dilution with donkey anti-mouse IgG Alexa Fluor® 568 (1:100). Images were observed with the Olympus BX40 microscope using an 100x objective lens (UPlanFI) and captured by the CoolSNAP cf (Photometrics) camera using

Metavue software. Images were merged and background was reduced using ImageJ software.

In vitro ubiquitylation assay

50-200μM of ubiquitin (Boston Biochem), 0.05-0.2μM of E1 enzyme, 1-5μM of E2 enzymes, and 1-12.5μM of E3 ligases were incubated together in reaction buffer (50mM Tris-HCl, pH 7.4; 1mM DTT and Re-energizing buffer (Boston Biochem)) for 2 hours at 37°C and then analyzed by SDS-PAGE and immunoblotting. Human recombinant UBE1 and UBC enzymes were purchased from Boston Biochem. Anti-ubiquitin immunoblots were probed with rabbit anti-ubiquitin antibodies (1:2500; Upstate) and goat anti-rabbit antibodies conjugated to HRP (1:5000; Pierce). Anti-GST immunoblots were probed with goat anti-GST antibodies (1:2500; GE Healthcare) and donkey anti-goat antibodies conjugated to HRP (1:5000; Jackson Immunoresearch).

Gene disruption experiments

Gene disruption plasmids were constructed with the 5' UTR and 3'UTR of PF14_0215 flanking a human dihydrofolate reductase (hDHFR) resistance cassette within a pCC1 vector. In addition, a *Saccharomyces cerevisiae* cytosine deaminase (ScCD) cassette resided outside the 5'UTR and 3'UTR regions, which would serve as the negative selection. Gene disruption vectors were transfected into 3D7 strains. Only strains that have undergone double recombination would be contain the

hDHFR (providing WR99210 resistsance) but not the ScCD cassette, which would kill the parasites still containing when treated with 5-fluorocytosine. Gene disruption and gene knockout vectors were transfected into 3D7 parasites. Transfectants were selected by WR99210 and 5-fluorocytosine drugging and screened with PCR.

Acknowledgements:

I would like to acknowledge Nadia Ponts for her bioinformatic work, Jacques Prudhomme for maintaining the parasite cultures and Elisandra Rodrigues for her assistance with purifying the PfHRD1 recombinant protein.

REFERENCES

Bagola K, Mehnert M, Jarosch E & Sommer T (2011) Protein dislocation from the ER. *Biochim. Biophys. Acta* **1808:** 925–936

Carvalho P, Goder V & Rapoport TA (2006) Distinct Ubiquitin-Ligase Complexes Define Convergent Pathways for the Degradation of ER Proteins. *Cell* **126:** 361–373

Certad G, Abrahem A & Georges E (1999) Cloning and partial characterization of the proteasome S4 ATPase from Plasmodium falciparum. *Exp. Parasitol.* **93:** 123–131

Chung D-WD & Le Roch KG (2010) Targeting the Plasmodium ubiquitin/proteasome system with anti-malarial compounds: promises for the future. *Infect Disord Drug Targets* **10:** 158–164

Deitsch K, Driskill C & Wellems T (2001) Transformation of malaria parasites by the spontaneous uptake and expression of DNA from human erythrocytes. *Nucleic Acids Res.* **29:** 850–853

Dick LR, Cruikshank AA, Grenier L, Melandri FD, Nunes SL & Stein RL (1996) Mechanistic studies on the inactivation of the proteasome by lactacystin: a central role for clasto-lactacystin beta-lactone. *J. Biol. Chem.* **271:** 7273–7276

Edelmann MJ, Nicholson B & Kessler BM (2011) Pharmacological targets in the ubiquitin system offer new ways of treating cancer, neurodegenerative disorders and infectious diseases. *Expert Rev Mol Med* **13:** e35

Fidock DA & Wellems TE (1997) Transformation with human dihydrofolate reductase renders malaria parasites insensitive to WR99210 but does not affect the intrinsic activity of proguanil. *Proc. Natl. Acad. Sci. U.S.A.* **94:** 10931–10936

Garcia LS (2010) Malaria. *Clinics in Laboratory Medicine* **30:** 93–129

Gardner MJ, Hall N, Fung E, White O, Berriman M, Hyman RW, Carlton JM, Pain A, Nelson KE, Bowman S, Paulsen IT, James K, Eisen JA, Rutherford K, Salzberg SL, Craig A, Kyes S, Chan M-S, Nene V, Shallom SJ, *et al* (2002) Genome sequence of the human malaria parasite Plasmodium falciparum. *Nature* **419:** 498–511

Hershko A & Ciechanover A (1998) THE UBIQUITIN SYSTEM. *Annu. Rev. Biochem.* **67:** 425–479

Klemba M & Goldberg DE (2005) Characterization of plasmepsin V, a membrane-bound aspartic protease homolog in the endoplasmic reticulum of Plasmodium falciparum. *Mol. Biochem. Parasitol.* **143:** 183–191

Kreidenweiss A, Kremsner PG & Mordmüller B Comprehensive study of proteasome inhibitors against Plasmodium falciparum laboratory strains and field isolates from Gabon. *Malar J* **7:** 187–187

Laney JD & Hochstrasser M (1999) Substrate Targeting in the Ubiquitin System. *Cell* **97:** 427–430

Li GD, Li JL, Mugthin M & Ward SA (2000) Molecular cloning of a gene encoding a 20S proteasome beta subunit from Plasmodium falciparum. *Int. J. Parasitol.* **30:** 729–733

Lindenthal C, Weich N, Chia YS, Heussler V & Klinkert MQ (2005) The proteasome inhibitor MLN-273 blocks exoerythrocytic and erythrocytic development of Plasmodium parasites. *Parasitology* **131:** 37–44

Pickart CM (2004) Back to the Future with Ubiquitin. *Cell* **116:** 181–190

Ponts N, Yang J, Chung D-WD, Prudhomme J, Girke T, Horrocks P & Le Roch KG (2008) Deciphering the Ubiquitin-Mediated Pathway in Apicomplexan Parasites: A Potential Strategy to Interfere with Parasite Virulence. *PLoS ONE* **3:** e2386

Prudhomme J, McDaniel E, Ponts N, Bertani S, Fenical W, Jensen P & Le Roch K (2008) Marine Actinomycetes: A New Source of Compounds against the Human Malaria Parasite. *PLoS ONE* **3:**

Ramasamy G, Gupta D, Mohmmed A & Chauhan VS (2007) Characterization and localization of Plasmodium falciparum homolog of prokaryotic ClpQ/HslV protease. *Mol. Biochem. Parasitol.* **152:** 139–148

Reynolds JM, El Bissati K, Brandenburg J, Günzl A & Mamoun CB Antimalarial activity of the anticancer and proteasome inhibitor bortezomib and its analog ZL3B. *BMC Clin Pharmacol* **7:** 13–13

Sato BK, Schulz D, Do PH & Hampton RY (2009) Misfolded Membrane Proteins Are Specifically Recognized by the Transmembrane Domain of the Hrd1p Ubiquitin Ligase. *Molecular Cell* **34:** 212–222

Sommer MS, Gould SB, Lehmann P, Gruber A, Przyborski JM & Maier U-G (2007) Der1-mediated Preprotein Import into the Periplastid Compartment of Chromalveolates? *Mol Biol Evol* **24:** 918–928

Spork S, Hiss JA, Mandel K, Sommer M, Kooij TWA, Chu T, Schneider G, Maier UG & Przyborski JM (2009) An Unusual ERAD-Like Complex Is Targeted to the Apicoplast of Plasmodium falciparum. *Eukaryotic Cell* **8:** 1134–1145

Sun Y (2003) Targeting E3 ubiquitin ligases for cancer therapy. *Cancer Biol. Ther.* **2:** 623–629

Tonkin CJ, van Dooren GG, Spurck TP, Struck NS, Good RT, Handman E, Cowman AF & McFadden GI (2004) Localization of organellar proteins in Plasmodium falciparum using a novel set of transfection vectors and a new immunofluorescence fixation method. *Mol. Biochem. Parasitol* **137:** 13–21

Trager W & Jensen J (1976) Human malaria parasites in continuous culture. *Science* **193:** 673–675

van Vugt M, van Beest A, Sicuri E, van Tulder M & Grobusch MP (2011) Malaria treatment and prophylaxis in endemic and nonendemic countries: evidence on strategies and their cost-effectiveness. *Future Microbiol* **6:** 1485–1500

Xie W & Ng DTW (2010) ERAD substrate recognition in budding yeast. *Seminars in Cell & Developmental Biology* **21:** 533–539

Yang Y, Kitagaki J, Dai R-M, Tsai YC, Lorick KL, Ludwig RL, Pierre SA, Jensen JP, Davydov IV, Oberoi P, Li C-CH, Kenten JH, Beutler JA, Vousden KH & Weissman AM (2007) Inhibitors of ubiquitin-activating enzyme (E1), a new class of potential cancer therapeutics. *Cancer Res.* **67:** 9472–9481

Yoo SJ, Seol JH, Shin DH, Rohrwild M, Kang MS, Tanaka K, Goldberg AL & Chung CH (1996) Purification and characterization of the heat shock proteins HslV and HslU that form a new ATP-dependent protease in Escherichia coli. *J. Biol. Chem.* **271:** 14035–14040

CHAPTER 3

Ubiquitin-dependent protein import into the *Plasmodium* apicoplast *via*

a duplicated ERAD-like system.

CHAPTER 3 PREFACE

The discovery that proteasome inhibitors could significantly inhibit *Plasmodium* parasites has ignited interest in the ubiquitin-dependent ER-associated protein degradation (ERAD) system. Though still early and yet relatively little investigation has been done so far, the 26S proteasome, the prokaryotic-like proteasome and also the more divergent E3 ligases involved in the ubiquitin proteasome system (UPS) are considered by some to be good candidates for antimalarial drug targeting (Chung & Le Roch, 2010). In our initial study of the *Plasmodium* ERAD system (Chapter 2), we were able to that the *Plasmodium* ERAD pathway is likely indispensible, which would make sense because it is generally accepted that regulated protein degradation is an essential part of eukaryotic biology. Thus, targeting the *Plasmodium* ERAD system, particularly the more diverse E3 ligases and DUBs for creating antimalarials would potentially create drugs that would greatly inhibit the parasites.

In this chapter, we explore a duplicated ERAD-like system that is localized to a parasite-specific organelle called the apicoplast. Similar to the ERAD pathway in the ER, the apicoplast ERAD-like system likely functions to translocate proteins across membranes. Being that the apicoplast is indispensible to the parasite, the apicoplast ERAD-like system presents itself as a potential antimalarial target.

INTRODUCTION

With the rise of drug-resistances by the malaria parasites to our currently available antimalarials (Koenderink *et al*, 2010), there can soon be a time where we will no longer have any pharmacological defense against the deadly malaria disease. Thus, new drugs and new drug-targets must be a priority.

A few years ago, we identified duplicated homologues of components that make up the ERAD system (Ponts *et al*, 2008), and are believed to localize to the apicoplast (Spork *et al*, 2009; Sommer *et al*, 2007; Hempel *et al*, 2010). These proteins are suspected to form an ERAD-like system in Apicomplexa that could be similar to the symbiont-specific ERAD-like machinery (SELMA) found in Chromalveolates, a supergroup of unicellular eukaryotes containing plastids related to red algae (Sommer *et al*, 2007; Hempel *et al*, 2010). Similarly to the classical ERAD machinery, the SELMA is implicated in the pre-protein import to the periplastid compartment (PPC) (Gallagher *et al*, 2011; Garcia, 2010), which is the space between the periplastid membrane (PPM) and the primary outermembrane (POM) – the second and third outermost membranes of the plastid, respectively. In *Plasmodium* and the apicomplexan *Toxoplasma gondii*, the ERAD-like system is strongly suspected to be involved in translocating proteins across the apicoplast membranes (Spork *et al*, 2009; Agrawal *et al*, 2009; Sommer *et al*, 2007; Hempel *et al*, 2009) in a fashion similar to the way the classical ERAD system retro-

translocates aberrant proteins from the ER lumen to the cytosol during protein quality control.

Like the ERAD system, initial studies indicate that the ERAD-like system may also be ubiquitin-dependent. In the Chromalveolate *Phaeodactylum tricornutum*, ERAD homologues of an E3 ligase and a DUB enzyme with verified *in vitro* activities were localized to the plastid (Hempel *et al*, 2010). In *P. falciparum* and five other *Plasmodia*, multiple components of a putative apicoplast ERAD-like system have been bioinformatically identified (Ponts *et al*, 2008; Spork *et al*, 2009; Sommer *et al*, 2007). However, unlike the *Phaeodactylum tricornutum* SELMA system, possible candidates for an E3 ligase and DUB in the *P. falciparum* ERAD-like system have yet to be identified. Given the central role of E3 ligases in both substrate recognition and ubiquitylation, an ERAD-like system without any E3 ligase seems unlikely.

In our present study, we used an extensive *in silico* search of ERAD-like proteins to identify all components of *P. falciparum*'s candidate proteins possibly involved in apicoplast trafficking. Biochemical and genetic experiments validate localization and function of the most central candidates of the system. In particular, we identify and characterize for the first time two E3 RING finger ubiquitin ligases (PFC0740c and PFC0510w) and one DUB (MAL8P1.126) targeting to the apicoplast. We find that the two E3 ligases have genuine *in vitro* ubiquitylation activities. In addition, we report that the combination of the apicoplast-targeted E1, E2, and our

identified E3 ligases successfully catalyze *in vitro* autoubiquitylation. Our results suggest that these enzymes may work together as a complete ubiquitylating system that has never been wholly identified previously. Furthermore, gene disruption experiments suggest that the ubiquitylating components of the ERAD-like system are essential to the parasite survival. Our analysis is the first functional characterization of the ubiquitylating components of the *P. falciparum* ERAD-like pathway and offers new avenues for antimalarial development strategies.

RESULTS

Computational discovery of ERAD-like components

We combined two powerful *in silico* strategies to identify all putative ubiquitin-associated proteins that may target to the apicoplast (See Materials and Methods). Briefly, the entire translated parasite genome was scanned to identify ubiquitylating proteins using a previously published approach (Ponts *et al*, 2008). The identified proteins were then analyzed with the apicoplast-targeting prediction tools PATS (Zuegge *et al*, 2001) and PlasmoAP (Foth *et al*, 2003). In the second approach, genes with a possible localization to the apicoplast were retrieved directly from PlasmoDB (473 genes) and analyzed for the presence of ubiquitylation-related domains. The combined results revealed 10 confident candidate ubiquitylating proteins likely to be targeted to the apicoplast (Table 3.1). We identified one single ubiquitin-like protein (PlasmoDB accession number, PF08_0067), two E1 ubiquitin activating proteins (PF13_0182 and PF13_0344), one E2 ubiquitin-conjugating enzyme (MAL13P1.227), two E3 ubiquitin ligases (PFC0510w, and PFC0740c) and three deubiquitylating enzymes (PF10_0308, PF10_0233, and MAL8P1.126) that are predicted to target to the apicoplast.

Table 1: Summary of *in silico* predicted apicoplast ubiquitylating proteins in *P. falciparum* and their homologues in other Apicomplexa.

Function	Accession number^	Signal Peptide	Transit Peptide	N-terminal leader	Predicted apicoplast targeting by PlasmoDB 8.0	orthologues and paralogues^						
		Prediction of a bi-partite apicoplast targeting signal	(PlasmoAP)	(PATS v1.2.1)		***Plasmodium* (duplication)**	***P. vivax***	***P. knowlesi***	***P. chabaudi***	***P. berghei***	***P. yoelii***	***T. gondii******
ubiquitin	PF08_0067	1..25	++	yes (p = 0.999)	yes	N/D	PVX_089620	PKH_051680	PCHAS_072420	PBANKA_071510	PY00539	-
E1 ubiquitin-activating enzyme	PF13_0182	1..26	++	yes (p = 0.979)	yes	N/D	PVX_082590	PKH_121970	PCHAS_135250	PBANKA_134790	PY01851, PY06413	TGME49_114890
E1 ubiquitin-activating enzyme	PF13_0344	1..35*	-	yes (p = 0.816)	no	N/D	PVX_115230	PKH_110530	PCHAS_114070	PBANKA_114120	PY02846	N/D
E2 ubiquitin-conjugating enzyme	MAL13P1.227	1..28	++	yes (p = 0.999)	yes	PF14_0128-b	PVX_085805; PVX_083175	PKH_120890; PKH_133470	PCHAS_103050; PCHAS_136300	PBANKA_102970; PBANKA_135840	PY00590; PY00590	TGME49_095990; TGME49_059090
E3 ring zf-C3H4 ubiquitin ligase	PFC0740c	1..51	-	no	yes	N/D	PVX_095385	PKH_082450	PCHAS_080860	PBANKA_080830	PY05143	TGME49_020570**
E3 ring zf-C3H4 ubiquitin ligase	PFC0510w	1..55	-	no	yes	N/D	PVX_119710	PKH_082790	PCHAS_041100	PBANKA_041010	PY01709	TGME49_067440**
OTU DUB	PF10_0308	1..22*	++	no	yes	N/D	PVX_111155	PKH_061570	-	PBANKA_051535	PY05983	TGME49_060510**
mov34 DUB	PF10_0233	1..22	++	yes (p = 0.929)	no	N/D	PVX_111515	PKH_060820	PCHAS_050830	PBANKA_050830	PY01461	N/D
mov34 DUB	MAL8P1.126	1..29	-	yes (p = 0.962)	no	N/D	PVX_088155	N/D	N/D	N/D	N/D	N/D

^PlasmoDB v7.1
*targetP (plant network)
++ very likely
+ likely
- unlikely
*** reviewed in Agrawal and Striepen 2010
**from Ponts et al 2008
N/D not detected

Table 3.1

PF08_0067 encodes a 373 amino acid (aa)-long protein. PF08_0067 is predicted to contain both a signal peptide and an apicoplast-specific transit peptide, consistent with the apicoplast targeting of GFP fused to PF08_0067's 100 N-terminal amino acids (Spork *et al*, 2009). PF08_0067 contains one single degenerated ubiquitin motif spanning from position 76 to 144. Using BLASTP (Altschul *et al*, 1997), close homologues of PF08_0067 were found in other *Plasmodium* species and *Theileria annulata*, a tick-born protozoan pathogen responsible for livestock disease. The characteristic di-glycine motif that is usually present at the C-terminal end of small modifiers is absent in the sequences from *Plasmodium* and *Theileria*. This di-glycine motif is usually involved in the formation of the isopeptide bond between the modifier and the target substrate. How this ubiquitin can covalently attach to other proteins remains unclear. Finally, TMHMM predicted a maximum of three putative transmembrane domains, all located on the last C-terminal quarter of PF08_0067 (Figure 3.1). Whether this protein is processed and subsequently released from the membrane by proteolytic cleavage remains to be seen.

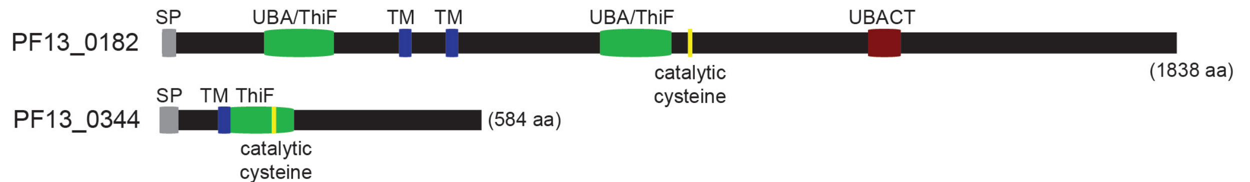
The two identified E1 UBAs, PF13_0182 and PF13_0344 have protein characteristics similar to that of UBA1 and UBA4, respectively. Both of them contain possible transmembrane domains and have catalytic cysteines that can covalently attach to ubiquitin found at position 957 and position 208 of PF13_0182 and PF13_0344 respectively (Figure 3.1). Recent analyses confirmed the apicoplast localization of PF13_0182 (Spork *et al*, 2009). The proposed E2 UBC protein,

encoded by MAL13P1.227, contains an N-terminal UBC-like domain with a consensus catalytic cysteine found at position 219 (Figure 3.1).

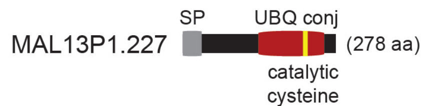
We further identified two potential apicoplast-targeted RING E3 ubiquitin ligase candidates, PFC0510w, and PFC0740c. Classically, HRD1 is a multiple transmembrane domain E3 ligase embedded within the ER membrane with its C-terminal end carrying the E3 RING zinc finger (zf-C3HC4) domain facing the cytosol. PFC0510w also has a C-terminal RING finger domain and four predicted transmembrane domains but, unlike the classical HRD1, the transmembrane domains are found at its C-terminal end (Figure 3.1). PFC0510w homology with canonical HRD1 is quite distant, even though the presence of multiple domains is compatible with an ability to form pores to participate in the recognition and translocation of pre-proteins across membranes in an ERAD-like manner. Considering these elements, PFC0510w is an excellent HRD1-like E3 ligase candidate of the apicoplast ERAD-like system. Finally, PFC0740c is predicted to be a soluble apicoplast-targeted protein with a C-terminal RING finger domain. PFC0740c does not show any clear homology with other proteins.

Ubiquitin

E1 ubiquitin activating enzyme



E2 ubiquitin conjugating enzyme



E3 ring finger ubiquitin ligase

Deubiquitinases

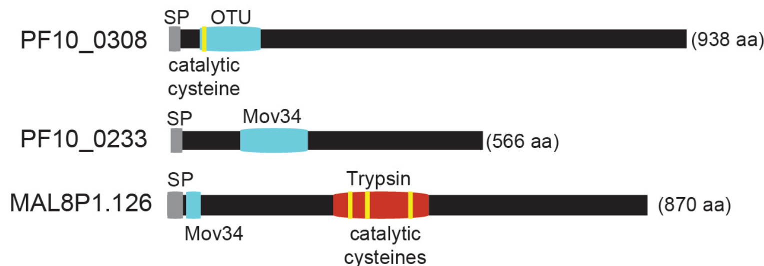


Figure 3.1: *In silico* domain architecture of putative *Plasmodium* ERAD-like ubiquitylating proteins that target the apicoplast

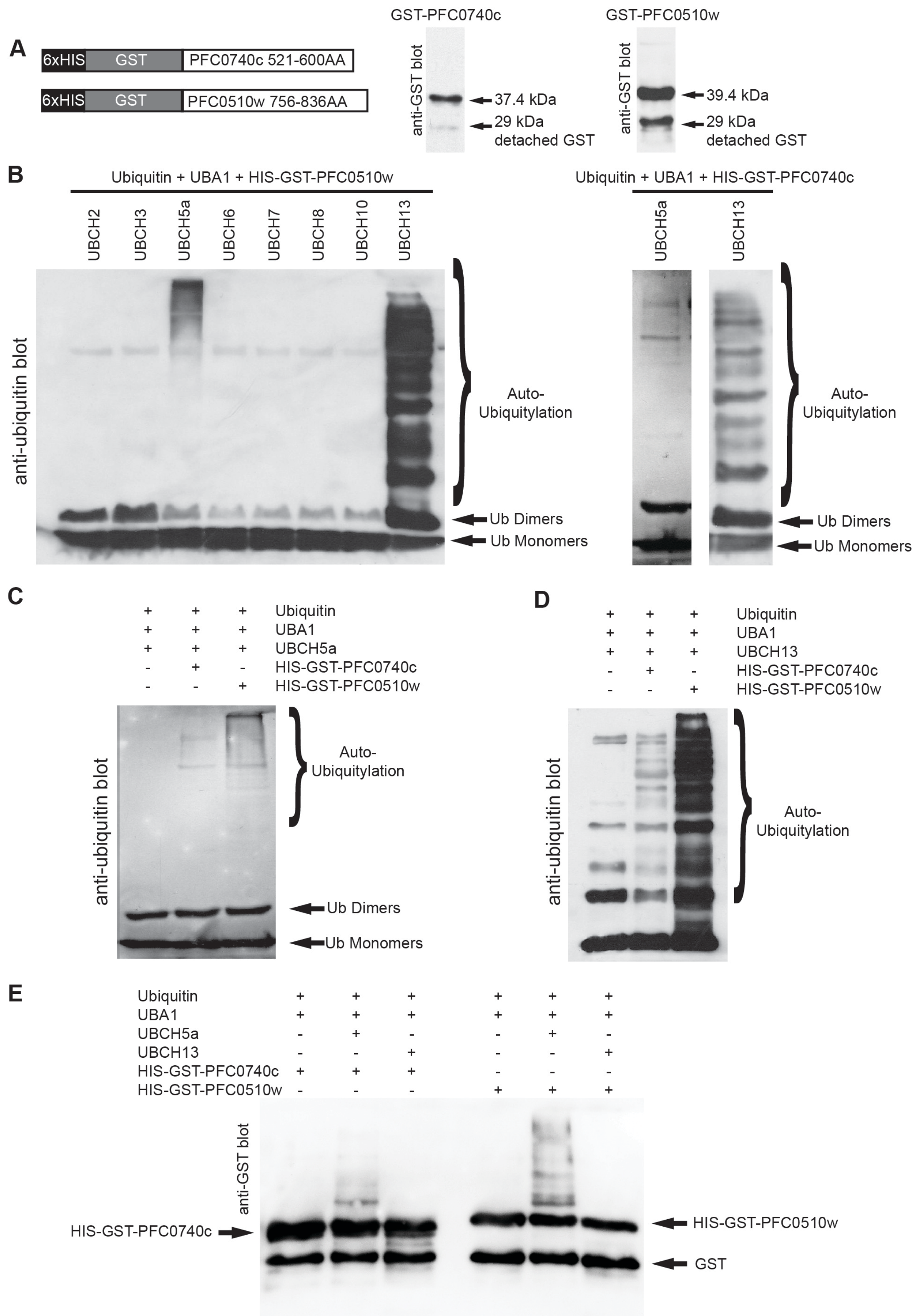
Domains were predicted with PFAMscan and pictures were generated with Matlab using an in-house script. Transmembrane (TM) domains were predicted with TMHMM. SP = signal peptide; TM = transmembrane domain; UBQ = ubiquitin domain; UBA/ThiF/UBACT = E1 ubiquitin activating domain; UBQ conj = E2 ubiquitin conjugating domain; zf-C3HC4 = E3 RING finger ubiquitin ligase domain; OTU refers to cysteine protease otubain deubiquitylating enzyme. Mov34 refers to metalloprotease deubiquitylating enzyme. Trypsin refers to trypsin-like domain.

Finally, we identified three apicoplast-targeted DUB candidates, PF10_0308, MAL8P1.126, and PF10_0233. None of them has clear homology with other known eukaryotic deubiquitylases. These predicted DUB enzymes complete the list, now whole, of candidate apicoplast core ubiquitylating components that constitute the *Plasmodium* ERAD-like system. The rest of the presented study focuses on validating the localization and function of the apicoplast ERAD-like system by experimental investigation at the genetic, biochemical and cell biology levels.

PFC0740c and PFC0510w have in vitro E3 RING finger ubiquitin ligase activity

In order to study the function of PFC0740c and PFC0510w, the RING domains of both proteins were cloned in *E. coli*, fused to tandem GST and 6xHIS tags, expressed, and purified (Figure 3.2A). *In vitro* ubiquitylation assays were performed by incubating the purified recombinant RING domains with commercially available human E1 (UBA1), various human E2 (UBC) enzymes, and commercial purified ubiquitin (Figure 3.2B). Anti-ubiquitin immunoblots show that different patterns of poly-ubiquitin tails were obtained when our recombinant E3 and the commercial E1 were mixed with two of the eight tested E2 conjugating enzymes, UBCH5a and $UBCH13_{complex}$. As a control, human E1 UBA1, human E2 UBC5Ha and ubiquitin were added together to test for *in vitro* ubiquitylation; however, poly-ubiquitin tails were not visible (Figure 3.2C). The appearance of *in vitro* poly-ubiquitylation appeared

only when recombinant PFC0740c and PFC0510w were added to the mixture, which demonstrates that these proteins mediate the formation of poly-ubiquitin tails. (Figure 3.2C). Furthermore, ubiquitylation assays with UBA1 and UBCH13$_{complex}$ alone showed low levels of auto-ubiquitylation (Figure 3.2D), which is consistent with previous studies (Doss-Pepe *et al*, 2005). However, with the addition of recombinant PFC0740c and PFC0510w, the intensity and diversity of poly-ubiquitin tails significantly increased. Anti-GST immunoblots revealed multiple shifts in molecular weight of PFC0740c and PFC0510w when incubated with the ubiquitin-activating enzyme UBCH5a (Figure 3.2E). These results indicate that UBCH5a mediates the attachment of ubiquitin to the RING domains of these proteins. On the other hand, our data indicates that E2 enzyme UBCH13 does not attach ubiquitin to the RING domains of either PFC0740c or PFC0510w; only free poly-ubiquitin tails were observed.



A
6xHIS
GST
PFC0740c 521-600AA
6xHIS
GST
PFC0510w 756-836AA
GST-PFC0740c
anti-GST blot
37.4 kDa
29 kDa
detached GST
GST-PFC0510w
anti-GST blot
39.4 kDa
29 kDa
detached GST
B
Ubiquitin + UBA1 + HIS-GST-PFC0510w
UBCH2
UBCH3
UBCH5a
UBCH6
UBCH7
UBCH8
UBCH10
UBCH13
anti-ubiquitin blot
Auto-
Ubiquitylation
Ub Dimers
Ub Monomers
Ubiquitin + UBA1 + HIS-GST-PFC0740c
UBCH5a
UBCH13
Auto-
Ubiquitylation
Ub Dimers
Ub Monomers
C
Ubiquitin
UBA1
UBCH5a
HIS-GST-PFC0740c
HIS-GST-PFC0510w
anti-ubiquitin blot
Auto-
Ubiquitylation
Ub Dimers
Ub Monomers
D
Ubiquitin
UBA1
UBCH13
HIS-GST-PFC0740c
HIS-GST-PFC0510w
anti-ubiquitin blot
Auto-
Ubiquitylation
E
Ubiquitin
UBA1
UBCH5a
UBCH13
HIS-GST-PFC0740c
HIS-GST-PFC0510w
anti-GST blot
HIS-GST-PFC0740c
HIS-GST-PFC0510w
GST

Figure 3.2: Recombinant RING domains of PFC0740c and PFC0510w have *in vitro* E3 ubiquitin ligase activity with human UbcH5a and UbcH13.

(**A**) Epitope-tagged recombinant PFC0740c and PFC0510w are depicted and anti-GST blots reveal purification. (**B**) *In vitro* ubiquitylation assays were performed with human UBA1, varying human UBCs and either recombinant PFC0510w (left panel) or recombinant PFC0740c (right two panels). Both PFC0740c and PFC0510w produced ubiquitylated products with UBCH5a and UBCH13. (**C**) Controls were done to show that human UBA1 and human UBCH5a by themselves did not produce ubiquitylated products; ubiquitylated products were detected only with the addition of either recombinant PFC0740c or PFC0510w. (**D**) Human UBA1 and human UBCH13 was able to produce some ubiquitylated products, though the range and magnitude of products increased with the addition of either PFC0740c or PFC0510w. (**E**) Indicated by the shifts of molecular weight of GST-fused proteins, GST-tagged PFC0740c and PFC0510w are attached with polymers of ubiquitin when incubated with human UBC5Ha but no attachment is detected with human UBCH13.

PFC0740c and PFC0510w localize to the apicoplast

In order to experimentally localize PFC0740c, we attempted to fuse various epitope tags (*i.e.,* 3xHA, STREP, FLAG, GFP) to the C-terminus for immunofluorescence microscopy. The attachment of the epitope tag to the C-terminal side of PFC0740c was designed to avoid interference with the N-terminal targeting signal peptide. Despite multiple attempts to fuse various epitope tags to PFC0740c, no successfully tagged stably viable strains were obtained. Indeed, even though epitope insertion in transfected parasites was confirmed by semi-quantitative PCR, signal intensity diminished over time until complete disappearance after several weeks of continuous culturing (Figure 3.3A). These results indicate that the successfully transfected parasites eventually die, leaving the surviving parasites with randomly integrated or episomal plasmids. It is highly probable that the epitope tag is interfering with the C-terminal active domains of the

protein. The delayed-lethality observed with successfully transfected parasites is consistent with the reports of a delayed-death phenomenon of parasites with disrupted apicoplasts (Sullivan *et al*, 2000). Moreover, gene knockout experiments failed to produce viable parasites further suggesting that PFC0740c may be essential for parasite survival (Figure 3.3B). This observation suggests that PFC0740c is essential.

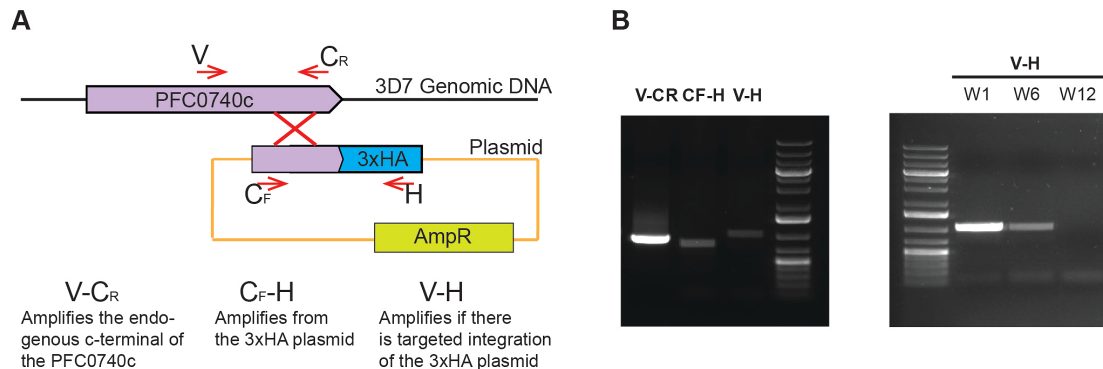


Figure 3.3: HA tagging strategy and eventual death of HA-tagged PFC0740c strains

(A) 3xHA tagging of PFC0740c strategy is shown in the left panel. V-H indicates primer pairs that amplifies only if there is targeted integration of the 3xHA plasmid to the PFC0740c gene. **(B)** After recovery, there was a diminishment of V-H PCR products over time (right panel), indicating that transfected strains exhibited a delayed-death effect of properly integrated vectors, leaving only recovered strains with randomly integrated plasmids. The number of weeks (W1, W6, W12) starts from the time we observed recovered strains.

By using an episomally expressing vector, we were able to circumvent the disruption of the endogenous gene, and found that first 167 amino acids of PFC0740c was sufficient to target GFP to the apicoplast at various stages of the

parasite's life cycle (Figure 3.4A). At the late ring and early trophozoite stages, PFC0740c(1-167)-GFP localizes to a small round area within the parasite, which is a classical trait of the apicoplast. As the parasite matures to a late trophozoite and early schizont, PFC0740c(1-167)-GFP branches out to form forked tubular structures and then eventually is divided up and separated into the budding merozoites as they begin to detach from the central residual body in the late schizont stages. PFC0740c(1-167)-GFP transfected parasites were immunostained with anti-PF14_0215 antibodies, which we previously shown to reside in the ER membrane. IFA microscopy images indicate no significant overlap between PFC0740c(1-167)-GFP and PF14_0215 (Figure 3.4B), showing that PFC0740c does not reside in the ER but targets to the apicoplast.

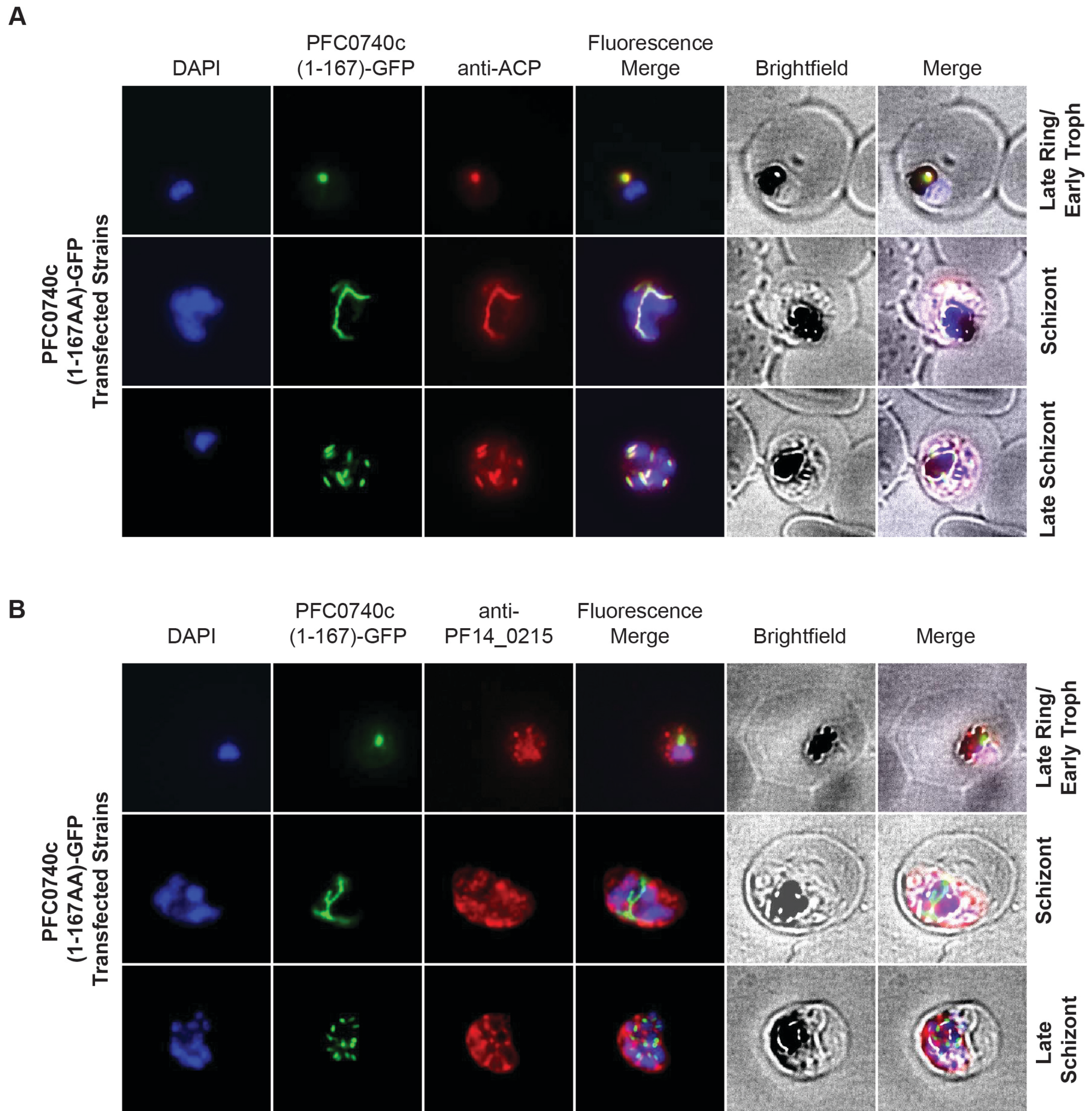


Figure 3.4: The N-terminus of PFC0740c is sufficient to target GFP to the apicoplast

(A) The first 167aa of PFC0740c were fused to GFP and episomally expressed in 3D7 parasites. IFA experiments using anti-ACP antibodies (ACP is an apicoplast stroma marker) indicate that PFC0740c targets to the apicoplast. **(B)** PFC0740c (1-167AA)-GFP transfected strains were stained with anti-PF14_0215 antibodies, which we previously shown to localize to the ER. Throughout the various parasite stages, PFC0740c maintains apicoplast like structures and did not colocalize with the PF14_0215.

In order to determine if PFC0510w has similar apicoplast localization as PFC0740c, a vector was designed to fuse GFP to the 3' end of the endogenous PFC0510w gene by single homologous crossover recombination (Figure 3.5A). Unlike PFC0740c, the endogenous PFC0510w protein was able to receive an epitope tag. After successful transfection and several rounds of drug cycling, PCR and RT-PCR of transfected cultures were implemented to confirm the targeted integration and transcription of the PFC0510w-GFP (Figure 3.5B). Immunofluorescence detection studies confirm that PFC0510w exhibits characteristics of apicoplast formations (Figure 3.5C-D). When immuno-stained against ACP, we observed significant colocalization, which validates PFC0510w as an apicoplast protein. However, colocalization with ACP, which resides specifically in the apicoplast stroma, was not completely tight. Previous analyses show that such a pattern is consistent with an apicoplast peripheral location rather than stromal (Kalanon *et al*, 2009). Therefore, PFC0510w may reside in the PPC as was seen in the SELMA E3 ligase of the *Phaeodactylum tricornutum* (Hempel *et al*, 2010) or it may be a membrane associated protein, which is consistent with it having multiple transmembrane domains. There is also a possibility that both scenarios are true and that PFC0510w is a membrane protein with its RING domain extending out into the PPC, consistent with the observed E3 ligase setup in the classical model of ERAD.

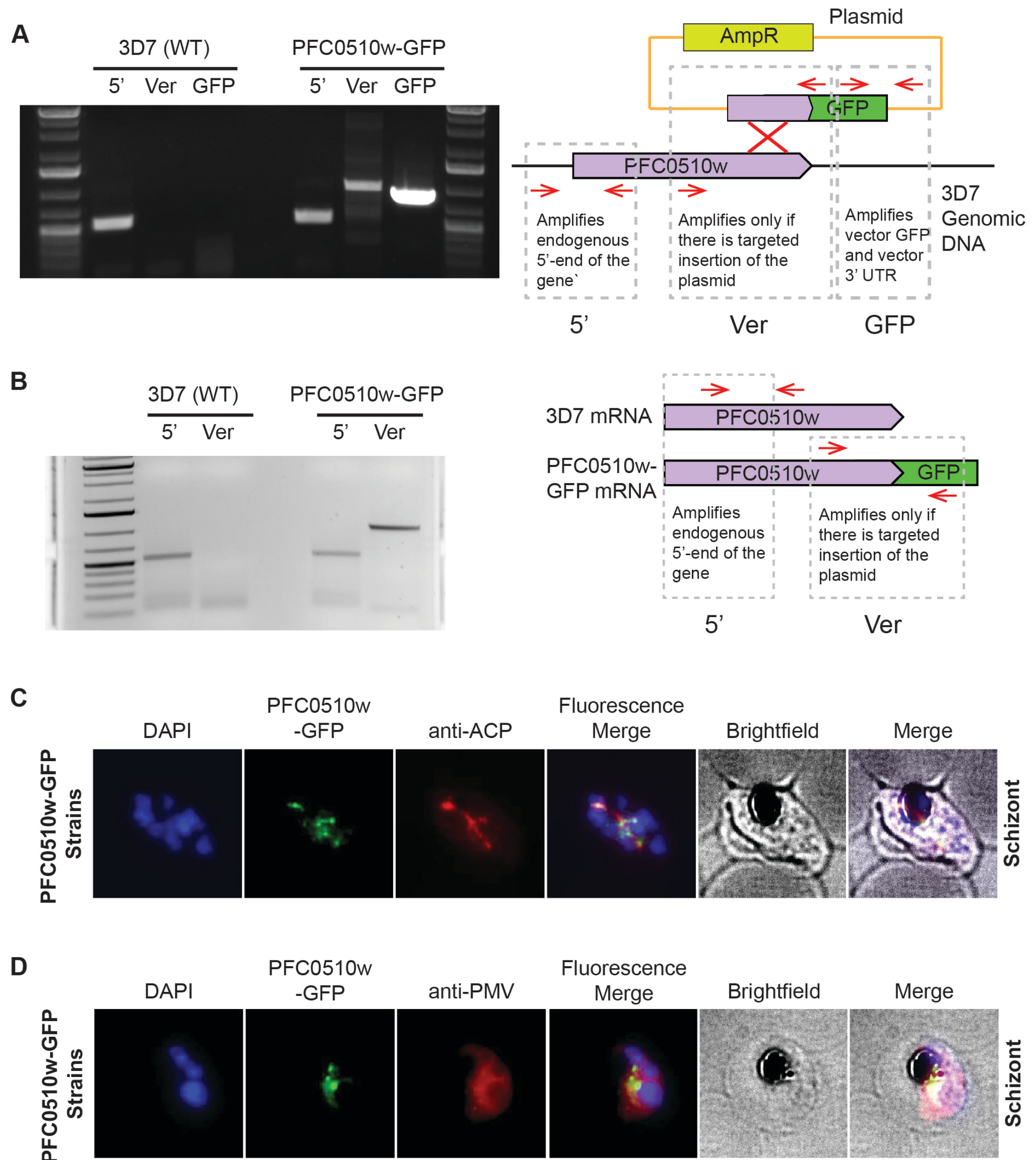


Figure 3.5: GFP tagging strategy and apicoplast localization of PFC0510w.

(A) 3D7 parasite strains were transfected with plasmids that had a GFP fused to the c-terminus of the PFC0510w for targeted integration by homologous recombination. Transfected strains (PFC0510w-GFP) were screened by PCR, where primers pairs (Ver) only amplified a product if proper integration had taken place. **(B)** RT-PCR reveals that GFP fused to the c-terminal of PFC0510w is being transcribed in the PFC0510w-GFP strains. **(C)** Endogenous PFC0510w was fused with GFP and showed high colocalization with anti-ACP antibodies and (**D**) low colocalization with PMV, an ER membrane marker.

Gene disruption studies suggest that components within the apicoplast ERAD-like system may be essential

As previously discussed, the endogenous PFC0740c gene when fused with epitope tags caused a delayed death effect. This led us to wonder if these putative ubiquitylating ERAD-like components were indeed essential. To answer this question we constructed gene disruption vectors that targeted the putative E1 ubiquitin activating PF13_0182 gene, the E3 ubiquitin ligase PFC0740c gene and the E3 ubiquitin ligase PFC0510w gene (see Figure 3.6A for gene disruption strategy). After multiple transfections, some parasites were able to grow under selective pressure for the human DHFR gene (Crabb & Cowman, 1996), the positive selection cassette. DNA samples from the drug-selected parasites were extracted and PCR experiments were performed to validate successful integration of the gene disruption vector (Figure 3.6A). Only the PFC0510w gene allowed for successful targeted integration of its disruption vector. PCR, RT-PCR, and comparative genomic hybridization (CGH) experiments validated the integration and partial gene disruption of PFC0510w (Figure 3.6A-B). Though we are certain that we achieved a targeted insertion of the hDHFR cassette into the PFC0510w gene, CGH data revealed deletions and also duplications of multiple genes, possibly leading to complementation (Figure 3.6C). Therefore, it is difficult to rule out that PFC0510w is not an essential *Plasmodium* gene. Additional gene essentiality studies must be performed for more confidence.

For the parasites that were transfected with gene disruption vectors targeting genes PF13_0182 and PFC0740c, no successful targeted integration was detected by PCR (Figure 3.6A). The parasites that came back under selective pressure likely had random integration of the vector elsewhere in the genome. The outcomes of the gene disruption studies by themselves are negative results. However, if you couple these negative results with the observation that fusing epitope tags to PFC0740c caused a diminishment (and eventual elimination) of the transfected strains, there is a stronger case that at least PFC0740c is an essential *Plasmodium* protein.

All together, these results show that PFC0510w and PFC0740c are functional apicoplast-targeted E3 ubiquitin ligases. Additionally, these findings suggest that PFC0740c is critical to the parasite development and that PFC0510w is an excellent ERAD-like HRD1 candidate due to its high homology to the classical HRD1 such as having membrane-bound features.

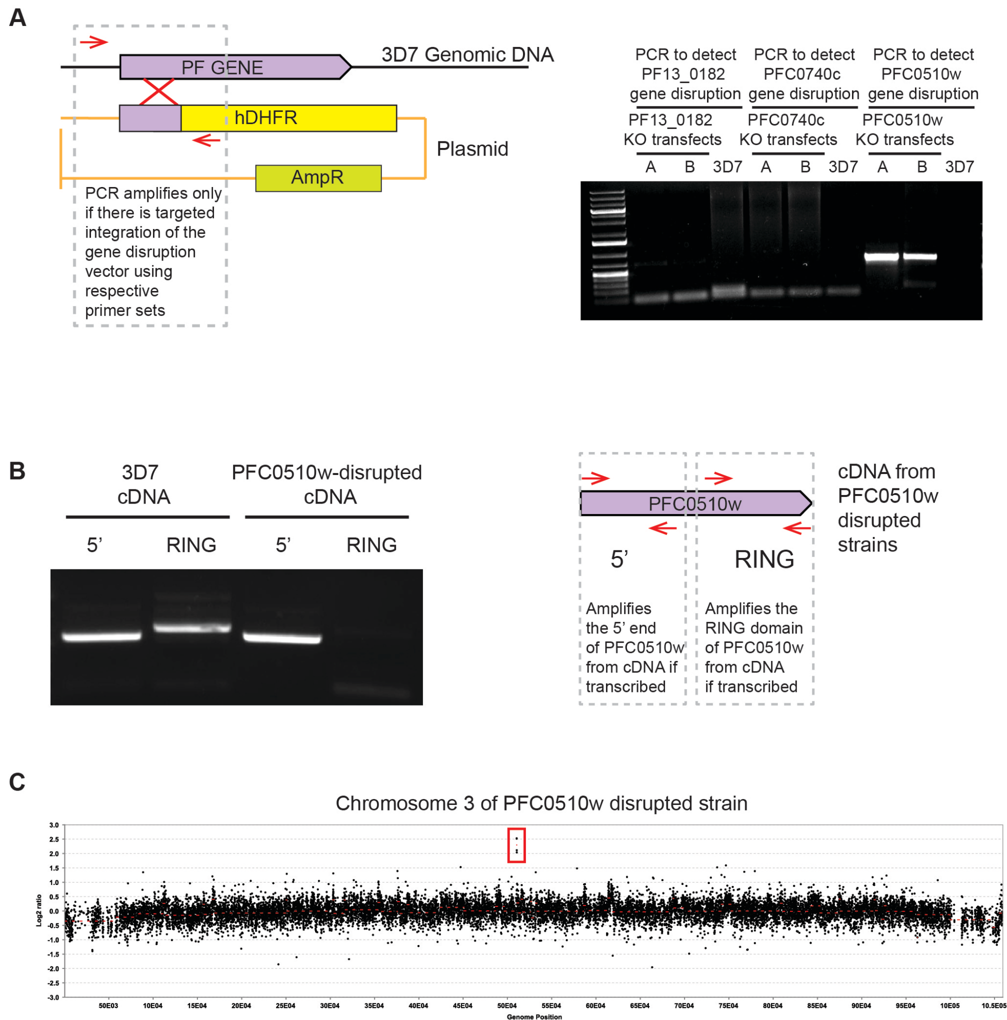


Figure 3.6: Gene disruption of ERAD-like ubiquitylating components

(A) Gene disruption strategy of PF13_0182, PFC0740c and PFC0510w. The human DHFR gene was used as the positive selection marker. PCR of transfected strains with disruption vectors show that only PFC0510w was able to be possibly disrupted. **(B)** RT-PCR of sub-cloned PFC0510w disrupted strains show no transcription of the RING domain of PFC0510w. **(C)** CGH microarray analysis show that the PFC0510w gene disruption vector was able to integrate into the genome of the transfected strain.

PF13_0344 is a functional ubiquitin activating enzyme and Mal13P1.227 is an apicoplast-targeted ubiquitin conjugating enzyme

We continued our investigation of the ERAD-like system components with the characterization of our candidate ERAD-like E1 and E2, PF13_0344 and MAL13P1.227, respectively. Again, we cloned and expressed recombinant MAL13P1.227 and PF13_0344 fused with a 6xHIS tag (Figure 3.7A). *In vitro* ubiquitylation assays were performed in the presence of our recombinant PFC0510w, PFC0740c or PF14_0215 (Figure 3.7B). When recombinant PF13_0344 and MAL13P1.227 are incubated with ubiquitin alone (lane 1), no auto-ubiquitylation was detected. However, when recombinant E3 ligases PFC0510w, PFC0740c or PF14_0215 were added, various patterns of auto-ubiquitylation were seen (lanes 2-4). These observations demonstrate that PF13_0344 and MAL13P1.227 are genuine and active E1 ubiquitin-activating and E2 ubiquitin-conjugating enzymes, respectively. Moreover, these E1 and E2 work in concert with our three recombinant ERAD/ERAD-like E3 ubiquitin ligases, *in vitro*. Given such a result, the cellular compartmentalization of these E1, E2 and E3s *in vivo* is crucial for accurate function.

We further tested MAL13P1.227 for E2 ubiquitin conjugating activity by immunoprecipitating it under native conditions using custom-made antibodies. In addition to commercially available UBA1, biotinylated-ubiquitin was used so that

only newly added ubiquitylated proteins could be detected by streptavidin affinity blotting. Auto-ubiquitylation was observed when the anti-MAL13P1.227 (E2) immunoprecipitate was used in the assay (Figure 3.7C). Since immunoprecipitations were performed under native conditions (necessary to preserve activities), the anti-MAL13P1.227 antibody probably co-precipitated interacting binding partners of Mal13P1.227 and possibly some of their targeted substrates.

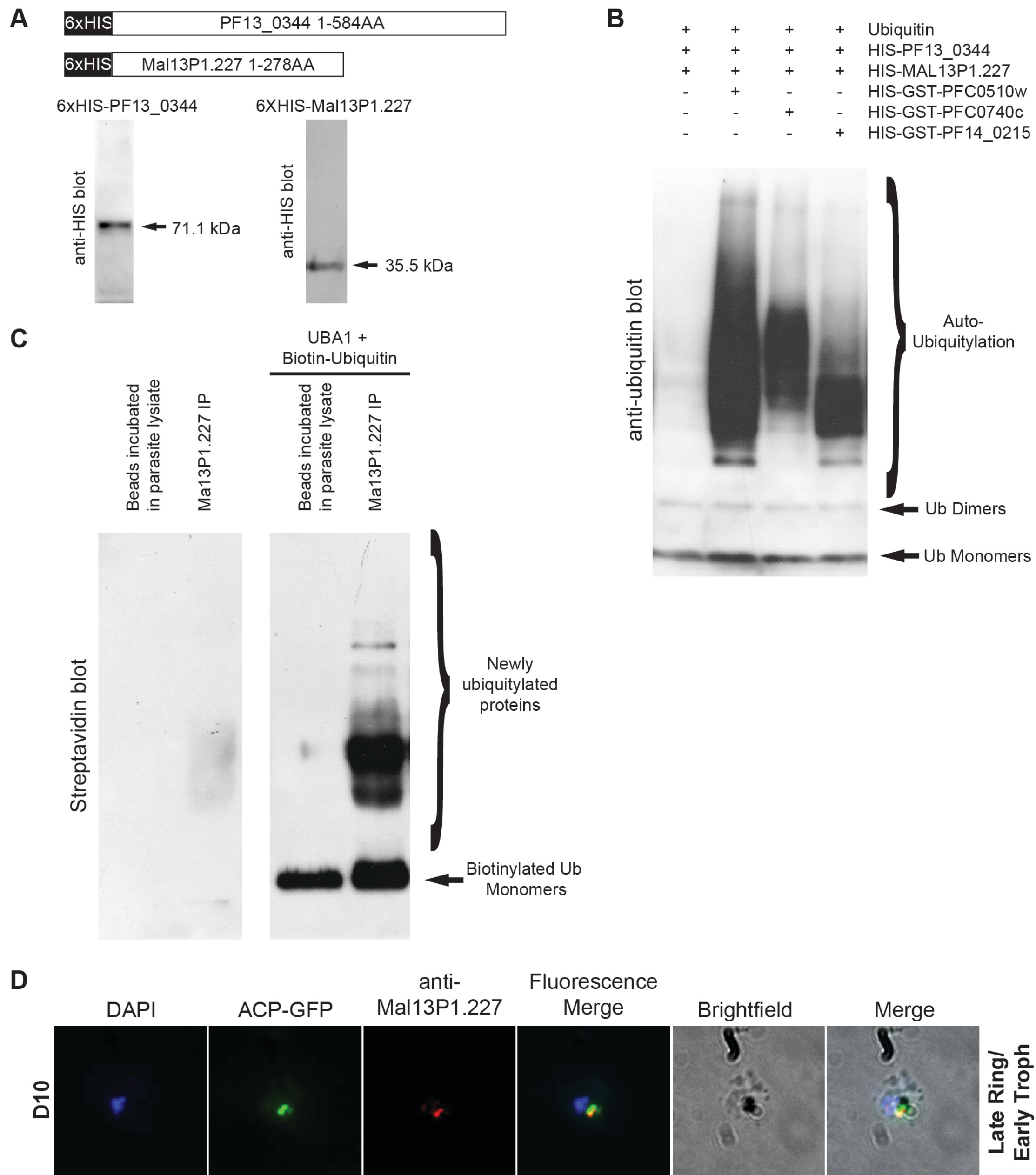


Figure 3.7: PF13_0344 and MAL13P1.227 have *in vitro* ubiquitin E1 activating and E2 conjugating activity, respectively.

(**A**) 6xHIS-tagged recombinant PF13_0334 and MAL13P1.227 are depicted and anti-HIS blots reveal purification. (**B**) Recombinant PF13_0344 and MAL13P1.227 produce ubiquitylated products *in vitro* when incubated with recombinant RING domains of E3 ligases PFC0510w, PFC0740c and PF14_0215. (**C**) Custom antibodies were used to immunoprecipitate MAL13P1.227 and its interacting proteins. *In vitro* ubiquitylation assays

were performed using the pulled-down proteins and biotinylated ubiquitin (which allows for the detection of newly ubiquitylated products). Though no ubiquitylation was observed with the eluted proteins from naked beads incubated in parasite lysates, a significant amount of ubiquitylated proteins was detected from the proteins eluted from the beads with attached anti-MAL13P1.227 antibodies that were also incubated in parasite lysates. (**D**) IFA experiments in D10 strains show MAL13P1.227 localizes to the apicoplast using anti-MAL13P1.227 antibodies

We examined the cellular localization of MAL13P1.227 using an epitope tagging strategy. Similar to what we observed with PFC0740c, this strategy was unsuccessful, indicating that MAL13P1.227 may also be essential for parasite survival. Thus indirect immunofluorescence microscopy using our custom antibody was performed. We found that MAL13P1.227 partially overlaps with the ACP signal, indicating an apicoplast localization, possibly recruited to a peripheral compartment (Figure 3.7D). Combined with our previous observation, our findings show that MAL13P1.227 is an active E2 ubiquitin-conjugating enzyme, components of the *P. falciparum* ERAD-like system, that resides at the apicoplast periphery, together with the E3 ligases PFC0510w and PFC0740c.

Characterizing putative deubiquitylating enzymes in the ERAD-like system

Deubiquitylating enzymes (DUB) that would deconjugate ubiquitin after transport through the second outermost apicoplast membrane logically appear as important components of the ERAD-like system. In addition to reversing the addition of ubiquitin, DUBs act to replenish the free ubiquitin pool. From the list of

the three *in silico* candidate apicoplast-targeted DUBs (Table 3.1), we were able to validate the *in vitro* DUB activity of PF10_0308. The PF10_0308 protein has a putative OTU domain. The OTU domain was cloned, expressed and tested in an *in vitro* deubiquitylating assay with ubiquitin conjugated to 7-amino-4-trifluoromethylcoumarin (AFC), a fluorophore that is capable of fluorescing when cleaved from the ubiquitin (Mason *et al*, 2004). When recombinant PF10_0308 was added together with ubiquitin-AFC, a significant level of relative fluorescence was detected which indicates that PF10_0308 was freeing the fluorophores from ubiquitin (Figure 3.8A). N-Ethylmaleimide (NEM), an alkylating agent and a highly potent DUB inhibitor, blocked the hydrolysis of ubiquitin-AFC by PF10_0308. These results strongly suggest PF10_0308 to be a bona fide DUB.

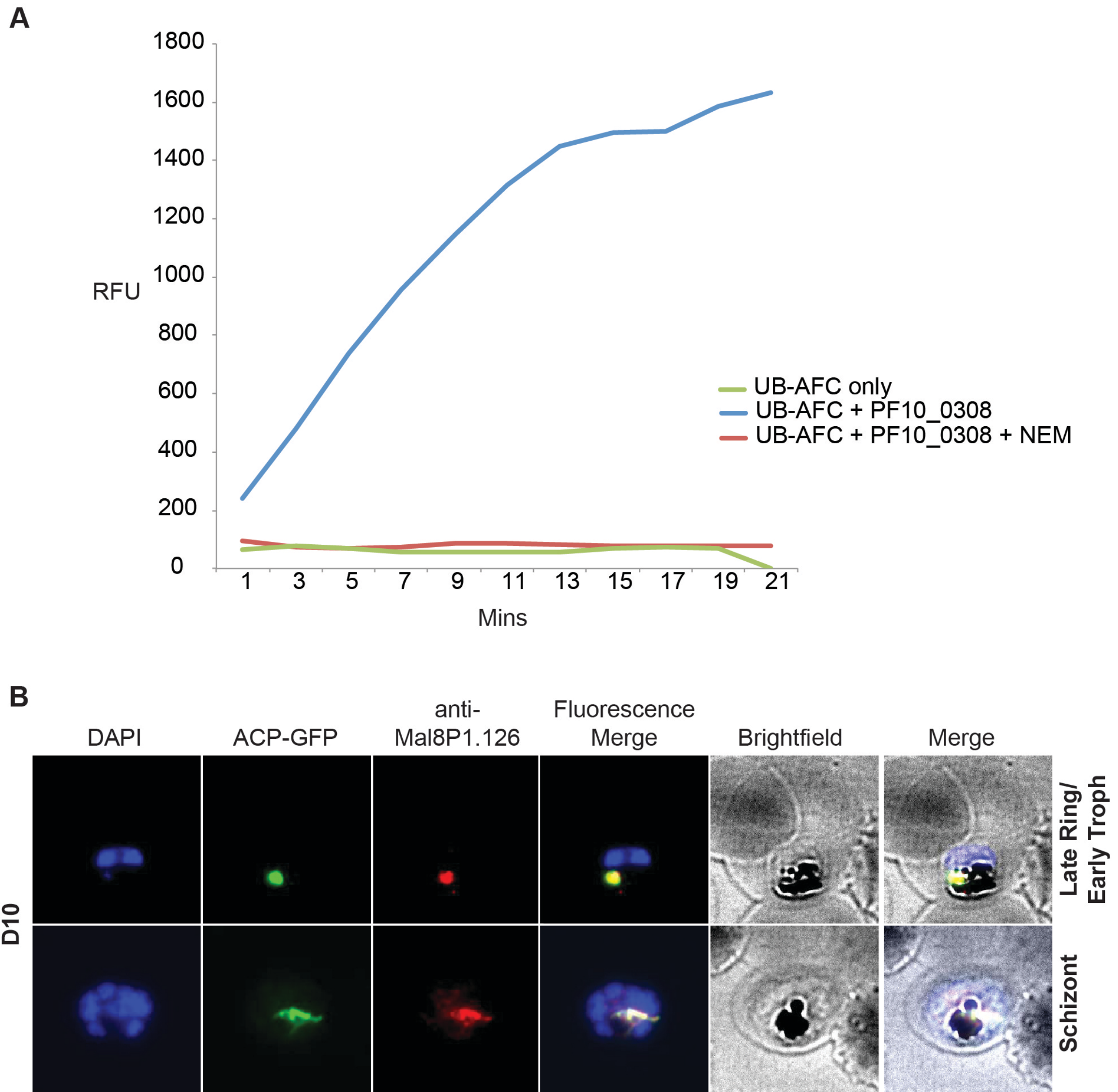


Figure 3.8: Putative deubiquitylating enzymes PF10_0308 and Mal8P1.126 mediate DUB activity and localize to the apicoplast, respectively.

(A) Ubiquitin-AFC incubated with recombinant PF10_0308 led to a rapid increase in relative fluorescence indicating the release of the ubiquitin moiety from the fluorophore. In the presence of the DUB inhibitor NEM, the hydrolysis of ubiquitin-AFC was blocked. **(B)** Using D10 strains and custom antibodies against Mal8P1.126, IFA experiments show colocalization of MAL8P1.126 with the apicoplast marker ACP.

In addition, we designed custom-made antibodies that are specific to MAL8P1.126. IFA experiments using D10 strains reveal that MAL8P1.126 colocalizes with ACP, confirming apicoplast localization (Figure 3.8B). Though the deubiquitylating activity of Mal8P1.126 has yet to be validated, the localization of the DUB candidate to the apicoplast suggests the presence of a complete ubiquitylating system within the *Plasmodium* plastid.

DISCUSSION

The described results provide the first characterization to date of the core ubiquitylating components of the *Plasmodium* apicoplast-specific duplicated ERAD-like system. While the ERAD system in *Plasmodium* is likely no different in function from the ones found in classical model organisms – to translocate misfolded proteins for proteasome dependent degradation – the ERAD-like system is believed to be the missing protein import system in the second outermost apicoplast membrane, the PPM. In this model, the secondary ERAD-like system acts as a PPM translocator, transporting plastid proteins from the ER lumen to the periplastid compartment (PPC) (Spork *et al*, 2009; Sommer *et al*, 2007; Kalanon *et al*, 2009).

Plasmodium's apicoplast is spatially-associated with the ER (Tonkin *et al*, 2006) though whether the apicoplast creates transient contacts with the ER or is situated terminally adjacent but separate from the ER is still unclear (Tonkin *et al*, 2008) (see Figure 3.9 for a schematic representation). In the first case, all secreted proteins wash over the PPM, and only proteins with the transit peptide, will be imported into the apicoplast. In the second scenario, ER-derived vesicles shuttle apicoplast proteins and fuse with the apicoplast outermembrane. In either case, apicoplast-targeted proteins have already crossed the most outermembrane when they are brought in to the ER lumen by co-translation through the Sec61 complex in the rough ER. However, how proteins are transported across the PPM, and eventually through the plastid's most inner membranes, remains elusive. The

identification and characterization of several apicoplast proteins that have homology to that of known ERAD components and other chloroplast translocons have given us possible clues to unraveling the mystery behind the translocation of proteins across the multiple inner membranes.

So far, preliminary studies have identified the plastid homologues to the ERAD ubiquitylating complex in a few apicomplexans (Spork *et al*, 2009; Sommer *et al*, 2007). Relict homologues of DER1 have been identified as integral proteins in the PPM with the C-terminus facing the PPC. In *Toxoplasma*, apicoplast DER1 knockdown leads to disrupted apicoplast import. In the diatom *Phaeodactylum tricornutum*, DER1 complexes are reported to bind with transit peptides of PPC-targeted plastid proteins but not stroma-targeted plastid proteins, suggesting that DER1 complexes may act to discriminate pre-proteins and help guide proteins to either the two outer-membranes or the apicoplast stroma (Hempel *et al*, 2009). In spite of no identified HRD1 homologue, Hempel *et al.* (Hempel *et al*, 2009) characterized an E3 RING finger ubiquitin ligase, *pt*E3P, that was specifically localized to the plastid PPM of *P. tricornutum*. By homology with the classical ERAD system, it is very likely that plastid ERAD-like ubiquitylating enzymes may serve to help recognize and translocate plastid pre-proteins with transit peptides across the PPM. Supporting this hypothesis, recent studies show that the transit peptide must be unstructured for proper transport to the apicoplast (Gallagher *et al*, 2011). While the ERAD system recognizes misfolded proteins, perhaps the apicoplast ERAD-like

system recognizes unfolded structures of the transit peptide for apicoplast translocation. In the ERAD-like pathway, we suggest that pre-proteins that have transit peptides are recognized by the DER1-like complex and ubiquitylated by ubiquitin ligases found at the PPM as a pre-requisite for plastid import.

In *Plasmodium*, duplicated DER1-like, E1 UBA1-like and ubiquitin-like proteins have been identified and localized to the apicoplast, consistent with the presence of an ERAD-like system to import nuclear-encoded proteins into the apicoplast (Spork *et al*, 2009). However, no apicoplast-targeted E3 ligase has been previously identified and it was proposed that the ERAD-like system may be independent of an E3 ligase activity (Spork *et al*, 2009; Hoeller *et al*, 2007). In our present analyses, we successfully identify and validate minimal necessary ubiquitylating components to make up a complete functioning system that targets the apicoplast (Figure 3.9).

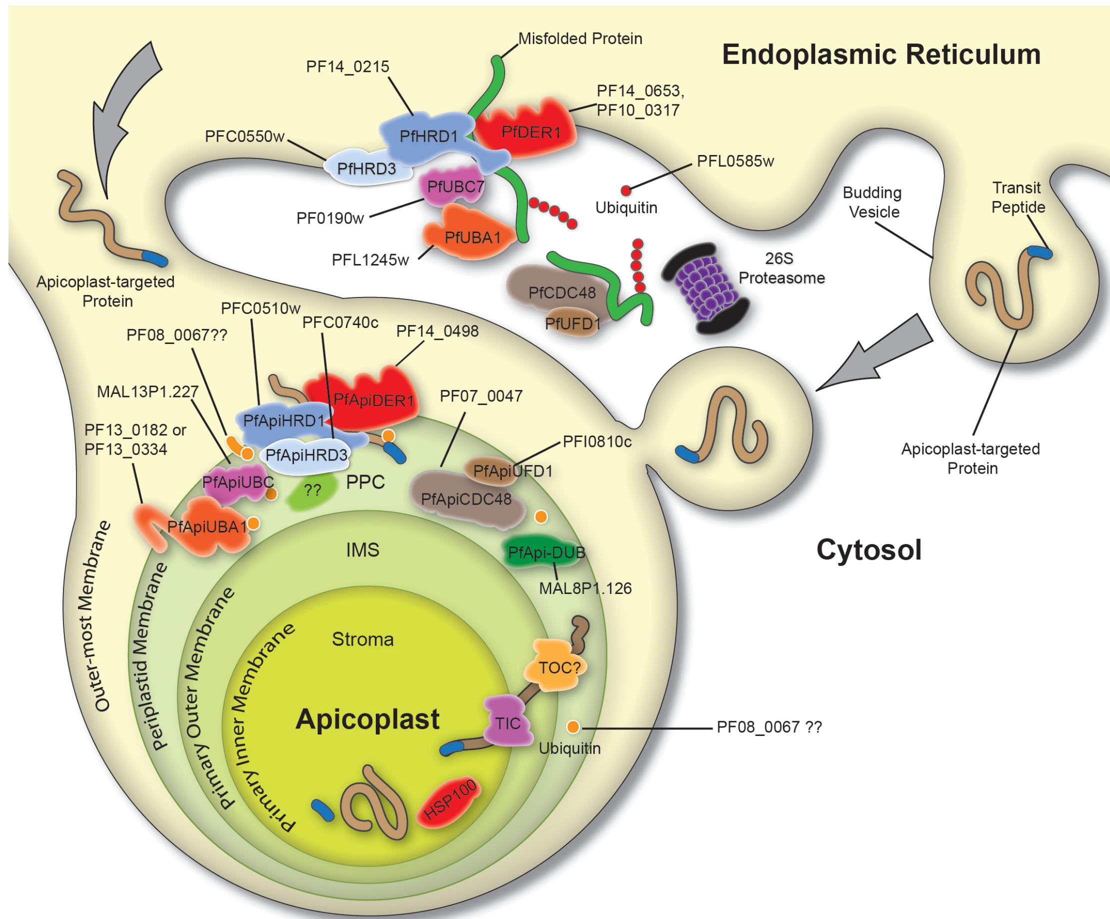


Figure 3.9: Overview model of the *Plasmodium* ERAD and ERAD-like systems

(Source: D. Chung)

Vesicular trafficking from the ER or by transient contact with the ER are the two main proposed routes of apicoplast protein import. In either case, a transmembrane protein import machinery must exist at the PPM. It is likely that the ERAD-like components that target to the apicoplast serve to function as this transmembrane import machinery. Similar to how the ERAD system recognizes misfolded proteins in the ER, the apicoplast ERAD-like system likely recognize the unstructured transit peptide of apicoplast preproteins and target them for translocation across the PPM. In the diagram above are some of the identified putative proteins that are believed to be involved in both the ERAD and ERAD-like systems of the *Plasmodium*, along with their proposed roles for either protein degradation or apicoplast import.

PPM, periplastid membrane; PPC, periplastid compartment; IMS, intermembranal space.

One of our characterized E3 ubiquitin ligases; PFC0510w, has distant homology with a classical HRD1 protein. Our fluorescence microscopy images indicate that PFC0510w reside at the periphery of the apicoplast and most likely localizes to the PPM with its RING domain within the PPC. It is very likely that PFC0510w works in conjunction with the apicoplast *Pf*DER1 to transport apicoplast proteins across the PPM. The nature of the E3 ligase PFC0740c is more mysterious, with no clear homology to other known proteins. Despite the absence of a predicted internal transmembrane domain, our localization experiments suggest that PFC0740c is associated with the apicoplast periphery. Close examination of the predicted signal peptide reveals that the last third of it could actually be an anchor signal (ranging from amino acid 39 to 61 according to TMHMMfix). In that case, PFC0740c would be anchored to the membrane *via* its N-terminal end in a way similar to the configuration of HRD3 in the classical ERAD complex. Likewise, with the absence of any predicted transmembrane domain, the ubiquitin-conjugating enzyme MAL13P1.227 is predicted to target to the stroma of the apicoplast. Nonetheless, our immunofluorescence imaging suggests a peripheral localization consistent with sequestration of the protein within the PPC and a role in the ERAD-like pathway.

While the role of ubiquitylation in the classical ERAD pathway is well documented in various eukaryotic model organisms, the role of ubiquitylation within the *Plasmodium* ERAD-like system remains unseen. As the apicoplast pre-

proteins are inserted into the PPM for translocations, these pre-proteins could be ubiquitylated to serve as the binding recognition substrate for a CDC48-like complex, which acts to fully dislodge the inserted pre-proteins from the PPM in an ATP-dependent manner. Following extraction, DUBs found at the PPC would cleave ubiquitin-like from the modified pre-protein to ensure proper maturation. PCC-targeted proteins would then be processed and folded, while stromal-targeted pre-proteins would continue their journey across the two most inner membranes of the plastid. The role of the specialized apicoplast-targeted ubiquitin-like (PF08_0067) in this process remains unclear. Despite a fairly long sequence (373 aa), PF08_0067 contains a single ubiquitin repeat at the N-terminus of the protein that ends around position 144. The rest of the protein represents a substantial portion of PF08_0067, containing a serine-rich repeat at its proximal side and three putative transmembrane domains at the C-terminus. The release of the ubiquitin moiety therefore would require proteolytic cleavage by a specific deubiquitylase or even by auto-catalysis. This processing can occur in any compartment along the protein's export pathway, such as the PPC. In any case, PF08_0067 lacks the di-glycine motif that is necessary for covalent attachment to target substrates. Perhaps, PF08_0067 may act as a chaperone instead of a covalently-bound modifier. Another possibility is that PF08_0067 may have similar functions as the HERP protein found in the ERAD of mammalian cells. Like PF08_0067, HERP has an ubiquitin-like domain at the N-terminus followed by transmembrane domains at the c-terminus (Schulze *et al*, 2005). It is reported that HERP is a membrane-bound protein that interacts with

HRD1 of the ERAD complex and positively regulates ubiquitylation of proteins needed for retro-translocation (Kny *et al*, 2011). If this was the case, an additional ubiquitin-like protein targeting the apicoplast will be needed.

The role of a complete and fully functional ERAD-like system in translocating apicoplast proteins through the PPM is now very likely, since the previous lack of an E3 ubiquitin ligase could have instead suggested a non-functional vestigial system. The ERAD-like system may however expand its function to the last two innermost membranes, POM and primary inner membrane (PIM). Though a translocon candidate for the POM has yet to be identified in *Plasmodium*, a homologue to the central component of the translocon of the outer chloroplast membrane (TOC), OMP85, has been identified in *Phaeodactylum tricornutum* and localized to its plastid (Bullmann *et al*, 2010). This suggests that a translocon similar to the TOC or OMP85 also is likely to exist in the *Plasmodium* apicoplast and function to translocate proteins across the apicoplast innermembranes. Furthermore, homologs for the translocon of the outer chloroplast membrane (TIC), *Pf*TIC22 and *Tg*TIC20, were found to localize to at least one of the apicoplast membranes in *Plasmodium* and *Toxoplasma*, respectively (Kalanon *et al*, 2009; van Dooren *et al*, 2008). Because of its evolutionary origins, it is believed that *Pf*TIC22 resides in the most inner membrane of the apicoplast (Lim *et al*, 2009) and may help in the translocation of proteins with a bipartite leader across the PIM into the apicoplast stroma.

The biology of the *Plasmodium* apicoplast is particularly complex. And because the apicoplast is an essential organelle that does not have a counterpart in its animal host, it is a promising target for therapeutic intervention. Thus, understanding the parasite's biology will allow for better developing disease strategies against malaria. Recently, it has been reported that the isoprenoid biosynthesis pathway is the sole essential function of the apicoplast within asexual stages (Yeh & Derisi, 2011). However, considering the maintenance of this large organelle throughout evolution, the apicoplast may serve other vital roles outside the asexual stages. In any case, the import of nuclear-encoded proteins involved in the essential pathways of the apicoplast remains unclear. The identification and characterization of an apicoplast-targeted ERAD-like system support the likelihood of an ubiquitin-dependent translocating system to regulate nuclear-encoded apicoplast proteins transport across the multiple membranes of the indispensible parasite apicoplast. Furthermore, although the apicoplast ERAD-like components already exhibit less homology to its more conserved ERAD counterpart, higher specificity for drug targeting can be achieved by focusing on the E3 ligases and DUBs, since they are more varied and diverse (Ponts *et al*, 2008). Altogether, our data highlight components of the ERAD-like system as excellent new antimalarial candidates.

CONCLUDING REMARKS FOR CHAPTER 3

If one cuts off the flow of proteins to the vital parasite organelle such as the apicoplast, then one also cuts off the livelihood of the entire parasite. The logic behind this statement appears fairly simple but the ability to carry out this statement a bit more difficult. However, with the identification and initial biological investigation of a duplicated ERAD-like system within the *Plasmodium* apicoplast, the ability to cut off the import of apicoplast proteins is now more tenable.

Granted, there is still much work to be done. We have yet begun to characterize the other ERAD-like components within the apicoplast and we still have very little knowledge of how proteins translocate across the two innermost membranes. Nonetheless, our early investigation of the *Plasmodium* ERAD-like system is an excellent opening step in identifying the essential functions of ubiquitylation within parasite biology. With more investigation, we may be able to uncover the full mechanisms of protein import to the apicoplast and hopefully utilize it for an effective therapeutic intervention.

MATERIALS AND METHODS

Bioinformatic analysis

Protein identification, analysis and protein-targeting predictions were conducted using PlasmoDB v8.0, hidden Markov models, Pfam domain searches, PATS v1.2.1N, PlasmoAP, TargetP v1.1, ChloroP v1.1, Signal3.0, BLASTP v2.2.24+, MUSCLE, T-COFFEE, InterProScan sequence search and TMHMM v2.0c. For a more detailed methodology, see supplemental materials.

Cloning and purification of recombinant proteins

A list of primers used for the cloning of recombinant proteins is given in Supplemental Table 1. The RING domains of PFC0740c and PFC0510w were codon optimized (Genscript) and cloned into pGS-21a (Genscript), which contains both a GST and 6xHIS tag. Cloned plasmids were then transfected into Bl21 DE3 RIL codon plus (Stratagene) *E. coli* cells. The RING domain of PF14_0215 and the OTU domain of PF10_0308 were amplified, cloned into pGS-21a, and transfected into Articexpress (Stratagene) *e.coli.*

6xHIS tagged recombinant proteins (Mal13P1.227, PF13_0334) were constructed by using a modified version of PGS21a (PGS-21aHIS), which has the GST coding region removed. GST removal was done by cutting PGS-21a with ClaI and NcoI and inserting a 6xHIS PCR fragment that was amplified using the primers

GATCGAGATCGATCTCGATC and ATCCATGGCCTTACCGCTGCTATGATGATGAT from the PGS-21a plasmid itself.

The primers used to amplify the full length Mal13P1.227 by PCR are AGGGGATCCCTTAAGCCGCGGATGTTCAACATAATGAGACCAAT and AGGAAGCTTTTAACTAGTGCTAGCTTAACATTTATCATCATGATATAGA. The primers used to amplify the E1 ubiquitin activating domain are AGGGGATCCCTTAAGCCGCGGATGTACCAAGTTGTCAAGGAAT and AGGAAGCTTTTAACTAGTGCTAGCTTATAAAAAGGGCAAATTTTTGAAT. The amplified regions were cut with SacII and SpeI, and then ligated into PGS-21aHIS vector.

Cloned expression *e.coli* cells were grown to an OD_{600} of >0.5. IPTG was added to a concentration of 1mM for induction and incubated overnight at 12°C. Cells were spun down and resuspended in lysis buffer (25mM Tris-HCl, pH 7.5, 500mM NaCl, and 1% (v/v) Triton X-100) with protease inhibitor cocktail (Roche) and 1mM AEBSF. Cells were sonicated and spun down.

GST purifications were performed with glutathione agarose (Sigma). Bound proteins were washed three times with GST wash buffer (25mM Tris-HCl, pH 7.5, 300mM NaCl and 1% Triton X-100). GST-tagged proteins were then eluted with GST elution buffer (25mM Tris-HCl, pH 7.5, 150mM NaCl, 15mM reduced glutathione and 0.01% Triton X-100 and 40% (v/v) glycerol). Anti-GST immunoblots were probed

with goat anti-GST antibodies (1:5000; GE Healthcare) and donkey anti-goat antibodies conjugated to horseradish peroxidase (HRP) (1:20,000; Jackson Immuno research).

HIS-tagged protein purifications were performed with Ni-NTA beads (Qiagen) and subsequently washed several times with a solution containing 25mM Tris-HCL pH 7.5, 500mM NaCl, 30mM imidazole and 5% glycerol. Purified proteins were eluted with 25mM Tris-HCl pH 7.5, 500mM NaCl, 250mM imidazole and 5% glycerol. Anti-HIS immunoblots were probed with mouse anti-HIS antibodies (1:2500; Millipore) and goat anti-mouse antibodies conjugated to horseradish peroxidase (HRP) (1:10,000; BioRad).

Cloning of GFP tagged PFC0740c and PFC0510w

The first 167AA of PFC0740c was amplified using 3D7 cDNA using primers CCCTCGAGATGAGTTTTATAGATGAATATGATT and CCGGTACCCCTAGGTATTTCTTCTAGTAACACACCAA and inserted into pARL2GFP (Przyborski *et al*, 2005) that has been cut with XhoI and KpnI.

The 3' end of PFC0510w was amplified from 3D7 genomic DNA using primers TGGGGCCCGTCGACACTAGTGAAGGTTGGGAACATATGCAAAGG and CCCGGTACCCTGCAGCTCGAGGGAAGAAAAGTTGGGGAGGGGACC and inserted into a modified PHH1 plasmid (Reed *et al*, 2000) with GFP, using Sal1 and Pst1.

Custom antibodies

We received an expression vector from the H. Zhu lab that had cloned a HIS tagged active domain of Mal13P1.227 (Vedadi *et al*, 2007). We expressed and purified the recombinant protein and used it for antibody production (Genscript) against Mal13P1.227. The following peptide sequences were used as antigens for the production of affinity-purified antibodies (Fisher Scientific or Genscript) in rabbits: EEIKKIYSITRDRYVY for Mal8P1.126.

Parasite culturing and transfection

The *P. falciparum* 3D7 and D10_ACP (leader) GFP (provided by MR4, MRA568) strains were grown in human O+ red blood cells according to standard protocols (Trager & Jensen, 1976) with the exception that cultures were incubated in gassed flasks. Transfection of parasite cultures were carried out by electroporation of infected human O+ red blood cells as described in (Deitsch *et al*, 2001). Transfectants were selected with WR99210 (Fidock & Wellems, 1997).

Immunofluorescence Assay

Immunofluorescence assay was performed using the protocol described in (Spork *et al*, 2009). Briefly, cells were resuspended in PBS and spun down for a few minutes at low speeds. With the supernatant discarded, the pellet was resuspended in 1 mL of PBS containing 4% paraformaldehyde and 0.0075% glutaraldehyde and

incubated at 37°C for 1 hour. After gentle spinning, the fixed pellet was washed (and incubated for 10 minutes) with 1.25M glycine/PBS. After another round of spinning, the pellet was then incubated in 0.1% TritonX-100/PBS solution (10 minutes) and then spun again and washed with 125mM glycine/PBS for another 10 minutes. After centrifugation, the pellet was blocked in 3% BSA/PBS for at least 1 hour at room temperature. Then the primary antibodies were added and incubated for at least 2 hours at room temperature (or overnight at 4°C). After washing with PBS, secondary antibodies were added and the cells were incubated for about two hours at room temperature and then subjected to a final round of washes. DAPI was added to visualize the nuclei.

Primary antibodies with their respective dilutions are as follows: anti-Mal13P1.227 (1:100). Secondary goat anti-rabbit IgG Alexa Fluor® 488 (1:100) or donkey anti-rabbit IgG Alexa Fluor® 568 (1:100) were then used for the respective primary antibodies. DAPI (100ng/mL final concentration) was added and slides were mounted with slow-fade mounting medium (Fluoromount-G; Southern Biotech) and viewed with fluorescence microscopy. Rabbit anti-ACP antibodies (a gift from the Geoffrey I. McFadden lab) were used at a 1:250 dilution (Tonkin *et al*, 2004). Mouse anti-plasmepsin V (PMV) antibodies (Klemba & Goldberg, 2005) were obtained from the Malaria Research and Reference Reagent Resource (MR4) center and used at a 1:20 dilution with donkey anti-mouse IgG Alexa Fluor® 568 (1:100).

Images were observed with the Olympus BX40 microscope using an 100x objective lens (UPlanFI) and captured by the CoolSNAP cf (Photometrics) camera using Metavue software. Images were merged and background was reduced using ImageJ software.

In vitro ubiquitylation assay

50-200μM of ubiquitin (Boston Biochem), 0.05-0.2μM of E1 enzyme, 1-5μM of E2 enzymes, and 1-12.5μM of E3 ligases were incubated together in reaction buffer (50mM Tris-HCl, pH 7.4; 1mM DTT and Re-energizing buffer (Boston Biochem)) for 2 hours at 37°C and then analyzed by SDS-PAGE and immunoblotting. Human recombinant UBE1 and UBC enzymes were purchased from Boston Biochem. Anti-ubiquitin immunoblots were probed with rabbit anti-ubiquitin antibodies (1:2500; Upstate) and goat anti-rabbit antibodies conjugated to HRP (1:5000; Pierce). Anti-GST immunoblots were probed with goat anti-GST antibodies (1:2500; GE Healthcare) and donkey anti-goat antibodies conjugated to HRP (1:5000; Jackson Immunoresearch).

Immunoprecipitation and *in vitro* ubiquitylation assay of pulled-down proteins

Isolated 3D7 parasites were resuspended in 20mM HEPES pH7.9, 10mM KCl, 1mM EDTA, 1 mM EGTA, 1mM DTT, 0.5mM AEBSF (Fisher Scientific), 0.65% Igepal v/v and cocktail protease inhibitor (Roche) (lysis buffer 1). Parasites were allowed

to lyse on ice for 10 minutes and then were spun down. Supernatant was collected. The pellet was then subsequently resuspended in 20mM HEPES pH 7.9, 0.1M NaCl, 0.1mM EDTA, 0.1mM EGTA, 1.5 mM $MgCl_2$, 1mM DTT, 1mM AEBSF and cocktail protease inhibitor (Roche) (lysis buffer 2), and then incubated at 4°C with vigorous shaking for 20 minutes. After spinning, the supernatant was collected. The remaining pellet was then resuspended in lysis buffer 1 and then sonicated and spun down. Supernatant was collected and pooled with the previous two collected supernatants to create a mixture of extracted proteins. Extracted proteins were precleared with magnetic Protein A beads (Millipore) and then incubated with respective antibodies for 2 hours at 4°C with gentle shaking. For a negative control, no antibodies (just beads) were added to a sample of protein extracts. Newly washed magnetic Protein A beads were added to each sample and allowed to incubate overnight at 4°C with gentle shaking. Samples were washed several times with lysis buffer 1. Pulled-down proteins were left bound to the beads.

Still bound to the beads, the pulled-down proteins were mixed together with 0.2M HEPES pH7.9, 50mM DTT, re-energizing buffer (Boston Biochem), 0.5μg/μL final concentration of biotin conjugated ubiquitin (Boston Biochem), 0.5μg/μL final concentration of ubiquitin (Boston Biochem), 5mM AEBSF (Fisher Scientific) and cocktail protease inhibitor (Roche). The reactions were allowed to incubate at 30°C with gentle agitation for two hours. After incubation, pulled-down proteins were

eluted with 4x Laemmli buffer and incubated at 95°C for 5 mins. Activity was visualized with affinity blots. Biotin affinity blots were probed with streptavidin conjugated with peroxidase (1:10,000; Jackson Immunoresearch).

Proteins and domain motifs datasets

The translated genome of *P. falciparum* v8.0 was obtained from PlasmoDB (www.plasmodb.org). A second *P. falciparum* proteins dataset was built and consists of 473 genes which products are predicted to be apicoplast-targeted by PlasmoAP (available on PlasmoDB).

Twenty-seven hidden Markov models (HMM) of domain motifs commonly found within ubiquitylating proteins were obtained from the PFAM database (Finn *et al*, 2010) (Pfam accession numbers PF04110, PF00888, PF05903, PF00646, PF00632, PF02099, PF02991. PF01398, PF02338, PF01088, PF02902, PF03416, PF00519, PF07525, PF00899, PF04564, PF09358, PF02134, PF00240, PF00443, PF03671, PF00179, PF09138, PF08325, PF00097, PF08746, PF02891).

Pfam domain search using probabilistic pattern recognition

The HMM search component of the HMMER 3.0 package was used to identify proteins that carry ubiquitylation-related domain motifs by iterative probabilistic pattern recognition (Johnson *et al*, 2010; Sonnhammer *et al*, 1998). Results were manually curated by visual inspection of protein sequence alignments using

hmmalign and mview (Brown *et al*, 1998) performed on the web bioinformatics portal Mobyle (Néron *et al*, 2009). Analyses were performed on both the *P. falciparum* translated genome and the 473 candidate apicoplast-targeted proteins to identify ubiquitylating proteins.

Identification of apicoplast-targeted ubiquitylating candidate proteins

The n-terminal side of the ubiquitylating proteins identified by HMM search was analyzed using PATS v1.2.1N (Zuegge *et al*, 2001), PlasmoAP (Foth *et al*, 2003), TargetP v1.1 (Emanuelsson *et al*, 2000) (with the prediction of the signal peptide cleavage site turned on (Nielsen & Krogh, 1998), ChloroP v1.1 (Emanuelsson *et al*, 1999), and SignalP3.0 (Bendtsen *et al*, 2004). Homologues were identified in the NCBI non-redundant protein database using BLASTP v2.2.24+ (Altschul *et al*, 1997) and aligned using MUSCLE (Edgar, 2004)and T-COFFEE (Di Tommaso *et al*, 2011).

Protein signature domain analysis and domain architecture

The domain architecture of the selected candidates was analyzed using the InterProScan Sequence Search from the InterPro database (Hunter *et al*, 2009). In addition, transmembrane domains were predicted by TMHMM v2.0c. Scenarios for spatial organization of the transmembrane domains were built using TMHMMfix (Melén *et al*, 2003).

Gene disruption experiments

Gene disruption plasmids were constructed for either a single or double homologous crossover strategy. In the single crossover strategy, two-GATEWAY (Invitrogen) cloning was implemented. The 5'end of the targeted gene and a hDHFR resistance cassette were amplified by PCR (primers are given in supplemental table 1) and inserted into pDONR221-P1P5r and pDONR221-P5P2 vectors, respectively, with BP clonase. The newly made vectors are now called ENTRY vectors. The reading frame cassette B (Invitrogen) was inserted into the pCC-1 vector by digestion with AVRII and AFLII followed by blunt end formation with klenow polymerase and ligation; this created the vector pDEST pCC-1 KO. The ENTRY vectors were incubated with the pDEST pCC-1 KO vector in a LR recombination reaction to produce the gene disruption vectors using a single homologous crossover strategy. In the double crossover strategy, the 5' UTR and 3'UTR of the targeted gene were cloned flanking a hDHFR resistance cassette within a pCC1 vector. In addition, a SCCD cassette resided outside the 5'UTR and 3'UTR regions, which would serve as the negative selection. Gene disruption vectors were transfected into 3D7 strains.

Gene disruption and gene knockout vectors were transfected into 3D7 parasites. Transfectants were selected by WR99210 drugging and screened with PCR.

The hDHFR gene was amplified using the primers GGGGACAACTTTGTATACAAAAGTTGTGATCAATTTATAGAAACAA and GGGGACCACTTTGTACAAGAAAGCTGGGTACTAGATTTAATAAATATG. The primers used to amplify the genes for disruption are as follows: PF13_0182, GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGAATATTTTGTGCATCCATAT and GGGGACAACTTTTGTATACAAAGTTGTCCTATTTCGTGAACACCA; PFC0740c, GGGGACAAGTTTGTACAAAAAAGTTGTCGTTTACATGACTTCTGC and GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGAGTTTTATAGATGAA; PFC0510w, GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGAGTGACAATATAGAA and GGGGACAACTTTTGTATACAAAGTTGTTTTCTTATTTTATAAGCTC.

Comparative genomic hybridization (CGH) experiments

CGH experiments were performed following standard NimbleGen protocol using a 385K *P. falciparum* CGH microarray (design: 080222_Plasmodium_3D7_WG_CGH) as previously described (Tan *et al*, 2009). Briefly, 1μg gDNA was labeled with 1 O.D. Cy3 or Cy5 labeled random 9-mers for 2 h at 37°C. Labeling reactions were terminated with 0.5M EDTA, precipitated with isopropanol, and resuspended in water. Six μg each of labeled test and reference sample were combined, dessicated and resuspended in hybridization buffer (Roche NimbleGen). Samples were denatured at 95°C for 5 mins, incubated at 42°C for 5 mins and loaded onto a microarray for overnight hybridization (16-20 h) at 42°C.

Microarrays were washed in washing buffers (Roche NimbleGen), dried in an array dryer (Array-It) and then scanned with a GenePix 4000B to acquire microarray images (Axon Instruments). CGH data was extracted and normalized using NimbleScan software (Roche NimbleGen).

Deubiquitylation assay

Purified GST-tagged PF10_0233 was added to a final working solution of 20mM Tris, pH 7.5, 10mM DTT, 5mM EDTA, 1% BSA and 5μM of ubiquitin-AFC (7-amino-4-trifluoro-methylcoumarin) (Boston Biochem). As one of the negative controls, 15mM of NEM was added to a subset of samples. The release of AFC fluorescence, indicating the hydrolysis of ubiquitin-AFC, was measured using Ex400nm and Em505nm. Over a period of 30 minutes, readings were taken every 15 seconds by a microplate reader (SpectraMax Gemini EM, Molecular Devices).

Acknowledgements:

I would like to acknowledge Nadia Ponts for her bioinformatic work, Jacques Prudhomme for maintaining the parasite cultures and John Tan and Michael Ferdig for their CGH analysis.

REFERENCES

Agrawal S, van Dooren GG, Beatty WL & Striepen B (2009) Genetic Evidence that an Endosymbiont-derived Endoplasmic Reticulum-associated Protein Degradation (ERAD) System Functions in Import of Apicoplast Proteins. *Journal of Biological Chemistry* **284:** 33683–33691

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W & Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25:** 3389–3402

Bendtsen JD, Nielsen H, von Heijne G & Brunak S (2004) Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.* **340:** 783–795

Brown NP, Leroy C & Sander C (1998) MView: a web-compatible database search or multiple alignment viewer. *Bioinformatics* **14:** 380–381

Bullmann L, Haarmann R, Mirus O, Bredemeier R, Hempel F, Maier UG & Schleiff E (2010) Filling the gap, evolutionarily conserved Omp85 in plastids of chromalveolates. *J. Biol. Chem.* **285:** 6848–6856

Chung D-WD & Le Roch KG (2010) Targeting the Plasmodium ubiquitin/proteasome system with anti-malarial compounds: promises for the future. *Infect Disord Drug Targets* **10:** 158–164

Crabb BS & Cowman AF (1996) Characterization of promoters and stable transfection by homologous and nonhomologous recombination in Plasmodium falciparum. *Proc. Natl. Acad. Sci. U.S.A.* **93:** 7289–7294

Deitsch K, Driskill C & Wellems T (2001) Transformation of malaria parasites by the spontaneous uptake and expression of DNA from human erythrocytes. *Nucleic Acids Res.* **29:** 850–853

van Dooren GG, Tomova C, Agrawal S, Humbel BM & Striepen B (2008) Toxoplasma gondii Tic20 is essential for apicoplast protein import. *Proc Natl Acad Sci U S A* **105:** 13574–13579

Doss-Pepe EW, Chen L & Madura K (2005) α-Synuclein and Parkin Contribute to the Assembly of Ubiquitin Lysine 63-linked Multiubiquitin Chains. *Journal of Biological Chemistry* **280:** 16619 –16624

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32:** 1792–1797

Emanuelsson O, Nielsen H, Brunak S & von Heijne G (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J. Mol. Biol.* **300:** 1005–1016

Emanuelsson O, Nielsen H & von Heijne G (1999) ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. *Protein Sci.* **8:** 978–984

Fidock DA & Wellems TE (1997) Transformation with human dihydrofolate reductase renders malaria parasites insensitive to WR99210 but does not affect the intrinsic activity of proguanil. *Proc. Natl. Acad. Sci. U.S.A.* **94:** 10931–10936

Finn RD, Mistry J, Tate J, Coggill P, Heger A, Pollington JE, Gavin OL, Gunasekaran P, Ceric G, Forslund K, Holm L, Sonnhammer ELL, Eddy SR & Bateman A (2010) The Pfam protein families database. *Nucleic Acids Res.* **38:** D211–222

Foth BJ, Ralph SA, Tonkin CJ, Struck NS, Fraunholz M, Roos DS, Cowman AF & McFadden GI (2003) Dissecting Apicoplast Targeting in the Malaria Parasite Plasmodium falciparum. *Science* **299:** 705 –708

Gallagher JR, Matthews KA & Prigge ST (2011) Plasmodium falciparum Apicoplast Transit Peptides are Unstructured in vitro and During Apicoplast Import. *Traffic* **12:** 1124–1138

Garcia LS (2010) Malaria. *Clinics in Laboratory Medicine* **30:** 93–129

Hempel F, Bullmann L, Lau J, Zauner S & Maier UG (2009) ERAD-Derived Preprotein Transport across the Second Outermost Plastid Membrane of Diatoms. *Mol Biol Evol* **26:** 1781–1790

Hempel F, Felsner G & Maier UG (2010) New mechanistic insights into pre-protein transport across the second outermost plastid membrane of diatoms. *Molecular Microbiology* **76:** 793–801

Hoeller D, Hecker C-M, Wagner S, Rogov V, Dötsch V & Dikic I (2007) E3-independent monoubiquitination of ubiquitin-binding proteins. *Mol. Cell* **26:** 891–898

Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, Bork P, Das U, Daugherty L, Duquenne L, Finn RD, Gough J, Haft D, Hulo N, Kahn D, Kelly E, Laugraud A, Letunic I, Lonsdale D, Lopez R, *et al* (2009) InterPro: the integrative protein signature database. *Nucleic Acids Res.* **37:** D211–215

Johnson LS, Eddy SR & Portugaly E (2010) Hidden Markov model speed heuristic and iterative HMM search procedure. *BMC Bioinformatics* **11:** 431

Kalanon M, Tonkin CJ & McFadden GI (2009) Characterization of Two Putative Protein Translocation Components in the Apicoplast of Plasmodium falciparum. *Eukaryotic Cell* **8:** 1146–1154

Klemba M & Goldberg DE (2005) Characterization of plasmepsin V, a membrane-bound aspartic protease homolog in the endoplasmic reticulum of Plasmodium falciparum. *Mol. Biochem. Parasitol.* **143:** 183–191

Kny M, Standera S, Hartmann-Petersen R, Kloetzel P-M & Seeger M (2011) Herp Regulates Hrd1-mediated Ubiquitylation in a Ubiquitin-like Domain-dependent Manner. *Journal of Biological Chemistry* **286:** 5151 –5156

Koenderink JB, Kavishe RA, Rijpma SR & Russel FGM (2010) The ABCs of multidrug resistance in malaria. *Trends Parasitol.* **26:** 440–446

Lim L, Kalanon M & McFadden GI (2009) New proteins in the apicoplast membranes: time to rethink apicoplast protein targeting. *Trends in Parasitology* **25:** 197–200

Mason DE, Ek J, Peters EC & Harris JL (2004) Substrate profiling of deubiquitin hydrolases with a positional scanning library and mass spectrometry. *Biochemistry* **43:** 6535–6544

Melén K, Krogh A & von Heijne G (2003) Reliability measures for membrane protein topology prediction algorithms. *J. Mol. Biol.* **327:** 735–744

Néron B, Ménager H, Maufrais C, Joly N, Maupetit J, Letort S, Carrere S, Tuffery P & Letondal C (2009) Mobyle: a new full web bioinformatics framework. *Bioinformatics* **25:** 3005–3011

Nielsen H & Krogh A (1998) Prediction of signal peptides and signal anchors by a hidden Markov model. *Proc Int Conf Intell Syst Mol Biol* **6:** 122–130

Ponts N, Yang J, Chung D-WD, Prudhomme J, Girke T, Horrocks P & Le Roch KG (2008) Deciphering the Ubiquitin-Mediated Pathway in Apicomplexan Parasites: A Potential Strategy to Interfere with Parasite Virulence. *PLoS ONE* **3:** e2386

Przyborski JM, Miller SK, Pfahler JM, Henrich PP, Rohrbach P, Crabb BS & Lanzer M (2005) Trafficking of STEVOR to the Maurer's clefts in Plasmodium falciparum-infected erythrocytes. *EMBO J* **24:** 2306–2317

Reed MB, Saliba KJ, Caruana SR, Kirk K & Cowman AF (2000) Pgh1 modulates sensitivity and resistance to multiple antimalarials in Plasmodium falciparum. *Nature* **403:** 906–909

Schulze A, Standera S, Buerger E, Kikkert M, van Voorden S, Wiertz E, Koning F, Kloetzel P-M & Seeger M (2005) The ubiquitin-domain protein HERP forms a complex with components of the endoplasmic reticulum associated degradation pathway. *J. Mol. Biol* **354:** 1021–1027

Sommer MS, Gould SB, Lehmann P, Gruber A, Przyborski JM & Maier U-G (2007) Der1-mediated Preprotein Import into the Periplastid Compartment of Chromalveolates? *Mol Biol Evol* **24:** 918–928

Sonnhammer EL, Eddy SR, Birney E, Bateman A & Durbin R (1998) Pfam: multiple sequence alignments and HMM-profiles of protein domains. *Nucleic Acids Res.* **26:** 320–322

Spork S, Hiss JA, Mandel K, Sommer M, Kooij TWA, Chu T, Schneider G, Maier UG & Przyborski JM (2009) An Unusual ERAD-Like Complex Is Targeted to the Apicoplast of Plasmodium falciparum. *Eukaryotic Cell* **8:** 1134–1145

Sullivan M, Li J, Kumar S, Rogers MJ & McCutchan TF (2000) Effects of interruption of apicoplast function on malaria infection, development, and transmission. *Molecular and Biochemical Parasitology* **109:** 17–23

Tan JC, Patel JJ, Tan A, Blain JC, Albert TJ, Lobo NF & Ferdig MT (2009) Optimizing comparative genomic hybridization probes for genotyping and SNP detection in Plasmodium falciparum. *Genomics* **93:** 543–550

Di Tommaso P, Moretti S, Xenarios I, Orobitg M, Montanyola A, Chang J-M, Taly J-F & Notredame C (2011) T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. *Nucleic Acids Res.* **39:** W13–17

Tonkin CJ, van Dooren GG, Spurck TP, Struck NS, Good RT, Handman E, Cowman AF & McFadden GI (2004) Localization of organellar proteins in Plasmodium falciparum using a novel set of transfection vectors and a new immunofluorescence fixation method. *Mol. Biochem. Parasitol* **137:** 13–21

Tonkin CJ, Kalanon M & McFadden GI (2008) Protein targeting to the malaria parasite plastid. *Traffic* **9:** 166–175

Tonkin CJ, Struck NS, Mullin KA, Stimmler LM & McFadden GI (2006) Evidence for Golgi-independent transport from the early secretory pathway to the plastid in malaria parasites. *Mol. Microbiol* **61:** 614–630

Trager W & Jensen J (1976) Human malaria parasites in continuous culture. *Science* **193:** 673–675

Vedadi M, Lew J, Artz J, Amani M, Zhao Y, Dong A, Wasney GA, Gao M, Hills T, Brokx S, Qiu W, Sharma S, Diassiti A, Alam Z, Melone M, Mulichak A, Wernimont A, Bray J, Loppnau P, Plotnikova O, *et al* (2007) Genome-scale protein expression and structural biology of Plasmodium falciparum and related Apicomplexan organisms. *Mol. Biochem. Parasitol* **151:** 100–110

Yeh E & Derisi JL (2011) Chemical Rescue of Malaria Parasites Lacking an Apicoplast Defines Organelle Function in Blood-Stage Plasmodium falciparum. *PLoS Biol* **9:** e1001138

Zuegge J, Ralph S, Schmuker M, McFadden GI & Schneider G (2001) Deciphering apicoplast targeting signals - feature extraction from nuclear-encoded precursors of Plasmodium falciparum apicoplast proteins. *Gene* **280:** 19–26

CONCLUSION

Some after-thoughts and perspectives

Though malaria is one of the oldest human diseases known to man, we have still yet to unlock its many secrets. Its greatest secret, of course, is how to ultimately defeat it. For several millenniums (Neghina *et al*, 2010), mankind has waged many battles with malaria that have resulted in countless human casualties. For most of those battles, human attempts to thwart and strike down the malicious and rampant disease have been blind and ineffectual at best, and even self-detrimental at worst. With remedies such as eating whole spiders wrapped in its own webbing or wearing a fish tooth amulet (Schlagenhauf, 2004), mankind was unaware of its enemy and hopelessly fighting in the dark.

It wasn't until the discovery of the causative agents, the *Plasmodium* spp., and their vectors in the late 19th century (Cox, 2010) that the first significant malarial mysteries were unraveled. With this newly obtained knowledge, the implementation of effectual measures to curb the spread of malaria (such as bed-nets, insecticides, and avoiding mosquito-ridden area) was made possible. Furthermore, in the mid 1900s, the synthesis of chloroquine has also greatly boosted mankind's fight against malaria (Petersen *et al*, 2011). Armed with an arsenal of effective antimalarial drugs and insecticides (DDT), along with a "global" push for malarial eradication, mankind was able to swing the war on malaria in its favor. As a result, malaria was considered eliminated in many developed countries (Nájera *et al*, 2011). However, because of various obstacles such as antimalarial

drug resistance, insecticide resistance, absence of community participation, regional wars, and lack of funding, the war on malaria is still far from over.

Today, malaria is still one of the deadliest infectious diseases in the world. It is mostly prevalent in sub-Saharan Africa, South East Asia and South America, where the resources and infrastructure to combat malaria are still lacking. Though we are now more knowledgeable about the malaria parasite and are better equipped against it, the parasite has also taken steps to build stronger defenses with its heightened drug resistances (O'Brien *et al*, 2011; Mita *et al*, 2009). In response to the surge of drug resistances by the malaria parasite, artemisinin combination therapies are now being implemented (Porter-Kelley *et al*, 2010). Though combinational therapies are more effectual, these treatments are significantly costlier than the conventional monotherapies, which is a daunting problem for those living in poverty-stricken areas where malaria is prevalent. Thus, the search for new cost-effective antimalarials and novel drug targets must be a priority.

About a decade ago, the genome of the deadliest human malarial species, *P. falciparum*, was sequenced (Gardner *et al*, 2002). Similar to the discovery of the malarial causative agent and vector, elucidating the *Plasmodium* genome sequence is a great milestone in understanding malaria and has led to many significant scientific discoveries and ventures. The *Plasmodium* genome has provided us with a comprehensive list of the parasite's putative proteins and also gives us good idea of

what types of cellular functions are important to the parasite biology based on the frequency of protein types. A previous study done by our lab (Ponts *et al*, 2008) revealed that there were over 100 ubiquitylating enzymes found in the *P. falciparum*, which led us to believe that ubiquitylation played a significant role. Furthermore, several studies (Lindenthal *et al*, 2005; Dick *et al*, 1996; Prudhomme *et al*, 2008; Czesny *et al*, 2009; Kreidenweiss *et al*) showed that a variety of proteasome inhibitors were capable of drastically inhibiting malarial parasites, thus strengthening our suspicions that ubiquitylation, at least within the context of protein degradation and probably also in other aspects, is an essential parasite process.

Our suspicions have led us to further investigate ubiquitylation in *Plasmodium* by identifying the *Plasmodium* ubiquitome with a multifaceted approach, employing both computational and biological tools (Chapter 1). In our search, we have found that ubiquitin conjugates are detected at every morphological stage of the parasite erythrocytic cycle, with more than half of the parasite's proteome representing possible targets for ubiquitylation. These protein targets were found to come from a wide range of cellular processes such as RNA metabolism, invasion, translation, and of course, protein degradation. Surprisingly, in spite of the numerous studies that have tested proteasome inhibitors against malarial parasites, there has been a relative dearth of research on the *Plasmodium* degradation pathways. That is why we have also undertaken efforts to elucidate the

Plasmodium ERAD system (Chapter 2) to see how it would compare to the more-studied human and yeast ERAD pathways. Though our initial findings of the *Plasmodium* ERAD pathway have shown highly similar attributes to their human counterparts, there is still much to explore. There may be yet a unique parasite attribute within its ERAD system that may be exploited for drug targeting. And at the very least, the more divergent E3 ligases and DUBs have presented themselves as good drug candidates.

During our investigation of *Plasmodium* ubiquitylation, we, along with other groups (Spork *et al*, 2009; Sommer *et al*, 2007), have identified a duplicated ERAD-like system that is likely localized to the apicoplast. We believe that this apicoplast ERAD-like system may serve to import preproteins across the outermembranes of the apicoplast. In chapter 3, we performed the first biological analysis of the putative ubiquitylating components of the duplicated ERAD-like pathway and validated that they localize to the apicoplast. However, there is still much work to be done in the future. Besides also investigating the other putative components of the ERAD-like pathway, functional experiments need to be performed to show that this parasite-specific pathway indeed operates to translocate proteins into the apicoplast. Because of the lack of genomic and functional tools in *Plasmodium*, it may be prudent to begin the functional analysis of these components in other plastid-containing parasites such as *T. gondii*, which allow for conditional mutation experiments and easier transfections.

There is another question of how the putative ubiquitin-like protein that is found to localize to the apicoplast (Spork *et al*, 2009) is able to modify proteins without a di-glycine motif. Perhaps it instead serves as a chaperone protein, assisting with the ubiquitylation of target proteins, as was seen with an ubiquitin-domain-containing HERP protein found in within the mammalian ERAD system. Or another possibility is that this ubiquitin-like protein serves to keep apicoplast-targeted proteins from folding during apicoplast import since proteins are reported to remain unfolded during membrane translocation. In either case, if the atypically long ubiquitin-like protein is serving as a chaperone protein, the more conserved ubiquitin (PFL0585w) must somehow find its way into the apicoplast to label apicoplast-targeted preproteins.

Since ubiquitin gene PFL0585w lacks the canonical ER targeting signal peptide that is recognized for co-translational import to the ER, the conserved *Plasmodium* ubiquitin is likely post-translationally expressed in the cytosol. Thus, if this ubiquitin protein is modifying apicoplast proteins for import, it would have to employ unconventional methods to find its way into the apicoplast. Though lacking any previous evidence, one possibility of apicoplast targeting by cytosolic ubiquitin is that transient membrane fusions could take place between the periplastid membrane (third outermost membrane) and the outermost membrane that would cause a temporary bridge between the cytosol and the periplastid compartment. This would allow the conserved cytosolic ubiquitin to flow into the periplastid

compartment where it has been reported in Cryptophytes that the active domains of plastid ubiquitylating proteins reside. Another possibility is that modification by ubiquitin is not a requirement for apicoplast import and that these ERAD-like ubiquitylating proteins simply retained their vestigial activity, as seen during the *in vitro* ubiquitylation experiments. Though there is still much to uncover about apicoplast import, there is also much promise. The ERAD-like system presents itself as an excellent antimalarial target candidate because of its essentiality and putative roles in parasite-specific functions.

While it is important to continually find new antimalarials and novel drug targets, a simple uniform strategy will not eradicate malaria. One day finding an effective antimalarial would clearly be a substantial step forward; however, if the infrastructure, funding and general support system were not available, the most potent antimalarial would be useless unless it can be delivered and administered to the people who need it. The war against malaria can only be won with a multi-level strategy that identifies and circumvents the biological, physical, social and cultural barriers that have consistently blocked the success of malaria control in the past. Only with a concerted effort with both vertical and horizontal approaches between the various disciplines (sociology, ecology, biology, chemistry, political science, etc.), organizations that are in the field, government agencies and even local communities can a successful malaria eradication campaign prevail.

REFERENCES

Cox FE (2010) History of the discovery of the malaria parasites and their vectors. *Parasit Vectors* **3:** 5

Czesny B, Goshu S, Cook JL & Williamson KC (2009) The proteasome inhibitor epoxomicin has potent Plasmodium falciparum gametocytocidal activity. *Antimicrob. Agents Chemother* **53:** 4080–4085

Dick LR, Cruikshank AA, Grenier L, Melandri FD, Nunes SL & Stein RL (1996) Mechanistic studies on the inactivation of the proteasome by lactacystin: a central role for clasto-lactacystin beta-lactone. *J. Biol. Chem.* **271:** 7273–7276

Gardner MJ, Hall N, Fung E, White O, Berriman M, Hyman RW, Carlton JM, Pain A, Nelson KE, Bowman S, Paulsen IT, James K, Eisen JA, Rutherford K, Salzberg SL, Craig A, Kyes S, Chan M-S, Nene V, Shallom SJ, *et al* (2002) Genome sequence of the human malaria parasite Plasmodium falciparum. *Nature* **419:** 498–511

Kreidenweiss A, Kremsner PG & Mordmüller B Comprehensive study of proteasome inhibitors against Plasmodium falciparum laboratory strains and field isolates from Gabon. *Malar J* **7:** 187–187

Lindenthal C, Weich N, Chia YS, Heussler V & Klinkert MQ (2005) The proteasome inhibitor MLN-273 blocks exoerythrocytic and erythrocytic development of Plasmodium parasites. *Parasitology* **131:** 37–44

Mita T, Tanabe K & Kita K (2009) Spread and evolution of Plasmodium falciparum drug resistance. *Parasitology International* **58:** 201–209

Nájera JA, González-Silva M & Alonso PL (2011) Some Lessons for the Future from the Global Malaria Eradication Programme (1955–1969). *PLoS Med* **8:**

Neghina R, Neghina AM, Marincu I & Iacobiciu I (2010) Malaria, a Journey in Time: In Search of the Lost Myths and Forgotten Stories. *The American Journal of the Medical Sciences* **340:** 492–498

O'Brien C, Henrich PP, Passi N & Fidock DA (2011) Recent clinical and molecular insights into emerging artemisinin resistance in Plasmodium falciparum. *Curr. Opin. Infect. Dis.* **24:** 570–577

Petersen I, Eastman R & Lanzer M (2011) Drug-resistant malaria: Molecular mechanisms and implications for public health. *FEBS Letters* **585:** 1551–1562

Ponts N, Yang J, Chung D-WD, Prudhomme J, Girke T, Horrocks P & Le Roch KG (2008) Deciphering the Ubiquitin-Mediated Pathway in Apicomplexan Parasites: A Potential Strategy to Interfere with Parasite Virulence. *PLoS ONE* **3:** e2386

Porter-Kelley JM, Cofie J, Jean S, Brooks ME, Lassiter M & Mayer DG (2010) Acquired resistance of malarial parasites against artemisinin-based drugs: social and economic impacts. *Infect Drug Resist* **3:** 87–94

Prudhomme J, McDaniel E, Ponts N, Bertani S, Fenical W, Jensen P & Le Roch K (2008) Marine Actinomycetes: A New Source of Compounds against the Human Malaria Parasite. *PLoS ONE* **3:**

Schlagenhauf P (2004) Malaria: from prehistory to present. *Infectious Disease Clinics of North America* **18:** 189–205

Sommer MS, Gould SB, Lehmann P, Gruber A, Przyborski JM & Maier U-G (2007) Der1-mediated Preprotein Import into the Periplastid Compartment of Chromalveolates? *Mol Biol Evol* **24:** 918–928

Spork S, Hiss JA, Mandel K, Sommer M, Kooij TWA, Chu T, Schneider G, Maier UG & Przyborski JM (2009) An Unusual ERAD-Like Complex Is Targeted to the Apicoplast of Plasmodium falciparum. *Eukaryotic Cell* **8:** 1134–1145

APPENDIX

APPENDIX A1:

Post-translational modifications in *Plasmodium*: more than you think.
Mol. Biochem. Parasitol. 168: 123–134 (2009)

Duk-Won Doug Chung, Nadia Ponts, Serena Cervantes and Karine G. Le Roch

Department of Cell Biology and Neuroscience, University of California, Riverside, 900 University Avenue, Riverside, CA 92521, USA.

Contents:

Introduction
1. Phosphorylation/dephosphorylation
    1.1. Kinases
        1.1.a. *Cell-cycle regulation*
        1.1.b. *Cell proliferation and differentiation*
        1.1.c. *Sexual differentiation*
        1.1.d. *Parasite egress and invasion*
        1.1.e. *Host-parasite interaction*
    1.2. Phosphatases
2. Acetylation
3. Methylation
4. Lipidation
    4.1. Glycosylation
    4.2. Prenylation
    4.3. Palmitoylation
5. Ubiquitylation
    5.1. Ubiquitin and ubiquitin-like proteins
    5.2. Ubiquitin enzymes
6. Protein cleavage and processing
    6.1. Hemoglobin degradation
    6.2. Parasite egress
    6.3. Parasite invasion
Concluding Remarks

Abstract

Recent evidences indicate that transcription in *Plasmodium* may be hard-wired and rigid, deviating from the classical model of transcriptional gene regulation. Thus, it is important that other regulatory pathways be investigated as a comprehensive effort to curb the deadly malarial parasite. Research in post-translational modifications in *Plasmodium* is an emerging field that may provide new venues for drug discovery and potential new insights into how parasitic protozoans regulate their life-cycle. Here, we discuss the recent findings of post-translational modifications in *Plasmodium*.

Keywords: Plasmodium; Post-translational modification; kinases; proteases; ubiquitin system; Life-cycle regulation

Introduction

Malaria is one of the deadliest infectious diseases of the world. Each year, malaria infects over 300 million people world-wide and causes an estimated one to two million deaths [1-3]. With the increase of drug resistance to most of the widely-used anti-malarial drugs, it is imperative to better understand key regulatory elements driving the *Plasmodium* life cycle, eventually leading to the discovery of new drug targets.

The sequencing of the *Plasmodium* genomes, along with subsequent comparative bioinformatics approaches, transcriptome and proteome analyses, have created a vast amount of information regarding protein prediction and their hypothetical functions in both the human host and mosquito vector [4-7]. While microarray analyses have demonstrated a remarkable change in steady-state mRNA levels during parasite development, only a relatively few regulatory motifs and transcription regulators have been uncovered so far [8]. In addition, unlike other organisms, there seems to be few transcriptional changes in *Plasmodium* following exposure to external stimuli [3, 9, 10]. These findings imply that parasite transcription could be hard-wired [10]. This rigidity in transcription suggests that post-transcriptional and post-translational regulations are likely to play major roles in regulating the parasite life cycle.

By definition, post-translational modifications (PTMs) must add or subtract a specific mass difference and not be particular to any one protein [11]. In addition to changing protein mass, PTMs can also alter the protein charge and conformation. Such changes modify the protein's enzyme activity, binding affinity and hydrophobicity [12]. The spectrum of the PTMs within a cell is immense and varies with respect to specificity and abundance. For example, some types of PTMs such as phosphorylation and ubiquitylation are universally employed to regulate a broad host of functions, have relatively high abundance, and have a wide range of target substrates. On the other hand, some PTMs such as acetylation are highly specific in their roles, have a relatively low abundance, and may target only a few proteins or even one target substrate at a time. Furthermore, PTMs can either modify a protein at one specific amino acid residue, or be associated with numerous different residues. The potential diversity of protein modifications seem limitless due to the fact that an individual protein can undergo a single modification or multiple types of modifications at several sites, possibly producing multiple protein isoforms, each with a unique biological activity. One can easily see why PTMs are responsible for a major increase in complexity from genome to proteome. For example, the human genome contains approximately 30,000 open reading frames, but is predicted to give rise to roughly 1.8 million different protein variants [13].

Furthermore PTMs are reversible and thus provide the flexibility and adaptability that are essential for mediating rapid cellular responses to the cell's

constantly changing conditions. Thus, post-translational modifications are vitally essential for the survival of all kinds of cells, with *Plasmodium* not being an exception. Due to their diversity, essentiality, and wide-spread roles, post-translational modifications could present new major targets for effective and specific therapeutic intervention against the malarial parasite. Here we present an overview of the major post-translational modifications found so far in *Plasmodium* and discuss the recent developments of this rapidly expanding field. A summary of the major proteins involved in regulating PTM in *Plasmodium* is also presented in Figure A1.1 and Table A1.1.

1. Phosphorylation/dephosphorylation

Phosphorylation involves the reversible esterification of a phosphate group to an amino acid residue by protein kinases (PKs) that transfer a phosphoryl group from an ATP to hydroxyamino acid residues, mostly serine, threonine and tyrosine [14]. As the most highly studied covalent modification of proteins in eukaryotic cells, phosphorylation events can be linked to practically most functions within a cell: cell growth, cell differentiation, receptor activations, metabolic pathways, enzyme activities, cytoskeletal organization, chromatin remodeling, protein activations/inhibitions and protein-protein interactions. Approximately 1.5% to 2% of genes in an eukaryotic genome are PK family genes, which reflects the importance of phosphorylation events. While protein kinases catalyze the phosphorylation of protein residues, phosphatases hydrolyze the phosphoester bond of the modified

amino acid, restoring the hydroxyamino acid to its unphosphorylated state [15]. In eukaryotes, the number of protein phosphatases is relatively small when compared to that of protein kinases. However, additional phosphatase regulatory proteins mediate specific regulation of these enzymes and it is speculated that the number of phosphatase complexes involved in regulatory pathways may exceed the protein kinases repertoires [16].

1.1. Kinases

Several kinase inhibitors have been shown to inhibit *Plasmodium* development at different stages of the life cycle [9, 17-20] validating the importance of phosphorylation in maintaining the parasite. Depending on the stringency applied to the computational tools, computational analysis studies retrieved 86 to 99 PK-related enzymes from the *Plasmodium falciparum* (*Pf*) genome [21, 22]. Phylogenetic studies have demonstrated that most of the eukaryotic families of PKs are present in the *Plasmodium* genome with the exception of two groups, ste-20 (STE) and the tyrosine protein kinases (TyrK) families. The STE family includes PKs involved in Mitogen-activated protein kinase (MAPK) cascades and plays a central role in transduction signals (see reference Ward et al. 2004 for further details). This group is evolutionarily conserved and their absence in the parasite genome may indicate that the mode of activation of the MAPK pathway in the parasite differs

from other eukaryotes. The TyrKs are known to function in hormone-response receptor-linked pathways essential for intercellular communication in multicellular organisms and are therefore not expected in the malaria parasite.

Several "orphan" PKs have also been identified in the *Plasmodium* kinome. They display only limited similarities with yeast or mammalian kinases. Examples of these atypical kinases include the NIMA-related kinase Nek1 (PFL1370W) [17], twenty FIKK PK-related proteins, found only in apicomplexan parasites [22, 23], and a family of calcium dependant kinases (CDPKs) that are usually found in plants and alveolate, but not in metazoans [24]. In the following section we briefly describe the key biological functions identified so far for a few selected *Plasmodium* kinases.

1.1.a. Cell-cycle regulation

In eukaryotes, kinases are known to play a major role in cell-cycle progression. This is particularly the case for cyclin-dependant kinases (CDKs), several of which have been identified in the *Pf* genome [22]. Two *Pf* CDKs, PfPK5 (MAL13P1.279) and Pfmrk (PF10_0141), have been shown to be positively regulated by the binding of *Plasmodium* and mammalian cyclins as well as negatively regulated by CDK inhibitors [25-27]. These results demonstrate that regulatory activities of CDKs are well-conserved across species. While the targets of these *Plasmodium* kinases have not yet been identified, it is speculated, that such proteins might play a major role in the regulation of the parasite schizogony and its

nuclear division cycles [28]. *In vitro* biochemical characterizations and crystal structure analyses have uncovered possible unique regulatory mechanisms of *Plasmodium* CDKs [27, 29] Such particularities indicate an atypical mode of DNA replication/mitosis in the parasite. [30]. Furthermore, an atypical *Pf* orphan protein kinase, PfPK7 (PFB0605w), that exhibits maximal homology to a MEKKK and a fungal PKA, has been shown to regulate parasite proliferation and development [31]. Discrepancies observed between parasite and host kinases may be exploited in the search of parasite-specific kinase inhibitors as potential anti-malarial drugs, but would need to be further validated *in vivo*.

1.1.b. Cell proliferation and differentiation

MAPKs are known to be key players in transduction signals, cell development and differentiation in response to a variety of stimuli [32]. Though no classical MAPK kinase has been identified in the *Plasmodium* genome, two atypical MAPK homologues have been characterized, Pfmap-1 (PF14_0294) and Pfmap-2 (PF11_0147). While the exact role of Pfmap-1 is still speculated, Pfmap-2 appears to be essential for the completion of the parasite asexual erythrocytic cycle [33]. However in *Plasmodium berghei* (*Pb*), the Pfmap-2 orthologue seems to be essential to parasite exflagellation in the mosquitos midgut [34]. Discrepancies observed between these two *Plasmodium* species will need to be further clarified. Nonetheless, it is possible to envision a complementation effect induced by species-specific *Plasmodium* kinases.

1.1.c. Sexual differentiation

Several identified *Plasmodium* kinases have been implicated in sexual differentiation. In *Pb*, a calcium-dependant kinase (CDPK4, PF07_0072 in *Pf*) has been shown to regulate gamete formation with the initiation of DNA replication, ookinate gliding mobilities, and mosquitos midgut invasion [35-37]. cGMP-dependant protein kinase (PFPKG, PF14_0346 in *Pf*) seems to be essential for mediating initiation of gametocytogenesis [18]. Finally, the *Pb* Pbnek-4 (MAL7P1.100 in *Pf*) has been shown to be critical for ookinate maturation [38]. The detection of an increased number of protein kinases involved in sexual differentiation demonstrates that PKs have a crucial role in regulating gametocytogenesis. These preliminary results provide a framework for identifying substrates of these *Plasmodium* enzymes to further comprehend the signal transduction involved in sexual differentiation.

1.1.d. Parasite egress and invasion

In contrast to other pathogens, apicomplexans such as *Plasmodium* can quickly and efficiently exit and enter the cell using their own machinery. Several steps involving a large collection of proteins have been shown to be required in erythrocytic egress and invasion [39]. From this collection, a few protein kinases have been shown to have a role in controlling these parasitic processes. The calmodulin-like PK (CDPK1, PFB0815w in *Pf*) has been shown to be required for the secretion of the parasite microneme contents and the formation of tight moving

junctions [40]. Two components of the acto-myosin motor complex, the myosin A tail domain interacting protein (MTIP) and the glideosome-associated protein 45 (GAP45) have been identified to be CDPK1 substrates. In addition, PfPKB (PFL2250c) [41], an important member of the phosphatidylinositol 3-kinase-dependent signaling pathway, has been implicated to regulate parasite gliding mechanisms [42, 43]. However, validating the role of PfCDPK1 and PfPKB *in vivo* remains to be established.

1.1.e Host-parasite interaction

Recent works have begun to reveal the importance of *Plasmodium* PKs in modulating the erythrocyte membranes of infected hosts. PfPKA (PFI1685w) has been shown to alter both the activity of an erythrocyte anion channel and the permeability of the host plasma membrane [44, 45]. Interestingly, the members of the *Plasmodium* FIKKs orphan kinases possess a *Plasmodium* export element (PEXEL) motif [46] that target parasite proteins to the host membrane [47, 48]. Immuno-microscopy reveals that many FIKK proteins can be found in the erythrocyte cytoplasm and colocalized with Maurer's clefts protein. This finding suggests the importance of these FIKKs in the remodeling of the infected erythrocyte membrane [46].

While it is increasingly apparent that a significant number of protein kinases regulate major parasite cell cycle and differentiation events, our understanding of their targeted substrates is extremely limited. In order to comprehend the function and essentiality of each kinase, a systematic biological approach involving reverse genetics, chemical genetics and phosphoproteome studies will need to be developed.

1.2. Phosphatases

A total of twenty-seven malaria protein phosphatases (PP) have been identified in the *Pf* genome (review in [49]). The identified *Plasmodium* PP clustered with the four major established eukaryotic PP families: the Metholophosphatases (PPP), the Serine/Threonine Phosphatases (PPM), the Protein Tyrosine Phosphatases (PTP) and the NLI Interacting Factor-like phosphatases (NIF). Protein sequence information as well as biochemical characterizations studies of several of these PPs in *Plasmodium* have validated their phosphatase activities and revealed significant differences with other eukaryotic organisms [49-52].

The use of phosphatase inhibitors have shown that *Plasmodium* PPs are essential and may be involved in invasion [53] and cell growth [54]. Protein-protein interaction and localization studies have highlighted a role of *Plasmodium* PPs in transcriptional elongation [55], nuclear protein activity [56] and the release of

infectious merozoites [57]. Interestingly, highly-conserved eukaryotic phosphatases seem to be either missing or too divergent to be detected in the *Plasmodium* genome. Among the missing phosphatases are the CDC25 homologue (known to play a major role in cell cycle control) [57, 58], the cdc14 phosphatase (known to regulate mitotic events) [59] and tyrosine phosphatases [60], though there is a report of a possible PRL tyrosine phosphatase [61]. These findings further validate the phylogenetic distance observed between *Plasmodium spp* and its vertebrate host. The identification of PP regulatory subunits, as well as their specific substrates, will need to be further investigated. However, it has become evident that these PPs are essential to the parasite and should be considered as potential targets for new anti-malarial strategies.

2. Acetylation

Acetylation is an addition of an acetyl functional group onto a protein substrate. Acetyltransferases transfer an acetyl group from acetyl coenzyme A onto conserved N-terminal lysine residues. The result of acetylation usually modifies DNA binding properties, protein stability, and protein-protein interactions.

Acetylation of actin at the N-terminal is a highly-conserved PTM that has also been reported in *Pf* [62]. In addition to actin, *Plasmodium* histones are found to be acetylated at their N-terminal lysine residues [63]. Histone PTMs are a vital part of the 'histone code' hypothesis that proposes specific combinations of PTMs that

modify chromatin structure and act as platforms for the binding of transcriptional regulators of gene expression. Consistent with other eukaryotes, *Plasmodium* histone acetylation is believed to play a major role in transcriptional regulation. In a genome-wide investigation of histone modifications and their relationship with transcriptional activation/silencing, Cui and colleagues [64] employed a combination of immunoprecipitation and DNA microarray detection (ChIP-chip) to show that acetylation of histone H3 (H3K9ac) was associated with active genes across the *Pf* genome. PfGCN5 was the first characterized histone acetyltransferase (HAT) subunit identified in *Plasmodium*. Recombinant *Pf*GCN5 displayed histone H3 acetylase activity *in vitro* and exists as a catalytic subunit of *P. falciparum* HAT complex. *Pf*ADA2, the yeast transcriptional coactivator homolog, has been characterized as another component of the trimeric catalytic core [65]. *In vitro* pull-down and yeast 2-hybrid experiments suggest that PfADA2 and PfGCN5 are present in complex(es) and may have conserved chromatin remodeling functions. Natural compounds that inhibit HAT activity are being investigated as potential antimalarials [66]. However, because HATs are conserved across species, toxicity studies will have to be further investigated.

Pf HDAC1 and *Pf*Sir2, two of the five putative *Plasmodium* Histone Deacetylases (HDACs), have been partially characterized in *Plasmodium* [67]. *Pf*Sir2, a yeast homologue of the silent information regulator 2 (SIR2) was found to bind to parasite telomeres and cause histone de-acetylation and silencing of the *var*

multi-gene family [68]. Inhibitors of HDAC activity have also been investigated for potential antimalarial drugs [69, 70].

3. Methylation

Methylation, is catalyzed by methyltransferases that attach a methyl group onto its substrate. Such modification increases lipophilicity and reduces the substrate solubility in water. Methylation is associated with the regulation of gene expression and protein activity. In *Plasmodium*, the most commonly methylated subtrates are the parasite histones. Genome-wide analysis of histone modification showed that H3K9me3 is associated with gene silencing [64]. Recently, chromatin immunoprecipitation of transcriptionally active *var* gene loci assessed the enrichment of histone H3 di- and trimethylation marks (H3K4me2 and H3K4me3) in the 5' flanking region [71].

At least nine *Plasmodium* histone methyltransferases (HMT) (containing a SET-domain, characteristic of histone lysine methyltransferase), and two *Plasmodium* demethylases (containing a JumonjiC-domain, characteristic of histone lysine demethylases (HDMs)) have been identified *in silico*. Phylogenetic analysis divided putative histone lysine methyltransferases (HKMTs) into five subfamilies with different putative substrate specificities. HKMTs and HDMs not only have substrate specificity, but also specificity for different methyl states (mono-, di-, tri-) [72].

4. Lipidation

Lipidation is the covalent binding of a lipid group to a peptide chain and can change the activity and/or cellular localization of the modified protein. Often times, the attachment of a hydrophobic chain can help to anchor soluble proteins, or proteins with weak membrane-affinity, to the inner face of a membrane. Some examples of lipidation include N-myristoylation, prenylation, GPI-anchor addition, and palmitoylation. Discussed below are a couple of post-translational modifications via lipidation that have been studied so far in *Plasmodium*.

4.1. GPI-anchoring

The glycosylphosphatidylinositol (GPI) anchor is a glycolipid moiety that is added to the C-terminal of proteins after translation. Such modified proteins are attached to the outer leaflet of the cell membrane. GPI-anchored proteins are ubiquitous among eukaryotic organisms and represent a very functionally diverse group that is being extensively studied in a wide-array of organisms.

Anchoring with GPI is the major type of glycosylation that are found in *Plasmodium* [73]. It is commonly postulated that *Plasmodium* proteins are glycosylated by the erythrocytic machinery (combinatorial metabolism) during the exchanges that occur between the parasitophorous vacuole and the erythrocyte cytoplasm [74]. In the human host, the parasite's GPI anchors have been linked to the pathobiology of the disease [75] and are associated with increased levels of

TNF-α (CD36-dependant) leading to a systemic inflammation reaction [76, 77]. It has been found that resistance to malaria in endemic areas is associated with the production of circulant antibodies directed against the parasite's GPI anchors [75] and are common targets for vaccine research. Also, most *P. falciparum* merozoite surface proteins are GPI-anchored and are involved in erythrocyte recognition and attachment to erythrocytes before invasion (reviewed by [78]. GPIs may also be involved in post-invasion processes. The *P. falciparum* GPI-anchored rhoptry protein Pf34 (PFD0955w) is suspected to play a role in the formation of the parasitophorous vacuole [79] [80].

In the mosquito host (*Anopheles spp.*), *P. falciparum* GPIs may play a direct role in mosquito immune response [81] [82]. The immune response in *Anopheles gambiae* is elicited by the parasite GPIs resulting in a significant reduction in fecundity (diminished egg production) [83]. The role of such effect on the host-parasite interactions remains open to discussion. The authors propose that the parasite may use its GPIs to trigger the mosquito immune response and "mis-direct" the specificity of the response against other pathogen-associated molecular patterns rather than ookinetes or microneme proteins essential for parasitic processes. Furthermore, the mosquito stage GPI anchored surface proteins P25 (PF10_0303) and P28 (PF10_0302) are shed during ookinete maturation and play an important role in midgut invasion [84-86]. Finally, the GPI anchor of the circumsporozoite protein (PFC0210c) plays a crucial role in sporogenesis [87]. Due to their general

importance, GPIs and their biosynthesis pathway are being investigated as potential targets to antimalarial strategies.

4.2. Prenylation

Prenylation, also known as isoprenylation, is the posttranslational modification of proteins by covalent attachment near the carboxyl terminal of isoprenyl lipids, a 15 carbon farnesyl or a 20 carbon geranylgeranyl group [88]. The attachment of isoprenyl lipids creates a hydrophobic tail that promotes membrane association, and plays an important role in cell signal transduction, vesicle trafficking, and cell cycle progression [89]. Unlike animals, fungi and archeabacteria, which use the classical mevalonate pathway for isoprenoid synthesis, the *Plasmodium* synthesizes its isoprenoid precursors via the 1-deoxy-D-xylulose 5-phosphate (DOXP) pathway within its apicoplast, a plastid-like organelle [90].

Prenylation is mediated by three enzymes: protein farnesyltransferase (PFT), protein geranylgeranyltransferase type I and type II. PFT in *Plasmodium* has been characterized by partial purification of protein farnesyltransferase, PfPFT, and radiolabeling of prenylated proteins [88]. Prenylation precursors are incorporated into asexual synchronized parasites in a stage-specific manner with the highest amount occurring from trophozoite to schizont, and schizont to ring transitions. Dolichylation, the addition of 11 isoprene units, have also been reported in *P. falciparum* during trophozoite and schizont stages.

Pharmaceutical companies have invested into the development of PFT inhibitors for the treatment of cancer with a few drugs in clinical trials. Interestingly, PFT inhibitors have been found to be potent antimalarials [91] and validate farnesylation as essential in the malaria parasite. Drug development for antimalarials are taking a 'piggy-back' approach, since the concentrations needed to inhibit parasites are significantly lower than that of mammalian cells [89].

4.3.Palmitoylation

Palmitoylation is the covalent attachment of fatty acids to cysteine residues of proteins, often giving soluble proteins (or proteins with weak-membrane avidity) a hydrophobic membrane anchor [92]. In *P. falciparum*, a 45kDa gliding-associated protein (GAP45) was found to be both palmitoylated and N-myristoylated [93]. It is hypothesized that GAP45 may play a role in binding acto-myosin motors to the outer face of the inner membrane complex, which is implicated as the underlying force driving both gliding motility and host cell invasion in *Plasmodium* merozoites.

5. Ubiquitination

Ubiquitin is a highly conserved 76 amino acid peptide found in eukaryotic organisms. Beyond the more familiar association with protein degradation, the modification of proteins by ubiquitin conjugation is known to serve as a regulatory signal for cell proliferation, cell-stress response, transcription, cell death, DNA repair, intracellular trafficking, endocytosis and signal transduction [94]. The

reversible conjugation of ubiquitin to the lysine residues in target proteins is controlled by a series of enzymes: ubiquitin activating enzymes (E1), ubiquitin-conjugating enzymes (E2), and ubiquitin ligases (E3) [95]. Ubiquitin is activated by E1 and transferred to E2. Then, ubiquitin is either transferred to a monomeric E3 that catalyzes ubiquitination of the target substrate or ubiquitinated E2 forms a complex with the E3 to catalyze ubiquitination of the substrate. Deubiquitinating enzymes (DUBs) serve to reverse ubiquitin-conjugation by removing ubiquitin from substrate proteins and also help to replenish the mono-ubiquitin pool.

Diverse forms of ubiquitin modifications have been reported, each potentially mediating a specific function. K48-linked poly-ubiquitin chains to substrates often serve as signals for targeted protein degradation via the ubiquitin/proteasome system [96]. In *Plasmodium,* several studies have validated the essentiality of the proteasome and protein turnover in regulating the cell-cycle progression. A collection of proteasome inhibitors has shown promising results. [97-102].

Apart from protein degradation, poly-ubiquitination *via* K63-linkages seem to play important roles in DNA damage tolerances, endocytosis, ribosomal protein synthesis, and inflammatory response [103]. In addition to poly-ubiquitination, proteins can either be mono-ubiquitinated, which is the attachment of a single ubiquitin to a protein, or multi-ubiquitinated, which is the attachment of individual

ubiquitin to a substrate at multiple sites. Both mono-ubiquitination and multi-ubiquitination are reported to have non-proteolytic roles such as endocytosis [104] and DNA repair [105]. In the following sections, we briefly describe what has been reported so far concerning ubiquitination in *Plasmodium*.

5.1. Ubiquitin and ubiquitin-like proteins

The *Plasmodium falciparum* ubiquitin gene, PfpUB (PFL0585w) is present as a single-copy on chromosome 12 with five tandem repeats of the ubiquitin open reading frame [106]. Translation of PfpUB comprises of five ubiquitin monomers that have sequence identities of no less than 94% to that of other eukaryotic species. RT-PCR and northern analysis revealed that steady-state transcript levels of PfpUB are expressed at all stages of the intrerythrocytic cycles, with significant increases at the late trophozoite and schizont stages. Under heat shock, polypeptide levels of PfpUB and ubiquitinated adducts showed dramatic increases without a significant increase in steady-state transcript levels. This observation suggests that the heat shock response appears to be maintained at the level of translation. [106].

Two additional ubiquitin moieties, Ub_{S27a} and Ub_{L40}, fused to the ribosomal proteins L40 (PF13_0346) and S27a (PF14_0027) have also been identified in the *P. falciparum genome* [107, 108]. Expression data indicates that these ubiquitin genes are expressed throughout the *P. falciparum* life cycle [5, 7]. Interestingly , though the

Pf Ub_{L40} shows high protein sequence fidelity with other eukaryotic species, *Pf* Ub_{S27a} harbors significant divergences within the ubiquitin domain [108].

In addition to ubiquitin, ubiquitin-like proteins (UBLps) have been identified as modifiers of cellular-processes. Though a number of UBLps (ISG15, FAT10, UFM1, FUB1) that are typical in higher eukaryotes are yet to be found in *Plasmodium*, gene expression data suggests that SUMO, NEDD8, HUB1, URM2 and ATG8 are expressed at all life cycle stages [107, 109].

Recently, *Issar et al.* [110] was the first to investigate and characterize SUMO within *P. falciparum*. Amino acid sequence comparisons revealed that *P. falciparum* gene PFE0285c (PfSUMO) has significant sequence homology to that of known SUMO orthologues from other eukaryotic organisms such as yeast, human, mouse and *S. pombe*, while immunoblot analysis confirmed the presence of SUMO in *P. falciparum*. In addition, enzyme homology searches have identified *P. falciparum* orthologues of all the necessary members of the SUMO pathway. Using LC-MS/MS analysis, more than 20 putative SUMO substrates of *P. falciparum* proteins were identified. These protein substrates varied from histones to transcription factors to RNA helicases, which implicates a wide-ranging regulatory scope of SUMO. Immunofluorescence assays indicate that PfSUMO localizes to distinctive subcellular compartments within the *P. falciparum* and also in the host cell cytoplasm within the parasite-derived structures called Maurer's clefts.

5.2. Ubiquitin enzymes

Computational studies have identified over a hundred proteins predicted to be involved in the reversible conjugation of ubiquitin or ubiquitin-like proteins (UbLps) in *P. falciparum* alone [107, 109].

Ubiquitin activating enzymes (E1)

Eight putative E1 have been identified in the parasite genome [107]. While primary sequence identity has been observed in the core ubiquitin activating enzyme domain, sequences outside of this core diverge rapidly as the functional requirements for these E1 enzymes change to specifically interact with their respective E2 conjugating enzymes. Sequence analysis indicates the existence of *Plasmodium* E1 paralogs for UBA1 and UBA1-like proteins, and UBA2, UBA3, UBA4, ATG7 proteins, which mediate the activation of ubiquitin-like proteins SUMO, NEDD8, URM1, and ATG8 respectively.

Ubiquitin conjugating enzymes (E2)

Fourteen putative E2 paralogs were found in *P. falciparum* [107]. They exhibited extensive conservation with other eukaryotic E2 proteins. Gene expression data for nine of the fourteen *Plasmodium* E2s reveal a diverse pattern of steady state mRNA at different stages of the intraerythrocytic cycle, suggesting the existence of a temporal profile of delivering ubiquitin or UbLps to different E3s,

indicating a potential additional level of temporal control in ubiquitination during the parasite's life cycle.

Recently, a *Pf* homolog (PfUBC13) of the E2 ubiquitin-conjugating enzyme 13 (UBC13) was characterized and found to be a substrate of the *Pf* protein kinase PfPK9 [111]. Reverse-phase HPLC and in-vitro ubiquitination assay show that PfPK9 phosphorylates PfUBC13 at S106 and suppresses ubiquitin-conjugating activity. Though the physiological role of PfUBC13 is unknown, the highly-conserved UBC13 (coupled with an ubiquitin E2-variant protein) assembles K63-linked ubiquitin chains [112], which mediate non-proteolytic pathways [113]. UBC13's conjugating activity regulates various cellular processes such as DNA repair [114], tumor suppressor p53 activity [115], and mitotic progression [116]

Ubiquitin ligating enzymes (E3)

Though *Plasmodium* E1 and E2 enzymes exhibited strong conserved homology with other E1 and E2 enzymes from other eukaryotes, *Plasmodium* E3 ligases were found to be highly divergent and the most abundant. Within the 54 putative E3 ligases identified, all superfamilies (HECT, RING, U-box, and cullin) of E3 ligases are represented within *P. falciparum* with E3 RING finger proteins making up the majority of the *Plasmodium* E3 ligases [107].

Functional annotation analysis reveals that these *Plasmodium* ligases have a wide-array of potential roles including cell-cycle regulation, trafficking, DNA repair, chromatin structure, and mRNA transport. However several of these proteins seem to be specific to the apicomplexan phylum. While the functional analysis of these parasite specific putative E3s will need to be further validated *in vitro* and *in vivo*, two-selected E3 ligases seem to be essential to the parasite erythrocytic cycle and have already shown *in vitro* ubiquitination activity validating further the importance of this pathway in parasite development (Chung and Le Roch, unpublished data).

Deubiquitinating enzymes

While ubiquitin ligases catalyze the attachment of ubiquitin and UbLs, DUBs hydrolyze ubiquitin and UbLps [117]. Depending on the computational tools employed, 18 or 29 *Plasmodium* DUBs were identified [107, 109]. Five distinct gene families were identified: the ubiquitin C-terminal hydrolases (UCHs); the ubiquitin-specific peptidases (USPs/UBPs); the ovarian tumor (OTU) domain proteins; the Josephin or Machado-Joseph disease (MJD) proteins and the JAMM (Jab1/MPN domain-associated metalloisopeptidase) domain proteins. In addition, DUBs for UbLps including SUMO, ATG8, and NEDD8 have also been identified.

PfUCH54 was the first DUB characterized in *P. falciparum* [118]. Using electrophilic probes that detect enzymes capable of removing ubiquitin and other ubiquitin-like proteins, PfUCH54 was found to possess both deubiquitination and deNeddylation activity. Using known active site residues and crystal structure of

homologous DUBs, PfUCH54 was found to have high homology to UCHL3 [118], an enzyme that has been reported to also have dual deubiquitinating and deNeddylating activity [119]. Though the function of PfUCH54 is not known, mouse UCHL3 is required to maintain a stable apical membrane epithelial sodium channel, facilitating the dynamic recycling of sodium channels at the apical surface [120].

Though functional analysis studies will need to further validate the roles of ubiquitination pathways in *Plasmodium*, E3 ligases and DUBs may represent a good target for new therapeutic interventions due to their diversity and potential involvement in parasite-specific pathways.

6. Protein cleavage and processing

Proteases are major virulence factors in parasitic diseases as largely reviewed by [121] and [122]. Five main classes of proteases have been identified: (1) cysteine proteases, (2) serine proteases, (3) threonine proteases, (4) aspartic proteases, and (5) metalloproteases. A computational analysis identified 92 proteases in the genome of *P. falciparum*, 83 of them being transcribed during the erythrocytic cell cycle and 67 being translated during the parasite's life cycle [123]. Proteases are commonly involved in a wide-array of biological processes such as lysosomal proteolysis, precursor protein processing or trafficking.

In *Plasmodium*, trafficking is a vital biological process, especially with regards to targeting proteins to the apicoplast, the essential chloroplast-like organelle of the parasite. Trafficking pathways leading to the apicoplast have not yet been elucidated but are known to involve the presence of both a signal peptide and a transit peptide that are processed along the pathway [124]. Wu et al identified a serine protease Signal Peptidase 1 (SP1, PF13_0118) that could be responsible for the cleavage of signal peptides [123]. In addition, metalloprotease falcilysin (PF13_0322 in *P. falciparum*) [125] has been recently implicated in transit peptide cleavage. [126].

Another important parasite-specific pathway is the PEXEL-mediated pathway that targets parasite proteins to the surface of infected erythrocytes. This pathway is thought to promote parasite evasion from the host immune system [127]. Chang et al. demonstrated that exported proteins are processed at their N-terminal end [128]. This processing involves acetylation and cleavage of the PEXEL motif in the endoplasmic reticulum. The authors suggest that this N-terminal processing may be more generally utilized for many exported soluble proteins. However, the protease responsible for such cleavage remains to be identified.

In *Plasmodium*, various proteases play key roles in hemoglobin degradation, egress and invasion during the parasite erythrocytic cycle. The roles of such proteases were usually determined from inhibitor studies, which emphasize the potential to target such enzymes by anti-malarial drugs.

6.1. Hemoglobin degradation

The degradation of hemoglobin involves various proteases called hemoglobinases [129] and provides the essential amino acids that are taken up by the parasite [121, 130, 131].

The cysteine proteases falcipain-2 (PF11_0165) and falcipain-3 (PF11_0162) are known to play a role in the early steps of hemoglobin degradation within the parasite food vacuole (see [132] for a review). In addition to these cysteine proteases, degradation of hemoglobin seems to involve metallo (falcilysin) and aspartic proteases such as the plasmepsins [130]. Four plasmepsins are present and active in the food vacuole of *P. falciparum*, plasmepsin I (PF14_0076), II (PF14_0077), IV (PF14_0075) and a histo-aspartic protease (HAP, PF14_0078) [133-135]. It was postulated that plasmepsins I and II are matured and released in the food vacuole by protein cleavage [133, 136]. A recent study demonstrated that falcipain-2 and falcipain-3 are responsible for plasmepsins preprocessing and that auto-processing can occur when falcipain activities are inhibited, providing an alternative pathway to activate plasmepsins [137].

6.2. Parasite egress

Parasite proteases are also known to trigger the degradation of parasite and host membranes, leading to the egress of infectious parasites. The role of these proteases in cell egress has been recently reviewed [138]. Briefly, the proteases

implicated in parasite egress are falcipain II, plasmepsin II and putative papain-like SERA proteases. SERA proteins are a family of nine members, which are activated by a subtilisin-like serine protease SUB1 (PFE0370c). Among them are SERA-4 (PFB0345c), SERA-5 (PFB0340c, [139]) and SERA-6 (PFB0335c), which are essential proteins that are expressed in the parasitophorous vacuole of the late trophozoite and schizont stage (erythrocytic cycle) [140]. Additionally, SERA-8 (PFB0325c), another member of the SERA protein family, is essential for sporozoite release from oocytes (mosquito stage) [141].

The involvement of the subtilisin-family serine protease PfSUB1 and the cysteine protease dipeptidyl peptidase 3 (DPAP3, PFD0230c) as primary regulators of parasite egress have been recently evidenced using serine and cysteine protease inhibitors [142]. Just prior to egress, the essential serine protease PfSUB1 is discharged from the exonemes into the parasitophorous vacuole space [143]. Inhibition of both DPAP3 and PfSUB1 blocks the processing of the serine repeat antigen protein SERA-5, which correlates with the inhibition of membrane rupture. In *Plasmodium*, two other subtilisin serine proteases of unknown function can be found: PfSUB2 (PF11_0381) and PfSUB3 (PFE0355c) (reviewed in [144]). PfSUB2 is believed to play an important role in the erythrocytic cycle while PfSUB3 appears to be more ubiquitous as it is expressed during the asexual blood stage, in gametocytes and in sporozoites [6].

6.3. Parasite invasion

The critical role of proteases in merozoite invasion of erythrocytes has been largely studied and reviewed [39, 145]. Briefly, invasion involves contact, interaction and junction between the merozoite and the red blood cell surface *via* adhesins and various GPI-anchored proteins, such as PfAMA-1 and MSPs followed by the active entry of the parasite within the host. In order for invasion to be completed, these interactions are interrupted by the shedding of the protein coat covering the merozoite surface made of various MSPs and adhesins. The proteases involved in such removal are called sheddases. For example, the adhesins Duffy binding ligand erythrocyte-binding antigen (DBL-EBP) 175 (EBA-175, MAL7P1.176) is shed from the merozoite at around the point of invasion by the rhomboid protease PfROM4 (PFE0340c) [146]. Rhomboid proteases are ubiquitous intramembrane serine proteases (see [147-149] for an extensive description of the different types of rhomboid proteases). Also, sporozoite invasion of hepatocytes is reported to be mediated by the shedding of PfAMA-1 and the thrombospondin-related adhesive protein (TRAP) by a serine protease [150] .

Double cleavage of MSPs is required for invasion and probably involves serine proteases. See Harris, Yeoh et al. 2005 [151] for further details. Very recently, it was suggested that MSP-1, MSP-6 and MSP-7 undergo proteolytic maturation catalyzed by PfSUB1 before egress, maturation that is essential for further processing of MSPs involved in invasion [152].

A role of cysteine proteases (falcipain) in erythrocyte invasion has also been proposed but remains elusive. Inhibitors of falcipain-1 (PF14_0553) block invasion of host red blood cells [153] whereas parasites with a disrupted falcipain-1 gene were not affected in terms of asexual growth [154]. More recently, the presence of an endogenous cysteine protease inhibitor in *Pf*, falstatin, has been identified [155]. The presence of such an endogenous inhibitor may be another level of proteolytic control of certain parasite/host proteases in order to facilitate erythrocyte invasion.

Proteolytic processing is finally involved in the maturation of proteins secreted by the merozoite-specific organelles, rhoptries. Rhoptry-associated proteins (RAP) are believed to play a role in invasion and are activated upon proteolytic cleavage. For example, the rhoptry-associated protein RAP-1 is maturated *in vivo* by multiple processing steps [156-158].

Concluding Remarks

In addition to the PTMs presented above, there are several other PTMs in *Plasmodium* that are being investigated but, due to space-constraints, are unable to be fully expanded upon in this review. However, we would like to briefly mention a few. For example, a *Plasmodium* cathepsin-C-like protein is reported to be modified *via* O-sulfonation [159]. Also, polyglutamylation of tubulin was found in the microtubule organizing centers and post-mitotic microtubular structures of *Plasmodium* [160]. Lastly, several chaperone proteins, involved in the proper folding of proteins, have been described in *Plasmodium* and are reported to be involved in roles such as trafficking [161-164].

Today, transcriptional studies make up a significant amount of publications that are dedicated to understanding mechanisms regulating the *Plasmodium* developmental cycle. However, recent evidences suggest that *Plasmodium* may not follow the classical transcriptional model and may have rigid transcription machinery, which may indicate the need to reallocate investigative efforts into other fields in order to better understand the parasite's life cycle regulation. With the exception of de-/phosphorylation, post-translational modifications in *Plasmodium* have been understudied, and are largely overshadowed by the classical view that proteins are mainly regulated at the transcriptional level. With further investigation, it is possible that post-translational regulation may reveal to be a bigger factor in parasite development than previously thought.

For example, though it is apparent that proteases are necessary to destabilize host cell membranes during invasion and egress, the mechanisms of how these proteases are activated and tightly-controlled temporally are yet unclear. However, there are increasing reports that PTMs may play a major role in both malarial invasion and egress. As already described above, during egress, it is suggested that additional proteases, such as SUB proteases, may act as regulators of egress by processing, and thereby activating, distinct effector proteins or signal transduction pathways [138, 143]. In addition, it is reported that *Pf* protein kinases, such as CDPK, may also be involved in regulating invasion and egress [19].

Besides having a major regulatory role within the parasite, PTMs may also play key roles outside the parasite by manipulating the host's signaling pathways, thereby usurping normal cellular processes for survival and escape from immune responses. Currently, there are an increasing number of reports that show PTMs to play an integral role in host-pathogen interactions, or cross-talk, in bacteria, viruses and protozoa [165-167]. For example, it has been shown that host organisms are known to employ a wide-range of post-translational modifications to initiate their immune responses in order to avoid invasion by pathogens. For instance, host cells use ubiquitination in defense strategies as a way to degrade parasitic proteins and also to activate inflammatory and anti-apoptotic genes in a non-degradative fashion [168]. As a countermeasure, it has been shown that pathogens are able to avoid and

exploit these ubiquitination defense strategies by secreting proteins of their own that either deubiquitinate or inhibit the host ubiquitinating enzymes [168, 169]. It is also reported that bacterial pathogens are able to provoke histone modifications (*via* acetylation and phosphorylation) and chromatin remodeling in host cells, thereby manipulating the host's transcriptional programming and diminish the host innate immune response [166]. The utilization of PTMs within host-pathogen interactions for the purposes of pathogenic growth and immune evasion is an emerging field, which may prove to be both widespread and diverse in the mechanisms at work.

The utilization of PTMs to manipulate host cell signaling and immune response may also be true in *Plasmodium*. In *P. falciparum*, it is reported that kinases and phosphatases were found among the 320+ proteins predicted to be secreted from the parasite [48]. More recently, over 30 proteins have been validated to be secreted from the *Pf*, of which 27 proteins are novel extracellular proteins [170]. Several post-translational modifiers, such as kinases, phosphatases, and proteases, were confirmed to be among the secreted proteins. Furthermore, preliminary functional analysis suggests that these secreted proteins are possibly involved in immune evasion and signaling [170].

Our current understanding of the roles and extent of PTMs within *Plasmodium* is still greatly limited due to the relative novelty and inherent

challenges of this particular field. More genome-wide approaches, along with their proper biological validation, will provide significant advances in filling the gaps of our present understanding of the regulatory mechanisms driving this deadly parasite. Though there is still much more to uncover, current reports collectively show that *Plasmodium* PTMs have vital roles in all aspects of the parasites' life cycle, including host-pathogen interactions. Because of the numerous types and virtually incalculable combinations of PTMs, the possibilities of regulation at the post-translational level are vast. In addition to being vitally essential, many *Plasmodium* PTM proteins are both highly divergent and specific to the parasite, making *Plasmodium* PTM proteins excellent candidates for drug targeting. Thus, *Plasmodium* PTMs may open new venues for drug discovery and may prove to be more significant in the regulation of this deadly parasite than previously acknowledged.

References:

1. Miller, L.H. and B. Greenwood, *Malaria--a shadow over Africa.* Science, 2002. **298**(5591): p. 121-2.

2. Breman, J.G., A. Egan, and G.T. Keusch, *The intolerable burden of malaria: a new look at the numbers.* Am J Trop Med Hyg, 2001. **64**(1-2 Suppl): p. iv-vii.

3. Guerra, C.A., P.W. Gikandi, A.J. Tatem, A.M. Noor, D.L. Smith, S.I. Hay*, et al.*, *The limits and intensity of Plasmodium falciparum transmission: implications for malaria control and elimination worldwide.* PLoS Med, 2008. **5**(2): p. e38.

4. Gardner, M.J., N. Hall, E. Fung, O. White, M. Berriman, R.W. Hyman*, et al.*, *Genome sequence of the human malaria parasite Plasmodium falciparum.* Nature, 2002. **419**(6906): p. 498-511.

5. Bozdech, Z., M. Llinas, B.L. Pulliam, E.D. Wong, J. Zhu, and J.L. DeRisi, *The transcriptome of the intraerythrocytic developmental cycle of Plasmodium falciparum.* PLoS Biol, 2003. **1**(1): p. E5.

6. Le Roch, K.G., Y. Zhou, P.L. Blair, M. Grainger, J.K. Moch, J.D. Haynes*, et al.*, *Discovery of gene function by expression profiling of the malaria parasite life cycle.* Science, 2003. **301**(5639): p. 1503-8.

7. Le Roch, K.G., J.R. Johnson, L. Florens, Y. Zhou, A. Santrosyan, M. Grainger*, et al.*, *Global analysis of transcript and protein levels across the Plasmodium falciparum life cycle.* Genome Res, 2004. **14**(11): p. 2308-18.

8. Coulson, R.M., N. Hall, and C.A. Ouzounis, *Comparative genomics of transcriptional control in the human malaria parasite Plasmodium falciparum.* Genome Res, 2004. **14**(8): p. 1548-54.

9. Le Roch, K.G., J.R. Johnson, H. Ahiboh, D.W. Chung, J. Prudhomme, D. Plouffe*, et al.*, *A systematic approach to understand the mechanism of action of the bisthiazolium compound T4 on the human malaria parasite, Plasmodium falciparum.* BMC Genomics, 2008. **9**: p. 513.

10. Ganesan, K., N. Ponmee, L. Jiang, J.W. Fowble, J. White, S. Kamchonwongpaisan*, et al.*, *A Genetically Hard-Wired Metabolic Transcriptome in Plasmodium falciparum Fails to Mount Protective Responses to Lethal Antifolates.* PLoS Pathogens, 2008. **4**(11): p. e1000214.

11. Hoffman, M.D., M.J. Sniatynski, and J. Kast, *Current approaches for global post-translational modification discovery and mass spectrometric analysis.* Analytica Chimica Acta, 2008. **627**(1): p. 50-61.

12. Clark, R.S., H. Bayir, and L.W. Jenkins, *Posttranslational protein modifications.* Crit Care Med, 2005. **33**(12 Suppl): p. S407-9.

13. Jensen, O.N., *Modification-specific proteomics: characterization of post-translational modifications by mass spectrometry.* Curr Opin Chem Biol, 2004. **8**(1): p. 33-41.

14. Reinders, J. and A. Sickmann, *State-of-the-art in phosphoproteomics.* Proteomics, 2005. **5**(16): p. 4052-61.

15. Sickmann, A. and H.E. Meyer, *Phosphoamino acid analysis.* Proteomics, 2001. **1**(2): p. 200-6.

16. Bollen, M., *Combinatorial control of protein phosphatase-1.* Trends Biochem Sci, 2001. **26**(7): p. 426-31.

17. Dorin, D., K. Le Roch, P. Sallicandro, P. Alano, D. Parzy, P. Poullet*, et al.*, *Pfnek-1, a NIMA-related kinase from the human malaria parasite Plasmodium falciparum Biochemical properties and possible involvement in MAPK regulation.* Eur J Biochem, 2001. **268**(9): p. 2600-8.

18. McRobert, L., C.J. Taylor, W. Deng, Q.L. Fivelman, R.M. Cummings, S.D. Polley*, et al.*, *Gametogenesis in malaria parasites is mediated by the cGMP-dependent protein kinase.* PLoS Biol, 2008. **6**(6): p. e139.

19. Kato, N., T. Sakata, G. Breton, K.G. Le Roch, A. Nagle, C. Andersen*, et al.*, *Gene expression signatures and small-molecule compounds link a protein kinase to Plasmodium falciparum motility.* Nat Chem Biol, 2008. **4**(6): p. 347-56.

20. Knockaert, M., N. Gray, E. Damiens, Y.T. Chang, P. Grellier, K. Grant*, et al.*, *Intracellular targets of cyclin-dependent kinase inhibitors: identification by affinity chromatography using immobilised inhibitors.* Chem Biol, 2000. **7**(6): p. 411-22.

21. Anamika, N. Srinivasan, and A. Krupa, *A genomic perspective of protein kinases in Plasmodium falciparum.* Proteins, 2005. **58**(1): p. 180-9.

22. Ward, P., L. Equinet, J. Packer, and C. Doerig, *Protein kinases of the human malaria parasite Plasmodium falciparum: the kinome of a divergent eukaryote.* BMC Genomics, 2004. **5**(1): p. 79.

23. Schneider, A.G. and O. Mercereau-Puijalon, *A new Apicomplexa-specific protein kinase family: multiple members in Plasmodium falciparum, all with an export signature.* BMC Genomics, 2005. **6**(1): p. 30.

24. Zhang, X.S. and J.H. Choi, *Molecular evolution of calmodulin-like domain protein kinases (CDPKs) in plants and protists.* J Mol Evol, 2001. **53**(3): p. 214-24.

25. Merckx, A., K. Le Roch, M.P. Nivez, D. Dorin, P. Alano, G.J. Gutierrez*, et al.*, *Identification and initial characterization of three novel cyclin-related proteins of the human malaria parasite Plasmodium falciparum.* J Biol Chem, 2003. **278**(41): p. 39839-50.

26. Li, Z., K. Le Roch, J.A. Geyer, C.L. Woodard, S.T. Prigge, J. Koh*, et al.*, *Influence of human p16(INK4) and p21(CIP1) on the in vitro activity of recombinant Plasmodium falciparum cyclin-dependent protein kinases.* Biochem Biophys Res Commun, 2001. **288**(5): p. 1207-11.

27. Le Roch, K., C. Sestier, D. Dorin, N. Waters, B. Kappes, D. Chakrabarti*, et al.*, *Activation of a Plasmodium falciparum cdc2-related kinase by heterologous p25 and cyclin H. Functional characterization of a P. falciparum cyclin homologue.* J Biol Chem, 2000. **275**(12): p. 8952-8.

28. Graeser, R., B. Wernli, R.M. Franklin, and B. Kappes, *Plasmodium falciparum protein kinase 5 and the malarial nuclear division cycles.* Mol Biochem Parasitol, 1996. **82**(1): p. 37-49.

29. Holton, S., A. Merckx, D. Burgess, C. Doerig, M. Noble, and J. Endicott, *Structures of P. falciparum PfPK5 test the CDK regulation paradigm and suggest mechanisms of small molecule inhibition.* Structure, 2003. **11**(11): p. 1329-37.

30. Schrevel, J., G. Asfaux-Foucher, and J.M. Bafort, *[Ultrastructural study of multiple mitoses during sporogony of Plasmodium b. berghei].* J Ultrastruct Res, 1977. **59**(3): p. 332-50.

31. Dorin-Semblat, D., A. Sicard, C. Doerig, L. Ranford-Cartwright, and C. Doerig, *Disruption of the PfPK7 gene impairs schizogony and sporogony in the human*

malaria parasite Plasmodium falciparum. Eukaryot Cell, 2008. **7**(2): p. 279-85.

32. Raman, M., W. Chen, and M.H. Cobb, *Differential regulation and properties of MAPKs.* Oncogene, 2007. **26**(22): p. 3100-12.

33. Dorin-Semblat, D., N. Quashie, J. Halbert, A. Sicard, C. Doerig, E. Peat*, et al.*, *Functional characterization of both MAP kinases of the human malaria parasite Plasmodium falciparum by reverse genetics.* Mol Microbiol, 2007. **65**(5): p. 1170-80.

34. Tewari, R., D. Dorin, R. Moon, C. Doerig, and O. Billker, *An atypical mitogen-activated protein kinase controls cytokinesis and flagellar motility during male gamete formation in a malaria parasite.* Mol Microbiol, 2005. **58**(5): p. 1253-63.

35. Billker, O., S. Dechamps, R. Tewari, G. Wenig, B. Franke-Fayard, and V. Brinkmann, *Calcium and a calcium-dependent protein kinase regulate gamete formation and mosquito transmission in a malaria parasite.* Cell, 2004. **117**(4): p. 503-14.

36. Siden-Kiamos, I., A. Ecker, S. Nyback, C. Louis, R.E. Sinden, and O. Billker, *Plasmodium berghei calcium-dependent protein kinase 3 is required for ookinete gliding motility and mosquito midgut invasion.* Mol Microbiol, 2006. **60**(6): p. 1355-63.

37. Ishino, T., Y. Orito, Y. Chinzei, and M. Yuda, *A calcium-dependent protein kinase regulates Plasmodium ookinete access to the midgut epithelial cell.* Mol Microbiol, 2006. **59**(4): p. 1175-84.

38. Reininger, L., O. Billker, R. Tewari, A. Mukhopadhyay, C. Fennell, D. Dorin-Semblat*, et al.*, *A NIMA-related protein kinase is essential for completion of the sexual cycle of malaria parasites.* J Biol Chem, 2005. **280**(36): p. 31957-64.

39. Cowman, A.F. and B.S. Crabb, *Invasion of red blood cells by malaria parasites.* Cell, 2006. **124**(4): p. 755-66.

40. Moskes, C., P.A. Burghaus, B. Wernli, U. Sauder, M. Durrenberger, and B. Kappes, *Export of Plasmodium falciparum calcium-dependent protein kinase 1 to the parasitophorous vacuole is dependent on three N-terminal membrane anchor motifs.* Mol Microbiol, 2004. **54**(3): p. 676-91.

41. Kumar, A., A. Vaid, C. Syin, and P. Sharma, *PfPKB, a novel protein kinase B-like enzyme from Plasmodium falciparum: I. Identification, characterization, and possible role in parasite development.* J Biol Chem, 2004. **279**(23): p. 24255-64.

42. Vaid, A. and P. Sharma, *PfPKB, a protein kinase B-like enzyme from Plasmodium falciparum: II. Identification of calcium/calmodulin as its upstream activator and dissection of a novel signaling pathway.* J Biol Chem, 2006. **281**(37): p. 27126-33.

43. Vaid, A., D.C. Thomas, and P. Sharma, *Role of Ca2+/calmodulin-PfPKB signaling pathway in erythrocyte invasion by Plasmodium falciparum.* J Biol Chem, 2008. **283**(9): p. 5589-97.

44. Merckx, A., G. Bouyer, S.L. Thomas, G. Langsley, and S. Egee, *Anion channels in Plasmodium-falciparum-infected erythrocytes and protein kinase A.* Trends Parasitol, 2009. **25**(3): p. 139-44.

45. Merckx, A., M.P. Nivez, G. Bouyer, P. Alano, G. Langsley, K. Deitsch*, et al.*, *Plasmodium falciparum regulatory subunit of cAMP-dependent PKA and anion channel conductance.* PLoS Pathog, 2008. **4**(2): p. e19.

46. Nunes, M.C., J.P. Goldring, C. Doerig, and A. Scherf, *A novel protein kinase family in Plasmodium falciparum is differentially transcribed and secreted to various cellular compartments of the host cell.* Mol Microbiol, 2007. **63**(2): p. 391-403.

47. Marti, M., R.T. Good, M. Rug, E. Knuepfer, and A.F. Cowman, *Targeting malaria virulence and remodeling proteins to the host erythrocyte.* Science, 2004. **306**(5703): p. 1930-3.

48. Hiller, N.L., S. Bhattacharjee, C. van Ooij, K. Liolios, T. Harrison, C. Lopez-Estrano*, et al.*, *A host-targeting signal in virulence proteins reveals a secretome in malarial infection.* Science, 2004. **306**(5703): p. 1934-7.

49. Wilkes, J.M. and C. Doerig, *The protein-phosphatome of the human malaria parasite Plasmodium falciparum.* BMC Genomics, 2008. **9**: p. 412.

50. Li, J.L. and D.A. Baker, *A putative protein serine/threonine phosphatase from Plasmodium falciparum contains a large N-terminal extension and five unique inserts in the catalytic domain.* Mol Biochem Parasitol, 1998. **95**(2): p. 287-95.

51. Mamoun, C.B., D.J. Sullivan, Jr., R. Banerjee, and D.E. Goldberg, *Identification and characterization of an unusual double serine/threonine protein phosphatase 2C in the malaria parasite Plasmodium falciparum.* J Biol Chem, 1998. **273**(18): p. 11241-7.

52. Dobson, S., T. May, M. Berriman, C. Del Vecchio, A.H. Fairlamb, D. Chakrabarti, *et al.*, *Characterization of protein Ser/Thr phosphatases of the malaria parasite, Plasmodium falciparum: inhibition of the parasitic calcineurin by cyclophilin-cyclosporin complex.* Mol Biochem Parasitol, 1999. **99**(2): p. 167-81.

53. Ward, G.E., H. Fujioka, M. Aikawa, and L.H. Miller, *Staurosporine inhibits invasion of erythrocytes by malarial merozoites.* Exp Parasitol, 1994. **79**(3): p. 480-7.

54. Yokoyama, D., A. Saito-Ito, N. Asao, K. Tanabe, M. Yamamoto, and T. Matsumura, *Modulation of the growth of Plasmodium falciparum in vitro by protein serine/threonine phosphatase inhibitors.* Biochem Biophys Res Commun, 1998. **247**(1): p. 18-23.

55. Mamoun, C.B. and D.E. Goldberg, *Plasmodium protein phosphatase 2C dephosphorylates translation elongation factor 1beta and inhibits its PKC-mediated nucleotide exchange activity in vitro.* Mol Microbiol, 2001. **39**(4): p. 973-81.

56. Kumar, R., A. Musiyenko, E. Cioffi, A. Oldenburg, B. Adams, V. Bitko, *et al.*, *A zinc-binding dual-specificity YVH1 phosphatase in the malaria parasite, Plasmodium falciparum, and its interaction with the nuclear protein, pescadillo.* Mol Biochem Parasitol, 2004. **133**(2): p. 297-310.

57. Blisnick, T., L. Vincensini, G. Fall, and C. Braun-Breton, *Protein phosphatase 1, a Plasmodium falciparum essential enzyme, is exported to the host cell and implicated in the release of infectious merozoites.* Cell Microbiol, 2006. **8**(4): p. 591-601.

58. Rudolph, J., *Cdc25 phosphatases: structure, specificity, and mechanism.* Biochemistry, 2007. **46**(12): p. 3595-604.

59. Trinkle-Mulcahy, L. and A.I. Lamond, *Mitotic phosphatases: no longer silent partners.* Curr Opin Cell Biol, 2006. **18**(6): p. 623-31.

60. Fauman, E.B. and M.A. Saper, *Structure and function of the protein tyrosine phosphatases.* Trends Biochem Sci, 1996. **21**(11): p. 413-7.

61. Pendyala, P.R., L. Ayong, J. Eatrides, M. Schreiber, C. Pham, R. Chakrabarti*, et al.*, *Characterization of a PRL protein tyrosine phosphatase from Plasmodium falciparum.* Mol Biochem Parasitol, 2008. **158**(1): p. 1-10.

62. Schmitz, S., M. Grainger, S. Howell, L.J. Calder, M. Gaeb, J.C. Pinder*, et al.*, *Malaria parasite actin filaments are very short.* J Mol Biol, 2005. **349**(1): p. 113-25.

63. Miao, J., Q. Fan, L. Cui, J. Li, J. Li, and L. Cui, *The malaria parasite Plasmodium falciparum histones: organization, expression, and acetylation.* Gene, 2006. **369**: p. 53-65.

64. Cui, L., J. Miao, T. Furuya, X. Li, X.Z. Su, and L. Cui, *PfGCN5-mediated histone H3 acetylation plays a key role in gene expression in Plasmodium falciparum.* Eukaryot Cell, 2007. **6**(7): p. 1219-27.

65. Fan, Q., L. An, and L. Cui, *Plasmodium falciparum histone acetyltransferase, a yeast GCN5 homologue involved in chromatin remodeling.* Eukaryot Cell, 2004. **3**(2): p. 264-76.

66. Cui, L., J. Miao, T. Furuya, Q. Fan, X. Li, P.K. Rathod*, et al.*, *Histone acetyltransferase inhibitor anacardic acid causes changes in global gene expression during in vitro Plasmodium falciparum development.* Eukaryot Cell, 2008. **7**(7): p. 1200-10.

67. Joshi, M.B., D.T. Lin, P.H. Chiang, N.D. Goldman, H. Fujioka, M. Aikawa*, et al.*, *Molecular cloning and nuclear localization of a histone deacetylase homologue in Plasmodium falciparum.* Mol Biochem Parasitol, 1999. **99**(1): p. 11-9.

68. Freitas-Junior, L.H., R. Hernandez-Rivas, S.A. Ralph, D. Montiel-Condado, O.K. Ruvalcaba-Salazar, A.P. Rojas-Meza*, et al.*, *Telomeric heterochromatin propagation and histone acetylation control mutually exclusive expression of antigenic variation genes in malaria parasites.* Cell, 2005. **121**(1): p. 25-36.

69. Andrews, K.T., T.N. Tran, N.C. Wheatley, and D.P. Fairlie, *Targeting histone deacetylase inhibitors for anti-malarial therapy.* Curr Top Med Chem, 2009. **9**(3): p. 292-308.

70. Dow, G.S., Y. Chen, K.T. Andrews, D. Caridha, L. Gerena, M. Gettayacamin*, et al.*, *Antimalarial activity of phenylthiazolyl-bearing hydroxamate-based histone deacetylase inhibitors.* Antimicrob Agents Chemother, 2008. **52**(10): p. 3467-77.

71. Lopez-Rubio, J.J., A.M. Gontijo, M.C. Nunes, N. Issar, R. Hernandez Rivas, and A. Scherf, *5' flanking region of var genes nucleate histone modification patterns linked to phenotypic inheritance of virulence traits in malaria parasites.* Mol Microbiol, 2007. **66**(6): p. 1296-305.

72. Cui, L., Q. Fan, L. Cui, and J. Miao, *Histone lysine methyltransferases and demethylases in Plasmodium falciparum.* Int J Parasitol, 2008. **38**(10): p. 1083-97.

73. Gowda, D.C., P. Gupta, and E.A. Davidson, *Glycosylphosphatidylinositol anchors represent the major carbohydrate modification in proteins of intraerythrocytic stage Plasmodium falciparum.* J Biol Chem, 1997. **272**(10): p. 6428-39.

74. Haldar, K., B.U. Samuel, N. Mohandas, T. Harrison, and N.L. Hiller, *Erythrocytic vacuolar rafts induced by malaria parasites.* Curr Opin Hematol, 2001. **8**(2): p. 92-7.

75. Naik, R.S., O.H. Branch, A.S. Woods, M. Vijaykumar, D.J. Perkins, B.L. Nahlen*, et al.*, *Glycosylphosphatidylinositol anchors of Plasmodium falciparum: molecular characterization and naturally elicited antibody response that may provide immunity to malaria pathogenesis.* J Exp Med, 2000. **192**(11): p. 1563-76.

76. Schofield, L. and F. Hackett, *Signal transduction in host cells by a glycosylphosphatidylinositol toxin of malaria parasites.* J Exp Med, 1993. **177**(1): p. 145-53.

77. Patel, S.N., Z. Lu, K. Ayi, L. Serghides, D.C. Gowda, and K.C. Kain, *Disruption of CD36 impairs cytokine response to Plasmodium falciparum glycosylphosphatidylinositol and confers susceptibility to severe and fatal malaria in vivo.* J Immunol, 2007. **178**(6): p. 3954-61.

78. Nasir ud, D., I. Ahmad, D.C. Hoessli, E. Walker-Nasir, and M.I. Choudhary, *Glycosylphosphatidylinositol (GPI) anchored proteins of Plasmodium falciparum: Antigenic determinants and role of sugar moieties in the GPI anchor.* Current Organic Chemistry, 2007. **11**(7): p. 609-618.

79. Proellocks, N.I., S. Kovacevic, D.J. Ferguson, L.M. Kats, B.J. Morahan, C.G. Black*, et al.*, *Plasmodium falciparum Pf34, a novel GPI-anchored rhoptry protein found in detergent-resistant microdomains.* Int J Parasitol, 2007. **37**(11): p. 1233-41.

80. Gilson, P.R., T. Nebl, D. Vukcevic, R.L. Moritz, T. Sargeant, T.P. Speed*, et al.*, *Identification and stoichiometry of glycosylphosphatidylinositol-anchored membrane proteins of the human malaria parasite Plasmodium falciparum.* Mol Cell Proteomics, 2006. **5**(7): p. 1286-99.

81. Lim, J., D.C. Gowda, G. Krishnegowda, and S. Luckhart, *Induction of nitric oxide synthase in Anopheles stephensi by Plasmodium falciparum: mechanism of signaling and the role of parasite glycosylphosphatidylinositols.* Infect Immun, 2005. **73**(5): p. 2778-89.

82. Akman-Anderson, L., M. Olivier, and S. Luckhart, *Induction of nitric oxide synthase and activation of signaling proteins in Anopheles mosquitoes by the malaria pigment, hemozoin.* Infect Immun, 2007. **75**(8): p. 4012-9.

83. Arrighi, R.B., F. Debierre-Grockiego, R.T. Schwarz, and I. Faye, *The immunogenic properties of protozoan glycosylphosphatidylinositols in the mosquito Anopheles gambiae.* Dev Comp Immunol, 2009. **33**(2): p. 216-23.

84. Blanco, A.R., A. Paez, P. Gerold, A.L. Dearsly, G. Margos, R.T. Schwarz*, et al.*, *The biosynthesis and post-translational modification of Pbs21 an ookinete-surface protein of Plasmodium berghei.* Mol Biochem Parasitol, 1999. **98**(2): p. 163-73.

85. del Carmen Rodriguez, M., P. Gerold, J. Dessens, K. Kurtenbach, R.T. Schwartz, R.E. Sinden*, et al.*, *Characterisation and expression of Pbs25, a sexual and sporogonic stage specific protein of Plasmodium berghei.* Molecular and Biochemical Parasitology, 2000. **110**(1): p. 147-159.

86. Baton, L.A. and L.C. Ranford-Cartwright, *Do malaria ookinete surface proteins P25 and P28 mediate parasite entry into mosquito midgut epithelial cells?* Malar J, 2005. **4**(1): p. 15.

87. Wang, Q., H. Fujioka, and V. Nussenzweig, *Mutational analysis of the GPI-anchor addition sequence from the circumsporozoite protein of Plasmodium.* Cell Microbiol, 2005. **7**(11): p. 1616-26.

88. Chakrabarti, D., T. Da Silva, J. Barger, S. Paquette, H. Patel, S. Patterson*, et al.*, *Protein farnesyltransferase and protein prenylation in Plasmodium falciparum.* J Biol Chem, 2002. **277**(44): p. 42066-73.

89. Eastman, R.T., F.S. Buckner, K. Yokoyama, M.H. Gelb, and W.C. Van Voorhis, *Thematic review series: lipid posttranslational modifications. Fighting*

parasitic disease by blocking protein farnesylation. J Lipid Res, 2006. **47**(2): p. 233-40.

90. Rohrich, R.C., N. Englert, K. Troschke, A. Reichenberg, M. Hintz, F. Seeber*, et al.*, *Reconstitution of an apicoplast-localised electron transfer pathway involved in the isoprenoid biosynthesis of Plasmodium falciparum.* FEBS Lett, 2005. **579**(28): p. 6433-8.

91. Fletcher, S., C.G. Cummings, K. Rivas, W.P. Katt, C. Horney, F.S. Buckner*, et al.*, *Potent, Plasmodium-selective farnesyltransferase inhibitors that arrest the growth of malaria parasites: structure-activity relationships of ethylenediamine-analogue scaffolds and homology model validation.* J Med Chem, 2008. **51**(17): p. 5176-97.

92. Baekkeskov, S. and J. Kanaani, *Palmitoylation cycles and regulation of protein function (Review).* Mol Membr Biol, 2009. **26**(1): p. 42-54.

93. Rees-Channer, R.R., S.R. Martin, J.L. Green, P.W. Bowyer, M. Grainger, J.E. Molloy*, et al.*, *Dual acylation of the 45 kDa gliding-associated protein (GAP45) in Plasmodium falciparum merozoites.* Mol Biochem Parasitol, 2006. **149**(1): p. 113-6.

94. Hershko, A. and A. Ciechanover, *The ubiquitin system.* Annu Rev Biochem, 1998. **67**: p. 425-79.

95. Laney, J.D. and M. Hochstrasser, *Substrate targeting in the ubiquitin system.* Cell, 1999. **97**(4): p. 427-30.

96. Kerscher, O., R. Felberbaum, and M. Hochstrasser, *Modification of proteins by ubiquitin and ubiquitin-like proteins.* Annu Rev Cell Dev Biol, 2006. **22**: p. 159-80.

97. Kreidenweiss, A., P.G. Kremsner, and B. Mordmuller, *Comprehensive study of proteasome inhibitors against Plasmodium falciparum laboratory strains and field isolates from Gabon.* Malar J, 2008. **7**: p. 187.

98. Reynolds, J.M., K. El Bissati, J. Brandenburg, A. Gunzl, and C.B. Mamoun, *Antimalarial activity of the anticancer and proteasome inhibitor bortezomib and its analog ZL3B.* BMC Clin Pharmacol, 2007. **7**: p. 13.

99. Lindenthal, C., N. Weich, Y.S. Chia, V. Heussler, and M.Q. Klinkert, *The proteasome inhibitor MLN-273 blocks exoerythrocytic and erythrocytic*

development of Plasmodium parasites. Parasitology, 2005. **131**(Pt 1): p. 37-44.

100. Mordmüller, B., R. Fendel, A. Kreidenweiss, C. Gille, R. Hurwitz, W.G. Metzger, *et al.*, *Plasmodia express two threonine-peptidase complexes during asexual development.* Molecular and Biochemical Parasitology, 2006. **148**(1): p. 79-85.

101. Prudhomme, J., E. McDaniel, N. Ponts, S.p. Bertani, W. Fenical, P. Jensen*, et al.*, *Marine Actinomycetes: A New Source of Compounds against the Human Malaria Parasite.* PLoS ONE, 2008. **3**(6): p. e2335.

102. Gantt, S.M., J.M. Myung, M.R. Briones, W.D. Li, E.J. Corey, S. Omura*, et al.*, *Proteasome inhibitors block development of Plasmodium spp.* Antimicrob Agents Chemother, 1998. **42**(10): p. 2731-8.

103. Mukhopadhyay, D. and H. Riezman, *Proteasome-Independent Functions of Ubiquitin in Endocytosis and Signaling.* Science, 2007. **315**(5809): p. 201-205.

104. Hicke, L., *Protein regulation by monoubiquitin.* Nat Rev Mol Cell Biol, 2001. **2**(3): p. 195-201.

105. Hofmann, K., *Ubiquitin-binding domains and their role in the DNA damage response.* DNA Repair, 2009. **In Press, Corrected Proof**.

106. Horrocks, P. and C.I. Newbold, *Intraerythrocytic polyubiquitin expression in Plasmodium falciparum is subjected to developmental and heat-shock control.* Mol Biochem Parasitol, 2000. **105**(1): p. 115-25.

107. Ponts, N., J. Yang, D.-W.D. Chung, J. Prudhomme, T. Girke, P. Horrocks*, et al.*, *Deciphering the Ubiquitin-Mediated Pathway in Apicomplexan Parasites: A Potential Strategy to Interfere with Parasite Virulence.* PLoS ONE, 2008. **3**(6): p. e2386.

108. Catic, A. and H.L. Ploegh, *Ubiquitin--conserved protein or selfish gene?* Trends Biochem Sci, 2005. **30**(11): p. 600-4.

109. Ponder, E.L. and M. Bogyo, *Ubiquitin-like modifiers and their deconjugating enzymes in medically important parasitic protozoa.* Eukaryot Cell, 2007. **6**(11): p. 1943-52.

110. Issar, N., E. Roux, D. Mattei, and A. Scherf, *Identification of a novel post-translational modification in Plasmodium falciparum: protein sumoylation in different cellular compartments.* Cell Microbiol, 2008. **10**(10): p. 1999-2011.

111. Philip, N. and T.A. Haystead, *Characterization of a UBC13 kinase in Plasmodium falciparum.* Proc Natl Acad Sci U S A, 2007. **104**(19): p. 7845-50.

112. Hofmann, R.M. and C.M. Pickart, *Noncanonical MMS2-encoded ubiquitin-conjugating enzyme functions in assembly of novel polyubiquitin chains for DNA repair.* Cell, 1999. **96**(5): p. 645-53.

113. Pickart, C.M., *Ubiquitin enters the new millennium.* Mol Cell, 2001. **8**(3): p. 499-504.

114. Brusky, J., Y. Zhu, and W. Xiao, *UBC13, a DNA-damage-inducible gene, is a member of the error-free postreplication repair pathway in Saccharomyces cerevisiae.* Curr Genet, 2000. **37**(3): p. 168-74.

115. Laine, A., I. Topisirovic, D. Zhai, J.C. Reed, K.L. Borden, and Z. Ronai, *Regulation of p53 localization and activity by Ubc13.* Mol Cell Biol, 2006. **26**(23): p. 8901-13.

116. Bothos, J., M.K. Summers, M. Venere, D.M. Scolnick, and T.D. Halazonetis, *The Chfr mitotic checkpoint protein functions with Ubc13-Mms2 to form Lys63-linked polyubiquitin chains.* Oncogene, 2003. **22**(46): p. 7101-7.

117. Singhal, S., M.C. Taylor, and R.T. Baker, *Deubiquitylating enzymes and disease.* BMC Biochem, 2008. **9 Suppl 1**: p. S3.

118. Artavanis-Tsakonas, K., S. Misaghi, C.A. Comeaux, A. Catic, E. Spooner, M.T. Duraisingh*, et al.*, *Identification by functional proteomics of a deubiquitinating/deNeddylating enzyme in Plasmodium falciparum.* Mol Microbiol, 2006. **61**(5): p. 1187-95.

119. Wada, H., K. Kito, L.S. Caskey, E.T. Yeh, and T. Kamitani, *Cleavage of the C-terminus of NEDD8 by UCH-L3.* Biochem Biophys Res Commun, 1998. **251**(3): p. 688-92.

120. Butterworth, M.B., R.S. Edinger, H. Ovaa, D. Burg, J.P. Johnson, and R.A. Frizzell, *The deubiquitinating enzyme UCH-L3 regulates the apical membrane recycling of the epithelial sodium channel.* J Biol Chem, 2007. **282**(52): p. 37885-93.

121. McKerrow, J.H., P.J. Rosenthal, R. Swenerton, and P. Doyle, *Development of protease inhibitors for protozoan infections.* Curr Opin Infect Dis, 2008. **21**(6): p. 668-72.

122. Armstrong, P.B., *Proteases and protease inhibitors: a balance of activities in host-pathogen interaction.* Immunobiology, 2006. **211**(4): p. 263-81.

123. Wu, Y., X. Wang, X. Liu, and Y. Wang, *Data-mining approaches reveal hidden families of proteases in the genome of malaria parasite.* Genome Res, 2003. **13**(4): p. 601-16.

124. Waller, R.F., M.B. Reed, A.F. Cowman, and G.I. McFadden, *Protein trafficking to the plastid of Plasmodium falciparum is via the secretory pathway.* Embo J, 2000. **19**(8): p. 1794-802.

125. Eggleson, K.K., K.L. Duffin, and D.E. Goldberg, *Identification and characterization of falcilysin, a metallopeptidase involved in hemoglobin catabolism within the malaria parasite Plasmodium falciparum.* J Biol Chem, 1999. **274**(45): p. 32411-7.

126. Ponpuak, M., M. Klemba, M. Park, I.Y. Gluzman, G.K. Lamppa, and D.E. Goldberg, *A role for falcilysin in transit peptide degradation in the Plasmodium falciparum apicoplast.* Mol Microbiol, 2007. **63**(2): p. 314-34.

127. Craig, A. and A. Scherf, *Molecules on the surface of the Plasmodium falciparum infected erythrocyte and their role in malaria pathogenesis and immune evasion.* Mol Biochem Parasitol, 2001. **115**(2): p. 129-43.

128. Chang, H.H., A.M. Falick, P.M. Carlton, J.W. Sedat, J.L. DeRisi, and M.A. Marletta, *N-terminal processing of proteins exported by malaria parasites.* Mol Biochem Parasitol, 2008. **160**(2): p. 107-15.

129. Goldberg, D.E., *Hemoglobin degradation.* Curr Top Microbiol Immunol, 2005. **295**: p. 275-91.

130. Francis, S.E., R. Banerjee, and D.E. Goldberg, *Biosynthesis and maturation of the malaria aspartic hemoglobinases plasmepsins I and II.* J Biol Chem, 1997. **272**(23): p. 14961-8.

131. Lew, V.L., T. Tiffert, and H. Ginsburg, *Excess hemoglobin digestion and the osmotic stability of Plasmodium falciparum-infected red blood cells.* Blood, 2003. **101**(10): p. 4189-94.

132. Rosenthal, P.J., *Cysteine proteases of malaria parasites.* Int J Parasitol, 2004. **34**(13-14): p. 1489-99.

133. Banerjee, R., J. Liu, W. Beatty, L. Pelosof, M. Klemba, and D.E. Goldberg, *Four plasmepsins are active in the Plasmodium falciparum food vacuole, including a protease with an active-site histidine.* Proc Natl Acad Sci U S A, 2002. **99**(2): p. 990-5.

134. Coombs, G.H., D.E. Goldberg, M. Klemba, C. Berry, J. Kay, and J.C. Mottram, *Aspartic proteases of Plasmodium falciparum and other parasitic protozoa as drug targets.* Trends Parasitol, 2001. **17**(11): p. 532-7.

135. Wyatt, D.M. and C. Berry, *Activity and inhibition of plasmepsin IV, a new aspartic proteinase from the malaria parasite, Plasmodium falciparum.* FEBS Lett, 2002. **513**(2-3): p. 159-62.

136. Francis, S.E., D.J. Sullivan, Jr., and D.E. Goldberg, *Hemoglobin metabolism in the malaria parasite Plasmodium falciparum.* Annu Rev Microbiol, 1997. **51**: p. 97-123.

137. Drew, M.E., R. Banerjee, E.W. Uffman, S. Gilbertson, P.J. Rosenthal, and D.E. Goldberg, *Plasmodium food vacuole plasmepsins are activated by falcipains.* J Biol Chem, 2008. **283**(19): p. 12870-6.

138. Blackman, M.J., *Malarial proteases and host cell egress: an 'emerging' cascade.* Cell Microbiol, 2008. **10**(10): p. 1925-34.

139. Pang, X.L., T. Mitamura, and T. Horii, *Antibodies reactive with the N-terminal domain of Plasmodium falciparum serine repeat antigen inhibit cell proliferation by agglutinating merozoites and schizonts.* Infect Immun, 1999. **67**(4): p. 1821-7.

140. Miller, S.K., R.T. Good, D.R. Drew, M. Delorenzi, P.R. Sanders, A.N. Hodder, *et al.*, *A subset of Plasmodium falciparum SERA genes are expressed and appear to play an important role in the erythrocytic cycle.* J Biol Chem, 2002. **277**(49): p. 47524-32.

141. Aly, A.S. and K. Matuschewski, *A malarial cysteine protease is necessary for Plasmodium sporozoite egress from oocysts.* J Exp Med, 2005. **202**(2): p. 225-30.

142. Arastu-Kapur, S., E.L. Ponder, U.P. Fonovic, S. Yeoh, F. Yuan, M. Fonovic, *et al.*, *Identification of proteases that regulate erythrocyte rupture by the malaria parasite Plasmodium falciparum.* Nat Chem Biol, 2008. **4**(3): p. 203-13.

143. Yeoh, S., R.A. O'Donnell, K. Koussis, A.R. Dluzewski, K.H. Ansell, S.A. Osborne, *et al.*, *Subcellular discharge of a serine protease mediates release of invasive malaria parasites from host erythrocytes.* Cell, 2007. **131**(6): p. 1072-83.

144. Withers-Martinez, C., L. Jean, and M.J. Blackman, *Subtilisin-like proteases of the malaria parasite.* Mol Microbiol, 2004. **53**(1): p. 55-63.

145. O'Donnell, R.A. and M.J. Blackman, *The role of malaria merozoite proteases in red blood cell invasion.* Curr Opin Microbiol, 2005. **8**(4): p. 422-7.

146. O'Donnell, R.A., F. Hackett, S.A. Howell, M. Treeck, N. Struck, Z. Krnajski, *et al.*, *Intramembrane proteolysis mediates shedding of a key adhesin during erythrocyte invasion by the malaria parasite.* J Cell Biol, 2006. **174**(7): p. 1023-33.

147. Baker, R.P., R. Wijetilaka, and S. Urban, *Two Plasmodium rhomboid proteases preferentially cleave different adhesins implicated in all invasive stages of malaria.* PLoS Pathog, 2006. **2**(10): p. e113.

148. Lemberg, M.K. and M. Freeman, *Functional and evolutionary implications of enhanced genomic analysis of rhomboid intramembrane proteases.* Genome Res, 2007. **17**(11): p. 1634-46.

149. Srinivasan, P., I. Coppens, and M. Jacobs-Lorena, *Distinct roles of Plasmodium rhomboid 1 in parasite development and malaria pathogenesis.* PLoS Pathog, 2009. **5**(1): p. e1000262.

150. Silvie, O., J.F. Franetich, L. Renia, and D. Mazier, *Malaria sporozoite: migrating for a living.* Trends Mol Med, 2004. **10**(3): p. 97-100; discussion 100-1.

151. Harris, P.K., S. Yeoh, A.R. Dluzewski, R.A. O'Donnell, C. Withers-Martinez, F. Hackett, *et al.*, *Molecular identification of a malaria merozoite surface sheddase.* PLoS Pathog, 2005. **1**(3): p. 241-51.

152. Koussis, K., C. Withers-Martinez, S. Yeoh, M. Child, F. Hackett, E. Knuepfer, *et al.*, *A multifunctional serine protease primes the malaria parasite for red blood cell invasion.* Embo J, 2009.

153. Greenbaum, D.C., A. Baruch, M. Grainger, Z. Bozdech, K.F. Medzihradszky, J. Engel, *et al.*, *A role for the protease falcipain 1 in host cell invasion by the human malaria parasite.* Science, 2002. **298**(5600): p. 2002-6.

154. Eksi, S., B. Czesny, D.C. Greenbaum, M. Bogyo, and K.C. Williamson, *Targeted disruption of Plasmodium falciparum cysteine protease, falcipain 1, reduces oocyst production, not erythrocytic stage growth.* Mol Microbiol, 2004. **53**(1): p. 243-50.

155. Pandey, K.C., N. Singh, S. Arastu-Kapur, M. Bogyo, and P.J. Rosenthal, *Falstatin, a cysteine protease inhibitor of Plasmodium falciparum, facilitates erythrocyte invasion.* PLoS Pathog, 2006. **2**(11): p. e117.

156. Bushell, G.R., L.T. Ingram, C.A. Fardoulys, and J.A. Cooper, *An antigenic complex in the rhoptries of Plasmodium falciparum.* Mol Biochem Parasitol, 1988. **28**(2): p. 105-12.

157. Howard, R.F. and R.T. Reese, *Plasmodium falciparum: hetero-oligomeric complexes of rhoptry polypeptides.* Exp Parasitol, 1990. **71**(3): p. 330-42.

158. Howard, R.F. and C.M. Schmidt, *The secretary pathway of plasmodium falciparum regulates transport of p82/RAP1 to the rhoptries.* Mol Biochem Parasitol, 1995. **74**(1): p. 43-54.

159. Medzihradszky, K.F., Z. Darula, E. Perlson, M. Fainzilber, R.J. Chalkley, H. Ball, *et al.*, *O-sulfonation of serine and threonine: mass spectrometric detection and characterization of a new posttranslational modification in diverse proteins throughout the eukaryotes.* Mol Cell Proteomics, 2004. **3**(5): p. 429-40.

160. Fennell, B.J., Z.A. Al-shatr, and A. Bell, *Isotype expression, post-translational modification and stage-dependent production of tubulins in erythrocytic Plasmodium falciparum.* Int J Parasitol, 2008. **38**(5): p. 527-39.

161. Banumathy, G., V. Singh, S.R. Pavithra, and U. Tatu, *Heat shock protein 90 function is essential for Plasmodium falciparum growth in human erythrocytes.* J Biol Chem, 2003. **278**(20): p. 18336-45.

162. Pavithra, S.R., R. Kumar, and U. Tatu, *Systems analysis of chaperone networks in the malarial parasite Plasmodium falciparum.* PLoS Comput Biol, 2007. **3**(9): p. 1701-15.

163. Mouray, E., M. Moutiez, S. Girault, C. Sergheraert, I. Florent, and P. Grellier, *Biochemical properties and cellular localization of Plasmodium falciparum protein disulfide isomerase.* Biochimie, 2007. **89**(3): p. 337-46.

164. Saridaki, T., C.P. Sanchez, J. Pfahler, and M. Lanzer, *A conditional export system provides new insights into protein export in Plasmodium falciparum-infected erythrocytes.* Cell Microbiol, 2008. **10**(12): p. 2483-95.

165. Randow, F. and P.J. Lehner, *Viral avoidance and exploitation of the ubiquitin system.* Nat Cell Biol, 2009. **11**(5): p. 527-34.

166. Hamon, M.A. and P. Cossart, *Histone modifications and chromatin remodeling during bacterial infections.* Cell Host Microbe, 2008. **4**(2): p. 100-9.

167. Stulemeijer, I.J. and M.H. Joosten, *Post-translational modification of host proteins in pathogen-triggered defence signalling in plants.* Mol Plant Pathol, 2008. **9**(4): p. 545-60.

168. Munro, P., G. Flatau, and E. Lemichez, *Bacteria and the ubiquitin pathway.* Curr Opin Microbiol, 2007. **10**(1): p. 39-46.

169. Edelmann, M.J. and B.M. Kessler, *Ubiquitin and ubiquitin-like specific proteases targeted by infectious pathogens: Emerging patterns and molecular principles.* Biochim Biophys Acta, 2008. **1782**(12): p. 809-16.

170. Singh, M., P. Mukherjee, K. Narayanasamy, R. Arora, S. Gupta, K. Natarajan*, et al.*, *Proteome analysis of plasmodium falciparum extracellular secretory antigens at asexual blood stages reveals a cohort of proteins with possible roles in immune modulation and signaling.* Mol Cell Proteomics, 2009.

Figure and Table legends:

Figure A1.1: General depiction of the roles of a few selected proteins involved in post-translational modifications during the life cycle of the malaria parasite

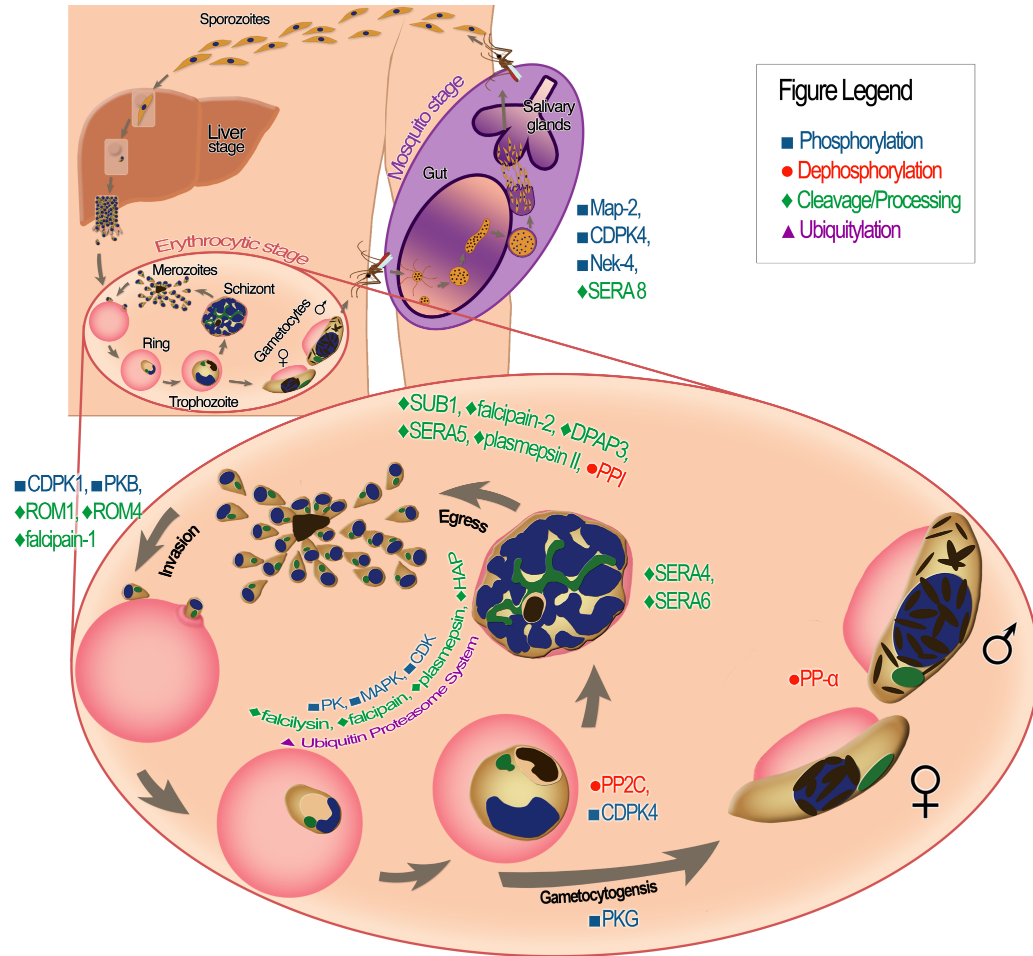


Table 1

Genes that are putatively involved in post-translational modifications in *Plasmodium*

Modification	Type of Protein	Protein	Gene ID	Putative Role
Phosphorylation	Casein kinase	CK1	PF11_0377	Unknown [20]
	Cyclin dependent kinase	PK5	MAL13P1.279	Cell-cycle Regulation [25-27]
	Cyclin dependent kinase	mrk	PF10_0141	Cell-cycle Regulation [25-27]
	Serine/threonine protein kinase	PK7	PFB0605w	Proliferation and development [31]
	Mitogen-activated protein kinase	map-1	PF14_0294	Unknown [33]
	Mitogen-activated protein kinase	map-2	PF11_0147	Asexual cycle regulation [33]
	Mitogen-activated protein kinase	map-2	PB000659.00.0	Exflagellation in mosquito midgut [34]
	Calcium-dependent kinase	CDPK4	PF07_0072	DNA replication, ookinete gliding mobility, mosquito midgut invasion [35-37]
	cGMP-dependant protein kinase	PKG	PF14_0346	Gametocytogenesis initiation [18]
	NIMA-related kinase	nek-4	MAL7P1.100	Ookinete maturation [38]
	Calmodulin-like protein kinase	CDPK1	PFB0815w	Secretion of microneme contents, formation of tight moving junctions [40]
	Protein kinase B	PKB	PFL2250c	Parasite gliding [41]
	Protein kinase A	PKA	PFI1685w	Mediates erythrocyte anion channels and permeability of host plasma membrane [44-45]
	FIKK orphan kinases	Multiple proteins	Multiple genes	Targets parasite proteins to host membranes [46]
Dephosphorylation	Mg^{2+}-dependent serine/threonine protein phosphatase 2C	PP2C	Multiple genes	Transcription elongation [55]
	VH1 family phosphatases	YVH1	Multiple genes	Nuclear protein activity [56]
	Protein phosphatase 1	PP1	Multiple genes	Release of infection merozoites [57]
	PPP-related protein serine/threonine phosphatase	PP-α	Multiple genes	Cell-cycle control and signal transduction in sexual stages [50]
Lipidation	Prenyl modifier	PFT α-subunit	PFL2050w	Unknown [88]
	Prenyl modifier	PFT β-subunit	PF11_0483	Unknown [88]
Ubiquitylation	Ubiquitin	pUB	PFL0585w	Polyubiquitin gene that provides the ubiquitin monomers that are covalently attached to proteins. Cell-cycle regulation [106]
	Ubiquitin	UbS27a	PF13_0346	Ubiquitin moiety that provide the ubiquitin monomers for attachment to proteins. Cell-cycle regulation [108]
	Ubiquitin	UbL40	PF14_0027	Ubiquitin moiety that provide the ubiquitin monomers for attachment to proteins. Cell-cycle regulation [108]
	SUMO	PfSUMO	PFE0285c	Ubiquitin-like peptide that is covalently attached to proteins for DNA repair, cell-cycle regulation, nuclear localization and protein stability [110]
	E2 conjugating enzyme	UBC13	PFE1350c	Was found to be phosphorylated by PfPK9. Role in cell-cycle has to be validated in *Plasmodium* [111]
	Deubiquitylating/DeNeddylating enzyme	UCH54	PF14_0576	Has dual deubiquitylating and deNeddylating activity. May have roles in maintaining stable apical membrane epithelial Na+ channels [118]
Cleavage/Processing	serine protease	subtilisin-1 (SUB1)	PFE0370c	Activates SERA proteins. Primary regulator of egress [143]
		subtilisin-2 (SUB2)	PF11_0381	Shedding of MSP-1 and PfAMA-1 after invasion [151]
		subtilisin-3 (SUB3)	PFE0355c	Ubiquitous role [144]
		ROM-1	PF11_0150	Shedding of adhesins [147]
		ROM-1	PB000352.00.0	May have roles in invasion [149]
		ROM-4	PFE0340c	Shedding of adhesins [146]
		SP1	PF13_0118	Cleavage of transit peptide [123]
	cysteine protease	falcipain-1	PF14_0553	Putative role in invasion [153]
		falcipain-2	PF11_0165	Hemoglobinase & role in egress [132]
		falcipain-3	PF11_0162	Hemoglobinase [132]
		SERA-4	PFB0345c	Expressed in the PV at the late schizont stage [140]
		SERA-5	PFB0340c	Role in merozoite egress [139]
		SERA-6	PFB0335c	Expressed in the PV at the late schizont stage [140]
		SERA-8	PFB0325c	Sporozoite release from oocystes [141]
		DPAP3	PFD0230c	Primary regulator of egress [142]
		calpain	MAL13P1.310	Unknown [132]
	metallo protease	falcilysin	PF13_0322	Hemoglobinase [125]
	aspartic protease	plasmepsin I	PF14_0076	Initiate degradation of hemoglobin [133]
		plasmepsin II	PF14_0077	Initiate degradation of hemoglobin & role in egress [133]
		plasmepsin IV	PF14_0075	Cleavage of denatured globin [133]
		histo-aspartic protease (HAP)	PF14_0078	Cleavage of denatured globin [133]

Table A1.1: Commonly studied post-translational modifications in *Plasmodium* and their putative biological functions.

APPENDIX A2:

Targeting the *Plasmodium* ubiquitin/proteasome system with anti-malarial compounds: promises for the future. *Infect Disord Drug Targets* 10: 158–164 (2010)

Duk-Won Doug Chung and Karine G. Le Roch

Copyright permission granted by Bentham Science Publishers

Abstract:

The human malarial parasite, *Plasmodium falciparum*, is responsible for one of the most infectious diseases of the world and is quickly gaining resistance to the commonly used antimalarial treatments. New data are continually reinforcing the idea that biological functions associated with the ubiquitin proteasome system (UPS) are not just limited to non-lysomal degradation of proteins but consist of a wide array of regulatory mechanisms such as cell cycle progression, transcriptional regulation, gene expression and trafficking. While there is much effort in understanding the UPS in many eukaryotic organisms, the *Plasmodium* UPS has been relatively understudied despite its potential as a therapeutic drug target. However, *in vitro* proteasome inhibitors studies have confirmed the essentiality of the UPS in *Plasmodia* with limited toxicity to human cell lines. In addition, computational studies have shown that there are a number of ubiquitinating proteins upstream of the proteasome that may serve as parasite-specific drug targets due to their variety and divergences from other eukaryotic species. In this

review, we highlight the major findings about *Plasmodium*'s UPS and discuss its possible implications as an effective and specific antimalarial target.

Introduction

Regulated protein degradation is an essential aspect of cell signaling and development in all eukaryotic cells. While transcriptional regulation can control cell progression in a temporal manner, post-transcriptional regulation of proteins can provide faster mechanisms of positive or negative regulation in most signaling pathways. One of the main post-transcriptional regulation mechanisms in signaling and cell progression in eukaryotic cells is the ubiquitin proteasome system (UPS). The UPS is known to regulate proteolysis and has been shown to be involved in a vast array of biological processes such as cell cycle progression, transcriptional regulation, quality control of newly synthesized proteins, gene expression, cell differentiation, and trafficking [1]. Although the ubiquitination pathway is more familiarly known for its association with the proteasome, it is now well documented that mono- and poly-ubiquitination of targeted substrates can mediate a wide array of cellular processes (*e.g.* cell proliferation, cell stress response, transcription, cell death, DNA repair, intracellular trafficking, endocytosis and signal transduction) in a proteasome-independent manner [2, 3] .

Over the past decades, our increased understanding of the importance of the UPS in many different human and infectious diseases has elicited a wide interest in deciphering the components of this pathway in many living organisms. While we have only begun to investigate and understand the hundreds of genes involved in this ubiquitous pathway, it is already becoming clear that many different

components of this system can be targets for inhibition of various human diseases, with pathogen-caused diseases not being an exception.

The first success in targeting the UPS against a human disease has been demonstrated by the inhibition of the proteasome for cancer treatment. Bortezomib (Velcade; Millennium), a dipeptidyl boronic acid inhibitor of the proteasome has been the first drug approved by the FDA for the treatment of relapsed or refractory multiple myeloma [4, 5]. While one can argue that the proteasome is ubiquitous to all human cell lines, the efficacy and limited toxicity observed so far by this proteasome inhibitor in cancer patients results from the fact that rapidly dividing cells are more sensitive to the inhibitor than non-dividing cells. The characterization of several more specific components in the UPS opens the door to new therapeutic interventions. In this review we will highlight our current knowledge of the ubiquitin/proteasome system and its importance in the human malaria parasite, *Plasmodium falciparum*, and describe its potential for new antimalarial therapies.

The *Plasmodium* proteasome system

The eukaryotic proteasome that catalyses proteolysis of targeted proteins is an intricate arrangement of many different protein complexes. At its core, the 20S proteasome is a threonine peptidase complex that self-compartmentalizes to form a barrel-like structure, with the inner cavity serving as the site of proteolytic activity.

In yeast, mammals and plants, the 20S proteasome consists of seven different α and β subunits. Capping the 20S proteasome core is the ATP-dependent 19S regulator (also known as PA700), which confers substrate specificity. Together, the 20S proteasome and the 19S regulator form the larger protein complex known as the 26S proteasome [1, 6], which is the downstream effector protein complex that catalyses degradation of targeted proteins.

Protein quality control in *Plasmodium falciparum* is particularly important. Indeed, the parasite's high replication rate, the generally large protein size, the abundance of low-complexity regions within globular domains and the accumulation of thermal-stress proteins induced by fever of the host during infection present additional challenges to proper protein folding and degradation in order to evade lethal accumulation of non-functional or misfolded proteins [7]. Previous studies have shown that there are two T1 threonine peptidase systems present in the *Plasmodium*: the 20S proteasome and *P. falciparum* hsIV (PfhsIV) (PFL1465c) [8, 9]. The 20S proteasome is mostly found in eukaryotes while Pfhs1V is an orthologue to the proteasome ancestor ClpQ/hsIV generally expressed in bacteria [9-11]. However, it has been reported that there are a wide range of eukaryotes known that also encode the proteasome-like hsIV in their genome [11] (also add the Ruiz-Gonzalez, 2006 paper). PfhsIV has been localized as a soluble protein to the cytosol and has been shown to have proteolytic characteristics [12]. In prokaryotes, it is known that hsIV by itself has limited peptidase activity and

requires the hydrolysis of ATP by hsIU for enhanced peptidase activity [13]. In *Plasmodium*, it has been shown that Pfhs1V and the *Plasmodium* hsIU (PfhsIU) (PFI0355c) also interact with each other [14] though it is still unclear whether or not PfhsIV requires PfhsIU for enhanced peptidase activity. PfhsIU and PfhsIV could be excellent antimalarial targets if found to be essential (Figure A2.1). Their bacterial characteristics allow for potential drugs that would be specific to the parasite and non-toxic to human cells.

Currently, our knowledge on the *Plasmodium* proteasome is still limited. The 20S eukaryotic proteasome is expressed and enzymatically active throughout the *Plasmodium* life cycle [15-17]. So far, only one of the beta subunits of the 20S proteasome (Mal8P1.142) along with a S4 ATPase subunit (PF10_0081) known to associate with the 20S proteasome have been partially characterized [18, 19]. To further characterize the role of the *Plasmodium* proteasome, immunoblot experiments have revealed an increased abundance of ubiquitinated proteins when parasites are incubated with proteasome inhibitors [17, 20]. These results suggest a role of the parasite proteasome in the degradation of ubiquitinated substrates.

Proteasome inhibitor studies

About a decade ago, due to promising results from proteasome inhibition studies in *Trypanosomes* and *Entamoeba spp.*, lactacystin, a proteasome inhibitor derived from *Streptomyces*, was used to test the effects of proteasome inhibition in

malaria parasites [21]. Lactacystin irreversibly binds to the catalytic threonines of the β subunits of the proteasome [22]. Gantt *et al.* showed that lactacystin can inhibit *P. falciparum, Plasmodium yoelii, and Plasmodium berghei*, the rodent malaria parasites, at low micromolar concentrations. In *P. falciparum*, synchronized parasites treated with lactacystin showed growth inhibition at the time of DNA replication initiation, which suggests that lactacystin is cell cycle specific. Shortly afterwards, additional studies showed that lactacystin was also effective against chloroquine-resistant and actinomycin D-resistant parasite strains, two anti-malarial compounds [19].

Following the lactacystin studies, additional proteasome inhibitors were pursued and tested against the *Plasmodium*. For instance, gliotoxin (GTX), a fungal metabolite that has a diverse range of biological activities including proteasome inhibitory properties [23] was tested against *P. falciparum* and found to have significant *in vitro* plasmodicidal activity within the low micromolar ranges [24].

As mentioned previously, bortezomib was the first proteasome inhibitor shown to have anti-cancer activity and demonstrate tolerable toxicities in patients with multiple myeloma in clinical trials [4]. Consequently, bortezomib along with its boronate analogs MLN-273 and ZL_3B were tested against *Plasmodia* to evaluate their antimalarial properties. In *P. berghei*, MLN-273 was found to inhibit exoerythrocytic development *in vitro* [25]. In *P. falciparum*, all three boronate

compounds arrested the parasite's erythrocytic cycle with 50% inhibitory concentrations (IC_{50}) within the low nanomolar range in both wild-type and drug-resistant strains [25, 26]. Results show that these boronate protease inhibitors induce cell cycle arrest at the ring stage prior to DNA synthesis. Though no data exists for the direct effect of these compounds on the *Plasmodium* proteasome, these studies suggest an integral role for the *P. falciparum* proteasome in cell cycle, DNA synthesis and parasite development.

Salinosporamide A (Sal A), a recently discovered proteasome inhibitor initially extracted from the marine actinomycete *Salinospora tropica* [27], has been shown to inhibit growth of many human malignant cell types [28, 29]. Sal A was identified as the most potent inhibitor of dividing melanoma cells. Now entering clinical trials (Neureus Pharmaceuticals) [30] for the treatment of multiple myeloma, the drug shows a unique ability to inactivate the three proteolytic activities of the 20S proteasome subunit without affecting other proteases [27] [31]. In *P. falciparum*, Sal A was found to have an IC_{50} of 11.4nM *ex vivo*, which is within the same range of inhibition activity than the current most effective anti-malarial artemisinin [17]. A phenotypic analysis of treated and untreated cultures showed that treatment with Sal A leads to similar morphological effects than the ones observed with lactasystin- or bortezomib-related compounds and further corroborate the proteasome commitment to control the erythrocytic cycle before and after DNA synthesis [17, 25, 32-34]. The efficacy of Sal A against the malaria

parasite has been further tested *in vivo* using the parasite mouse model *P. yoelii.* Experiments have shown that *P. yoelii*-infected mice have a significant decrease in parasitemia when treated with Sal A at 130μg/kg (in comparison, treatment with artemisinin derivitives requires doses in the mg/kg range (Vivas et al. 2007) In agreement with the *in vitro* data, Sal A significantly decreased parasite growth *in vivo* at a very low concentration.

Even though highly comparable results may be achieved through various growth detection methods, factors such as assay parameters (*e.g.* initial parasitemia) and variances in *Plasmodium* laboratory isolates may affect IC_{50} values. These variable factors make it difficult to create a fair evaluation of potential anti-malarial compounds based on published results deriving from different laboratories. In order to effectively compare the plasmodicidal activity of the known proteasome inhibitors, Kreidenweiss *et al.* [35] tested a dozen proteasome inhibitors (epoxomicin, YU101, YU102, MG132, MG115, Z-L_3-VS, Ada-Ahx_3-L_3-VS, bortezomib, lactacystin, gliotoxin, PR39 and PR11) against chloroquine-susceptible strains (3D7 and D10), chloroquine-resistant strains (Dd2) and field isolates. Using a uniform drug susceptibility assay, they concluded that most of these compounds conferred inhibition against *P. falciparum* with epoxomicin as the most effective against all three parasite strains. The mode of action by which any of these proteasome inhibitors arrest *Plasmodium*'s erythrocytic development is not precisely known. However, the increased detection of ubiquitin conjugates after

drug incubation, when compared to untreated cultures, validates the likely inhibition of the *Plasmodium* proteasome complex [17, 20].

Despite the high conservation between human, yeast and *Plasmodium* proteasomes, proteasome inhibitors have effectively exhibited antimalarial properties with limited host toxicity. This observation may be linked to the fact that malaria parasites, like cancer cells, have a high rate of replication and thus are more sensitive to proteasome inhibitors. Furthermore, an amino acid (Y168G) divergence, found in the catalytic domain of the proteasome ß subunit, may possibly be exploited to increase drug specificity against the parasite proteasome [17]. Though the effectiveness of proteasome inhibitors against malaria in human clinical trials remains to be evidenced, targeting the *Plasmodium* proteasome has so far shown much promise (Figure A2.1).

Ubiquitinating enzymes of the *Plasmodium* UPS

While the proteasome complexes have been the first to be considered for drug targeting and development, other components of the UPS could be targeted in a more specific manner. As mentioned previously, proteins targeted for degradation *via* the proteasome system in eukaryotic cells are tagged with a polymer chain made up of highly-conserved 76 amino acid-peptides called ubiquitin. The reversible conjugation of ubiquitin to the lysine residues of substrate proteins is mediated by a hierarchical series of enzymes: ubiquitin activating enzyme (E1), ubiquitin-

conjugating enzymes (E2), and ubiquitin ligases (E3) [36, 37]. First, ubiquitin is both activated by and attached to E1 to form an E1-ubiquitin thioester. Next, the activated ubiquitin is transferred to E2 to form an E2-ubiquitin thioester. Then finally, in the presence of E3, which interacts with E2 and target protein substrates, ubiquitin is attached with an isopeptide bond at a lysine residue of the target protein substrate. It is important to highlight that the organization of the ubiquitin cascade is strictly hierarchical. Only a few E1s activate ubiquitin for the entire cellular ubiquitination network. These few E1s can interact with a variety of multiple E2s and potentially several hundred E3s catalyze substrate ubiquitination in a target specific manner (Figure A2.1). To further complete the complexity of the mechanism regulating the UPS, deubiquitinating enzymes (DUBs) serve to reverse ubiquitin conjugation by removing ubiquitin from substrate proteins and replenish the mono-ubiquitin pool. Several components of this pathway have been predicted to be parasite-specific enzymes or adaptor proteins that may possibly be specifically targeted by small molecule inhibitors with minimal human toxicity [38]. The following section summarizes the current knowledge of the different components of the ubiquitination pathway in *Plasmodia* and their potential for the development of new antimalarial strategies.

Ubiquitin

In eukaryotes, various forms of ubiquitin modifications have been reported, each potentially mediating a specific cellular function. For example, in the case of

poly-ubiquitination, which is the attachment of a chain of ubiquitin to a protein, since ubiquitin contains seven internal lysine residues (K6, K11, K27, K29, K33, K48 and K63) there can be at least seven potential linkages between each ubiquitin. It has been found that covalent conjugation of K48-linked poly-ubiquitin chains to protein substrates acts as signals for targeted protein degradation *via* the UPS [39]. On the other hand, poly-ubiquitination by K63 linkages seems to play important roles in DNA damage tolerances, endocytosis, ribosomal protein synthesis, and inflammatory response [40]. Over the past few years, it has been demonstrated that proteins can be mono-ubiquitinated (attachment of a single ubiquitin) or multi-ubiquitinated (attachment of individual ubiquitin to a substrate at multiple sites) but be involved in many different cellular processes such as DNA repair [41], endocytosis and trafficking [42] without any proteolytic role.

There are two known sources *de novo* of monomeric ubiquitin proteins: (i) the proteolytic cleavage of polyubiquitin proteins, and (ii) the proteolytic cleavage of ubiquitin-fusion proteins. Horrocks *et al.* [43] identified and characterized the polyubiquitin gene from *P. falciparum*. They show that the *P. falciparum* polyubiquitin gene (PfpUB) (PFL0585w) is present as a single-copy on chromosome 12 with five tandem repeats of the ubiquitin open reading frame and exhibits sequence identities of at least 94% to that of other eukaryotic species. Expression data show that steady-state transcript levels of PfpUB are expressed at all stages of the intraerythrocytic cycle, with significant elevations at the late trophozoite and

schizont stages [15, 16]. Ub_{S27a} and Ub_{L40} are two additional ubiquitin moieties that are fused to the ribosomal proteins L40 (PF13_0346) and S27a (PF14_0027) have also been identified in the *P. falciparum* genome [38, 44]. It is reported that these ubiquitin genes are expressed throughout the *P. falciparum* life cycle [15, 16]. *P. falciparum* Ub_{L40} shows high protein sequence fidelity with other eukaryotic species whereas *P. falciparum* Ub_{S27a} harbor significant divergences in the ubiquitin domain, a feature that was also found in most protists [44]. Though PfpUB and Ub_{L40} are likely poor parasite-specific drug targets due to their high conservation with human ubiquitin, the divergence found in malarial Ub_{S27a} may possibly be exploited for antimalarial drug research. However, more investigation of this ubiquitin moiety needs to be done in order to validate it as an essential drug target.

Ubiquitin activating enzymes (E1)

Eight putative E1 and E1-like enzymes have been identified in the parasite genome [38]. At the core of the ubiquitin activating enzyme domain, a high level of primary sequence identity is found when compared to other eukaryotic models. However, outside this core, sequences diverge significantly as the functional requirements for these E1 enzymes change in order to specifically interact with their respective E2 conjugating enzymes. Sequence analysis indicates the existence of *Plasmodium* E1 paralogs for UBA1 and UBA1-like proteins, which are responsible for the activation of ubiquitin. These ubiquitin activating enzymes contain a domain related to the bacterial thiamine biosynthesis protein Thif which contributes to the

ATP-binding site required for the adenylation of ubiquitin. In addition to this ATP-binding site, a conserved cysteine residue, responsible for the formation of the thiol ester with ubiquitin, is often referred as the catalytic site. Small inhibitors could target both of these domains. However, because E1 enzymes are well conserved across species, the discovery of parasite-specific inhibitors could be a challenging task.

Ubiquitin conjugating enzymes (E2)

Fourteen putative E2 paralogs were found in *P. falciparum* [38] and exhibited extensive conservation with other eukaryotic E2 proteins. Gene expression data for nine of the fourteen *P. falciparum* E2s showed a diverse pattern of steady state mRNA levels at different stages of the intraerythrocytic cycle. This pattern suggests the existence of a temporal profile of delivering ubiquitin to different E3s, highlighting a potential additional level of temporal control in ubiquitination during the parasite's life cycle.

A *P. falciparum* homolog, PfUBC13 (PFE1350c), of the E2 ubiquitin-conjugating enzyme 13 (UBC13) was characterized and revealed to be a substrate of the *P. falciparum* protein kinase PfPK9 [45]. Reverse-phase HPLC and *in vitro* ubiquitination assay show that PfPK9 phosphorylates PfUBC13 at S106 and suppresses ubiquitin-conjugating activity. Though the physiological role of PfUBC13 is unknown, the highly-conserved UBC13 (coupled with an ubiquitin E2-variant

protein) assembles K63-linked ubiquitin chains [46], which mediates non-proteolytic pathways [47]. It is reported that UBC13's conjugating activity regulates various cellular processes such as DNA repair [48], tumor suppressor p53 activity [49], mitotic progression [50], and immune receptor signaling [51]. Though E2s are essential enzymes in the UPS pathway, because of the high conservation of the active domains and the absence of defined catalytic pockets [52], E2s most likely are not suitable drug target candidates.

Ubiquitin ligating enzymes (E3*)*

Regulating the pattern of interaction with ubiquitin substrates and providing the specificity necessary in this degradation pathway, *Plasmodium* E3 ligases are found to be highly divergent and the most abundant in the *Plasmodium* genome compared to E1s and E2s. Within the 54 putative E3 ligases found in *P. falciparum*, all superfamilies (HECT, RING and U-box and cullin) of E3 ligases are represented with E3 RING finger proteins making up the majority of the *Plasmodium* E3 ligases [38]. Functional annotation analysis reveals that *Plasmodium* E3 ligases have a wide-array of potential roles including cell cycle regulation, trafficking, DNA repair, chromatin structure, and mRNA transport. While the functional analysis of these putative E3s will need to be further validated *in vivo*, three parasite-specific E3 ligases have already shown *in vitro* ubiquitination activity (Chung D. and Le Roch K., unpublished data). Furthermore, knock out experiments strongly suggest that

these proteins are essential to the parasite erythrocytic cycle (Chung D. and Le Roch K., unpublished data).

Because E3 ligases are highly divergent, numerous in kind and ubiquitinate substrates in a target-specific manner, creating small molecule inhibitors against these E3 ligases may be one of the most specific ways to interfere with the parasite UPS pathway. Recent studies between the human MDM2 RING finger protein and its P53 target indicate that the creation of small molecule inhibitors against E3 ligases can be successfully achieved [53].If the synthesis of small inhibitors against *Plasmodium* ligases can be achieved, this may possibly lead to a parasite-specific therapeutic intervention that will eliminate the risk of side effects in the human host.

Deubiquitinating enzymes (DUBS)

While E3 ligases can catalyze the attachment of ubiquitin to proteins, DUBs acts to remove ubiquitin [54]. Depending on the computational tool employed, 18 or 29 *Plasmodium* DUBs were identified [38, 55]. Collectively, these searches were able to identify DUB enzymes from all five gene families of DUBs: the ubiquitin C-terminal hydrolases (UCHs); the ubiquitin-specific peptidases (USPs/UBPs); the ovarian tumor (OTU) domain proteins; the Josephin or Machado-Joseph disease (MJD) proteins and the JAMM (Jab1/MPN domain-associated metalloisopeptidase) domain proteins. PfUCH54 (PF14_0576) was the first DUB identified and

characterized in *P. falciparum* [56]. Using electrophilic probes that detect enzymes capable of removing ubiquitin and other ubiquitin-like proteins, PfUCH54 was found to possess both deubiquitination and deneddylation activity. Using known active site residues and crystal structure of homologous DUBs, PfUCH54 was found to have high homology to UCHL3 [56], an enzyme that has been reported to also have dual deubiquitinating and deneddylating activity [57]. Though the function of PfUCH54 is not known, mouse UCHL3 is required to maintain a stable apical membrane epithelial sodium channel, facilitating the dynamic recycling of sodium channels at the apical surface [58]. Though not as abundant as E3 ligases, because of their divergent properties and their known roles in regulating proper cellular functions, DUBs may also serve as an effective drug target candidate (Figure A2.1).

Concluding remarks

Investigators are continually discovering that the UPS plays an integral role in a vast and diverse range of biological functions within many eukaryotic model systems. Though there is still relatively little information about the *Plasmodium* UPS, our current knowledge suggests that the malarial UPS is also a key player in the biological maintenance and propagation of this deadly parasite. For instance, the essentiality of the malarial 26S proteasome to the parasite has been proven multiple times with the use of different proteasome inhibitor assays, achieving parasite inhibitory concentrations within the nanomolar range. Thus the malarial proteasome would be, so far, the most conventional drug target in the malarial UPS pathway. The conserved sequence homology between parasite and human host proteasome proteins can be overcome by the fact that the parasite cells divide more rapidly than host cells.

Moreover, the Y168G divergence found at the β subunit of the catalytic site of the malarial proteasome may give rise to a parasite-specific drug design and limit potential side effects. Finally, the pfhsIU/pfhsIV proteasome may also present itself as a possible drug target with high specificity because of its bacterial nature. However, the essentiality of this bacterial proteasome within the *Plasmodium* has yet to be tested.

Among all components of the UPS, E3 ligases are the most abundant and diverse enzymes in the *Plasmodium* UPS pathway, followed by DUBs. *Plasmodium* E3 ligases would certainly have a high amount of selectivity for drug discovery. Finding a small molecule inhibitor against the unconventional *Plasmodium* E3 ligases will not be an easy undertaking. On the other hand, an increased understanding of the role and essentiality of parasite-specific E3s will significantly reduce the difficulty of this task. A combination of different global mass spectrometry and functional genomic approaches to reveal the parasite ubiquitome, along with target validation experiments, will certainly make a major impact in the field.

All in all, despite being a relatively new and understudied field, the *Plasmodium* ubiquitin/proteasome system presents itself as an excellent host of multiple drug targets due to having both highly specific and vitally essential enzymes within its pathway. Only with further investigation and heightened effort can we both determine and utilize the malarial ubiquitin/proteasome system as viable antimalarial targets.

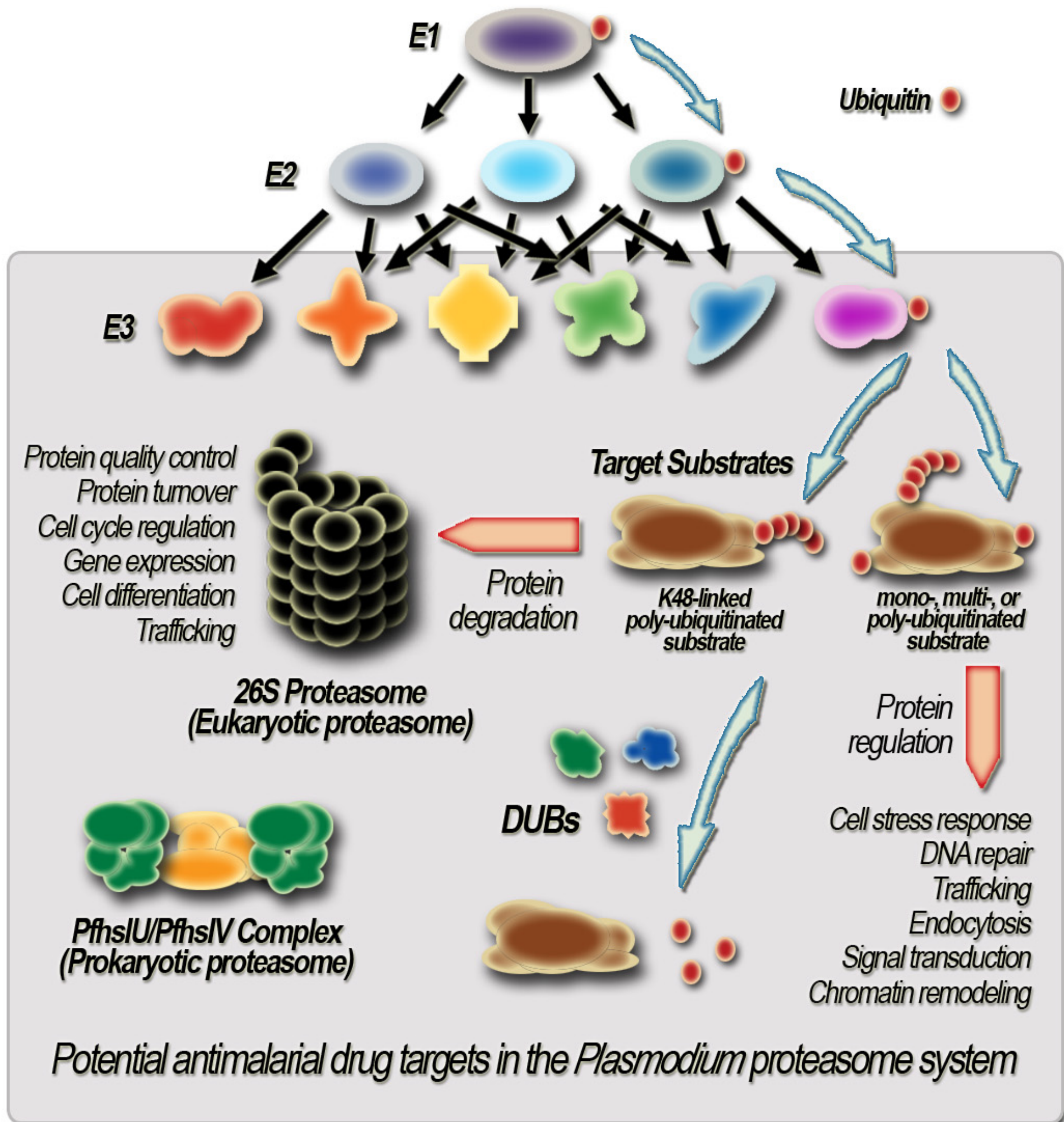

E1
Ubiquitin
E2
E3
Protein quality control
Protein turnover
Cell cycle regulation
Gene expression
Cell differentiation
Trafficking
Target Substrates
Protein degradation
K48-linked poly-ubiquitinated substrate
mono-, multi-, or poly-ubiquitinated substrate
26S Proteasome
(Eukaryotic proteasome)
Protein regulation
DUBs
Cell stress response
DNA repair
Trafficking
Endocytosis
Signal transduction
Chromatin remodeling
PfhslU/PfhslV Complex
(Prokaryotic proteasome)
Potential antimalarial drug targets in the Plasmodium proteasome system

Figure legend:

Figure A2.1: Graphical depiction of the possible drug targets within the *Plasmodium* UPS. **26S proteasome:** proteasome inhibitor assays have validated the 26S proteasome as an excellent drug target, conferring parasite inhibition in the nM range with limited host cell toxicity. **PfhslU/pfhslV proteasome complex:** because of its bacterial characteristics, this complex may provide a parasite-specific drug target, though the essentiality of this prokaryotic proteasome still needs to be validated. **E3 ligases and E3 target substrates**: as the most abundant and diverse group of enzymes within the UPS, E3 ligases potentially can provide multiple drug targets that are unique to *Plasmodium*. In addition, their protein target substrates may also potentially be excellent drug targets because they lead to a variety of cellular functions. **DUBs**: similar to the E3 ligases, in terms of their relative divergence and variety, DUBs present themselves as potential drug targets.

References

[1]. Jung, T.; Catalgol, B.; Grune, T. The proteasomal system. *Mol Aspects Med*, **2009**. 30(4): p. 191-296.

[2]. Hershko, A.; Ciechanover, A. The ubiquitin system. *Annu Rev Biochem*, **1998**. 67: p. 425-79.

[3]. Chen, Z.J.; Sun, L.J. Nonproteolytic functions of ubiquitin in cell signaling. *Mol Cell*, **2009**. 33(3): p. 275-86.

[4]. Roccaro, A.M.; Hideshima, T.; Richardson, P.G.; Russo, D.; Ribatti, D.; Vacca, A.; Dammacco, F.; Anderson, K.C. Bortezomib as an antitumor agent. *Curr Pharm Biotechnol*, **2006**. 7(6): p. 441-8.

[5]. Adams, J.; Palombella, V.J.; Sausville, E.A.; Johnson, J.; Destree, A.; Lazarus, D.D.; Maas, J.; Pien, C.S.; Prakash, S.; Elliott, P.J. Proteasome inhibitors: a novel class of potent and effective antitumor agents. *Cancer Res*, **1999**. 59(11): p. 2615-22.

[6]. Besche, H.C.; Peth, A.; Goldberg, A.L. Getting to first base in proteasome assembly. *Cell*, **2009**. 138(1): p. 25-8.

[7]. Kreidenweiss, A.; Kremsner, P.; Mordmuller, B., *Comprehensive study of proteasome inhibitors against Plasmodium falciparum laboratory strains and field isolates from Gabon*. 2008. p. 187.

[8]. Gille, C.; Goede, A.; Schloetelburg, C.; Preissner, R.; Kloetzel, P.M.; Gobel, U.B.; Frommel, C. A comprehensive view on proteasomal sequences: implications for the evolution of the proteasome. *J Mol Biol*, **2003**. 326(5): p. 1437-48.

[9]. Mordmüller, B.; Fendel, R.; Kreidenweiss, A.; Gille, C.; Hurwitz, R.; Metzger, W.G.; Kun, J.F.J.; Lamkemeyer, T.; Nordheim, A.; Kremsner, P.G. Plasmodia express two threonine-peptidase complexes during asexual development. *Molecular and Biochemical Parasitology*, **2006**. 148(1): p. 79-85.

[10]. Missiakas, D.S., F; Betton, JM; Georgopoulos, C; Raina, S Identification and characterization of HsIV HsIU (ClpQ ClpY) proteins involved in overall proteolysis of misfolded proteins in Escherichia coli. *Embo J*, **1996**. 15(24): p. 6899-6909.

[11]. Gille, C.; Goede, A.; Schlöetelburg, C.; Preißner, R.; Kloetzel, P.-M.; Göbel, U.B.; Frömmel, C. A Comprehensive View on Proteasomal Sequences: Implications

for the Evolution of the Proteasome. *Journal of Molecular Biology*, **2003**. 326(5): p. 1437-1448.

[12]. Ramasamy, G.; Gupta, D.; Mohmmed, A.; Chauhan, V.S. Characterization and localization of Plasmodium falciparum homolog of prokaryotic ClpQ/HslV protease. *Mol Biochem Parasitol*, **2007**. 152(2): p. 139-48.

[13]. Yoo, S.J.; Seol, J.H.; Shin, D.H.; Rohrwild, M.; Kang, M.S.; Tanaka, K.; Goldberg, A.L.; Chung, C.H. Purification and characterization of the heat shock proteins HslV and HslU that form a new ATP-dependent protease in Escherichia coli. *J Biol Chem*, **1996**. 271(24): p. 14035-40.

[14]. Subramaniam, S.; Mohmmed, A.; Gupta, D. Molecular modeling studies of the interaction between Plasmodium falciparum HslU and HslV subunits. *J Biomol Struct Dyn*, **2009**. 26(4): p. 473-9.

[15]. Le Roch, K.G.; Zhou, Y.; Blair, P.L.; Grainger, M.; Moch, J.K.; Haynes, J.D.; De La Vega, P.; Holder, A.A.; Batalov, S.; Carucci, D.J.; Winzeler, E.A. Discovery of gene function by expression profiling of the malaria parasite life cycle. *Science*, **2003**. 301(5639): p. 1503-8.

[16]. Bozdech, Z.; Llinas, M.; Pulliam, B.L.; Wong, E.D.; Zhu, J.; DeRisi, J.L. The transcriptome of the intraerythrocytic developmental cycle of Plasmodium falciparum. *PLoS Biol*, **2003**. 1(1): p. E5.

[17]. Prudhomme, J.; McDaniel, E.; Ponts, N.; Bertani, S.p.; Fenical, W.; Jensen, P.; Le Roch, K. Marine Actinomycetes: A New Source of Compounds against the Human Malaria Parasite. *PLoS ONE*, **2008**. 3(6): p. e2335.

[18]. Li, G.D.; Li, J.L.; Mugthin, M.; Ward, S.A. Molecular cloning of a gene encoding a 20S proteasome beta subunit from Plasmodium falciparum. *Int J Parasitol*, **2000**. 30(6): p. 729-33.

[19]. Certad, G.; Abrahem, A.; Georges, E. Cloning and partial characterization of the proteasome S4 ATPase from Plasmodium falciparum. *Exp Parasitol*, **1999**. 93(3): p. 123-31.

[20]. Lindenthal, C.; Weich, N.; Chia, Y.S.; Heussler, V.; Klinkert, M.Q., *The proteasome inhibitor MLN-273 blocks exoerythrocytic and erythrocytic development of Plasmodium parasites*. 2005, Cambridge Journals Online. p. 37-44.

[21]. Gantt, S.M.; Myung, J.M.; Briones, M.R.; Li, W.D.; Corey, E.J.; Omura, S.; Nussenzweig, V.; Sinnis, P. Proteasome inhibitors block development of Plasmodium spp*. Antimicrob Agents Chemother*, **1998**. 42(10): p. 2731-8.

[22]. Fenteany, G.; Standaert, R.F.; Lane, W.S.; Choi, S.; Corey, E.J.; Schreiber, S.L. Inhibition of proteasome activities and subunit-specific amino-terminal threonine modification by lactacystin*. Science*, **1995**. 268(5211): p. 726-31.

[23]. Kroll, M.; Arenzana-Seisdedos, F.; Bachelerie, F.; Thomas, D.; Friguet, B.; Conconi, M. The secondary fungal metabolite gliotoxin targets proteolytic activities of the proteasome*. Chem Biol*, **1999**. 6(10): p. 689-98.

[24]. Hatabu, T.; Hagiwara, M.; Taguchi, N.; Kiyozawa, M.; Suzuki, M.; Kano, S.; Sato, K. Plasmodium falciparum: the fungal metabolite gliotoxin inhibits proteasome proteolytic activity and exerts a plasmodicidal effect on P. falciparum*. Exp Parasitol*, **2006**. 112(3): p. 179-83.

[25]. Lindenthal, C.; Weich, N.; Chia, Y.S.; Heussler, V.; Klinkert, M.Q. The proteasome inhibitor MLN-273 blocks exoerythrocytic and erythrocytic development of Plasmodium parasites*. Parasitology*, **2005**. 131(Pt 1): p. 37-44.

[26]. Reynolds, J.M.; El Bissati, K.; Brandenburg, J.; Gunzl, A.; Mamoun, C.B. Antimalarial activity of the anticancer and proteasome inhibitor bortezomib and its analog ZL3B*. BMC Clin Pharmacol*, **2007**. 7: p. 13.

[27]. Dick, L.R.; Cruikshank, A.A.; Grenier, L.; Melandri, F.D.; Nunes, S.L.; Stein, R.L. Mechanistic studies on the inactivation of the proteasome by lactacystin: a central role for clasto-lactacystin beta-lactone*. J Biol Chem*, **1996**. 271(13): p. 7273-6.

[28]. Williams, P.G.; Buchanan, G.O.; Feling, R.H.; Kauffman, C.A.; Jensen, P.R.; Fenical, W. New cytotoxic salinosporamides from the marine Actinomycete Salinispora tropica*. J Org Chem*, **2005**. 70(16): p. 6196-203.

[29]. Feling, R.H.; Buchanan, G.O.; Mincer, T.J.; Kauffman, C.A.; Jensen, P.R.; Fenical, W. Salinosporamide A: a highly cytotoxic proteasome inhibitor from a novel microbial source, a marine bacterium of the new genus salinospora*. Angew Chem Int Ed Engl*, **2003**. 42(3): p. 355-7.

[30]. Fenical, W.; Jensen, P.R.; Palladino, M.A.; Lam, K.S.; Lloyd, G.K.; Potts, B.C. Discovery and development of the anticancer agent salinosporamide A (NPI-0052)*. Bioorg Med Chem*, **2009**. 17(6): p. 2175-80.

[31]. Adams, J.; Kauffman, M. Development of the proteasome inhibitor Velcade (Bortezomib). *Cancer Invest*, **2004**. 22(2): p. 304-11.

[32]. Gonzalez, J.; Ramalho-Pinto, F.J.; Frevert, U.; Ghiso, J.; Tomlinson, S.; Scharfstein, J.; Corey, E.J.; Nussenzweig, V. Proteasome activity is required for the stage-specific transformation of a protozoan parasite. *J Exp Med*, **1996**. 184(5): p. 1909-18.

[33]. Gonzalez, J.; Bai, G.; Frevert, U.; Corey, E.J.; Eichinger, D. Proteasome-dependent cyst formation and stage-specific ubiquitin mRNA accumulation in Entamoeba invadens. *Eur J Biochem*, **1999**. 264(3): p. 897-904.

[34]. de Diego, J.L.; Katz, J.M.; Marshall, P.; Gutierrez, B.; Manning, J.E.; Nussenzweig, V.; Gonzalez, J. The ubiquitin-proteasome pathway plays an essential role in proteolysis during Trypanosoma cruzi remodeling. *Biochemistry*, **2001**. 40(4): p. 1053-62.

[35]. Kreidenweiss, A.; Kremsner, P.G.; Mordmuller, B. Comprehensive study of proteasome inhibitors against Plasmodium falciparum laboratory strains and field isolates from Gabon. *Malar J*, **2008**. 7: p. 187.

[36]. Laney, J.D.; Hochstrasser, M. Substrate targeting in the ubiquitin system. *Cell*, **1999**. 97(4): p. 427-30.

[37]. Pickart, C.M. Back to the future with ubiquitin. *Cell*, **2004**. 116(2): p. 181-90.

[38]. Ponts, N.; Yang, J.; Chung, D.-W.D.; Prudhomme, J.; Girke, T.; Horrocks, P.; Le Roch, K.G. Deciphering the Ubiquitin-Mediated Pathway in Apicomplexan Parasites: A Potential Strategy to Interfere with Parasite Virulence. *PLoS ONE*, **2008**. 3(6): p. e2386.

[39]. Kerscher, O.; Felberbaum, R.; Hochstrasser, M. Modification of proteins by ubiquitin and ubiquitin-like proteins. *Annu Rev Cell Dev Biol*, **2006**. 22: p. 159-80.

[40]. Mukhopadhyay, D.; Riezman, H. Proteasome-Independent Functions of Ubiquitin in Endocytosis and Signaling. *Science*, **2007**. 315(5809): p. 201-205.

[41]. Hofmann, K. Ubiquitin-binding domains and their role in the DNA damage response. *DNA Repair*, **2009**. In Press, Corrected Proof.

[42]. Hicke, L. Protein regulation by monoubiquitin. *Nat Rev Mol Cell Biol*, **2001**. 2(3): p. 195-201.

[43]. Horrocks, P.; Newbold, C.I. Intraerythrocytic polyubiquitin expression in Plasmodium falciparum is subjected to developmental and heat-shock control. *Mol Biochem Parasitol*, **2000**. 105(1): p. 115-25.

[44]. Catic, A.; Ploegh, H.L. Ubiquitin--conserved protein or selfish gene? *Trends Biochem Sci*, **2005**. 30(11): p. 600-4.

[45]. Philip, N.; Haystead, T.A. Characterization of a UBC13 kinase in Plasmodium falciparum. *Proc Natl Acad Sci U S A*, **2007**. 104(19): p. 7845-50.

[46]. Hofmann, R.M.; Pickart, C.M. Noncanonical MMS2-encoded ubiquitin-conjugating enzyme functions in assembly of novel polyubiquitin chains for DNA repair. *Cell*, **1999**. 96(5): p. 645-53.

[47]. Pickart, C.M. Ubiquitin enters the new millennium. *Mol Cell*, **2001**. 8(3): p. 499-504.

[48]. Brusky, J.; Zhu, Y.; Xiao, W. UBC13, a DNA-damage-inducible gene, is a member of the error-free postreplication repair pathway in Saccharomyces cerevisiae. *Curr Genet*, **2000**. 37(3): p. 168-74.

[49]. Laine, A.; Topisirovic, I.; Zhai, D.; Reed, J.C.; Borden, K.L.; Ronai, Z. Regulation of p53 localization and activity by Ubc13. *Mol Cell Biol*, **2006**. 26(23): p. 8901-13.

[50]. Bothos, J.; Summers, M.K.; Venere, M.; Scolnick, D.M.; Halazonetis, T.D. The Chfr mitotic checkpoint protein functions with Ubc13-Mms2 to form Lys63-linked polyubiquitin chains. *Oncogene*, **2003**. 22(46): p. 7101-7.

[51]. Yamamoto, M.; Okamoto, T.; Takeda, K.; Sato, S.; Sanjo, H.; Uematsu, S.; Saitoh, T.; Yamamoto, N.; Sakurai, H.; Ishii, K.J.; Yamaoka, S.; Kawai, T.; Matsuura, Y.; Takeuchi, O.; Akira, S. Key function for the Ubc13 E2 ubiquitin-conjugating enzyme in immune receptor signaling. *Nat Immunol*, **2006**. 7(9): p. 962-70.

[52]. Winn, P.J.; Religa, T.L.; Battey, J.N.; Banerjee, A.; Wade, R.C. Determinants of functionality in the ubiquitin conjugating enzyme family. *Structure*, **2004**. 12(9): p. 1563-74.

[53]. Canner, J.A.; Sobo, M.; Ball, S.; Hutzen, B.; DeAngelis, S.; Willis, W.; Studebaker, A.W.; Ding, K.; Wang, S.; Yang, D.; Lin, J. MI-63: A novel small-molecule

inhibitor targets MDM2 and induces apoptosis in embryonal and alveolar rhabdomyosarcoma cells with wild-type p53. *Br J Cancer*. 101(5): p. 774-781.

[54]. Singhal, S.; Taylor, M.C.; Baker, R.T. Deubiquitylating enzymes and disease. *BMC Biochem*, **2008**. 9 Suppl 1: p. S3.

[55]. Ponder, E.L.; Bogyo, M. Ubiquitin-like modifiers and their deconjugating enzymes in medically important parasitic protozoa. *Eukaryot Cell*, **2007**. 6(11): p. 1943-52.

[56]. Artavanis-Tsakonas, K.; Misaghi, S.; Comeaux, C.A.; Catic, A.; Spooner, E.; Duraisingh, M.T.; Ploegh, H.L. Identification by functional proteomics of a deubiquitinating/deNeddylating enzyme in Plasmodium falciparum. *Mol Microbiol*, **2006**. 61(5): p. 1187-95.

[57]. Wada, H.; Kito, K.; Caskey, L.S.; Yeh, E.T.; Kamitani, T. Cleavage of the C-terminus of NEDD8 by UCH-L3. *Biochem Biophys Res Commun*, **1998**. 251(3): p. 688-92.

[58]. Butterworth, M.B.; Edinger, R.S.; Ovaa, H.; Burg, D.; Johnson, J.P.; Frizzell, R.A. The deubiquitinating enzyme UCH-L3 regulates the apical membrane recycling of the epithelial sodium channel. *J Biol Chem*, **2007**. 282(52): p. 37885-93.

APPENDIX A3:

Deciphering the Ubiquitin-Mediated Pathway in Apicomplexan Parasites: A Potential Strategy to Interfere with Parasite Virulence *PLoS ONE* **3:** e2386 (2008)

Nadia Ponts[1], Jianfeng Yang[1], **Duk-Won Doug Chung[1]**, Jacques Prudhomme[1], Thomas Girke[2], Paul Horrocks[3], Karine G. Le Roch[1]

[1]Department of Cell Biology and Neurosciences, University of California at Riverside, Riverside, California, United States of America, [2]Center for Plant Cell Biology (CEPCEB), University of California at Riverside, Riverside, California, United States of America, [3]Department of Medicine, Institute for Science and Technology in Medicine, Keele University, Keele, United Kingdom, 4 Department of Life Sciences, Institute for Science and Technology in Medicine, Keele University, Keele, United Kingdom

Abstract

Background: Reversible modification of proteins through the attachment of ubiquitin or ubiquitin-like modifiers is an essential post-translational regulatory mechanism in eukaryotes. The conjugation of ubiquitin or ubiquitin-like proteins has been demonstrated to play roles in growth, adaptation and homeostasis in all eukaryotes, with perturbation of ubiquitin- mediated systems associated with the pathogenesis of many human diseases, including cancer and neurodegenerative disorders.

Methodology/Principal Findings: Here we describe the use of an HMM search of functional Pfam domains found in the key components of the ubiquitin-mediated pathway necessary to activate and reversibly modify target proteins in eight

apicomplexan parasitic protozoa for which complete or late-stage genome projects exist. In parallel, the same search was conducted on five model organisms, single-celled and metazoans, to generate data to validate both the search parameters employed and aid paralog classification in Apicomplexa. For each of the 13 species investigated, a set of proteins predicted to be involved in the ubiquitylation pathway has been identified and demonstrates increasing component members of the ubiquitylation pathway correlating with organism and genome complexity. Sequence homology and domain architecture analyses facilitated prediction of apicomplexan-specific protein function, particularly those involved in regulating cell division during these parasite's complex life cycles.

Conclusions/Significance: This study provides a comprehensive analysis of proteins predicted to be involved in the apicomplexan ubiquitin-mediated pathway. Given the importance of such pathway in a wide variety of cellular processes, our data is a key step in elucidating the biological networks that, in part, direct the pathogenicity of these parasites resulting in a massive impact on global health. Moreover, apicomplexan-specific adaptations of the ubiquitylation pathway may represent new therapeutic targets for much needed drugs against apicomplexan parasites.

Introduction

Apicomplexans are obligate protozoa intracellular parasites responsible for several major human diseases prevalent in the developing world. These include organisms belonging to the genera *Plasmodium*, *Toxoplasma* and *Cryptosporium*. *Toxoplasma gondii* and *Cryptosporium parvum* are the etiological agents of toxoplasmosis and cryptosporidiosis, respectively, which are predominantly opportunistic infectious agents responsible for severe mortality amongst immuno-suppressed patients such as those infected with HIV. The human malarial parasite *Plasmodium falciparum*, which is responsible for over a million deaths annually [1], is perhaps the most significant apicomplexan parasitic organism. The global impact, both in terms of mortality and morbidity, of apicomplexan parasites is currently on the rise, principally due to the increase of drug resistant strains. For example, *P. falciparum* has evolved resistance to many front-line antimalarial drugs [2] and with apparently limited prospects in the delivery of new safe, effective and cheap antimalarial drugs, little immediate respite is likely. There is clearly an urgent need to characterize and validate new drug targets, effective not only against *P. falciparum* but other apicomplexan parasites as well.

Genome sequencing projects are available for several apicomplexan parasites, with many of them completed. The full genome sequence of the human malarial parasite *P. falciparum* and the rodent malaria parasites *P. yoelii*, *P. berghei* and *P. chabaudi* have been published [3,4], with that of the human malarial parasite *P.*

vivax well underway. In addition, the complete annotated genomes of *T. gondii*, *C. parvum* and *Cryptosporium hominis* have been recently released [5,6]. Post genomic technologies, such as comparative bioinformatic approaches, global microarray and proteome analyses have created a vast amount of information pertaining to gene and protein sequence/structure prediction, interspecies identification of ortholog or paralog genes as well as temporal and developmentally associated patterns of mRNA and protein accumulation [7–13]. Together these studies have greatly advanced our understanding of gene expression throughout these parasites' complex life cycles in various host cells and insect vectors. Moreover, comparative analyses may provide key data regarding protein networks, and their potential as novel drug targets. For example, apicomplexan parasites, unlike higher eukaryotes, utilize the non-mevalonate pathway to synthesize isoprenoids [14]. Inhibitors of one of the key initial enzymes in this pathway, 1-deoxy-D-xylulose 5-phosphate, such as the herbicide fosmidomycin in combination with clindamycin, are currently being evaluated for treatment of uncomplicated *P. falciparum* malaria [15]. These data indicate the importance of comparative genomics in evaluating the potential for novel drug targets in apicomplexan parasites.

Here we describe a comparative analysis of one of the essential post-translational regulatory networks commonly found in eukaryotic cells–the ubiquitin/proteasome system (UPS). Modification of proteins via covalent conjugation to ubiquitin (or more often polyubiquitin chains) is a well-established

signal for proteosomal destruction [16]. In the early 1980s, the key role of ubiquitin in the selective pathway for degradation of proteins was demonstrated, which was followed over the next two decades by additional roles in a wide range of cellular processes. In addition to ubiquitin, ubiquitin-like proteins (UBLps) have also been identified as modifiers of cellular processes [17,18]. Together, ubiquitin and UBLps provide a reversible modification that regulates a wide range of cellular activities including DNA repair, transcription, cellular division, endocytosis, intracellular trafficking and the immune response. Importantly, defects in this pathway are associated with human diseases, including cancer and neurodegenerative disorders such as Parkinson's disease. By targeting disease-specific components of the UPS, several potential new drugs for cancer and neurodegenerative are currently under development. The potential to chemically target the UPS in the treatment of P. falciparum has been established [19–21]. However, this work focuses on the inhibition of the proteasome and the therapeutic window between the apicomplexa and the host proteasomes may be limited. Inhibition of apicomplexan-specific components of the enzymatic cascade that process, activate and transfer ubiquitin and UBLps to their various protein targets may offer attractive alternative targets.

Specificity in the conjugation of ubiquitin and ubls to their final target is elegantly achieved via an activation and transfer cascade [22] (Figure A3.1). Ubiquitin-activating enzymes (termed E1) exist for ubiquitin and each UBLp. These typically adenylate the terminal glycine residue of ubiquitin/UBLp and transfers it

to an internal cysteine residue with the formation of a thioester bond. The activated ubiquitin/UBLp is trans-esterified to an ubiquitin conjugating protein (termed E2). Whereas several E2 proteins are capable of accepting an activated ubiquitin molecule, typically only one E2 exists for each of the UBLps characterized thus far. Finally, ubiquitin ligases (termed E3) catalyze the transfer of ubiquitin/ubl from E2 to a lysine side chain on a specific target protein (this may occur directly or indirectly via conjugation to the E3) to form an isopeptide bond. Since ubiquitin contains several lysine residues, it can itself be ubiquitinylated, leading to the formation of polyubiquitin chains. Differences in affinity for ubiquitin/UBLp by the component parts of the cascade, as well as a hierarchical increase in the numbers of these proteins (e.g. there is one E1 for ubiquitin, several E2s and an increasing number of characterized E3s), drive the transfer of ubiquitin/UBLp through the cascade with the final target specificity mediated through the E3 complex.

Ubiquitin, a highly conserved 76 amino acid peptide, was first described in 1974 [16]. From the late 1970's onwards, a number of UBLps have also been described. These proteins do not share extensive primary sequence homology with ubiquitin, but rather share a common tertiary structure (the ubiquitin fold) and activation/conjugation mechanism through variant E1, E2 and E3 proteins. To date, within mammalian systems, over 10 UBLps have been described, including interferon-stimulated gene 15 (ISG-15), neuronal precursor cell expressed developmentally down regulated 8 (NEDD8), and small ubiquitin-related modifier

(SUMO) [23].

Analysis of the E1-activating enzymes indicates that they share sequence homology to MoeB/ThiF domains of prokaryotic biosynthetic proteins involved in sulfur donor systems [24]. These proteins similarly rely on an initial adenylation of a peptide with a C-teminal diglycine motif. E1 proteins either have two MoeB/ThiF domains necessary for the adenylation and subsequent internal transfer to form a thiolester bond or are a complex of two heterodimers that each contains one MoeB/ThiF domain. Selection of an E2 for transfer the UBL modifier via a transesterification reaction relies on additional motifs present in E1. The E2 protein contains a single motif that mediates interaction with both E1 and E3, signifying the ‘‘shuttle’’ status of E2 in the transfer of ubiquitin/UBLps between activation and subsequent ligation to their final target. E2s are present as multiple isoforms, each with distinct roles. E2s exist for each UBL modifier, with multiple E2s capable of accepting ubiquitin. However, even within the ubiquitin E2 isoforms, there is functional divergence in the specific E3s they interact with, and thus the cellular processes they are involved in. For example, Rad6p and Cdc34p E2 isoforms deliver ubiquitin to E3s that ultimately target proteins involved in the regulation of DNA repair and cell cycle progression, respectively [25].

E3 ubiquitin ligases are very diverse. They have been classified into three main classes according to the presence of specific domain motifs: Homologous to E6-

associated protein C-terminus (HECT), Really Interesting New Proteins (RING, e.g. MDM2 known to target p53) and U-box. Two sub-classes of RINGs have further been defined: RING in between RING-RING (RIR) and Cullin-RING ligases (CRL), which are multi-protein complex E3s (see [26] for a review). These CRLs are associated with proteins carrying F-box domains, which are involved in substrate recognition (see [27] for a review). The CRL anaphase-promoting complex, involved in cell cycle progression, is a typical example of an SCF-type ligase (Skp1-Cullin-Fbox). Except for the HECT family that has a direct role in catalyzing ubiquitylation, E3s are adaptors molecules that bring the E2 enzyme and the target substrate into close proximity to promote ubiquitylation. The RING finger family represents the largest group of E3s and is characterized by a cysteine/histidine-rich/zinc chelating domain that specifically promotes protein-protein interaction, as well as protein-DNA binding. In eukaryotes, RING fingers have been shown to be the key regulator of polyubiquitylation and protein degradation. However, they have also been shown to play a pivotal role in monoubiquitylation of substrates, independent of degradation. Monoubiquitylation has been shown to regulate events such as the endocytosis of cell receptors (e.g. the ring finger c-Cbl is required for the endocytosis of the Epidermal Growth factor Receptor, EGFR), DNA-repair (ubiquitylation of p53 by the RING finger MDM2) and transcriptional regulation (activation of NF-kB by the RING finger TRAF6).

Conversely, de-ubiquitinylation enzymes (deubiquitinases or DUBs)

specifically remove ubiquitin/UBLps. DUBs are a large group of cysteine proteases or zinc-dependent metalloproteases that specifically cleave after the terminal carbonyl of the last residue of ubiquitin adducts. Compared to the proteins involved in the activation, conjugation and ligation of ubiquitin/UBLps relatively little is known about the functional role of DUBs. However, evidence suggests that DUBs are key regulators of the ubiquitin system; DUBs are functionally similar to protein phosphatases in the phosphorylation system. Based on their sequences similarities, structural studies and potential mechanism of action, DUBs fall into at least six distinct subfamilies: the ubiquitin C-terminal hydrolases (UCH-Peptida- se_C12), the ubiquitin specific proteases (USP-UCH), otubains (OTU), the ataxin-3/Josephin ubiquitin protease (MJD), the JAMM isopeptidase (Mov34) and the recent *in silico* prediction of the permuted papain fold peptidase (PPPDE) [28,29]. In addition to these DUBs subfamilies, three distinct families of deubiquitinating- like enzymes (DUBLs) are detected in eukaryotes: the SUMO- specific proteases (SENPs-Peptidase_C48), the autophagins (Peptidase_C54), and the newly predicted WLM family of zinc-dependant peptidases (WLM) mostly found in plants and fungi but apparently absent in animals [29].

The UPS is known to play important roles in modulation of immune and inflammatory responses. Deregulation of the UPS can lead to the development of inflammatory and autoimmune diseases, such as inflammatory arthritis, psoriasis, allergy and asthma (see [30] for review). Proteasome inhibitors have been

developed as therapeutic molecules, principally as anticancer drugs [31–34]. In the context of host-pathogen interactions, both bacteria and viruses were shown to use components of their UPS as virulence factors. The E3 ubiquitin ligase from Pseudomonas syringae has been shown to induce sensitivity in tomato plants by targeting a host kinase, Fen, to the proteasome, which leads to the inhibition of the Fen-activated immunity-associated programmed cell death [35]. The DUB SseL (from Salmonella enterica), which causes gastroenteritis in humans, has similarly been implicated in its virulence [36]. Components of the UPS have been shown to be involved in many aspects of viral pathogenesis (see [37] for a review). Two RING-finger E3 ubiquitin ligases, K3 and K5, from herpes virus promote immune evasion by targeting MHC class 1 to ubiquitylation and endolysosomal degradation. The human papillomavirus E6 protein interacts with the cellular E3 ubiquitin ligase E6-associated protein. This complex mediates the proteasome-dependant degradation of the key tumor suppressor protein p53. DUBs have also been shown to be involved in viral pathogenesis. In Epstein-Barr virus infection of B cells, a group of cellular DUBs are activated, which include UCH-L1 and UCH- L5. In adenovirus infection, the viral proteinase L3 23K is responsible for the cleavage of viral precursor polyproteins, and may function as a DUB [38].

These data implicate the UPS in roles from colonization, infection, immune evasion and virulence for a range of pathogens. To date, potential roles for the UPS in mediating similar roles for apicomplexan parasites have yet to be explored. Here

we describe an *in silico* proteomic analysis of UPS from eight apicomplexan parasites: *P. falciparum*, *P. vivax*, *P. yoelii*, *P. berghei*, *P. chabaudi*, *C. parvum*, *C. hominis*, and *T. gondii*. Five other eukaryotic model organisms, including *Homo sapiens*, *Saccharomyces cerevisiae*, *Caenorhabditis elegans*, *Drosophila melanogaster* and *Arabidopsis thaliana*, were analyzed in parallel for comparative purposes. We aim here to identify and describe the most complete ubiquitylation pathway in apicomplexan parasites, with a particular focus on *Plasmodium falciparum*, highlighting those components that are specific to apicomplexan parasites. Our results open new research perspectives and are expected to pilot the development of new strategies in the battle against these devastating apicomplexan diseases.

Results and Discussion

In silico prediction of ubiquitylation pathway components in apicomplexan genomes

We selected 24 Pfam domains that are known to be related to the UPS (see materials and methods section). These 24 Pfam domains are commonly found in ubiquitin and UBLps, E1 and E1-like enzymes, E2 enzymes, E3 enzymes and DUBs. Each Pfam domain family was used in an hmmsearch application of the translated genomes of *Plasmodium* spp. *falciparum*, *vivax, yoelii*, *berghei* and *chabaudi*, *T. gondii*, *Cryptosporidium spp*. *parvum* and *hominis*, *S. cerevisiae*, *C. elegans*, *D. melanogaster*, *H. sapiens* and *A. thaliana*. HMM searches were run using a series of increasingly stringent threshold E-values, from E-value #1 to E-value #0.1 (data not shown). With regards to the five eukaryotic model organisms that were used, the threshold E-value #0.5 gave the most consistent results when compared to previously published results. The number of UPS- related proteins in *A. thaliana* and the other model organisms has previously been analyzed, particularly the number of E2 and E3 enzymes that are found in *A. thaliana*, *H. sapiens*, *C. elegans* and *S. cerevisiae* (see [39–42] for reviews). The observation that our results (Table A3.1) are consistent with these existing data sets would appear to validate both the HMM search strategy with a threshold E- value set at 0.5, as well as providing standard datasets in these model organisms for subsequent comparative analysis of the Apicomplexa data sets.

Amongst the 13 proteomes investigated in this study, a total of 4453 proteins were identified as carrying one or more of the 24 selected Pfam domains (Table A3.1 and supplemental Table S1). For example, 114 proteins were found in *P. falciparum*, 145 in *T. gondii*, 114 in *C. parvum* and 127 in *S. cerevisiae*. In each case, these numbers of UPS component proteins represent approximately 2.5% of their respective proteomes. Given the good correlation between numbers of proteins identified in each apicomplexan with that of the single celled model eukaryote S. cerevisiae, these data would appear to suggest that the apicomplexan datasets are relatively complete. For those *Plasmodium* species such as *P. chabaudi* or *P. berghei*, the relative under-representation of identified proteins would more likely reflect the completeness of the respective genome project rather than an absolute reduction in UPS components. The number of UPS components increases considerably in multi-cellular organisms with increased genome complexity (e.g. 678 proteins were identified in *C. elegans* and 883 in *H. sapiens* while some 1452 proteins were identified in *A. thaliana*). In *H. sapiens*, 162 DUBs/DUBLs were found although a previous publication only identified 95 putative DUBs/DUBLs from which 79 exhibited conserved catalytic residues [28]. Such a difference can be explained by the fact that proteomes extracted from *H. sapiens* and *D. melanogaster* genomes contain multiple isoforms for some families of DUBs and DUBLs. Furthermore, the Hidden Markov Model that we used to search for UPS components compiled more complete datasets than many other search approaches would do. For example, while our HMM search identified domain OTU-carrying proteins in

apicomplexan parasites (OTU is a major sub-class of DUB) none was reported in a recent publication on parasitic protozoa deconjugating enzymes where the authors used a more selective BLASTP homology search [43]. This observation further highlights the exhaustiveness of the HMM search.

With regards to the relative abundance of each domain family, a striking observation is that a high proportion of F-box-carrying proteins are present in multi-cellular organisms (e.g. 43% in *A. thaliana*) while only few of them were identified in apicomplexan organisms. F-box-containing proteins are adaptor proteins in Cullin-RING-Ligase complexes (CRLs), and are involved in direct and specific substrate recognition. Previous authors have hypoth- esized that the very high number of F-box proteins in *A. thaliana* suggests that plants can assemble numerous CRLs, which could control a wide array of substrates [27]. The low number of F-box proteins detected in apicomplexan parasites could indicate that there is no need for these specific adaptors, or that their amino acid sequences are highly divergent from other eukaryotic cells and could not be detected using our standard HMM search. An alternate hypothesis is that a different family of proteins in apicomplexa could carry out the role of adaptor.

To investigate the global degree of conservation of the predicted UPS proteins, an all-against-all blast search was performed for each domain studied and between the 13 genomes analyzed. The bit scores obtained were reported as a color scale

(from red, ''highly divergent''; to blue ''highly conserved'') in triangular distance matrices. Results of this analysis are shown in Figure A3.2. Using this methodology it is particularly straightforward to realize that ubiquitin/ubiquitin-like activating enzymes and ubiquitin- conjugating enzymes are conserved in all eukaryotic cells including apicomplexan parasites. HECT-ubiquitin ligases and Cullin- ubiquitin ligases are also well-conserved, while RING and RING-like ubiquitin ligases, U-box ubiquitin ligases and ubiquitin ligase adaptors F-box show more diversity. This is particularly striking with regards to RING/RING-like ubiquitin ligases and F- box adaptor proteins where almost every protein considered in our study is divergent from all others, with the exception of U-box, RING/RING-like-containing proteins from *A. thaliana*. In this case, diversity intra-species is much lower than in any other species. The biological significance of this observation remains to be elucidated.

With regards to the eight subclasses of DUBs analyzed, the Josephins (MJD), UCHs (peptidase_C12), autophagins (peptida- se_C54) and deSUMOylases (peptidase_C48) families are well conserved within and between species with the exception of a clear differential expansion in the *A. thaliana* UCHs and deSUMOylases. In the other subclasses, relative divergences exist within and between species. An increased divergence can be observed in the metalloprotease (JAMM/Mov 34). The function of the WLM family, usually found only in plant and fungus, in *Plasmodium* and *Toxoplasma* thus deserves to be fully investigated.

When possible, the complete dataset was used to predict apicomplexan functional homologs of known UPS components from data available from the five model organisms investigated here. For each domain, dendrogram trees were built with all the 13 species that we used in this study. For purpose of clarity, only apicomplexan data are presented here. However, our complete results are available for download on the laboratory website (http://lerochlab.ucr.edu/UPS_prediction_data). Figure A3.3, A3.4, A3.5 and A3.6 show the apicomplexan data for ubiquitin/UBLps, E1, E2, and E3-Ubox enzymes respectively. The rest of the apicomplexan data for the other domains are given in supplemental figure S1.

Ubiquitin and ubiquitin-like proteins

Ubiquitin is a 76 amino acid protein extensively conserved between all eukaryotic sequences, with similarities in excess of 98% between humans, yeast and apicomplexan parasites [43,44]. UBLPs generally bear little primary sequence identity to ubiquitin (Figure A3.2), however they share two principle features. First is a compact tertiary structural motif consisting of five beta-sheets and a single alpha helix, termed the ubiquitin fold, in addition to a low complexity C-terminus available for activation and conjugation. Second is a shared biological function through activation, conjugation and reversible modification of a target protein's activity. Since first identification of the first UBLp (the interferon stimulated gene

15, ISG15) in 1979 additional UBLps (at least 10 to date) have been described with an escalating frequency and evidence of UBLps ever-widening role in the modification of cellular processes [17,23]. Paralogs for polyubiquitin, two ubiquitin-ribosomal protein fusions (Ub-S27a and Ub-52), neural precursor cell expressed developmentally and down-regulated 8 (NEDD8), small ubiquitin-related modifier (SUMO), homologous to ubiquitin 1 (HUB1), ubiquitin-related modifier 1 (URM1) and autophagy 8 (ATG8) have been described in most of the apicomplexans investigated here (Figure A3.3) [43,44]. The fact that there are missing paralogs more likely reflects the quality of the genome sequence available for the different apicomplexan species (depending upon the status and fold-coverage of their genome projects) rather than absolute absence from the genome. For example, several incomplete sequences from the murine malarial parasites (*P. chabaudi* and *P. yoelii*) contain partial ubiquitin sequences; however it was impossible to definitively assign the final ubiquitin gene based on the sequence available. In addition to ubiquitin/UBLps, several genes were identified by the HMM search that contain the highly related ubiquitin- conjugated role in processes such as signal transduction and proteasomal delivery [45]. These proteins (including paralogs of yeast RAD23 and DSK2) were manually edited from the list of ubiquitin/UBLps for all the organisms investigated.

As in all eukaryotes, ubiquitin is encoded by one of three types of fusion-protein precursors in the apicomplexans investigated here (Figure A3.3). Although

multiple copies of these genes may exist in higher eukaryotes, only single copies were identified in apicomplexans. The first type, polyubiquitin, consists of three to five direct repeats of the ubiquitin coding sequence. The two remaining ubiquitin-fusion genes encode N-terminal ubiquitin fused to one of two ribosomal proteins (S27a and S52). In all cases, subsequent proteolytic cleavage of the polypeptide encoded releases the ubiquitin monomers. A suitable steady state level of available ubiquitin monomers is provided by de novo synthesis of ubiquitin and recycling of ubiquitin following cleavage from their target proteins. During periods of stress, elevated demands for ubiquitin are met, in part, by increased levels of polyubiquitin expression [46]. Expression data available for *P. falciparum* indicates all three ubiquitin genes are expressed throughout the parasite's life cycle [7,9], while polyubiquitin also appears to be induced during a heat-shock response [44,47].

Reversible modification by SUMO is generally associated with processes involving nuclear integrity and function, and more specifically with nuclear transport, subnuclear targeting and genome stability (for review see [48]). More recently, conjugation by SUMO has been suggested to play an additional role of antagonizing the effect of ubiquitin conjugation–an evolution that appears to suggest a complex interplay of protein modification above that of simply activating and inactivating a protein [49,50]. Although not identified in all *Plasmodium spp*., single copies of genes encoding SUMO have been identified across all the apicomplexan organisms investigated here (Figure A3.3). While higher eukaryotes

typically have three to four variants of SUMO, like most single- celled eukaryotes apicomplexans only have one SUMO variant. These appear most similar to SUMO-1 in that they lack anbinding domain (UBD). These typically N-terminal located domains are found in proteins that have evolved to adopt the ubiquitin domain in a non- intrinsic sumoylation motif within the N-terminus (yKXE, where y represents a hydrophobic residue and K the targeted lysine), suggesting that polysumoylation does not have a functional role in apicomplexans. Thus, only ubiquitin, by virtue of multiple internal modifiable lysines is capable of forming conjugated polymeric chains on target proteins.

NEDD8, also termed related to ubiquitin 1 (RUB1), is most similar to ubiquitin at the primary sequence level. NEDD8 typically accumulates in the nucleus where its only known target, cullin, is found [51]. As described later, cullins form the scaffold for the SCF (Skp-Cul1-F-box) E3 ubiquitin ligase complexes [52]. NEDD8 appears to play an essential role in cell cycle control in actively proliferating cells and is down-regulated during cell differentiation [53]. Single genes encoding NEDD8 were identified in all apicomplexan families investigated here (Figure A3.3), branching closely with all the ubiquitin-fusion genes as would be expected given the higher primary sequence similarity of this UBL modifier.

Single copies of genes encoding the less characterized UBLps URM1 and HUB1 are found throughout the apicomplexan lineages investigated here (Figure A3.3).

Both UBLps have only been recently discovered [54,55], and little is known about their biological roles. HUB1 is noteworthy for the absence of the typical di-glycine C-terminal motif common to most UBLps, rather having a di-tyrosine motif. As yet, E1 and E2 proteins that would activate and conjugate HUB1 have not been characterized and recent reports suggest that a more ‘‘hormonal’’ role may exist in higher eukaryotes [56]. However, in *S. cerevisiae*, conjugation to proteins involved in mRNA and pre-mRNA splicing have been described, and may more likely reflect the role of HUB 1 in apicomplexans [55]. The second UBLP, URM1, shares very little homology to ubiquitin, but appears more closely related to the *Escherichia coli* sulphur carring proteins ThiS and MoaD involved in thiamin and molybdopterin synthesis, respectively [54]. In *S. cerevisiae*, URM1 has only been found to conjugate to alkyl hydroperoxide reductase 1 (AHP1), suggesting some role in adaptation to oxidative stress may similarly operate in apicomplexans [57].

The autophagy system facilitates degradation of the cytoplasm following engulfment in a vesicle followed by fusion to lysosomes, a process necessary for both cell differentiation and response to starvation. Analysis of mutations in autophagy in *S. cerevisiae* identified two UBLps involved in this system, termed ATG8 and ATG12 [58]. Previous analysis of several apicomplexan and kinetoplast genomes highlighted that while a gene encoding ATG8 could be readily identified across a range of protozoa [43], no evidence exists for the gene encoding ATG12 (see Figure A3.3). ATG12 plays a key role in the initial formation of the

autophagosome, while ATG8 is conjugated to the amide group of phosphatidylethanolamine in the membrane, altering the membrane dynamics; thus, ATG8 is unique amongst UBLps in not conjugating a protein. Interestingly, while ATG12 has not been found in kinetoplastids, autophagy has been demonstrated to be active in *Leishmania spp*. and play a key role in parasite virulence [59]. The Pfam search described here identified a single gene in *P. falciparum* as being an ATG12 paralog (Table A3.1, PF14_0779). However, though the predicted polypeptide shares some primary sequence homology to ATG12 from C. elegans, it lacks a C-terminal glycine. Further, the cognate E2 and target proteins for ATG12, ATG10 and ATG5, respectively, are absent from *P. falciparum* (as well as the other apicomplexans investigated).

A number of UBLps typical of higher eukaryotes (ISG15, FAT10, UFM1, FUB1) have not been found in this analysis, nor that previously described by Ponder and Bogyo (2007). Although some UBLps may not be expected based on their predicted roles in immune system regulation in higher eukaryotes, their absence, coupled with that of SUMO variants and ATG12 in apicomplexans suggest a more restricted role for UBLps in apicomplexan cell biology. However, analysis of gene expression data (microarray and proteomics) for SUMO, NEDD8, HUB1, URM1 and ATG8, where available (particularly for *P. falciparum* and *T. gondii*), suggests that these UBLps are expressed at all the life stages investigated. These data suggest that ubiquitin/UBLps are essential compo- nents in controlling cellular processes

throughout apicomplexans complex parasitic life cycles.

Ubiquitin/UBL activating enzymes (E1)

The first step in the ubiquitin/UBLps activation and conjuga- tion cascade is mediated via E1 proteins. A number of isoforms of E1 exist, each responsible for the activation of different ubiquitin/ UBLps (for review see [60]). All E1s, however, share a common mechanism of action. The initial step is the ATP-dependent adenylation of the C-terminus of the cognate ubiquitin/UBLp, which is then held in a non-covalent interaction until subsequent attack by an active site cysteine resulting in covalent attachment of the ubiquitin/UBLp via a thioester bond. The final step in the mechanism is the transfer of the activated ubiquitin/UBLp to E2 via a transesterification reaction.

E1 proteins are characterized by the presence of the ubiquitin activating (UBA) Pfam domain. Additional motifs in E1 are responsible for the correct selection of ubiquitin/UBLp for activation and subsequent E2 to which transfer the activated ubiquitin/UBLp [61]. Whereas the E1 responsible for ubiquitin activation (homologous to UBA1 of *S. cerevisiae*) can deliver activated ubiquitin to several E2 isoforms, the E1s responsible for activating the UBLps SUMO and NEDD8, termed UBA2 and UBA3, respectively, only transfer to a single cognate E2 (see below). UBA1 has two UBA domains on a single polypeptide. UBA2 and UBA3, each only

have one UBA domain containing the active site cysteine required for the covalent attachment of activated SUMO and NEDD8, and actually represent one part of a E1 heterodimer complex with AOS1 or APPBp1, respectively [60]. Both AOS1 and APPBp1 each contain one UBA domain, thus resulting in an E1 complex with two UBA domains. The analysis presented here indicates the presence of paralogs for UBA1, UBA2 and UBA3 in all the apicomplexan lineages (Figure A3.4). High level of primary sequence identity in the core of the UBA domains present in UBA1-3 is conserved across all thirteen eukaryotes described in this analysis (Figure A3.2). However, outside of this core homology, sequences diverge rapidly as the functional requirements for these sequences in specifically interacting with different ubiquitin/UBLps and E2 proteins alter. Assigning paralogs for AOS1 and APPBp1 has not been possible here–although unassigned proteins containing UBA domains are present in all apicomplexans investigated here and may well represent functional paralogs for these proteins. Analysis of transcriptional patterns in P. falciparum of UBA2 and UBA3 with those of the unassigned UBA containing proteins, as well as searching for existing characterized yeast two hybrid interactions, did not provide any additional clues in defining AOS1 or APPBp1 paralogs [62].

In addition to those described above, three additional E1 proteins are indicated in Figure A3.4. The first, UBA4, is responsible for activation of URM1. Interestingly, UBA4 has only a single UBA motif, but does have a rhodanese

homology domain (RHD) [63]. Rhodanese and RHD containing enzymes are responsible for sulphur transfer reactions and form a persulphide bond on their active site cysteines. Interestingly, an E2 for URM1 has not been identified to date, and it is suggested that the RHD may act as a substitute in-built E2 in the transfer of URM1. The second E1, ATG7, is the only E1 that is capable of activating more than one UBL modifier–ATG8 and ATG12 [64], however, as described above, ATG12 does not appear to be present in the protozoa lineage. ATG7 is characterized by a single C-terminal UBA domain with a large N-terminal ATG7 specific motif. The final E1 isoform, termed UBA1-like (due to presence of two UBA motifs), tend to be larger than Uba1. In several eukaryotes, two or more isoforms of ubiquitin E1 exist, however, whether these UBA1-like E1s represent a second ubiquitin E1 or are required for transfer of a different UBL modifier (note no E1 for HUB1 has been assigned should it actually have a conjugating role) remains to be determined.

Proteomic and transcriptomic profiling data available for *P. falciparum* provide extensive evidence for the ubiquitous expression of E1s throughout the parasite's life cycle [7,9]. Interestingly, detailed analysis of transcription during intraerythrocytic development suggests a temporal pattern of transcript accumulation in the early trophozoite stages when the parasite becomes more metabolically active. Similar data available for *T. gondii* similarly suggest constitutive expression of E1s throughout apicomplexan life cycles.

Ubiquitin/UBL conjugating enzymes (E2)

Eukaryotes express a number of E2 isoforms, typically of between 17–22kDa (see [65] for review). E2s are readily identified by the presence of a conserved central 150-residue domain that forms a tertiary structure where the cysteine in the active site, which accepts the activated ubiquitin/UBLp from E1 via a transesterification reaction, is buried in a shallow groove [66]. The extensive conservation exhibited by the E2 proteins identified in the search of the 13 genomes described here is readily exemplified in Figure A3.2.

The apicompexan parasites investigated here have eight to fourteen E2 proteins, similar to the 14 described for the only other single cell eukaryote S. cerevisiae (Table A3.1). The number of E2 isoforms tends to increase with increasing genome complexity. Given the relative completeness of the *Cryptosporidium spp*. genomes, the relative small number of E2 identified in *C. hominis* and *C. parvum*, eight and eleven, respectively, may suggest that these represent a true variation from the mean of 13 to 14 E2s found in *Plasmodium spp*. and *T. gondii*. Specifically, two E2 variant paralogs immediately adjacent to the UEV branch in Figure A3.5 (containing the *P. falciparum* genes PF14_0128 and MAL13P1.227) are only found in *Plasmodium spp*. and *T. gondii* in this analysis and are atypical E2s of up to 54kDa with a long N-terminal extension. N- and C-terminal extensions in E2 are thought to play key roles in recognition and association with

E3s and their subsequent protein target and thus these atypical E2s may reflect a specific adaptation in the *Plasmodium* and *Toxoplasma* lineages.

Different isoforms of E2 have distinct roles in regulating downstream functions through specific interaction with distinct E3s ([22] for review and [67]). While several E2s are capable of cascading activated ubiquitin through to different E3s, only single E2 isoforms conjugate to SUMO and NEDD8; UBC9 and UBC12, respectively. Paralogs for both UBC9 and UBC12 are present in all the apicomplexan lineages investigated here (Figure A3.5). One isoform of E2, termed the Ub-E2 variant (UEV), lacks both a key HPN amino acid motif and the active site cysteine in the E2 core and is incapable of conjugating ubiquitin. UEVs instead form a heterodimer with the UBC13 E2 isoform and direct a subset of E3s to conjugate ubiquitin to its target through the side chain of Lys63 (as opposed to more typical conjugation through the side chain of Lys48) [50]. Paralogs for both UBC13 and a UEV are present in all the apicomplexan lineages investigated here (Figure A3.5). Conjugation of ubiquitin through Lys63 generally acts as non- proteolytic signals for processes such as DNA repair [49]. Thus, proteins may be conjugated by polyubiquitin chains, single ubiquitin molecules through more than one lysine side chain and even competitively with SUMO. This diversity of conjugation has important implications in post-translational modifications directing a diverse response in the target protein. Interestingly, strong yeast two-hybrid data in *P. falciparum* indicates a clear association of the UBC13 and UEV paralogs in this

organism [62]. One E2 molecule not reported in this analysis is ATG3, which is responsible for conjugation to the UBLp ATG8. This E2 exhibits extreme diversity to that of other E2s and lacks the core E2 Pfam motif used in this analysis. Paralogs exist in all apicomplexans investigated here (PFI0280c, PB000344.03.0, PC000563.02.0, Pv098725, PY04567, chro.80308, cgd8_2650 and 46.m01688).

Extensive gene expression data for nine of the fourteen *P. falciparum* E2s suggest a diverse pattern of steady state mRNA accumulation at different stages of intraerythrocytic development. The fact that different E2 isoforms are expressed at distinct stages in the parasite's life cycle suggests that a temporal profile of delivering ubiquitin/UBLps to different E3s exists, which highlights a potential additional level of temporal control in the UPS system during apicomplexan parasite's life cycles.

Ubiquitin/UBL ligases (E3)

E3 ubiquitin/UBL ligases are a very diverse group of proteins involved in specifically transferring ubiquitin/UBLps to a given substrate. In all organisms, 48% of the predicted UPS components identified belong to the E3 ubiquitin/UBL ligase family. This high percentage of E3 reflects the specificity that is required for specific substrate recognition. Table A3.2 summarizes all potential E3 ubiquitin/UBL ligases that have been found in *P. falciparum*, and their homologs in *T. gondii, C. parvum*, and yeast. There are three superfamilies of E3 ubiquitin/UBL ligases. HECT

ubiquitin ligases have a direct role in catalysis during ubiquitylation, whereas RING (Really Interesting New Gene) finger and U-box E3s are involved in multi-protein complexes. RING finger E3s are the most abundant ubiquitin/UBL ligases.

Our search identified four HECT domain-containing proteins in *P. falciparum*, and other apicomplexans. Three of them have a homolog in S. cerevisiae: TOM1, UFD4, and HUL5 (HUL5 has unknown functions). The fourth HECT-domain protein that we identified in apicomplexans does not match any protein from yeast but is similar to UPL5 in *A. thaliana* (see table A3.2). UPL5 has an unknown function, but is annotated as potentially involved in cell proliferation. TOM1 (MAL8P1.23 in *P. falciparum*, and 86.m00385 in *T. gondii*) has been recently described as being involved in cell cycle arrest after DNA damage, mediating CDC6 ubiquitylation, a protein essential to initiation of DNA replication [68]. UFD4 (MAL7P1.19 in *P. falciparum*, and 80.m02344 in *T. gondii*) is involved in the ubiquitin fusion degradation pathway (UFD pathway) [69], which results in polyubiquitylation of ubiquitin fusion proteins that do not fall into the N-end rule pathway (the N- end rule relates the in vivo half-life of a protein to the identity of its N-terminal residue; for example, in eukaryotes, a protein with an isoleucine at its N-terminal end will be targeted to the proteasome more rapidly than a protein with a glycine, which is a stabilizing residue; see [70] for a review). Little is known about the UFD pathway, and its physiological functions remain unknown. Recent works indicate that UFD4 is involved in controlling the degradation of RAD4, a nucleotide excision repair protein

[71].

U-box proteins are another family of ubiquitin ligases that are structurally similar to RING finger proteins but lack the metal binding sites (see [26] for a review). A sub-group of U-box proteins is also termed E4 ubiquitin chain assembly factor, and is known for its ability to add a polyubiquitin chain on a substrate already primed for degradation by oligoubiquitylation [72]. We identified two U-box domain-containing proteins that are present in all apicomplexan parasites: the *S. cerevisiae* UFD2 (PF08_0020 in *P. falciparum*, and 72.m00386 in *T. gondii*) and PRP19 (PFC0365w in *P. falciparum*, and 641.m01564 in *T. gondii*) homologs (Figure A3.6). UFD2 is an E4 ubiquitin ligase, involved in the UFD pathway [72]. In yeast, UFD2 interacts with the AAA ATPase (ATPase associated with various activities) CDC48, which possesses a chaperone-like activity and mediates ubiquitin-dependant endoplasmic reticulum associated degradation (ERAD) pathway [73]. Richly et al. [74] proposed a short ubiquitylation chain-dependant escort pathway to the proteasome that would involve UFD2, which adds ubiquitin proteins to a substrate already mono- or di- ubiquitinated. In this escort pathway, the Cdc48-Npl4-Ufd1 would act to restrict chain length to four to six ubiquitin. CDC48 (p97 in mammalians) is involved in several different functions in cell, such as retrotranslocation from the ER to the cytosol (quality control), transcriptional control or cell cycle regulation (see [75] for a review). CDC48 is the cyclin-dependant kinase (CDK) that promotes cell cycle progression in yeast. Previous

works indicate that CDC48 regulates the stability of several cell-cycle regulators in a UPS-dependant manner. CDC48 was also found to regulate the stability of proteins involved in controlling the expression of genes involved in fatty acid metabolism in yeast (see [75] for a review). We used BLASTP to identify homologs of proteins involved in the escort pathway and do not carry UPS-related domains (and are thus absent from our dataset). We predicted CDC48, Ufd1 and Npl4 in *P. falciparum*, PFF0940c, PF14_0178 and PFE0380c respectively (see supplemental Table S2). All these proteins seem to be well conserved in *P. falciparum*. Furthermore, clear orthologs of CDC48, Ufd1 and Npl4 are predicted in *P. yoelii*, *C. parvum*, *C. hominis*, and *T. gondii* (from OrthoMCL-DB, [76], see supplemental Table S2). These findings could indicate that the Ufd2-dependant Cdc48-Npl4-Ufd1 escort pathway exist in *P. falciparum* and other Apicomplexa, with functions similar to the ones observed in yeast.

PRP19 is an oligomeric U-box-containing E3 ligase [77] that plays a role in mRNA splicing [78], spliceosome activation and recycling [79,80], and DNA damage response [81]. PRP19 is part of a complex consisting of at least eight units, of which CDC5 and PLRG1 (Pleiotropic regulator 1) [78]. Lu and Legerski [81] demonstrated that, in DNA damage conditions, PRP19 is ubiquitinated. The authors showed that ubiquitinated PRP19 fails to interact with either CDC5 or PLRG1, and over expression of PRP19 reduces the levels of apoptosis after exposure of cells to DNA damage. Little is known about apoptosis or programmed cell death in Apicomplexa.

Apoptosis-like events have been described in *P. falciparum* and *P. berghei* [82,83]. Al-Olayan and co-workers suggested that apoptosis could be a possible mechanism for limiting intensity of infection in the mosquito by *P. berghei* (see [84] for review). Unfortunately no data is currently available about the function or the expression pattern of PRP19 homolog (PFC0365w) in *P. falciparum* ookinetes. In *T. gondii*, the homolog of PRP19 (641.m01564) contains WD40 repeats, which are known to be involved in a wide array of cellular processes ranging from signal transduction and transcription regulation to cell cycle control and apoptosis [85,86]. However, the functions of PFC0365w and 61.m01564 remain to be elucidated.

A third U-box containing ubiquitin ligase, that is absent in yeast, has been identified in Plasmodium species (PF07_0026 in *P. falciparum*) and shares extensive homology with the human protein CHIP (C-terminal of Hsp70-interacting protein). Like UFD2, CHIP has been described as being an E4 ubiquitin ligase in human. A particular feature of CHIP is that its catalytic activity requires its homodimerization through the U-box domain (see [87] for review). CHIP is known to be involved in protein quality control by promoting ubiquitylation of denatured proteins in an Hsp70/Hsp90-dependant manner. CHIP is also involved in heat shock response and prevention of apoptosis. However, the physiological substrates of CHIP still remain unidentified. Previous experiments suggested that CHIP might have multiple functions that could be proteolytic either dependent or independent from proteasome degradation [88]. However, the physiological role(s) of CHIP remain(s)

unknown. In *P. falciparum*, the CHIP homolog PF07_0026 is mainly expressed at the sporozoite stage (mosquito stage). Interestingly we also identified CHIP homologs in *P. berghei*, *P. chabaudi*, *P. vivax* and *P. yoelii* (respectively PB001535.02.0, PC000957.01.0, Pv087910, PY00139) whereas no homolog was found in yeast *Cryptosporidium spp*. and *T. gondii*. Thus, it can be hypothesized that CHIP homologs in *Plasmodium* are involved in the hepatic infection stage, although a precise role remains to be determined.

Cullin-containing proteins belong to the E3 ubiquitin ligase family. In association with a RING finger E3 and a substrate recognition protein such as F-box proteins, they form Cullin- RING-Ligases (CRLs). The most famous CRLs are the Skp1- Cullin1-Fbox (SCF) complex, containing the cullin protein CDC53 in yeast, and the Anaphase Promoting Complex/ Cyclosome (APC/C), containing the cullin protein APC2 in yeast. Both the SCF and the APC/C are involved in cell-cycle progression (see [89] for review). A homolog to CDC53 has been found in each apicomplexan (PF08_0094 in *P. falciparum*, and 80.m02207 in *T. gondii*). The SCF contains at least four subunits: CDC53 (Cullin1) which is stabilized by the ubiquitin-like modifier NEDD8, the RING E3 RBX1, the adaptor protein Skp1 and a F- box protein for substrate recognition (see [39] for review). The present study allowed us to confirm the presence of F-box proteins as well as NEDD8 and RBX1 homologs in apicomplexan genomes (table A3.1) (MAL13P1.64 and PFC0845c respectively in *P. falciparum*, see Figure A3.3). Using BLASTP, we identified homologs of Skp1 in the

genomes of each apicomplexan (MAL13P1.337 in *P. falciparum*, see supplemental table S3). The minimum components that are required for the cell-cycle regulator SCF are therefore all present in apicomplexan genomes. The situation is different as far as the APC/C is concerned. While a homolog to CUL8 and the RING finger APC11 involved in the APC/C were recognized by our analysis (respectively PFF1445c and PFF1180w in *P. falciparum*), the cullin classically involved in the APC/C (APC2) was not found in apicomplexan parasites. The role of CUL8 is not well known. Previous data suggest that it may be involved in anaphase progression [90] in a CRL other than the classic APC/C. The presence of a slightly different APC/C in *P. falciparum* is not entirely unexpected. Cell division during schizogony is apparently asynchronous in *P. falciparum*, that is, several rounds of DNA replication/DNA division occur before final cytokinesis, instead of the paradigm of successive cycles of alternating DNA replication/ DNA division/cytokinesis ([91–93]).

RING finger and RING-like E3 ubiquitin ligases are the largest group of E3s. They do not have a direct catalytic role in linking ubiquitin and ubiquitin-like modifiers. They rather act as adaptor partners: RING E3s interact with both an E2 ubiquitin- conjugating enzyme (that is carrying ubiquitin) and its given specific substrate, bringing the substrate in close proximity with ubiquitin. Since RING and RING-like E3s are involved in substrate recognition, and given the diversity of proteins that are targeted by the UPS, a large diversity of RING and RING-like E3s is expected (homologs of RBX1 and APC11 are only two of the numerous RING and

RING-like proteins, up to 55 in *T. gondii*, that were functionally identified by our analysis). For example, we have found potential RAD16, RAD5, and TFB3 homologs (PFL2440w, MAL13P1.216, and PFE0610c repectively in *P. falciparum*, and 42.m00128, 641.m01484, and 641.m02557 repectively in *T. gondii*), which are known to be involved in nucleotide excision repair (NER) [94]. Previous work showed that RAD16 mediates histone H3 acetylation before global nucleotide excision repair [95]. Interestingly, several of the RING E3s that were found in apicomplexan parasites are potentially involved in mRNA turnover and stability (PFI0470w in *P. falciparum*) or pre-mRNA maturation (PF14_0139 and PF14_0416 in *P. falciparum*, and 50.m03082 and 42.m00073 in *T. gondii*) [96]. Consistent with previous work, this observation suggests that regulation of mRNA stability is a major mechanism of gene regulation in *P. falciparum* [10]. Several others of the identified RING E3s are potentially involved in histone ubiquitination (e.g. PfBre1 PFF0165c in *P. falciparum*, 35.m01589 in *T. gondii*), or chromatin remodeling and silencing (PF10_0046 in *P. falciparum* and 76.m01590 in *T. gondii*). Recent data have increasing shown that epigenetic mechanisms play a key role in the control of gene expression in Apicomplexa. In *Plasmodium*, chromatin modeling is one of the proposed mechanisms that control allelic exclusion of the var virulence genes [97,98]. In *T. gondii*, acetylation of histone H4 or acetylation on lysine 9 of histone H3 as well as trimethylation of lysine 4 of histone H3 were shown to be located close to the 59 UTR of the active genes [99]. Increasingly, evidence suggests that the precepts of the universal ''histone code'' and the molecular mechanisms that

underpin reversible histone modification similarly applies to Apicomplexa, although much yet remains to be elucidated.

It is interesting to notice that several RING-domain proteins found in apicomplexans do not cluster with any other known protein. In *P. falciparum*, PF14_0054 and PF13_0188 are two examples of proteins with no evident homologs in the organisms analyzed, besides their apicomplexan counterparts. With regards to their expression profiles [7,9], PF14_0054 is expressed at the sporozoite and the late schizont stages of *Plasmodium*'s life cycle, which suggests that this protein may be specific to apicomplexan processes such as parasite invasion. However, functions of several of the predicted parasitic RING and RING-like proteins remain to be elucidated, either because there are no known homologs in other model organisms, or because the function of the matching homolog remain unknown.

Whilst prediction of function based on sequence homology can be misleading, a study of the domain architecture of RING and RING-like E3 ligases may highlight additional features. *P. falciparum*'s RING and RING-like E3 ligases were investigated using SMART to predict all protein motifs present([100], see material and methods). Results are shown in Figure A3.7. This analysis revealed that *P. falciparum* RING and RING-like E3 ligases possess major domain architectures found in E3 ubiquitin/UBL ligases from other model organisms: an N-terminal or C-terminal RING domain, a RING domain associated with a zinc finger domain such as C2H2 domain, the

architecture RING/RNA recognition motif (RRM), RING/helicase, RING/forkhead-associated do- main (FHA), and RING/in between RING (IBR). It is interesting to observe that predicted *Plasmodium* RING E3s with a single RING domain appear more abundant than those of yeast. Several of these proteins carry several coiled-coil regions, e.g. up to six in PFF0165c. Coiled-coil domains are known to be involved in regulation of gene expression and many other biological processes (see [101] for a review). The role of coiled-coil domains in host-pathogen interactions has particularly been studied. The molecular cross talk that occurs between a pathogen and its given host is complex. The pathogen has to collect and process diverse host signals in order to modulate the expression of its virulence genes. In gram-negative bacteria, coiled-coil proteins are involved in type III secretion systems that are used to deliver virulence effector proteins into, or close to, the host cell (see [102] for a review). In *Agrobacterium tumefaciens*, the histidine kinase VirA activates the expression of virulence genes in response to multiple wound- derived plant signals, *via* the involvement of coiled-coil structures [103]. The HIV protein gp41 contains a conserved coiled-coil domain that is critical for the entry of the virus into the host cell [104]. Thus, the apparent abundance of coiled-coil-containing RING E3 ligases in *P. falciparum* may provide a link to the particular virulence of this parasite and would appear to warrant further investigation.

RING E3 ubiquitin ligases with one or multiple predicted transmembrane domains are also more abundant in *P. falciparum* than in yeast (ten and three

proteins, respectively). Membrane bound E3 ubiquitin ligases have been shown implicated in the regulation of immune recognition during virus infections (see [105] for a review). The abundance of such RING E3 ubiquitin ligases would be of obvious interest in understanding how chronic infections are established during infection by these parasites.

DUBs and DUBLs enzymes

DUBs are responsible for generating ubiquitin precursors from fusion protein and polyubiquitin chains. They are involved in the disassembly of polyubiquitin chains for the maintenance of the available ubiquitin within the cell. DUBs are also responsible for the editing of ubiquitin conjugates to alter their commitment for proteolysis [28,106]. In these roles, DUBs have been shown to be associated with proteasome subunits [107,108]. A second important pathway associated with deubiquitylation is the control and degradation of membrane protein trafficking to the vacuole/ lysosome [109–112]. Furthermore, DUBLs have been shown to play additional key roles in eukaryotic cells. For example, the SUMO deconjugating enzymes (also termed SENPs) play a role in the control of transcription [113], the autophagy-related DUBLs similarly regulate vacuole targeting pathways [114,115] and the WLM family is thought to be specifically involved in desumoylation and chromatin structure maintenance [29]. Increasing data indicate that most DUBs/DUBLs regulate a limited number of specific substrates, where they interact either directly with ubiquitin/UBLp, or with the target protein to which ubiquitin/

UBLp is conjugated.

Between 20 to 40 DUBs and DUBLs were found in apicomplexan genomes (table 1), while 165 were detected in *A. thaliana*. Moreover, most of the major DUBs and DUBLs subfamilies were found within each of the apicomplexan genomes. The few exceptions in *P. chabaudi* and *P. berghei* are probably the consequences of weak genomic sequences coverage for these two apicomplexan species.

UCH-L3, which was identified in all Apicomplexa, is known to cleave ubiquitin and the UBLp NEDD8. The first active DUBs (UCH-L3 and UCH-L5) were recently identified in P. falciparum and T. gondii, and showed to both react with ubiquitin and NEDD8 [116,117]. SUMOylases (SENPs), of which at least two are present in apicomplexan genomes, are known to be SUMO-specific proteases but have also been shown in some cases to cleave NEDD8. UBLps, such as ATG8 involved in autophagy, are also processed by distinct proteases, such as the Autophagin peptidase C54, of which one isoform has been identified in all Apicomplexa investigated here.

Determination of target specificity in vivo has been challenging in other model organisms. It is possible that *in vivo* protein localization and the presence of adaptors such as E3 ligases can increase specific interactions [118]. Despite the challenge of target identification, the function of several DUBs have been clearly

identified and implicated in critical cellular process. The fact that all DUBs subclasses are present in apicomplexans suggests that several of their functions are conserved. For example, deubiquitination of proteins at the proteasome lid is necessary for protein degradation and recycling of ubiquitin. Various DUBs, such as the metalloprotease JAMM (RPN11 and RPN8 in yeast) the UBP6 in yeast or UCH-L5 in mammalian, have been found in complex with proteasome subunits.

Increasing evidence suggests that several DUBs are implicated in remodeling chromatin structure, transcriptional regulation and gene silencing. In yeast, UBP8 and UBP10 have been implicated in the dynamic histone monoubiquitination of H2B. The presence of UBP8 correlates with changes in transcriptional regulation [118] whilst UBP10 is required for the recruitment of the silencing factor Sir2 [119]. Both of these DUBs have homologs in Apicomplexa (eg. PFI0225w and PF14_0145 in *P. falciparum*, respectively) and could be potentially involved in the regulation of parasite virulence genes, principally the var gene family in *P. falciparum*.

DUBs are also implicated in the endocytic pathway and intracellular traffic [112]. JAMM and USP proteases (AMSH and USP8 in human) as well as the human OTU (VCIP135) protease have all been shown to have a role in endocytosis and vesicle assembly, respectively. Whether their homologs share the same function in trafficking in Apicomplexa warrants further interest. However, recent studies in antimalarial drug resistance have identified a strong genetic association between

drug resistance in *Plasmodium* and a locus containing a DUB related to USP7 [120]. Further studies will be required to validate the potential role of DUBs in the evolution of parasite drug resistance.

Expression profiles analysis [7,9] reveals that all the putative DUBs/DUBLs identified here were expressed in at least one stage of the *P. falciparum* life-cycle, providing additional evidence for a functional role in the parasite's life cycle.

Conclusion

The present study allowed for the identification of up to 114 proteins that are predicted to be involved in the UPS of *P. falciparum* and other Apicomplexa. All apicomplexans possess the complete machinery that is required to ubiquitylate proteins (i.e. ubiquitin and ubiquitin-like modifiers, E1 enzymes, E2 enzymes, E3 enzymes, and deubiquitinases). Ubiquitin and the common UBL modifiers SUMO, NEDD8, HUB1, URM1, and ATG8 were identified in apicomplexans. However, several UBLps are missing, such as SUMO variants and ATG12, although it has been suggested that autophagy is one of the parasite's death pathway [121]. Further investigations are required to elucidate the role of UBLps in apicomplexan biology. Our results also highlighted apicomplexans-specific features in enzymes involved in transferring ubiquitin and UBL modifiers to a target substrate. For example, two E2 variants found only in *P. falciparum* and *T. gondii* have up to triple the molecular weight of E2s found in other organisms, which could reflect adaptation of *Plasmodium* and *Toxoplasma* lineages. Finally, the superfamily of E3 ubiquitin ligases is very diverse, and several of the E3s predicted in apicomplexans do not find homolog in other eukaryotic organisms. Such proteins could have a cellular role directly linked to parasitic processes, such as invasion. The over-representation of RING E3 ligases that contain coiled-coil domains, known to play important role in host- pathogen interactions, supports this hypothesis.

An example of apicomplexan adaptation in the UPS is an apparently modified

APC/C (called APC-related). In yeast, SCF and APC are known to play a fundamental role in cell-cycle control. All proteins known to be involved in the SCF machinery are present in the *Plasmodium* genome. The presence of a modified APC complex in the atypical erythrocytic *Plasmodium* cycle is not surprising. The cell cycle in *Plasmodium* can be closely related to the early embryogenesis division observed in *D. melanogaster* with a depletion of key cell cycle regulators. During the first 13 cell divisions of a *D. melanogaster* zygote, the divisions are rapid, synchronous and occur in the absence of cytokinesis and detectable gap phase. These divisions result in the formation of a syncytial cell containing a large number of nuclei in a single cytoplasm. In *Plasmodium*, divisions during schizogony are rapid and asynchronous with up to four or five rounds (four rounds equals 16n and five rounds equals 32n) of DNA synthesis and mitosis during the trophozoite and early schizont stages. Cytokinesis takes place late in the cycle during the mature schizont stage. This is a major divergence from the classical cell- cycle events that consists of a linear succession of G1/S/G2/M phases. A proposed model for the *Plasmodium* cell-cycle control is presented in Figure A3.8 where the classical SCF complex could be involved in the control of the checkpoint between the gap G1 and S phase. The expression profile of the cullin analogue (highly expressed in the trophozoite and schizont stages) potentially involved in the parasite SCF complex reinforces this hypothesis. It will then be expected that the modified APC complex could regulate the atypical parasite cytokinesis. Interestingly, proteasome inhibitors have been shown to block parasite cell-cycle progression at the ring and late schizont stage

[122]. These observations help validate the hypothesis of two major parasite cell-cycle checkpoints at the G1/S and the cytokinesis phases.

The number of genes that our work predicted to encode components of the UPS in *Plasmodium* and other apicomplexa is a good estimate when compared to other eukaryotic organisms. Our results demonstrate that E3 ubiquitin/ubl ligases remain one of the most specific components of the UPS. The presence of numerous and diverse E3s, and in particular RING domain-containing proteins, suggests that target-specific ubiquitylation *via* E3 ligases is a complex and important part of cellular regulation in eukaryotic cells, including apicomplexan parasites. In terms of sequence homology, when counterparts exist, most of the ubiquitin ligases identified in Apicomplexa display divergences from their human- host counterpart. The domain architecture analysis for *P. falciparum*'s RING finger E3s revealed an abundance of coiled- coil domain. Since such domains have been shown to play a major role in signal transduction during the molecular cross-talk that occurs during a viral or a bacterial infection, our hypothesis is that E3 ubiquitin/UBL ligases in *P. falciparum* and other Apicomplexa are involved in pathogen virulence and/or pathogenicity.

The potential to treat apicomplexan parasites *via* drugs that target the proteasome component of the UPS has been established in *P. falciparum*. However, this potential to exploit these drugs is limited. The latest generations of drugs that

target the UPS have focused more on the specificity of the action of E3 ligases and DUB/DUBLs. Given the apparent apicomplexan diversity in these proteins, opportunities to develop small molecule inhibitors specific against the the apicomplexan-specific E3 ligase and DUB/ DUBLs offer some possibility for new, much needed, therapeutics for these devastating global diseases.

Materials and Methods

Proteome and domain motifs datasets

Full proteome datasets from the following thirteen organisms were downloaded*: P. falciparum*, *P. vivax*, P. *yoelii*, *P. berghei*, *P. chabaudi, C. parvum C. hominis, Toxoplasma gondii, S. cerevisiae, C. elegans, D. melanogaster, A. thaliana* and *H. sapiens*. Additional information regarding data sources and release information are given in table S4. Twenty-four Pfam motifs, found throughout the UPS, were selected (focusing on ubiquitin and UBLp, E1 and E1- like enzymes, E2 enzymes, E3 and E3-like enzymes and DUB/ DUBL enzymes). Their Pfam HMM profiles were downloaded from the Pfam_ls HMM library version 22.0 [123] at http://pfam. janelia.org (Pfam accession numbers: ubiquitin, PF00240; APG12, PF04110; MAP1_LC3, PF02991; UPF0185, PF03671; Urm1, PF09138; UBACT, PF02134; ThiF, PF00899; UQ_con, PF00179; zf-C3HC4, PF00097; zf-RING-like, PF08746; zf-MIZ, PF02891; Cullin, PF00888; U-box, PF04564; F-box, PF00646; HECT, PF00632; Josephin, PF02099; Mov34, PF01398; OTU, PF02338; DUF862, PF05903; WLM, PF08325; Peptidase C12, PF01090; Peptidase C48, PF02902; Peptidase C54, PF03416; UCH, PF00443).

Protein identification using HMM search

Components of the HMMER 2.3.2 package (released in October 2003) were used throughout the study [124]. This software can be freely downloaded from http://hmmer.janelia.org. The program hmmsearch was used to analyze domain

distribution among all proteome datasets and to extract sequences that carry the Pfam domains described above. HMM searches were run using a series of incrementally increasing threshold E- values, from E-value #1 to E-value #0.1, and results were checked for false positives. Threshold E-value #0.5 gave the best quality results, and thus was used in the present study.

Building colored distance matrices

Proteins carrying same domain motifs were pair aligned by BLASTP. Output bits scores were used as protein distance values: the higher the values are, the more proteins are similar and therefore could be evolutionary closed. All scores below 20 were set to 20. Values were scaled in the range [0,1] by normalizing using a global transformation (d = minimum score/score). A color- mapping matrix, from red to blue, was built, using MATLABH 7, from the normalized scale. The red color means ''highly divergent'', and blue means ''highly conserved''.

Building dendrogram trees

Protein sequences were aligned using ClustalW [125]. For each protein family, bootstrap parameters were set to 10 times the number of proteins in the set; for example, a set containing 100 proteins was boostrapped 1000 times. Dendrogram trees were generated from ClustalW multiple alignments, using the maxi- mum likelihood method in the PHYLIP package [126]. Trees were visualized with MEGA4 [127] and manually annotated.

Domain architecture of the predicted RING and RING-like E3 ubiquitin ligases

Our datasets of *P. falciparum* and yeast RING and RING-like E3 ubiquitin ligases were searched for functional domains using SMART [100]. Our dataset from yeast was used as a reference.

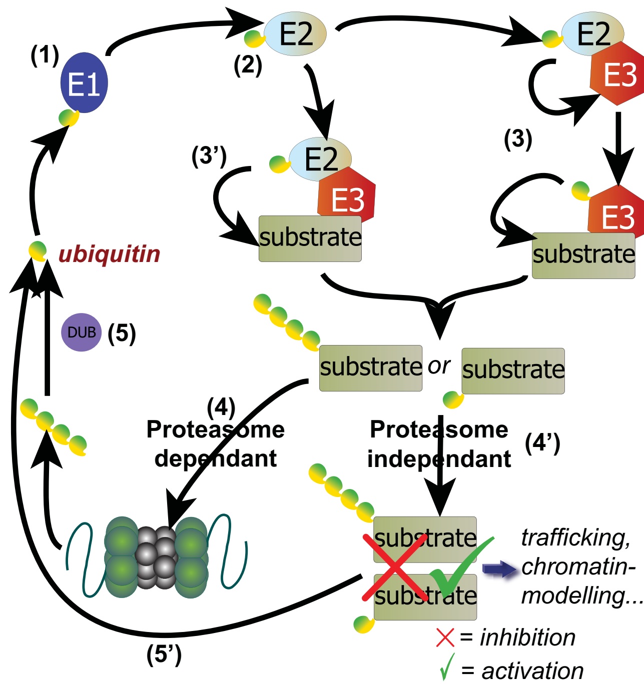


Figure A3.1. Representation of the ubiquitin-mediated pathways. (1) Ubiquitin is activated by E1 ubiquitin-activating enzyme, and (2) transferred to E2 ubiquitin-conjugating enzyme. Then, ubiquitin is either transferred to a monomeric E3 ubiquitin ligase that catalyzes ubiquityla- tion of the target substrate (3), or ubiquitinated E2 forms associates with the E3 to catalyze ubiquitylation of the substrate (39). Polyubiquitinated substrate can be targeted to the proteasome and destroyed (4). Poly or monoubiquitylation can also be an activation/repression signal (49) that modulates the substrate activity in several cellular processes such as trafficking or chromatin modeling. Finally, deubiquitinating enzymes (DUB) finally recycle ubiquitin proteins (5 and 59).

Table 1. Predicted number of UPS components in the 13 analyzed genomes.

Domains/Genomes	*P. falciparum*	*P. vivax*	*P. yoelii*	*P. chabaudi*	*P. berghei*	*T. gondii*	*C. parvum*	*C. hominis*	*S. cerevisiae*	*C. elegans*	*D. melanogaster*	*H. sapiens*	*A. thaliana*
Ubiquitin and Ubiquitin like													
Ubiquitin	6	6	3	2	3	6	7	5	7	8	10	25	31
APG12	1	0	0	0	0	0	0	0	1	1	1	1	2
MAP1_LC3	1	1	1	1	1	1	1	1	1	2	2	11	9
UPF0185	0	0	0	0	0	1	1	1	0	1	0	1	1
Urm1	1	1	1	1	1	1	1	1	1	0	1	2	1
Ubiquitin activating enzymes													
ThiF	8	8	8	9	9	11	8	6	8	8	10	16	14
UBACT													
Ubiquitin conjugating enzymes													
UQ_con	14	13	11	13	15	13	11	8	14	23	47	57	43
Ubiquitin ligases													
RING finger & RING-like	42	40	34	36	34	55	50	46	43	162	221	451	490
HECT	4	4	4	5	4	8	4	3	5	9	20	38	7
cullin	2	2	2	1	2	3	3	3	4	6	11	10	6
U-box	3	3	3	3	3	2	2	2	2	5	9	15	63
F-box	3	2	0	0	0	4	3	0	12	400	44	94	620
De-ubiquitinases													
OTU	3	3	1	0	1	10	2	1	2	5	6	15	12
Josephin	2	2	1	1	2	1	1	1	0	2	1	6	2
Mov34	6	6	6	4	5	7	4	3	4	8	11	19	15
DUF862	3	3	3	3	3	3	2	2	0	1	2	3	10
WLM	1	1	2	1	1	1	0	0	1	0	0	0	2
UCH	9	9	9	8	7	12	9	8	18	26	42	93	46
Peptidase_C12	2	2	2	2	2	2	2	2	1	4	5	4	17
Peptidase_C48	2	2	2	2	2	3	2	2	2	5	9	7	59
Peptidase_C54	1	1	1	0	0	1	1	1	1	2	2	15	2
Total	**114**	**109**	**94**	**92**	**95**	**145**	**114**	**96**	**127**	**678**	**454**	**883**	**1452**

doi:10.1371/journal.pone.0002386.t001

Table A3.1

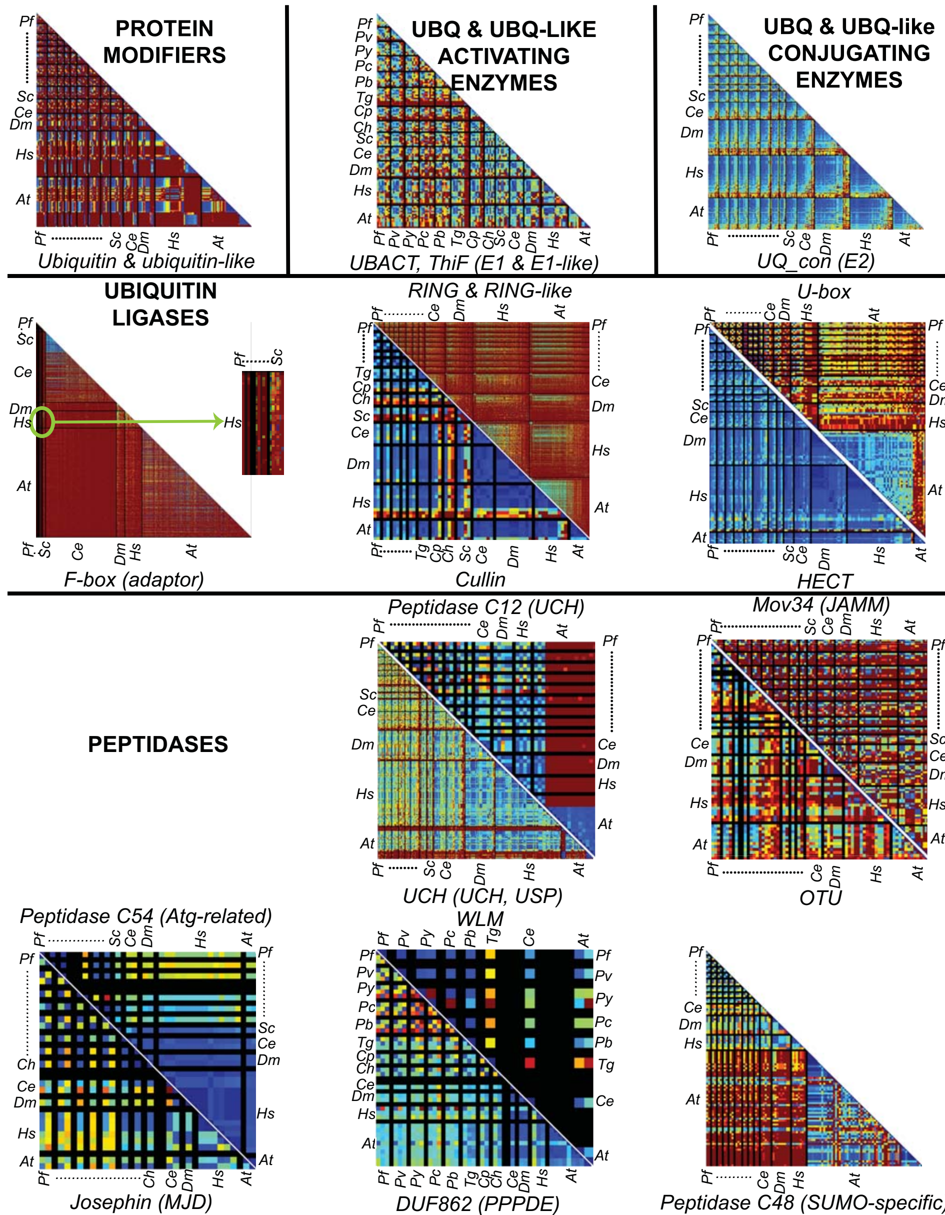


Figure A3.2. Color matrix representation of by-domain diversity for the 13 proteomes. For each domain, BLASTALL (BLASTP) was run with data from the 13 genomes. Normalized bit scores were plotted following a color scale ranging from ''0 = red = very different'' to ''1 = blue = identical''. All matrices are triangular. Black lines delimit species, with their respective initials are written on each side of the matrix. For each matrix, the order of the species is the following (from left to right or top to bottom): P. falciparum (Pf), P. vivax (Pv), P. yoelii (Py), P. chabaudi (Pc), P. berghei (Pb), T. gondii (Tg), C. parvum (Cp), C. hominis (Ch), S. cerevisiae (Sc), C. elegans (Ce), D. melanogaster (Dm), H. sapiens (Hs), A. thaliana (At). When the space did not allow writing initials for all species, the first and the last in the succession were indicated separated by dots.

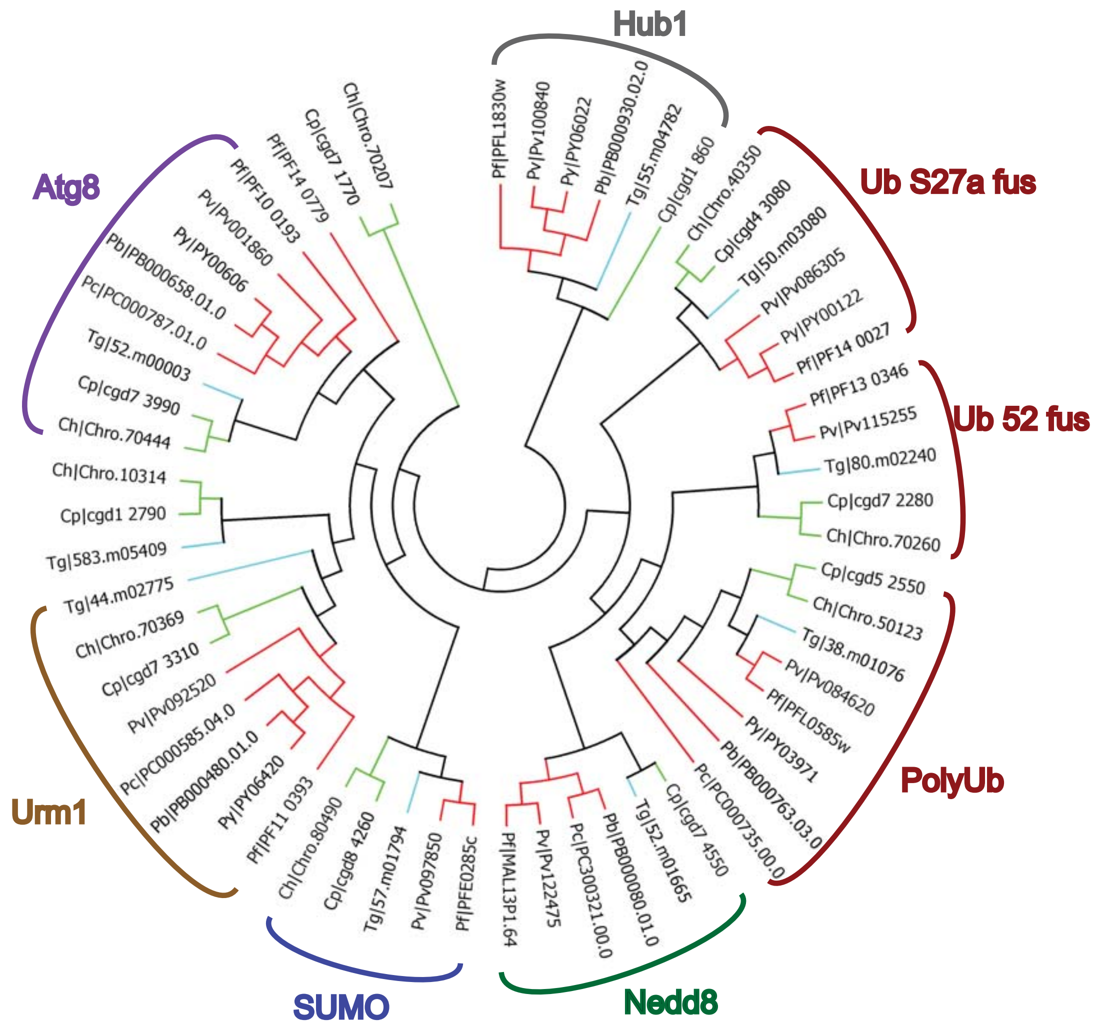


Figure A3.3. Dendrogram tree of ubiquitin and ubiquitin-like modifiers in *Plasmodium spp.*, *Cryptosporidium spp.* and *T. gondii.*

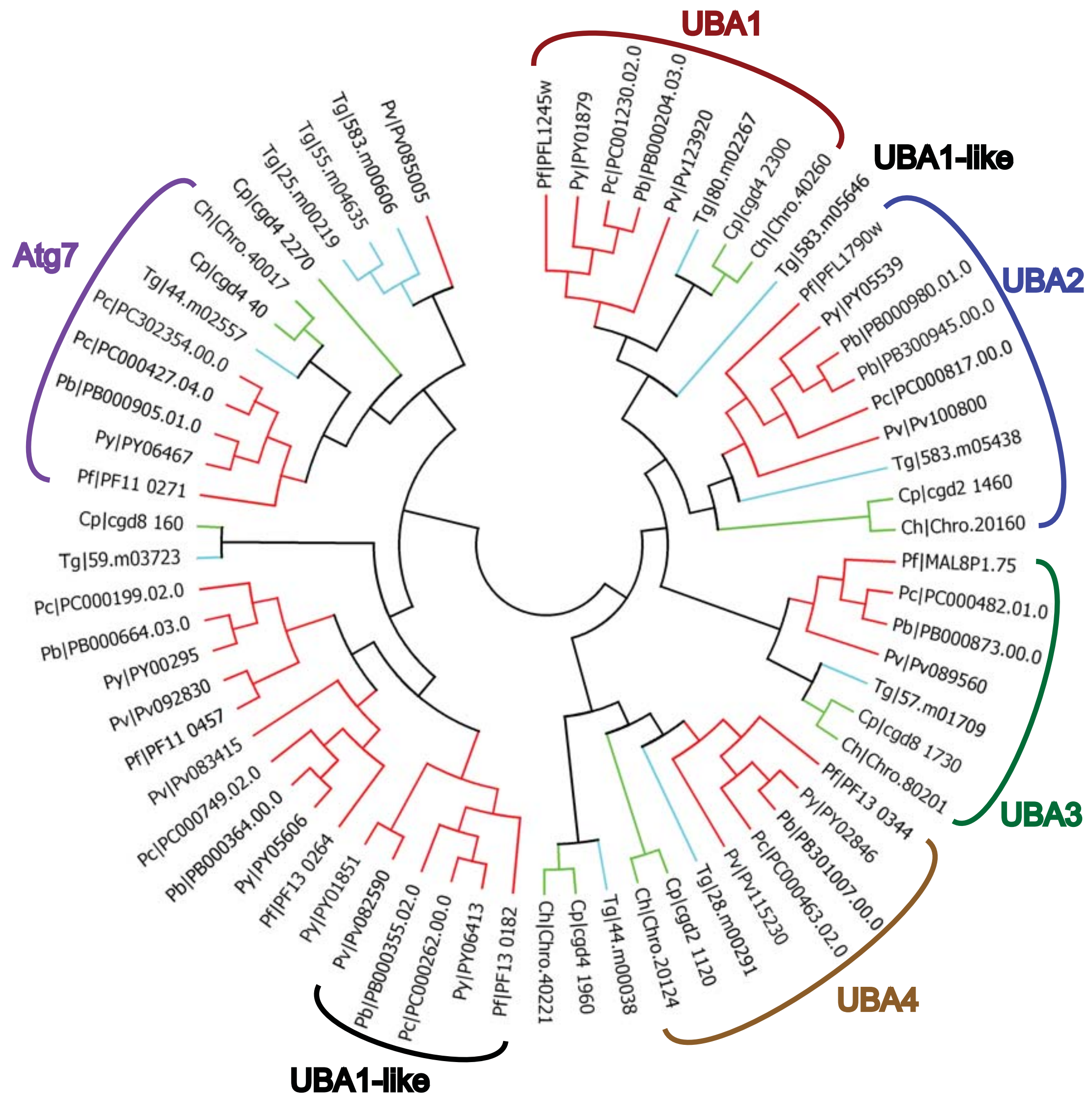


Figure A3.4. Dendrogram tree of ubiquitin and ubiquitin-like activating enzymes in *Plasmodium spp.*, *Cryptosporidium spp.* and *T. gondii*.

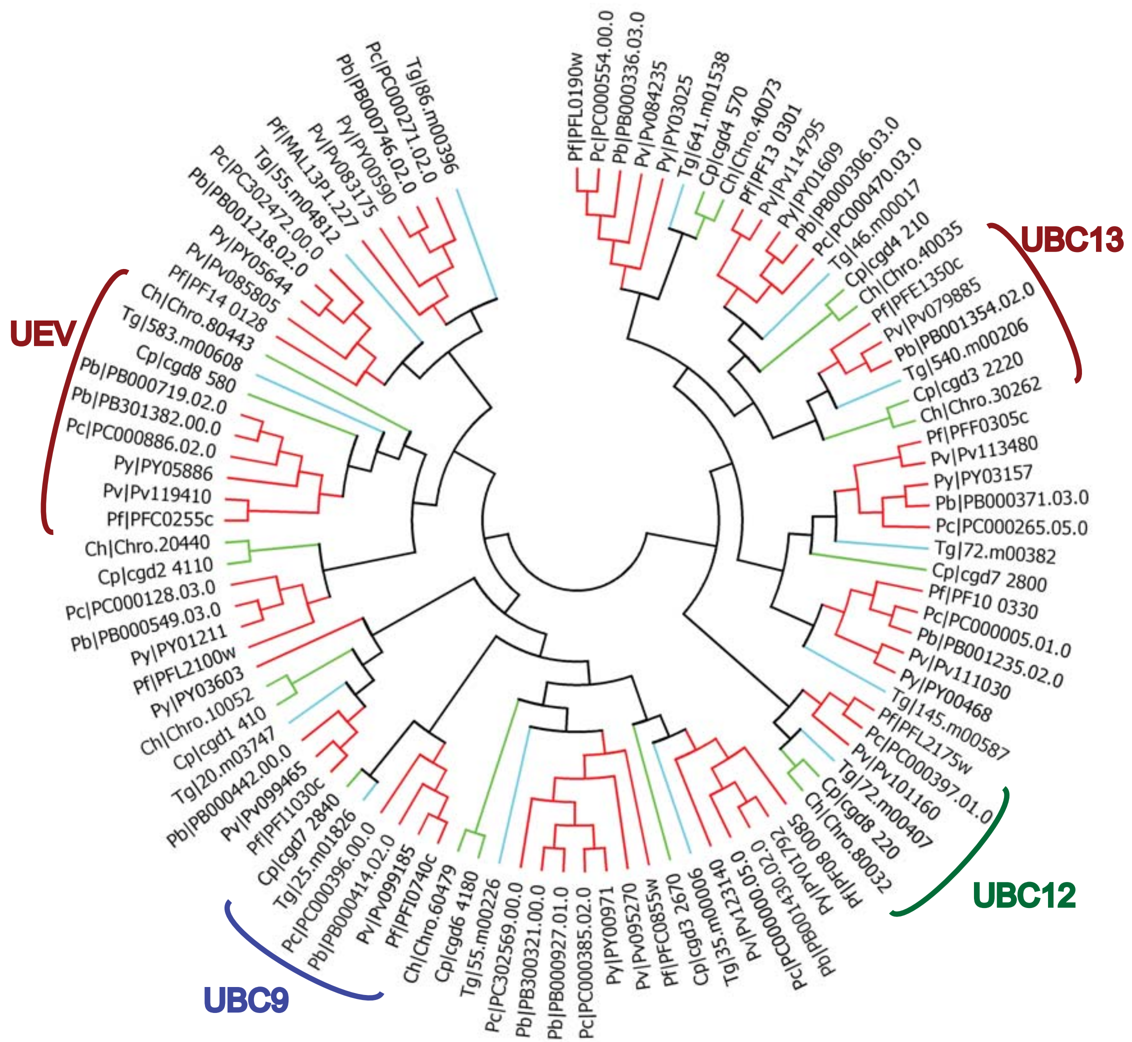


Figure A3.5. Dendrogram tree of ubiquitin and ubiquitin-like conjugating enzymes in *Plasmodium spp., Cryptosporidium spp.* and *T. gondii.*

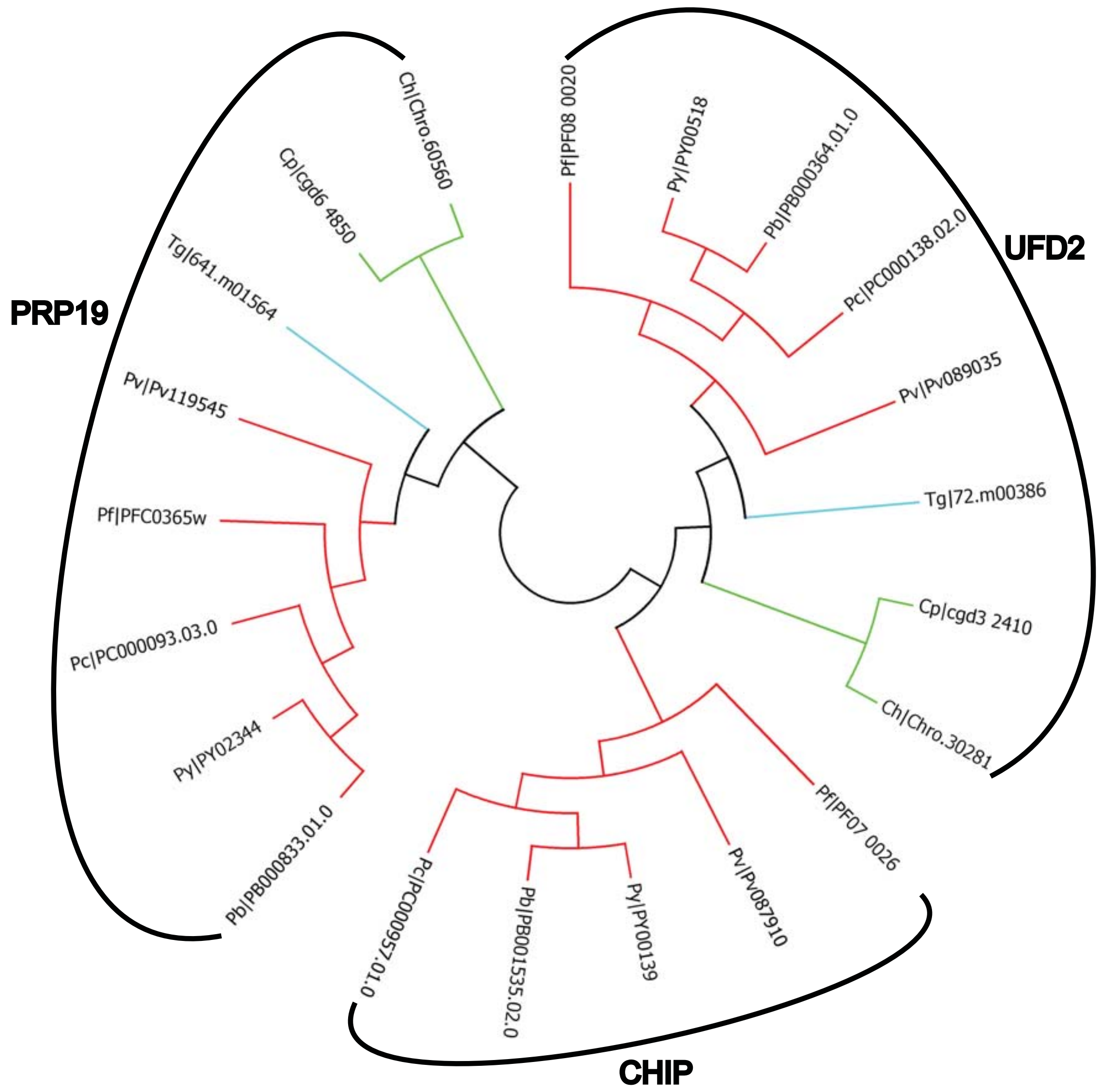


Figure A3.6. Dendrogram tree of ubiquitin and ubiquitin-like conjugating enzymes in *Plasmodium spp.*, *Cryptosporidium spp.* and *T. gondii*.

Table A3.2

Table 2. Annotated list of E3 ubiquitin and ubiquitin-like ligases in *P. falciparum*, with their homologs in *T. gondii*, *C. parvum* and *S. cerevisiae.*

Domain	*P. falciparum*	*T. gondii*	*C. parvum*	*S. cerevisiae*	Annotation
HECT	MAL8P1.23	86.m00385	none	TOM1	GO: mRNA transport; similar to UPL1/UPL2 in *A. thaliana*
HECT	PF11_0201	64.m00324	cgd8_1200	HUL5	similar to UPL6/UPL7 in *A. thaliana*
			cgd1_1920		
HECT	MAL7P1.19	80.m02344	none	UFD4	cytoplasmic E3 for degradation of ubiquitin fusion protein
HECT	PFF1365c	25.m01837	cgd7_4990	none	GO: cell proliferation; UPL5 in *A. thaliana*
		59.m03523			
		72.m00400			
Cullin	PF08_0094	80.m02207	cgd4_3150	CDC53	structural protein of SCF complexes
Cullin	PFF1445c	none	none	CUL8	possible role in anaphase progression;
U-box	PF08_0020	72.m00386	cgd3_2410	UFD-2	ubiquitin chain assembly factor E4; ubiquitin fusion degradation protein
U-box	PFC0365W	641.m01564	cgd6_4850	PRP19	splicing factor associated with the spliceosome
U-box	PF07_0026	none	none	none	similar to CHIP in *A. thaliana* and *H. sapiens*
RING finger	PFI0470w	none	cgd2_2410	SSM4	Deg1 signal-mediated degradation pathway; GO: mRNA turnover and stability
RING finger	MAL13P1.405	20.m03749	cgd8_4800	none	have homologs in *A. thaliana* only; unknown function
RING finger	PF14_0215	540.m00334	cgd8_2560	none	similar to HRD1-like; GO: ERAD pathway
RING finger	PFC0510w	50.m05636	cgd7_4170	HRD1	involved in the ERAD pathway
			cgd1_1790		
RING finger	PFE1490c	46.m00026	cgd8_3470	none	similarities with RIE1 in *A. thaliana* (seed development)
RING finger	PF10_0276	42.m00120	cgd7_4910	none	unknown function
		44.m02707			similar to ATL4 in *A. thaliana*
RING finger	PFF0355c	74.m00769	none	none	unknown function; found in Apicomplexa only
	PF14_0054				
RING finger	PF10_0072	80.m03951	cgd1_1950	none	unknown function
RING finger	PFF0755c	57.m01707	cgd2_2950	RKR1	GO: chromatine structure
RING finger	PFC0740c	57.m01858	cgd4_1360	none	unknown function; GO: cell growth regulation
			cgd3_1260		
RING finger	PFL0440c	540.m00204	cgd5_3990	ASI1	with ASI2 and ASI3 ensures the fidelity of SPS-sensor signalling
			cgd5_3970		
RING finger	PFE100w	none	cgd5_3900	PEX2	component of the CORVET complex
	PFI0805w				
RING finger	PFL1620w	none	none	DMA1	spindle position and orientation
				DMA2	
RING finger	PFC0175w	none	none	YKR017C	homologous to ariadne ubiquitin conjugating enzyme binding protein in *H. sapiens*
RING finger	PFF1325c	20.m03922	cgd2_1820	none	unknown function
RING finger	PFC0610c	583.m00699	cgd4_4310	PIB1	GO: endosomal trafficking, vacuolar trafficking
RING finger	PFF0165c	35.m01589	cgd2_880	BRE1	involved in histone H2B ubiquitination
RING finger	MAL7P1.155	none	none	none	unknown function; possible cytoskeleton-related
RING finger	PF14_0416	42.m00073	cgd6_3300	CWC24	element of the spliceosome
RING finger	PF14_0139	50.m03082	cgd8_3720	none	GO: cell proliferation
RING finger	PFL1705w	49.m03145	cgd7_4960	YER068W	Not-like; component of the CCR4-Not complex
RING finger	PFC0425w	none	none	none	unknown function; found in Apicomplexa only
	MAL13P1.224				
RING finger	PFL0275w	20.m03824	cgd7_3320	none	possible topoisomerase 1
RING finger	PFD0765w	none	none	none	unknown function
RING finger	PF11_0244	20.m03803	none	none	unknown function; found in Apicomplexa only
RING finger	PF10_0046	76.m01590	cgd3_2060	none	similar to CIP8 in *A. thaliana*
RING finger	PFF1180w	none	cgd1_2640	APC11	element of the anaphase promoting complex/cyclosome
RING finger	PFC0845c	none	cgd8_930	RBX1	element of the Skp1-Cullin-Fbox complex

Table A3.2 (cont.)

Table 2. cont.

Domain	*P. falciparum*	*T. gondii*	*C. parvum*	*S. cerevisiae*	Annotation
RING finger	PFB0440c	583.m05584	cgd3_3460	none	unknown function
RING finger	MAL13P1.216	641.m01484	none	RAD5	component of the SWI/SNF pathway
RING finger	PFL2440w	42.m00128	cgd4_140	RAD16	component of the SWI/SNF pathway
RING finger	PF10_0117	none	cgd2_1750	none	unknown function
RING finger	PFE0610c	641.m02557	cgd1_3300	TFB3	component of the nucleotide excision repair pathway
RING finger	PF13_0188	none	cgd5_1200	none	unknown function; found in Apicomplexa only
RING finger	MAL13P1.122	none	cgd5_400	none	unknown function; found in Apicomplexa only
RING finger	PFC0690c	none	cgd7_1170	YDR266C	role in partionning of cytoplasm
RING finger	PF11_0330	59.m03727	none	none	possible SUMO ligase

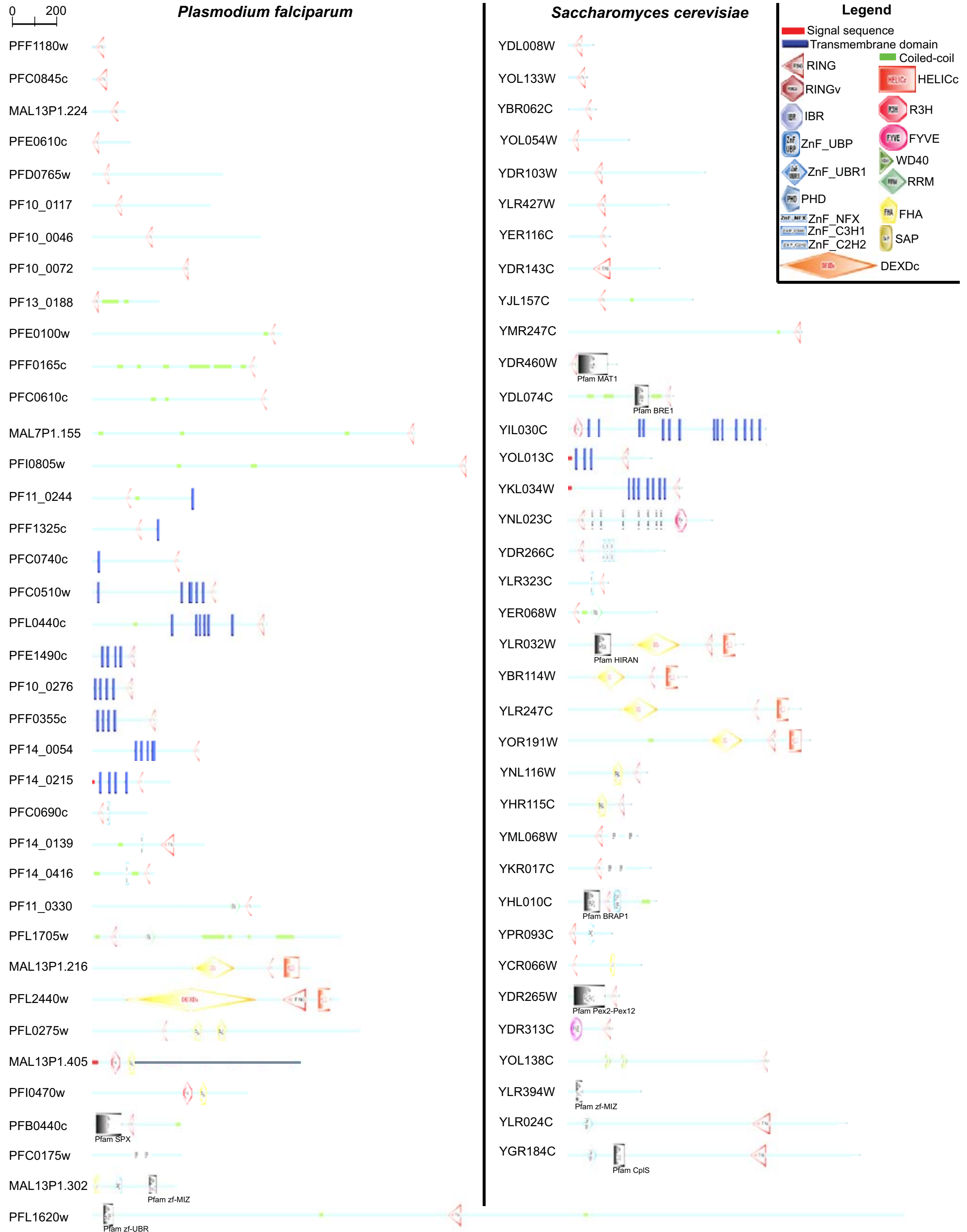

0
200
Plasmodium falciparum
PFF1180w
PFC0845c
MAL13P1.224
PFE0610c
PFD0765w
PF10_0117
PF10_0046
PF10_0072
PF13_0188
PFE0100w
PFF0165c
PFC0610c
MAL7P1.155
PFI0805w
PF11_0244
PFF1325c
PFC0740c
PFC0510w
PFL0440c
PFE1490c
PF10_0276
PFF0355c
PF14_0054
PF14_0215
PFC0690c
PF14_0139
PF14_0416
PF11_0330
PFL1705w
MAL13P1.216
PFL2440w
PFL0275w
MAL13P1.405
PFI0470w
PFB0440c
Pfam SPX
PFC0175w
MAL13P1.302
Pfam zf-MIZ
PFL1620w
Pfam zf-UBR
Saccharomyces cerevisiae
YDL008W
YOL133W
YBR062C
YOL054W
YDR103W
YLR427W
YER116C
YDR143C
YJL157C
YMR247C
YDR460W
Pfam MAT1
YDL074C
Pfam BRE1
YIL030C
YOL013C
YKL034W
YNL023C
YDR266C
YLR323C
YER068W
YLR032W
Pfam HIRAN
YBR114W
YLR247C
YOR191W
YNL116W
YHR115C
YML068W
YKR017C
YHL010C
Pfam BRAP1
YPR093C
YCR066W
YDR265W
Pfam Pex2-Pex12
YDR313C
YOL138C
YLR394W
Pfam zf-MIZ
YLR024C
YGR184C
Pfam CplS
Legend
Signal sequence
Transmembrane domain
Coiled-coil
RING
HELICc
RINGv
IBR
R3H
ZnF_UBP
FYVE
WD40
ZnF_UBR1
RRM
PHD
ZnF_NFX
FHA
ZnF_C3H1
ZnF_C2H2
SAP
DEXDc

Figure A3.7. Domain architecture of RING and RING-like E3 ligases in *P. falciparum* compared to *S. cerevisiae*. Domain architectures were retrieved using batch access to SMART database (http://smart.embl-heidelberg.de/smart/batch.pl). Domains in black and grey were retrieved from the Pfam database (http://pfam.sanger.ac.uk). Transmembrane segments are predicted with the TMHMM2 program, coiled coil regions are predicted with the Coils2 program, and signal peptides are predicted with the SignalP program. Legend: RING=RING finger; RINGv=C4HC3 RING-variant; IBR=in between RING fingers; ZnF_UBP=Ubiquitin Carboxyl-terminal Hydrolase-like zinc finger; ZnF_UBR1=Putative zinc finger in N-recognin, a recognition component of the N-end rule pathway; PHD = plant homeodomain zinc finger; ZnF_NFX = zinc finger domain repressor of transcription; ZnF_C3H1 = zinc finger domain; ZnF_C2H2 = zinc finger domain; DEXDc = DEAD-like helicases superfamily; HELICc = helicase superfamily c-terminal domain; R3H=Putative single-stranded nucleic acids-binding domain; FYVE=zinc finger present in Fab1, YOTB, Vac1, and EEA1; WD40=WD40 repeats; RRM = RNA recognition motif; FHA = Forkhead associated domain; SAP = Putative DNA-binding (bihelical) motif predicted to be involved in chromosomal organisation; Pfam MAT1=CDK-activating kinase assembly factor MAT1; Pfam BRE1=CDK-activating kinase assembly factor MAT1; Pfam HIRAN = HIP116, Rad5p N-terminal domain, found in the N-terminal regions of the SWI2/SNF2 proteins; Pfam BRAP2 = BRCA1-associated protein 2; Pfam Pex2-Pex12 = Pex2/Pex12 amino terminal region; Pfam ClpS = ATP-dependent Clp protease adaptor protein ClpS; Pfam SPX = SYG1, Pho81 and XPR1 domain; Pfam zf-UBR = Putative zinc finger in N-recognin (UBR box); Pfam zf-MIZ = MIZ/SP-RING zinc finger. The grey bar represents the primary structure of proteins.

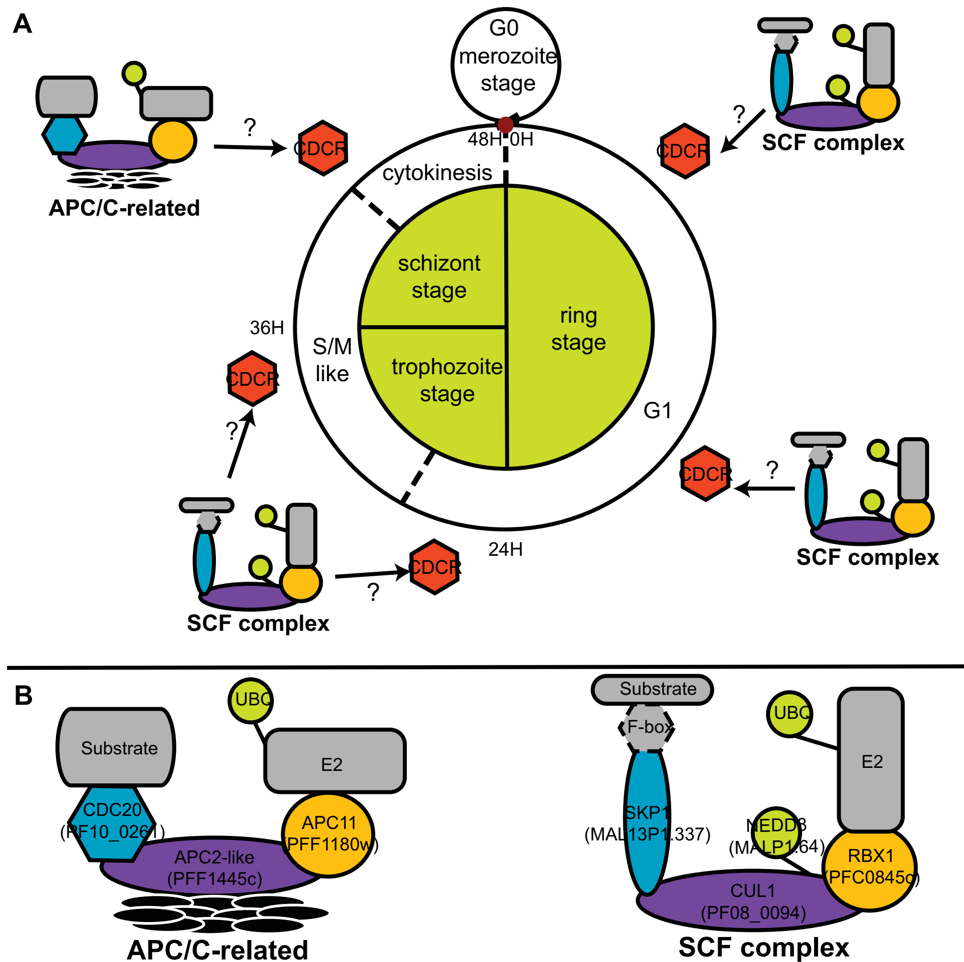


Figure A3.8. Proposed representation of *Plasmodium falciparum* erythrocytic cell cycle and two predicted regulatory complexes, a Skp1- Cullin-F-box (SCF) complex and an anaphase-promoting complex-related (APC/C-related). A: the 48 hours erythrocytic cell cycle of *P. falciparum*. Morphological stages are given in the inner circle; the outer circle proposes corresponding classical cell cycle phases. H=hours; CDCR = cell division cycle related. The red dot represents when invasion of red blood cells by merozoites occurs. B: detailed APC/C-related and SCF complexes in *P. falciparum.* Identifiers from PlasmoDB are given in parenthesis.

References

1. WHO, UNICEF (2005) World Malaria Report 2005. Geneva (Switzerland) and New York (USA).

2. Linares GE, Rodriguez JB (2007) Current status and progresses made in malaria chemotherapy. Curr Med Chem 14: 289–314.

3. Carlton JM, Angiuoli SV, Suh BB, Kooij TW, Pertea M, et al. (2002) Genome sequence and comparative analysis of the model rodent malaria parasite Plasmodium yoelii yoelii. Nature 419: 512–519.

4. Gardner MJ, Hall N, Fung E, White O, Berriman M, et al. (2002) Genome sequence of the human malaria parasite Plasmodium falciparum. Nature 419: 498–511.

5. Abrahamsen MS, Templeton TJ, Enomoto S, Abrahante JE, Zhu G, et al. (2004) Complete genome sequence of the apicomplexan, Cryptosporidium parvum. Science 304: 441–445.

6. Xu P, Widmer G, Wang Y, Ozaki LS, Alves JM, et al. (2004) The genome of Cryptosporidium hominis. Nature 431: 1107–1112.

7. Bozdech Z, Llinas M, Pulliam BL, Wong ED, Zhu J, et al. (2003) The transcriptome of the intraerythrocytic developmental cycle of Plasmodium falciparum. PLoS Biol 1: E5.

8. Le Roch KG, Johnson JR, Florens L, Zhou Y, Santrosyan A, et al. (2004) Global analysis of transcript and protein levels across the Plasmodium falciparum life cycle. Genome Res 14: 2308–2318.

9. Le Roch KG, Zhou Y, Blair PL, Grainger M, Moch JK, et al. (2003) Discovery of gene function by expression profiling of the malaria parasite life cycle. Science 301: 1503–1508.

10. Shock J, Fischer K, DeRisi J (2007) Whole-genome analysis of mRNA decay in Plasmodium falciparum reveals a global lengthening of mRNA half-life during the intra-erythrocytic development cycle. Genome Biology 8: R134.

11. Radke JR, Behnke MS, Mackey AJ, Radke JB, Roos DS, et al. (2005) The transcriptome of Toxoplasma gondii. BMC Biol 3: 26.

12. Gajria B, Bahl A, Brestelli J, Dommer J, Fischer S, et al. (2008) ToxoDB: an integrated Toxoplasma gondii database resource. Nucleic Acids Res 36: D553–556.

13. Florens L, Washburn MP, Raine JD, Anthony RM, Grainger M, et al. (2002) A

proteomic view of the Plasmodium falciparum life cycle. Nature 419: 520–526.

14. Jomaa H, Wiesner J, Sanderbrand S, Altincicek B, Weidemeyer C, et al. (1999) Inhibitors of the nonmevalonate pathway of isoprenoid biosynthesis as antimalarial drugs. Science 285: 1573–1576.

15. Wiesner J, Jomaa H (2007) Isoprenoid biosynthesis of the apicoplast as drug target. Curr Drug Targets 8: 3–13.

16. Ciechanover A (2005) Proteolysis: from the lysosome to ubiquitin and the proteasome. Nat Rev Mol Cell Biol 6: 79–87.

17. Hochstrasser M (1996) Ubiquitin-dependent protein degradation. Annu Rev Genet 30: 405–439.

18. Hochstrasser M (2000) Biochemistry. All in the ubiquitin family. Science 289: 563–564.

19. Gantt SM, Myung JM, Briones MR, Li WD, Corey EJ, et al. (1998) Proteasome inhibitors block development of Plasmodium spp. Antimicrob Agents Chemother 42: 2731–2738.

20. Lindenthal C, Weich N, Chia YS, Heussler V, Klinkert MQ (2005) The proteasome inhibitor MLN-273 blocks exoerythrocytic and erythrocytic development of Plasmodium parasites. Parasitology 131: 37–44.

21. Gonzalez J, Ramalho-Pinto FJ, Frevert U, Ghiso J, Tomlinson S, et al. (1996) Proteasome activity is required for the stage-specific transformation of a protozoan parasite. J Exp Med 184: 1909–1918.

22. Laney JD, Hochstrasser M (1999) Substrate targeting in the ubiquitin system. Cell 97: 427–430.

23. Herrmann J, Saguner AM, Versari D, Peterson TE, Chade A, et al. (2007) Chronic proteasome inhibition contributes to coronary atherosclerosis. Circ Res 101: 865–874.

24. Lake MW, Wuebbens MM, Rajagopalan KV, Schindelin H (2001) Mechanism of ubiquitin activation revealed by the structure of a bacterial MoeB-MoaD complex. Nature 414: 325–329.

25. Hershko A, Ciechanover A (1998) The ubiquitin system. Annu Rev Biochem 67: 425–479. 26. Hatakeyama S, Nakayama K-iI (2003) U-box proteins as a new

family of ubiquitin ligases. Biochemical and Biophysical Research Communications 302: 635–645.

27. Lechner E, Achard P, Vansiri A, Potuschak T, Genschik P (2006) F-box proteins everywhere. Curr Opin Plant Biol 9: 631–638.

28. Nijman SM, Luna-Vargas MP, Velds A, Brummelkamp TR, Dirac AM, et al. (2005) A genomic and functional inventory of deubiquitinating enzymes. Cell 123: 773–786.

29. Iyer LM, Koonin EV, Aravind L (2004) Novel predicted peptidases with a potential role in the ubiquitin signaling pathway. Cell Cycle 3: 1440–1450.

30. Wang J, Maldonado MA (2006) The ubiquitin-proteasome system and its role in inflammatory and autoimmune diseases. Cell Mol Immunol 3: 255–261.

31. Nalepa G, Rolfe M, Harper JW (2006) Drug discovery in the ubiquitin-proteasome system. Nat Rev Drug Discov 5: 596–613.

32. Richardson P, Mitsiades C, Schlossman R, Ghobrial I, Hideshima T, et al. (2007) The Treatment of Relapsed and Refractory Multiple Myeloma. Hematology 2007: 317–323.

33. Srikanth M, Davies FE, Morgan GJ (2008) An update on drug combinations for treatment of myeloma. Expert Opinion on Investigational Drugs 17: 1–12.

34. Uy GL, Trivedi R, Peles S, Fisher NM, Zhang QJ, et al. (2007) Bortezomib inhibits osteoclast activity in patients with multiple myeloma. Clin Lymphoma Myeloma 7: 587–589.

35. Rosebrock TR, Zeng L, Brady JJ, Abramovitch RB, Xiao F, et al. (2007) A bacterial E3 ubiquitin ligase targets a host protein kinase to disrupt plant immunity. Nature 448: 370–374.

36. Rytkonen A, Poh J, Garmendia J, Boyle C, Thompson A, et al. (2007) SseL, a Salmonella deubiquitinase required for macrophage killing and virulence. Proc Natl Acad Sci U S A 104: 3502–3507.

37. Gao G, Luo H (2006) The ubiquitin-proteasome pathway in viral infections. Can J Physiol Pharmacol 84: 5–14.

38. Balakirev MY, Jaquinod M, Haas AL, Chroboczek J (2002) Deubiquitinating function of adenovirus proteinase. J Virol 76: 6323–6331.

39. Willems AR, Schwab M, Tyers M (2004) A hitchhiker's guide to the cullin ubiquitin ligases: SCF and its kin. Biochim Biophys Acta 1695: 133–170.

40. Wong BR, Parlati F, Qu KB, Demo S, Pray T, et al. (2003) Drug discovery in the ubiquitin regulatory pathway. Drug Discovery Today 8: 746–754.

41. Vierstra RD (2003) The ubiquitin/26S proteasome pathway, the complex last chapter in the life of many plant proteins. Trends Plant Sci 8: 135–142.

42. Kipreos ET (2005) Ubiquitin-mediated pathways in C. elegans. WormBook. pp 1–24.

43. Ponder EL, Bogyo M (2007) Ubiquitin-like modifiers and their deconjugating enzymes in medically important parasitic protozoa. Eukaryot Cell 6: 1943–1952.

44. Horrocks P, Newbold CI (2000) Intraerythrocytic polyubiquitin expression in Plasmodium falciparum is subjected to developmental and heat-shock control. Mol Biochem Parasitol 105: 115–125.

45. Buchberger A (2002) From UBA to UBX: new words in the ubiquitin vocabulary. Trends Cell Biol 12: 216–221.

46. Ciechanover A, Schwartz AL (1994) The ubiquitin-mediated proteolytic pathway: mechanisms of recognition of the proteolytic substrate and involvement in the degradation of native cellular proteins. Faseb J 8: 182–191.

47. Oakley MS, Kumar S, Anantharaman V, Zheng H, Mahajan B, et al. (2007) Molecular factors and biochemical pathways induced by febrile temperature in intraerythrocytic Plasmodium falciparum parasites. Infect Immun 75: 2012–2025.

48. Gill G (2004) SUMO and ubiquitin in the nucleus: different functions, similar mechanisms? Genes Dev 18: 2046–2059.

49. Haracska L, Torres-Ramos CA, Johnson RE, Prakash S, Prakash L (2004) Opposing effects of ubiquitin conjugation and SUMO modification of PCNA on replicational bypass of DNA lesions in Saccharomyces cerevisiae. Mol Cell Biol 24: 4267–4274.

50. Hoege C, Pfander B, Moldovan GL, Pyrowolakis G, Jentsch S (2002) RAD6-

dependent DNA repair is linked to modification of PCNA by ubiquitin and SUMO. Nature 419: 135–141.

51. Hori T, Osaka F, Chiba T, Miyamoto C, Okabayashi K, et al. (1999) Covalent modification of all members of human cullin family proteins by NEDD8. Oncogene 18: 6829–6834.

52. Wu JT, Lin HC, Hu YC, Chien CT (2005) Neddylation and deneddylation regulate Cul1 and Cul3 protein accumulation. Nat Cell Biol 7: 1014–1020.

53. Kumar S, Yoshida Y, Noda M (1993) Cloning of a cDNA which encodes a novel ubiquitin-like protein. Biochem Biophys Res Commun 195: 393–399.

54. Furukawa K, Mizushima N, Noda T, Ohsumi Y (2000) A protein conjugation system in yeast with homology to biosynthetic enzyme reaction of prokaryotes. J Biol Chem 275: 7462–7465.

55. Wilkinson CR, Dittmar GA, Ohi MD, Uetz P, Jones N, et al. (2004) Ubiquitin- like protein Hub1 is required for pre-mRNA splicing and localization of an essential splicing factor in fission yeast. Curr Biol 14: 2283–2288.

56. Brailoiu GC, Dun SL, Chi M, Ohsawa M, Chang JK, et al. (2003) Beacon/ ubiquitin-like 5-immunoreactivity in the hypothalamus and pituitary of the mouse. Brain Res 984: 215–223.

57. Goehring AS, Rivers DM, Sprague GF Jr (2003) Attachment of the ubiquitin-related protein Urm1p to the antioxidant protein Ahp1p. Eukaryot Cell 2: 930–936.

58. Mizushima N, Noda T, Yoshimori T, Tanaka Y, Ishii T, et al. (1998) A protein conjugation system essential for autophagy. Nature 395: 395–398.

59. Besteiro S, Williams RA, Morrison LS, Coombs GH, Mottram JC (2006) Endosome sorting and autophagy are essential for differentiation and virulence of Leishmania major. J Biol Chem 281: 11384–11396.

60. Huang DT, Walden H, Duda D, Schulman BA (2004) Ubiquitin-like protein activation. Oncogene 23: 1958–1971.

61. Walden H, Podgorski MS, Schulman BA (2003) Insights into the ubiquitin transfer cascade from the structure of the activating enzyme for NEDD8. Nature 422: 330–334.

62. LaCount DJ, Vignali M, Chettier R, Phansalkar A, Bell R, et al. (2005) A protein

interaction network of the malaria parasite Plasmodium falciparum. Nature 438: 103–107.

63. Hofmann K, Bucher P, Kajava AV (1998) A model of Cdc25 phosphatase catalytic domain and Cdk-interaction surface based on the presence of a rhodanese homology domain. J Mol Biol 282: 195–208.

64. Komatsu M, Tanida I, Ueno T, Ohsumi M, Ohsumi Y, et al. (2001) The C- terminal region of an Apg7p/Cvt2p is required for homodimerization and is essential for its E1 activity and E1-E2 complex formation. J Biol Chem 276: 9846–9854.

65. Hershko A, Ciechanover A, Varshavsky A (2000) Basic Medical Research Award. The ubiquitin system. Nat Med 6: 1073–1081.

66. Hamilton KS, Ellison MJ, Barber KR, Williams RS, Huzil JT, et al. (2001) Structure of a conjugating enzyme-ubiquitin thiolester intermediate reveals a novel role for the ubiquitin tail. Structure 9: 897–904.

67. Pickart CM (2001) Mechanisms underlying ubiquitination. Annu Rev Biochem 70: 503–533.

68. Hall JR, Kow E, Nevis KR, Lu CK, Luce KS, et al. (2007) Cdc6 stability is regulated by the Huwe1 ubiquitin ligase after DNA damage. Mol Biol Cell 18: 3340–3350.

69. Johnson ES, Ma PC, Ota IM, Varshavsky A (1995) A proteolytic pathway that recognizes ubiquitin as a degradation signal. J Biol Chem 270: 17442–17456.

70. Varshavsky A (1997) The N-end rule pathway of protein degradation. Genes to Cells 2: 13–28.

71. Ju D, Xie Y (2006) A synthetic defect in protein degradation caused by loss of Ufd4 and Rad23. Biochem Biophys Res Commun 341: 648–652.

72. Koegl M, Hoppe T, Schlenker S, Ulrich HD, Mayer TU, et al. (1999) A novel ubiquitination factor, E4, is involved in multiubiquitin chain assembly. Cell 96: 635–644.

73. Ye Y, Meyer HH, Rapoport TA (2003) Function of the p97-Ufd1-Npl4 complex in retrotranslocation from the ER to the cytosol: dual recognition of nonubiquitinated polypeptide segments and polyubiquitin chains. J Cell Biol 162: 71–84.

74. Richly H, Rape M, Braun S, Rumpf S, Hoege C, et al. (2005) A series of ubiquitin binding factors connects CDC48/p97 to substrate multiubiquityla- tion and

proteasomal targeting. Cell 120: 73–84.

75. Ye Y (2006) Diverse functions with a common regulator: ubiquitin takes command of an AAA ATPase. J Struct Biol 156: 29–40.

76. Chen F, Mackey AJ, Stoeckert CJ Jr, Roos DS (2006) OrthoMCL-DB: querying a comprehensive multi-species collection of ortholog groups. Nucleic Acids Res 34: D363–368.

77. Vander Kooi CW, Ohi MD, Rosenberg JA, Oldham ML, Newcomer ME, et al. (2006) The Prp19 U-box crystal structure suggests a common dimeric architecture for a class of oligomeric E3 ubiquitin ligases. Biochemistry 45: 121–130.

78. Ajuh P, Sleeman J, Chusainow J, Lamond AI (2001) A direct interaction between the carboxyl-terminal region of CDC5L and the WD40 domain of PLRG1 is essential for pre-mRNA splicing. J Biol Chem 276: 42370–42381.

79. Chen CH, Kao DI, Chan SP, Kao TC, Lin JY, et al. (2006) Functional links between the Prp19-associated complex, U4/U6 biogenesis, and spliceosome recycling. Rna 12: 765–774.

80. Chan SP, Kao DI, Tsai WY, Cheng SC (2003) The Prp19p-associated complex in spliceosome activation. Science 302: 279–282.

81. Lu X, Legerski RJ (2007) The Prp19/Pso4 core complex undergoes ubiquitylation and structural alterations in response to DNA damage. Biochem Biophys Res Commun 354: 968–974.

82. Al-Olayan EM, Williams GT, Hurd H (2002) Apoptosis in the malaria protozoan, Plasmodium berghei: a possible mechanism for limiting intensity of infection in the mosquito. Int J Parasitol 32: 1133–1143.

83. Picot S, Burnod J, Bracchi V, Chumpitazi BF, Ambroise-Thomas P (1997) Apoptosis related to chloroquine sensitivity of the human malaria parasite Plasmodium falciparum. Trans R Soc Trop Med Hyg 91: 590–591.

84. Hurd H, Grant KM, Arambage SC (2006) Apoptosis-like death as a feature of malaria infection in mosquitoes. Parasitology 132 Suppl: S33–47.

85. Li D, Roberts R (2001) WD-repeat proteins: structure characteristics, biological function, and their involvement in human diseases. Cell Mol Life Sci 58: 2085–2097.

86. Smith TF, Gaitatzes C, Saxena K, Neer EJ (1999) The WD repeat: a common

architecture for diverse functions. Trends Biochem Sci 24: 181–185.

87. Murata S, Chiba T, Tanaka K (2003) CHIP: a quality-control E3 ligase collaborating with molecular chaperones. Int J Biochem Cell Biol 35: 572–578.

88. Zhang M, Windheim M, Roe SM, Peggie M, Cohen P, et al. (2005) Chaperoned ubiquitylation–crystal structures of the CHIP U box E3 ubiquitin
ligase and a CHIP-Ubc13-Uev1a complex. Mol Cell 20: 525–538.

89. Nakayama KI, Nakayama K (2005) Regulation of the cell cycle by SCF-type ubiquitin ligases. Semin Cell Dev Biol 16: 323–333. 90. Michel JJ, McCarville JF, Xiong Y (2003) A role for Saccharomyces cerevisiae Cul8 ubiquitin ligase in proper anaphase progression. J Biol Chem 278:22828–22837.

91. Arnot DE, Gull K (1998) The Plasmodium cell-cycle: facts and questions. Ann Trop Med Parasitol 92: 361–365.

92. Doerig C, Chakrabarti D, Kappes B, Matthews K (2000) The cell cycle in protozoan parasites. Prog Cell Cycle Res 4: 163–183.

93. Read M, Sherwin T, Holloway SP, Gull K, Hyde JE (1993) Microtubular organization visualized by immunofluorescence microscopy during erythrocytic schizogony in Plasmodium falciparum and investigation of post-translational modifications of parasite tubulin. Parasitology 106 (Pt 3): 223–232.

94. Glassner BJ, Mortimer RK (1994) Synergistic interactions between RAD5, RAD16 and RAD54, three partially homologous yeast DNA repair genes each in a different repair pathway. Radiat Res 139: 24–33.

95. Teng Y, Liu H, Gill HW, Yu Y, Waters R, et al. (2007) Saccharomyces cerevisiae Rad16 mediates ultraviolet-dependent histone H3 acetylation required for efficient global genome nucleotide-excision repair. EMBO Rep.

96. Goldfeder MB, Oliveira CC (2007) CWC24p, a novel Saccharomyces cerevisiae nuclear RING-finger protein, affects pre-snoRNA U3 splicing. J Biol Chem.

97. Duraisingh MT, Voss TS, Marty AJ, Duffy MF, Good RT, et al. (2005) Heterochromatin silencing and locus repositioning linked to regulation of virulence genes in Plasmodium faiciparum. Cell 121: 13–24.

98. Freitas-Junior LH, Hernandez-Rivas R, Ralph SA, Montiel-Condado D, Ruvalcaba-Salazar OK, et al. (2005) Telomeric heterochromatin propagation and histone acetylation control mutually exclusive expression of antigenic variation genes in

malaria parasites. Cell 121: 25–36.

99. Gissot M, Kelly KA, Ajioka JW, Greally JM, Kim K (2007) Epigenomic modifications predict active promoters and gene structure in Toxoplasma gondii. Plos Pathogens 3: 709–719.

100. Schultz J, Milpetz F, Bork P, Ponting CP (1998) SMART, a simple modular architecture research tool: Identification of signaling domains. Proceedings of the National Academy of Sciences 95: 5857–5864.

101. Mason JM, Arndt KM (2004) Coiled coil domains: stability, specificity, and biological implications. Chembiochem 5: 170–176.

102. Delahay RM, Frankel G (2002) Coiled-coil proteins associated with type III secretion systems: a versatile domain revisited. Mol Microbiol 45: 905–916.

103. Gao R, Lynn DG (2007) Integration of rotation and piston motions in coiled- coil signal transduction. J Bacteriol 189: 6048–6056.

104. He Y, Liu S, Jing W, Lu H, Cai D, et al. (2007) Conserved residue Lys574 in the cavity of HIV-1 Gp41 coiled-coil domain is critical for six-helix bundle stability and virus entry. J Biol Chem 282: 25631–25639.

105. Ohmura-Hoshino M, Goto E, Matsuki Y, Aoki M, Mito M, et al. (2006) A novel family of membrane-bound E3 ubiquitin ligases. J Biochem 140: 147–154.

106. Wing SS (2003) Deubiquitinating enzymes–the importance of driving in reverse along the ubiquitin-proteasome pathway. Int J Biochem Cell Biol 35: 590–605.

107. Borodovsky A, Kessler BM, Casagrande R, Overkleeft HS, Wilkinson KD, et al. (2001) A novel active site-directed probe specific for deubiquitylating enzymes reveals proteasome association of USP14. Embo J 20: 5187–5196.

108. Verma R, Aravind L, Oania R, McDonald WH, Yates JR 3rd, et al. (2002) Role of Rpn11 metalloprotease in deubiquitination and degradation by the 26S proteasome. Science 298: 611–615.

109. Haglund K, Sigismund S, Polo S, Szymkiewicz I, Di Fiore PP, et al. (2003) Multiple monoubiquitination of RTKs is sufficient for their endocytosis and degradation. Nat Cell Biol 5: 461–466.

110. Hicke L, Dunn R (2003) Regulation of membrane protein transport by ubiquitin

and ubiquitin-binding proteins. Annu Rev Cell Dev Biol 19: 141–172.

111. Kato M, Miyazawa K, Kitamura N (2000) A deubiquitinating enzyme UBPY interacts with the Src homology 3 domain of Hrs-binding protein via a novel binding motif PX(V/I)(D/N)RXXKP. J Biol Chem 275: 37481–37487.

112. Mizuno E, Iura T, Mukai A, Yoshimori T, Kitamura N, et al. (2005) Regulation of epidermal growth factor receptor down-regulation by UBPY- mediated deubiquitination at endosomes. Mol Biol Cell 16: 5163–5174.

113. Uzunova K, Gottsche K, Miteva M, Weisshaar SR, Glanemann C, et al. (2007) Ubiquitin-dependent proteolytic control of SUMO conjugates. J Biol Chem 282: 34167–34175.

114. Burnett B, Li F, Pittman RN (2003) The polyglutamine neurodegenerative protein ataxin-3 binds polyubiquitylated proteins and has ubiquitin protease activity. Hum Mol Genet 12: 3195–3205.

115. Scheel H, Tomiuk S, Hofmann K (2003) Elucidation of ataxin-3 and ataxin-7 function by integrative bioinformatics. Hum Mol Genet 12: 2845–2852.

116. Frickel EM, Quesada V, Muething L, Gubbels MJ, Spooner E, et al. (2007) Apicomplexan UCHL3 retains dual specificity for ubiquitin and Nedd8 throughout evolution. Cell Microbiol 9: 1601–1610.

117. Artavanis-Tsakonas K, Misaghi S, Comeaux CA, Catic A, Spooner E, et al. (2006) Identification by functional proteomics of a deubiquitinating/deNeddy- lating enzyme in Plasmodium falciparum. Mol Microbiol 61: 1187–1195.

118. Gardner RG, Nelson ZW, Gottschling DE (2005) Ubp10/Dot4p regulates the persistence of ubiquitinated histone H2B: distinct roles in telomeric silencing and general chromatin. Mol Cell Biol 25: 6123–6139.

119. Emre NC, Ingvarsdottir K, Wyce A, Wood A, Krogan NJ, et al. (2005) Maintenance of low histone ubiquitylation by Ubp10 correlates with telomere-proximal Sir2 association and gene silencing. Mol Cell 17: 585–594.

120. Hunt P, Afonso A, Creasey A, Culleton R, Sidhu AB, et al. (2007) Gene encoding a deubiquitinating enzyme is mutated in artesunate- and chloroquine- resistant rodent malaria parasites. Mol Microbiol 65: 27–40.

121. Totino PR, Daniel-Ribeiro CT, Corte-Real S, Ferreira-da-Cruz MD (2007) Plasmodium falciparum: Erythrocytic stages die by autophagic-like cell death under

drug pressure. Exp Parasitol.

122. Prudhomme J, McDaniel E, Ponts N, Bertani S, Fenical W, et al. (2008) Marine actinomycetes: a new source of compounds against the human malaria parasite. PLoS One.

123. Finn RD, Mistry J, Schuster-Bockler B, Griffiths-Jones S, Hollich V, et al. (2006) Pfam: clans, web tools and services. Nucleic Acids Research 34: D247–251.

124. Durbin R, Eddy S, Krogh A, Mitchison G (1998) Biological sequence analysis: probabilistic models of proteins and nucleic acids: Cambridge University Press.

125. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Research 22: 4673–4680.

126. Felsenstein J (1993) PHYLIP (Phylogeny Inference Package). 3.5c ed: Distributed by the author. Department of Genetics, University of Washington, Seattle.

127. Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) Software Version 4.0. Molecular Biology and Evolution 24: 1596–1599.

APPENDIX A4:

Genomics and integrated systems biology in Plasmodium falciparum: a path to malaria control and eradication. *Parasite Immunology*. **34:** 50–60. (2012)

Karine G. Le Roch, **Duk-Won Doug Chung** & Nadia Ponts

Department of Cell Biology and Neuroscience, University of California Riverside, Institute for Integrative Genome Biology, and Center for Disease Vector Research, Riverside, CA, USA

SUMMARY:

The first draft of the human malaria parasite's genome was released in 2002. Since then, the malaria scientific community has witnessed a steady embrace of new and powerful functional genomic studies. Over the years, these approaches have slowly revolutionized malaria research and enabled the comprehensive, unbiased investigation of various aspects of the parasite's biology. These genome-wide analyses delivered a refined annotation of the parasite's genome, delivered a better knowledge of its RNA, proteins and metabolite derivatives, and fostered the discovery of new vaccine and drug targets. Despite the positive impacts of these genomic studies, most research and investment still focus on protein targets, drugs and vaccine candidates that were known before the publication of the parasite genome sequence. However, recent access to next-generation sequencing technologies, along with an increased number of genome-wide applications, is expanding the impact of the parasite genome on biomedical research, contributing to a paradigm shift in research activities that may possibly lead to new optimized

diagnosis and treatments. This review provides an update of Plasmodium falciparum genome sequences and an overview of the rapid development of genomics and system biology applications that have an immense potential of creating powerful tools for a successful malaria eradication campaign.

Keywords: epigenetics, genomics, malaria, proteomics, systems biology, transcriptomics

INTRODUCTION

Malaria is a mosquito-borne disease caused by a eukaryotic protozoan parasite of the genus Plasmodium. With up to one million deaths per year, malaria remains one of the deadliest infectious diseases in the world and has been recognized as one of the strongest forces driving evolutionary selection in the human genome. There are five different species of Plasmodium that can infect humans; *P. falciparum, P. vivax, P. malariae, P. ovale* and more recently *P. knowlesi. P. falciparum* is responsible for the most severe malignant malaria leading to death, especially in children under 5 years old in sub-Saharan African countries. In addition to its deleterious effects on human health, malaria has a significant impact on poverty and is a major impediment to economic development.

Despite the success of an eradication campaign after the Second World War in developed countries (Europe and North America) and a significant reduction of cases in developing parts of the world, malaria is still widespread in all tropical and subtropical areas and can still affect more than 40% of the world population. Recent advances in treatments – these include the development of new combinational therapies, the increased use of bed nets and improved insecticides – have contributed to the reduction of detected infections in select African countries and revived hope that malaria is a disease that can be eradicated.

While there is still no approved vaccine, malaria is a curable disease. Since ancient times, traditional medicinal plants have been used to treat malaria. Currently,

quinine and artemisinin, isolated from the bark of the Peruvian Cinchona succirubra tree and the Chinese plant Artemisia annua, respectively, remain the most effective available antimalarials. Unfortunately, artemisinin-based combination therapies (ACTs), recently adopted as our last resort in combating malaria infection, are already challenged by ACT-resistant strains detected in south-east Asia. With the spread of parasite resistance to all current antimalarial drugs, successful control and eradication will only be achieved if new efficient tools and cost-effective antimalarial strategies are developed.

When the near-completed sequence of the genome of the human malaria parasite P. falciparum was first published (1), the scientific community predicted that it would accelerate the discovery of new drug targets and vaccine strategies. Almost a decade later, this is still a work-in-progress. The genome sequence of the malaria parasite has nonetheless provided the foundation for modern biomedical research. The goal is now to transform our increasing knowledge of the parasite's biology into actual improvements of human health. Such achievement requires an integrated understanding of both the pathogen's and the host's responses to infection. In this review, we present an over- view of the P. falciparum genome as well as recent advances in genomics and systems biology that have led to major improvements in the understanding of the pathogen. We discuss the impact of these approaches on the development of new therapeutic strategies as well as exploring the long- term goal of global malaria eradication.

THE MALARIA PARASITE GENOME

The first draft of the P. falciparum genome was published after 7 years of international effort. The genome was sequenced using the Sanger method and chromosome shotgun strategy (1). The size of the genome was initially estimated at 22Æ8 Mb separated into 14 chromosomes and 5300 protein-encoding predicted genes. In addition to its nuclear genome, the parasite contains 6- and 35-kb circular DNA found in its mitochondria and plant- related apicoplast, respectively. Today, the P. falciparum genome remains to be the most AT-rich genome. The overall (A + T) composition is 80.6% and can rise to 95% in introns and intergenic regions. After almost 9 years of coordinated genome curation efforts, the complete genome sequence is defined as haploid and 23.26 Mb in size. It contains 6372 genes and 5524 protein-coding genes (genome version: 06-01-2010, http:// plasmodb.org/plasmo/). Approximately half of these genes have no detected sequence homology with any other model organism. Despite recent access to comparative and functional genomics studies and the completion of genome sequencing of more than eight Plasmodium species, the cellular function of most of the parasite genes remains obscure.

Over the past few years, extensive resequencing efforts have been successfully undertaken to identify genes and genetic traits associated with parasite's drug resistance and severity of the clinical outcomes. Initial sequencing surveys of genetic variation across the P. falciparum genomes were published in 2007 (2–4).

Volkman et al. (2) sequenced high-quality draft genomes of three parasite laboratory clones (the reference sequenced as 3D7, HB3 and Dd2) isolated from different parts of the world. Their work alone identified 26845 single-nucleotide polymorphisms (SNPs) at a frequency of one SNP every 780 bases between the three clones and an additional 37 039 insertion–deletions (indels) between 3D7 and HB3. They further extended their genotyping to 12 P. falciparum strains and 20 genomic regions from 54 worldwide P. falciparum isolates. Results were consistent with initial genetic diversity studies that were performed using whole-genome microarray analysis (5). All together, they identified more than 46937 SNPs (one every 446 bases in average) across the whole genome. High levels of SNPs were detected in genes involved in antigenic variation as well as genes involved in drug resistance. These data were further con- firmed by the survey of approximately 60% of P. falciparum predicted genes (3) and a shotgun sequencing strategy of a Ghanaian clinical isolate (4). Taken together, these reports identified a high number of rare SNP variants and suggested that most SNPs have yet to be discovered. As a whole, these results underscore the importance of creating comprehensive maps of genetic diversity in P. falciparum field isolates. These SNPs are strongly suspected to be markers for various phenotypic traits such as virulence or resistance to drugs.

Recent advances in next-generation sequencing (NGS) technologies are enabling fast and affordable production of large amounts of genome sequence information. These technologies are already opening new perspectives of functional genomics in the field of primary, applied and clinical malaria research.

HIGH-THROUGHPUT SEQUENCING IN MALARIA RESEARCH

After 30 years of dominance of first-generation 'Sanger' dideoxy sequencing, the past 5 years have seen the explosion of NGS methods. Next-generation sequencing has transformed the field of whole-genome sequencing and analysis. Unlike Sanger sequencing, NGS avoids the need for bacterial cloning and therefore bypasses associated biases. For example, AT- or GC-rich regions are often toxic to bacteria and difficult to reliably read with cloning-based sequencing. This issue is of major importance in the case of the P. falciparum's extremely AT-rich genome. The major leap forward from NGS is the ability to pro- duce an enormous amount of data within small volumes; a tremendous number of DNA fragments, up to 2 billion short reads per instrument run, can be sequenced in parallel.

Three main NGS platforms have been commercialized over the past 5 years: the Roche 454 (Roche Life Sciences, Branford, CT, USA), the Applied Biosystems SOLiD (Applied Biosystems , Carlsbad, CA, USA) and finally the Illumina (formally known as Solexa) Genome Analyzer and Hi-Seq platforms. The Illumina sequencers are currently capable of producing up to 2 billion reads per run – this value is continually increasing as a result of constant improvements of reagents and consumables – with a recommended read length of 35–100 bp. This technique is by far the most successful NGS method to sequence the P. falciparum genome. Many variations of the technique were developed specifically for the sequencing of the (A

+ T)-rich genome of the malaria parasite (6–8) (Figure A4.1). Over the last couple of years only, many studies have used Illumina's NGS technology to identify SNPs and other mutations linked to drug resistance in the murine malaria parasite P. chabaudi (9,10) and the human malaria parasite P. vivax (11). Other analyses have contributed to the characterization of the P. falciparum transcriptome with the discovery of new splicing events (12–14) and transcription start sites (15). Finally, Illumina's NGS technology was used to discover atypical features of P. falciparum's chromatin (6,16) and various epigenetic events (7).

Currently, the future of high-throughput sequencing seems to be leaning towards single-cell sequencing applications. Going further, third-generation sequencing (TGS) technologies propose to use single molecules as direct tem- plates for sequencing (techniques so far under development at Helicos Biosciences and Pacific Biosciences). These TGS technologies should simplify the sample preparation procedure, avoid the bias introduce by DNA amplification and library preparation and be even more affordable than their predecessors. Nevertheless, the power of high-throughput sequencing also represents one of the major pitfalls for the analysts. The high-throughput and depth of quantitative measurements produced by NGS and TGS technologies come at the cost of producing sophisticated algorithms and software tools capable of accurately examining millions to billions of reads. The data generated by these methods are complex, novel and abundant. The computational and statistical analysis of raw outputs is the tricky step where

incorrect normalization and processing can yield misleading conclusions. Novel methods of quantitative analysis are constantly under development and testing. There is yet no consensus on which analytical approach is the most accurate, particularly for the Plasmodium genome.

The avalanche of whole-genome data over the past few years generated an immense source of knowledge that still requires maturing and processing. Nevertheless, in the near future, these powerful genomic approaches will certainly catalyse the transformation of this biological knowledge into viable therapeutic strategies. Single-cell sequencing will accelerate the genotyping of strains in patients' blood sample or other field isolates. Comparative genomics then will be an important source of information regarding the evolution and dynamics of malaria parasites' populations. Ultimately, such knowledge could be used for accurate diagnosis and targeted treatment of patients.

FROM GENOME TO FUNCTION

The ultimate goal of functional genomics and systems biology in the malaria field is to discover the properties of the parasite to pinpoint its weaknesses and identify long- lasting antimalarial strategies. The amalgamation of large- scale genome-wide analyses (microarrays, deep sequencing, quantitative mass spectrometry, epigenome mapping, computational modelling, etc.) has been used to mine Plasmodium's genome in an unbiased manner and identify the genetic

elements that may be targeted in the fight against malaria (Figure A4.2). Here, we present major contributions of the main 'omics' to the malaria field.

Transcriptomics

Microarray-based large-scale analyses of P. falciparum's transcripts led to the discovery of expressed genes, their functional association with the various stages of the parasite life cycle and their involvement in particular bio- logical processes with a high degree of accuracy (17–20). More recent sequencing-based studies such as RNA-seq confirmed these initial microarray experiments and showed promising results on the prediction of new splicing events. These studies also allowed the identification of new open reading frames with their untranslated flanking regions (12–14,21). Moreover, transcriptome analyses in P. falciparum field isolates identified previously unknown factors involved in pathogenesis and immune evasion (22–26). Finally, analyses of transcription profiles of variant surface antigens identified patterns that are specific to the parasite's sexual stages and could be relevant for new vaccine interventions (27,28). In addition to mRNA-related transcriptomics, noncoding protein RNA (ncpRNA) transcriptome has been analysed (29). In eukaryotes, structural ncpRNA is known to participate in the regulation of diverse bio- chemical pathways, e.g. transcription, translation, epigenetic regulations, cell differentiation and proliferation. In P. falciparum, 604 putative ncpRNAs were detected (30–32) and were showed to form a complex regulatory network. All together, these latest analyses suggest that P. falciparum

ncpRNAs may play a critical role in determining antigenic variation and virulence mechanisms (29).

Proteomics

Previous proteomics (33–35) and interactomics (36) studies have confirmed and complemented the functional annotations proposed based on transcriptome profiling. Numerous proteomics analyses surveyed stage-specific proteins and investigated as potential drug targets, including sex- specific proteins in male and female gametocytes that could be utilized for transmission blocking strategies (37). Parasite surface proteins (parasite proteins that are exported to the surface of the infected red blood cells) also represent new potential antigens for rational vaccine development (33–35,38,39). Genomics, cell biology and proteomics studies identified a conserved protein export motif, the PEXEL motif, which has been reported in as many as 400 proteins. Most of these proteins are expressed during the erythrocytic stages. They are thought to be involved in creating knob structures on the surface of red blood cells and are often associated with cytoadherence and antigenic variation. A complete understanding of their function and regulation will therefore be critical to disrupt one of the most pathological effects of Plasmodium infections.

Metabolomics

In an effort to improve functional annotation and increase our understanding of the parasite's biology, a number of research groups have been leveraging biochemical metabolic profiling and metabolomics strategies (40). Metabolomics is the study of the entire repertoire of metabolites, i.e. small molecules such as amino acids, sugars and fatty acids that are known to perform critical functions in various biological processes. Correlation analyses of transcriptomics, proteomics and metabolomics data are a powerful way to identify new metabolic pathways as well as genes that encode for specific enzymatic functions (41,42). While the study of metabolomics in Plasmodium is still in its infancy, it has already uncovered important bio- logical insights with possible implications in terms of adaptation, evolution and host–pathogen interactions (43– 45).

Miscellaneous functional genomics

Functional genomics suffers from the lack of tools to analyse the malaria parasite's genome. For example, gene silencing using RNAi cannot be used in Plasmodium because the machinery does not exist in the parasite; gene knockout experiments are time-consuming processes not compatible with large-scale high-throughput analyses. However, in the past few years, a transposon-based mutagenesis approach in Plasmodium has been developed (46). A Plasmodium-specific selection cassette was added to the lepidopteran transposon piggyBac and transfected in parasites together with a transposase-containing helper plasmid (47).

Random insertional mutants are obtained by multiple integrations of the transposon at TTAA recognition sites. Recent studies used piggyBac-based approaches to validate candidate parasite-specific secreted proteins (48) or identify genes that are essential for the parasite's proliferation (49). Used in combination with other genomics and proteomics analyses, piggyBac-based strategies could provide a better understanding of the parasite's biology and its interactions with its hosts.

Accessibility and usability of knowledge

The data of large-scale and functional genomic analyses must be accessible and intelligible for practical and efficient usage. The task belongs to the informatics and bioinformatics fields that can provide the necessary tools. Up to now, data depositary banks and the Web-based data- bases such as PLASMODB (http://plasmodb.org/plasmo/) have greatly facilitated the access, the comprehensive visualization and the analysis of large data sets. Gene predictions and annotations, new drug target identifications and discoveries of vaccine candidates all resulted from various genome-wide analyses. However, it is critical that such resources remain well maintained and free for maximized accessibility. Indeed, a systemic view of the malaria parasite's biology can only be achieved with the successful integration and accessibility of the data from various origins.

GENOMICS AND GENE REGULATION

Genome-wide transcriptomic and proteomic approaches have been extensively utilized to identify regulatory elements that control gene expression in the parasite's genome. For example, one approach consisted of a DNA motif discovery framework based on the detection of dependencies between microarray-based transcriptomic data and the presence of DNA motifs within the 5¢ untranslated regions of genes (50). This approach identified *in silico* 21 potential motifs found in approximately 2700 genes expressed in P. falciparum. The method, how- ever, may not perform very well on highly degenerated or atypical motifs. Another approach consists of identifying quantitative trait loci that are involved in gene expression variations (eQTLs) in various clones of P. falciparum (51). Using tiling arrays, Gonzales et al. identified hot spots of sequence polymorphisms spread throughout the entire genome that control the expression of nearly 18% of the genes from a distance. More recently, potential regulatory sequences found at nucleosome-free regions of DNA have been identified using formaldehyde-assisted isolation of regulatory elements (FAIRE) coupled with NGS at high resolution and large scale (13). In addition, ChIP-on-chip experiments using histone H4-specific antibodies were used to discover nucleosome-bound sequences and also suggest the potential presence of nucleosome-free regulatory elements (52). These kinds of studies have provided a considerable amount of data in just a few years. The mechanisms that P.falciparum uses to regulate gene expression remain nonetheless elusive. Indeed, the remark- able changes in steady-state mRNA levels, with a tightly

coordinated cascade of transcripts throughout the parasite life cycle, remain challenging to comprehend.

Transcriptional vs. post-transcriptional mechanisms of gene regulation

The core transcriptional machinery that drives RNA polymerase II-dependent transcription (53) and 27 Apicomplexan AP2 (ApiAP2) plant-related transcription factors (54,55) have been identified as major regulators of parasite gene expression. All together, the proteins involved in the transcriptional machinery (including general transcription factors), along with ApiAP2-specific transcription factors, represent <2% of the total genome. Considering the P. falciparum's genome size, twice this amount is required for a classical 'transcription factor-mediated' model of gene regulation (53,56,57). Thus, either more atypical and elusive regulators remain to be discovered, or gene regulation in Plasmodium is not so classically based on the coordinated action of specific positive / negative regulators only.

The initial characterization of the ApiAP2 transcription factor family was a major step forward understanding key regulators in Plasmodium (58). However, their exact role in the parasite's biology remains to be determined. Further- more, recent studies have started to underline that the malaria parasite may have adapted and optimized its mechanisms of transcriptional regulation for its lifestyle. Indeed, in contrast to what is observed in other eukaryotes, environmental perturbations and small molecule inhibitors do not cause specific transcriptional changes in P.

falciparum, as revealed by genome-wide analyses of parasite expression profiles in response to stress (59–61). The concept of transcriptional rigidity in Plasmodium was recently conceived (59). Parasites subjected to chemical or environmental stresses do not specifically compensate for the stress-targeted pathways at the transcriptional level; instead, they exhibit a strong cell cycle arrest and an induction of genes involved in general (nonspecific) stress responses and sexual differentiation. Taken together, these studies highlight an unusual method of transcriptional regulation with a limited capacity for positive or negative feedback mechanisms.

Additional analyses of mRNA vs. protein profiles show significant varying time shifts between transcript and protein levels. These data enforce that extensive post-transcriptional mechanisms of gene regulation may have important roles during parasite development (38,62,63). Following these latest observations, the characterization of protein complexes involved in translational repression (64) and whole-genome analysis of mRNA decay rates strongly supports the idea that post-transcriptional regulation may be an important mechanism for gene regulation in P. falciparum (65).

Genome-wide chromatin structure

Recent studies highlight the importance of key chromatin components that regulate parasite development (53,66,67). A large number of chromatin-modifying

complexes have recently been identified [reviewed in (68)] leading to the hypothesis that malaria parasites may, in large part, be subject to epigenetic mechanisms that control gene expression. Epigenetic modifications involve reversible modifications to DNA or proteins that do not affect the genome sequence but are inheritable and modulate gene expression as well as other biological processes (69). In the human malaria parasite, heterochromatic silencing was shown to control mutually exclusive expression of antigenic variation genes in the parasite (66,67,70). More recently, several studies investigated the genome-wide distribution of various euchromatic / heterochromatic histone marks. Lopez-Rubio et al. (71) used high-resolution ChIP-on-chip to map the positions of trimethylated lysine 9 histone H3 (H3K9me3), trimethylated lysine 4 histone H3 (H3K4me3) and acetylated lysine 9 histone H3 (H3K9ac) in P. falciparum. They showed that H3K9me3, a silencing mark, has an atypical distribution in the P. falciparum genome; H3K9me3 is indeed confined within the subtelomeric and limited chromosome internal regions that are closely associated with genes involved in antigenic variation. On the contrary, the active marks, H3K4me3 and H3K9ac, dis- play a broad distribution across the genome. Additional ChIP-on-chip profiling of histones H3, H3K4me3, H3K9me3 and H3K9ac confirmed that heterochromatin is restricted to the regions of the genome that contain the variant surface antigen families in P. falciparum (72). Further analyses also confirmed the colocalization of the hetero- chromatin protein 1 to H3K9me3, along with their association with regions of the genome that code for Plasmodium virulence factors

(73,74). Global histone mass spectrometry analysis also confirmed the prevalence of active acetylated histone marks compared with inhibitory methylated ones (75). All together, these results suggest an atypical euchromatin / heterochromatin structure in the malaria parasite; active chromatin is prevalent genome-wide, whereas silencing marks are less frequent although they seem to play a significant role in transcriptional control of genes involved in phenotypic variation and pathogenesis.

Chromatin remodelling across the parasite's cycle

Upon transcriptional activation, eukaryotic promoter nucleosomes are partially removed by sliding or disassembly, allowing DNA to become directly accessible to transcription factors (76,77) and other DNA-binding proteins. Indeed, various genome-wide analyses provided evidence that active regulatory regions and gene promoters of highly expressed genes are, at least partially, nucleosome-depleted (78,79). Nucleosome positioning is typically driven by active remodelling complexes or dictated by the sequence of the binding DNA itself (80). In particular, poly(dA:dT) tracks are harder to bend around histones, and nucleosomes have a lower affinity for such sequences (81,82). Considering the extremely high AT content of P. falciparum's genome, this latest observation may have important consequences for the parasite's biology.

Recently, the nucleosome landscape of P. falciparum was investigated in reference to gene regulation by using two genome-wide methods, both coupled to

NGS: (i) FAIRE to isolate protein-free DNA; and (ii) micrococcal nuclease- assisted isolation of mononucleosomal elements (MAINE) to isolate DNA fragments associated with histones (13). The combined use of both methods provides a comprehensive view of the chromatin structure across P. falciparum's genome. Complementary opposite results were obtained by both methods (nucleosome-bound regions were identified with MAINE, and interspacing nucleosome-free regions were identified with FAIRE) as reflected by a high negative correlation coefficient. Nucleosomes were pre- dominantly found within coding sequences, which have a higher GC content relative to noncoding regions. Similar results were obtained using an anti-histone H4 ChIP-on- chip (52) and are consistent with three recent analyses of nucleosome distribution in human, worms and flies, demonstrating a marked preference of nucleosomes for exons (83–85). Moreover, Ponts et al. demonstrated the occurrence of massive and atypical genome-wide nucleosome depletion at the early trophozoite stage ('open' transcriptionally active state) before a progressive repacking of chromatin, while the cycle progresses towards the schizont stage ('closed' transcriptionally silent stage) of the intra-erythrocytic cycle. These results are consistent with a dip in canonical histone abundance at trophozoite stage that was previously observed using mass spectrometry (33,38). Nonetheless, those changes in chromatin structure do not fully explain the changes of mRNA steady-state levels across the intra-erythrocytic cycle, with the exception of ring stage- or exo-erythrocytic-specific genes (13,14). Such observations are consistent with recent data, demonstrating that mRNA

steady-state levels and transcription rate do not correlate for about half of the parasite's genes (86). In that case, genes could be massively transcribed at the trophozoite stage followed by major regulations at the post-transcriptional level. This hypothesis finds support in the fact that the parasite's preinitiation complex interacts with both stage-specific 'active' and 'inactive' promoters (87) and that mRNA decay rates are significantly lengthened during the intra-erythrocytic cycle suggesting major post-transcriptional regulations (65).

To further complement these data, Bartfai et al. used a ChIP-seq approach to show that, unlike in other eukaryotes, the histone H2A variant H2A.z is a constant and ubiquitous feature of all intergenic regions throughout the parasite erythrocytic cycle (7). As H2A.z is usually involved in chromatin destabilization and active transcription in eukaryotes (88–90), these results are consistent with a transcriptionally permissive state of P. falciparum's chromatin during the asexual cycle. In addition, previous mass spectrometry studies showed that, unlike the abundant and more variable canonical histones, H2A.z is present at low and constant level throughout the parasite's cycle (33,38). This observation, combined with the high sensitivity of H2A.z to MNase digestion (88,89), is consistent with the relative nucleosome depletion that was observed by MAINE-seq and ChIP-on-chip in noncoding regions of the genome (6,52). Given the low levels of H2A.z and its extreme sensitivity to MNase digestion, H2A.z-containing nucleosomes can mostly be detected by targeted and specific immunoprecipitation-based sample

enrichments. Quantitative measurements in such experiments, however, imply a careful normalization of histone variant levels vs. canonical histones.

All together, these data confirm an unusual parasite chromatin structure and speculate an active transcriptional state during most of the erythrocytic cycle with a few exceptions such as clonally variant genes as well as genes known to be essential to early erythrocytic and sexual stage differentiation. It is therefore possible that part of transcriptional regulation in P. falciparum could occur during elongation rather than initiation. This hypothesis is supported by the recent observation that H2A.z seems to facilitate the passage of the RNA polymerase II (90).

As a whole, it is likely that a complex model that integrates major transcription factors, histone positioning and turnover, histone variants and modifications, and post- transcriptional regulations will be required to fully comprehend the Plasmodium's unusual gene expression machinery. Given the major differences observed in parasite epigenetic features compared with all other eukaryotic organisms, inhibitors developed against Plasmodium-specific epigenetic enzymes have a strong potential for new therapeutic strategies against P. falciparum.

GENOMICS FOR THE DISCOVERY OF NEW ANTIMALARIAL STRATEGIES

Many of the current drug therapies are based on chemically engineered variants of already known antimalarial compounds (e.g. aminoquinolines and /

or peroxides). Intensive exploration of the P. falciparum genome has lead to the identification of parasite-specific essential genes or metabolic pathways that could be targeted for rational drug designs (18,23,60,62,91–93). For example, a fosmidomycin-sensitive mevalonate-independent pathway of isoprenoid biosynthesis, absent from higher eukaryotes and located in the plant plastid-like parasite organelle namely the apicoplast, was identified in P. falciparum (94). Along with the discovery of new drug targets, the discovery of mechanisms of drug resistance has been significantly refined using genome-wide analysis. Typically, mechanisms of drug resistance are determined by examining the genetic differences between sensitive and resistant strains. The best-studied case of drug resistance in P. falciparum is chloroquine resistance (CQR). Chloroquine resistance is mediated by a transporter gene (Pfcrt) and by the multi- drug resistance gene (Pfmdr1). The discovery of the genes associated with CQR took years of heavy molecular, epidemiology and genetic studies. Research is still ongoing to fully comprehend CQR in the parasite. Today, whole-genome analytic tools provide the capability of analysing rap- idly the genetic changes that occur in the genome of a resistant strain. Whole-genome scanning using tiling microarrays has already been used for this purpose. For example, initial analyses found relatively abundant copy number variations in P. falciparum -resistant strains (5). Point mutations in the apicoplast were recently associated with resistance to clindamycin, a drug used in combination with quinine for the treatment of malaria in pregnant women and infants (95). Another striking example of the power of genomics in drug discovery is the

identification of a potent drug by cell proliferation–based compound screening (96) followed by the discovery of one of its targets using high-density microarrays and sequencing (97). Without the advent of genomics, such a process would have required many years. All together, it is likely that these genome-wide approaches will soon uncover mechanism of drug resistance including emerging resistance of artemisinin. To further highlight the power of genomic studies for the discovery of new effective antimalarial strategies, a recent genome-wide SNP analysis identified regions of high and low recombination frequencies (hot spots and cold spots). The study found that loci associated with known drug resistance are likely found in hot spots (98). A practical consequence of these observations for a long-term antimalarial strategy is that drug targets should be encoded by genes located in cold spots rather than hot spots.

Genome-wide proteomic analyses have generated a high number of potential new vaccine candidates. Several new parasite surface antigens have recently been discovered throughout the malaria parasite life cycle (33–35,38,39). The availability of the P. falciparum genome has also allowed the development of new genome-wide protein microarrays to probe human plasma from individuals before and after malaria season. These novel genome-wide methods have already delivered important insights into parasite proteins associated with immunoreactivity in an unbiased manner (99–101). It is highly probable that these studies will soon improve our understanding of the molecular basis of protective immunity and

facilitates the discovery of new efficient vaccine strategies.

All together, the increasing number and performances of genome-wide technologies is transforming the scientific field. Genomics and systems biological studies have already contributed significantly to a better understanding of the malaria parasite's biology. Most importantly, they have generated an exceptional pipeline of new drugs tar- gets and vaccine candidates. The challenge today will be to bring these achievements to efficient and affordable antimalarial products.

WHAT FUTURE FOR MALARIA ERADICATION?

Constantly diminishing costs of high-throughput genomics and DNA sequencing technologies have dramatically changed the way science is being done over the past few years. These changes should soon transform the way we assess genetic risk factors and the way we think about medicine, treatments and possible disease eradication in developing countries.

Genomics has already greatly contributed to our under- standing of the malaria parasite and the human genetic factors that influence the susceptibility and the response to both malaria and antimalarial drugs / vaccines. The full integration of the newly acquired knowledge to the disease strategy will undoubtedly provide bases to prevent the resurgence of malaria [e.g. Peru (95)] and the arising and spread of resistances by analysing parasites' population dynamics

and evolution (e.g. resistances to artemisinin in south-east Asia). The catalogue of putative drugs and drug targets has already increased together with the panel of candidates for vaccination strategies. Beyond drug discovery, genomics was recently proven to be particularly efficient in the discovery of a drug mechanism of action within a 2-year time span by coupling drug screening and genomics (97). Ultimately, diagnostic and curative treatment could be improved by genotyping both the host and the infecting parasite. Such optimized treatment would contribute to a better use of drugs and a better management of the spread of resistances.

Integrated genomics approaches, where all aspects of the disease management are considered, have the power to lead to better malaria strategy. However, it is only with free and open access to genome databases, continuing technology development, accurate identification of genetic and environmental factors, continuing financial investments, development of close private–public partnerships, collaboration of governmental and nongovernmental organizations, academic institutions, individuals and good pol- icy decisions that the benefits of genomics and systems biological studies can be fully utilized and manifested to achieve new drug and vaccine targets that have emerged from genomic analyses and bring us closer to the eradication of malaria.

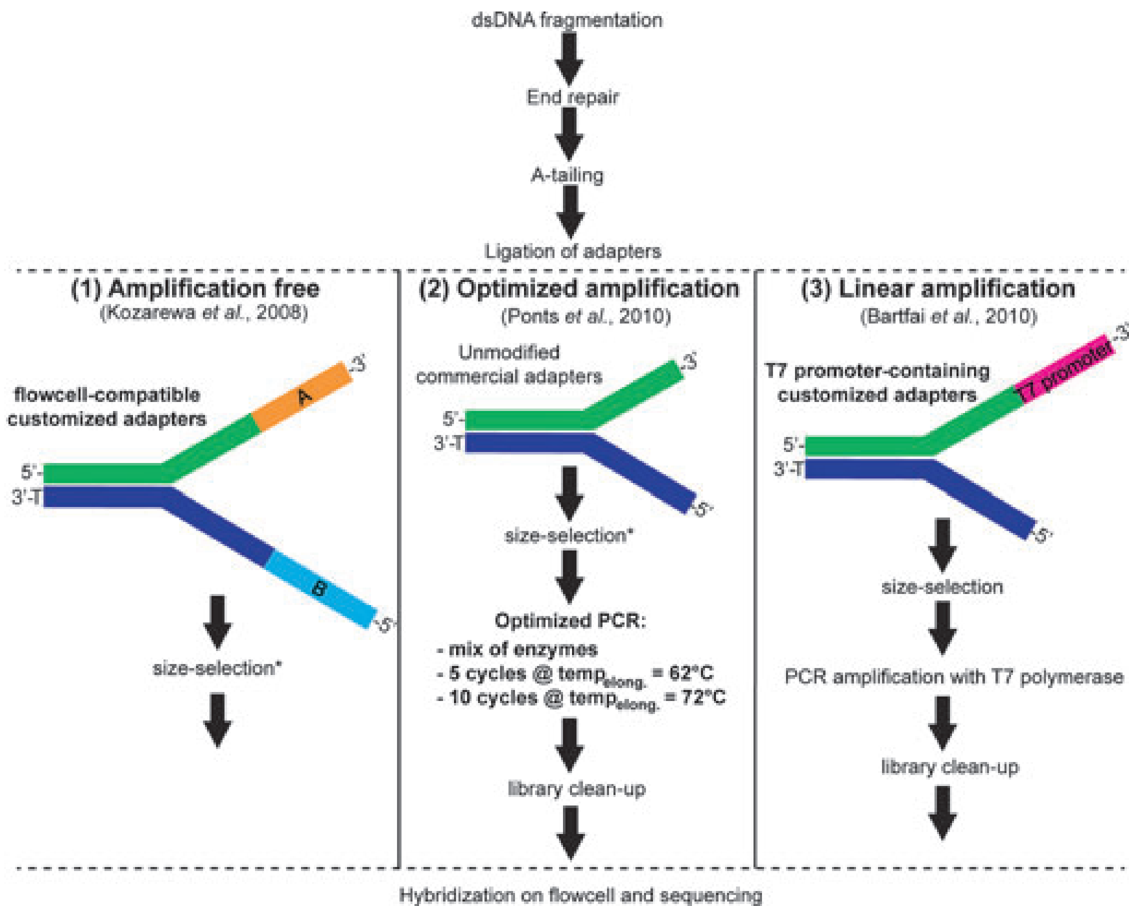


Figure A4.1 Methods for next-generation sequencing (NGS) of the human malaria parasite's genome. Library preparation for sequencing on Illumina NGS platforms usually involves the fragmentation of the dsDNA by a method of choice (e.g. sonication), end repair (fill in) and the addition of a single A tail on the 3¢ end of each DNA strand before specific and oriented double-stranded adapters can be ligated. These chimera fragments are then PCR-amplified using specific primers that contain sequences compatible with the oligos attached to the Illumina sequencing flowcell. In the amplification-free protocol (8), the PCR step is entirely skipped. The adapters are thus customized to contain the flowcell-compatible sequences in the absence of PCR. After adapter ligation, the library can be eventually size selected and purified prior to hybridization on the flowcell and sequencing. In both the optimized amplification (6) and the linear amplification proto- cols (7), the PCR step is performed with a polymerase that is different from the original Illumina protocol. The linear amplification proto- col uses the T7 polymerase and therefore requires that one of the adapters contains the sequence corresponding to the T7 promoter. In the case of the optimized amplification protocol, a mix of enzymes is used to perform the PCR [for example, a combination of the NEB Phusion and the Takara Ex Taq has been successfully used in the past (6,16)]. The parameters for PCR cycling have also been adapted to the very high AT content of the genome.

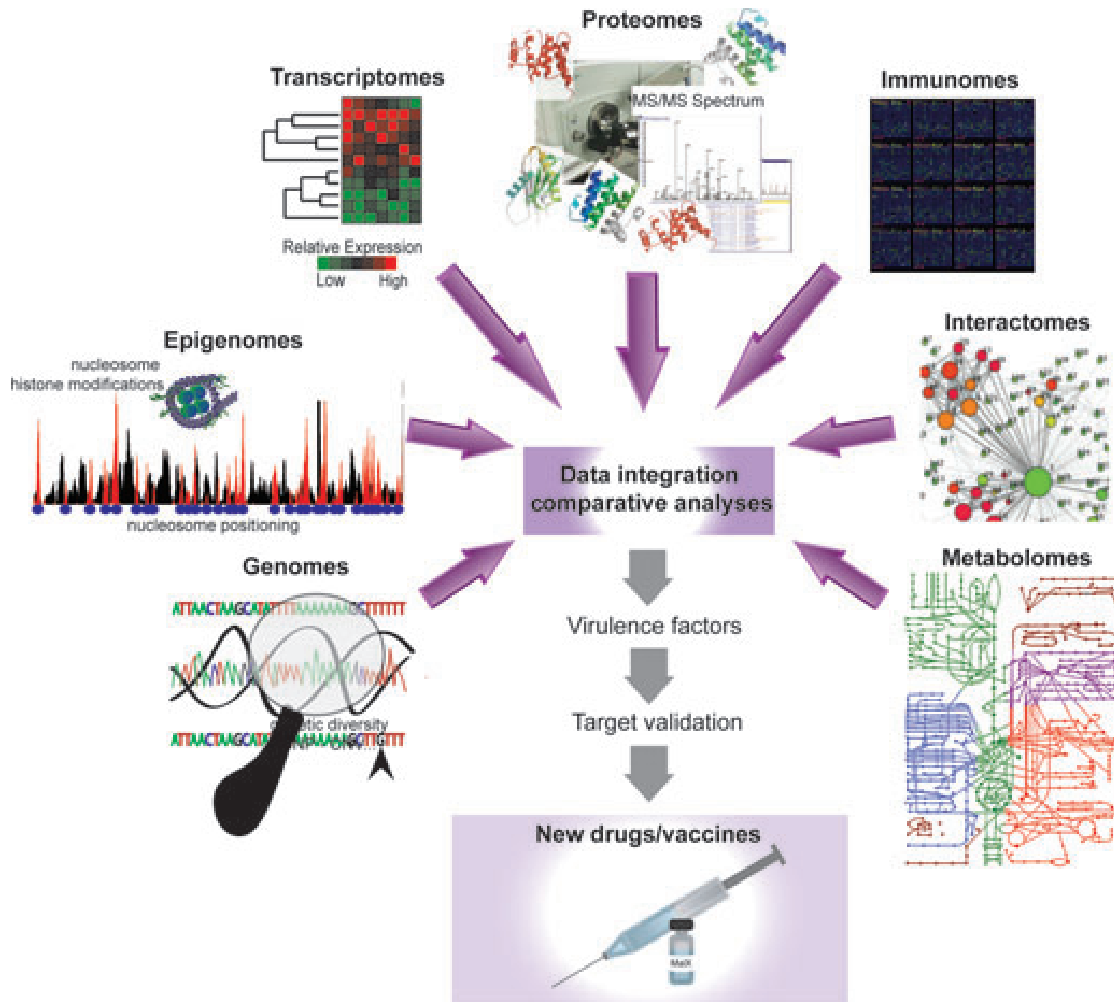


Figure A4.2 A system biology approach to understand the parasite biology for the design of new drug and vaccines strategies. Genomes to identify genetic diversity and drug resistances, Epigenomes to understand transcriptional regulation, Transcriptomes to evaluate mRNA steady state, Proteomes to evaluate protein levels, Immunomes to discovered protective antigens, Interactomes to comprehend protein–protein interactions and Metabolomes to evaluate metabolites.

REFERENCES

1. Gardner MJ, Hall N, Fung E, et al. Genome sequence of the human malaria parasite Plasmodium falciparum. Nature 2002; 6906: 498–511.

2. Volkman SK, Sabeti PC, DeCaprio D, et al. A genome-wide map of diversity in Plasmo- dium falciparum. Nat Genet 2007; 39: 113– 119.

3. Mu J, Awadalla P, Duan J, et al. Genome- wide variation and identification of vaccine targets in the Plasmodium falciparum genome. Nat Genet 2007; 39: 126–130.

4. Jeffares DC, Pain A, Berry A, et al. Genome variation and evolution of the malaria para- site Plasmodium falciparum. Nat Genet 2007; 39: 120–125.

5. Kidgell C, Volkman SK, Daily J, et al. A systematic map of genetic variation in Plasmodium falciparum. PLoS Pathog 2006; 2: e57.

6. Ponts N, Harris EY, Prudhomme J, et al. Nucleosome landscape and control of transcription in the human malaria parasite. Genome Res 2010; 20: 228–238.

7. Bµrtfai R, Hoeijmakers WAM, Salcedo- Amaya AM, et al. H2A.Z demarcates inter- genic regions of the Plasmodium falciparum epigenome that are dynamically marked by H3K9ac and H3K4me3. PLoS Pathog 2010; 6: e1001223.

8. Kozarewa I, Ning Z, Quail MA, Sanders MJ, Berriman M & Turner DJ. Amplification-free Illumina sequencing-library prepa- ration facilitates improved mapping and assembly of (G + C)-biased genomes. Nat Methods 2009; 6: 291–295.

9. Hunt P, Martinelli A, Modrzynska K, et al. Experimental evolution, genetic analysis and genome re-sequencing reveal the mutation conferring artemisinin resistance in an isogenic lineage of malaria parasites. BMC Genomics 2010; 11: 499.

10. Martinelli A, Henriques G, Cravo P & Hunt P. Whole genome re-sequencing identifies a mutation in an ABC transporter (mdr2) in a Plasmodium chabaudi clone with altered sus- ceptibility to antifolate drugs. Int J Parasitol 2011; 41: 165–171.

11. Dharia NV, Bright AT, Westenberger SJ, et al. Whole-genome sequencing and micro- array analysis of ex vivo Plasmodium vivax reveal selective pressure on putative drug resistance genes. Proc Natl Acad Sci USA 2010; 107: 20045–20050.

12 Dimon MT, Sorber K & DeRisi JL.HMMSplicer: a tool for efficient and sensitive discovery of known and novel splice junctions in RNA-Seq data. PLoS ONE 2010; 5:

e13875.

13. Otto TD, Wilinski D, Assefa S, et al. New insights into the blood-stage transcriptome of Plasmodium falciparum using RNA-Seq. Mol Microbiol 2010; 76: 12–24.

14. Sorber K, Dimon MT & DeRisi JL. RNA- Seq analysis of splicing in Plasmodium falciparum uncovers new splice junctions, alternative splicing and splicing of antisense transcripts. Nucleic Acids Res 2011; 39: 3820–3835.

15. Tuda J, Mongan AE, Tolba MEM, et al. Full-parasites: database of full-length cDNAs of apicomplexa parasites, 2010 update. Nucleic Acids Res 2011; 39(Database issue):D625–D631.

16. Ponts N, Harris EY, Lonardi S & Le Roch KG. Nucleosome occupancy at transcription start sites in the human malaria parasite: a hard-wired evolution of virulence? Infect Genet Evol 2011; 11: 716–724.

17. Le Roch KG, Zhou Y, Blair PL, et al. Discovery of gene function by expression profil- ing of the malaria parasite life cycle. Science 2003; 301: 1503–1508.

18. Llinµs M, Bozdech Z, Wong ED, Adai AT & DeRisi JL. Comparative whole genome tran- scriptome analysis of three Plasmodium falciparum strains. Nucleic Acids Res 2006; 34: 1166–1173.

19. Young JA, Fivelman QL, Blair PL, et al. The Plasmodium falciparum sexual develop- ment transcriptome: a microarray analysis using ontology-based pattern identification. Mol Biochem Parasitol 2005; 143: 67–79.

20. Bozdech Z, Mok S, Hu G, et al. The transcriptome of Plasmodium vivax reveals diver- gence and diversity of transcriptional regulation in malaria parasites. Proc Natl Acad Sci USA 2008; 105: 16290–16295.

21. Sorber K, Chiu C, Webster D, et al. The long march: a sample preparation technique that enhances contig length and coverage by high-throughput short-read sequencing. PLoS ONE 2008; 3: e3495.

22. Siau A, Toure FS, Ouwe-Missi-Oukem-Boyer O, et al. Whole-transcriptome analysis of Plasmodium falciparum field isolates: identification of new pathogenicity factors. J Infect Dis 2007; 196: 1603–1612.

23. Daily JP, Scanfeld D, Pochet N, et al. Distinct physiological states of Plasmodium falciparum in malaria-infected patients. Nature 2007; 450: 1091–1095.

24. Daily JP, Le Roch KG, Sarr O, et al. In vivo transcriptional profiling of Plasmodium falci- parum. Malar J 2004; 3: 30.

25. Daily JP, Le Roch KG, Sarr O, et al. In vivo transcriptome of Plasmodium falciparum reveals overexpression of transcripts that encode surface proteins. J Infect Dis 2005; 191: 1196–1203.

26. Mackinnon MJ, Li J, Mok S, et al. Comparative transcriptional and genomic analysis of Plasmodium falciparum field isolates. PLoS Pathog 2009; 5: e1000644.

27. Petter M, Bonow I & Klinkert M-Q. Diverse expression patterns of subgroups of the rif multigene family during Plasmodium falciparum gametocytogenesis. PLoS ONE 2008; 3: e3779.

28. Wang CW, Mwakalinga SB, Sutherland CJ, et al. Identification of a major rif transcript common to gametocytes and sporozoites of Plasmodium falciparum. Malar J 2010; 9: 147.

29. Raabe CA, Sanchez CP, Randau G, et al. A global view of the nonprotein-coding tran- scriptome in Plasmodium falciparum. Nucleic Acids Res 2010; 38: 608–617.

30. Chakrabarti K, Pearson M, Grate L, et al. Structural RNAs of known and unknown function identified in malaria parasites by comparative genomics and RNA analysis. RNA 2007; 13: 1923–1939.

31. Li F, Sonbuchner L, Kyes SA, Epp C & Dei- tsch KW. Nuclear non-coding RNAs are transcribed from the centromeres of Plasmodium falciparum and are associated with centromeric chromatin. J Biol Chem 2008; 283: 5692–5698.

32. Mourier T, Carret C, Kyes S, et al. Genome- wide discovery and verification of novel structured RNAs in Plasmodium falciparum. Genome Res 2008; 18: 281–292.

33. Florens L, Washburn MP, Raine JD, et al. A proteomic view of the Plasmodium falciparum life cycle. Nature 2002; 419: 520–526.

34. Florens L, Liu X, Wang Y, et al. Proteomics approach reveals novel proteins on the surface of malaria-infected erythrocytes. Mol Biochem Parasitol 2004; 135: 1–11.

35. Lasonder E, Ishihama Y, Andersen JS, et al. Analysis of the Plasmodium falciparum prote- ome by high-accuracy mass spectrometry. Nature 2002; 419: 537–542.

36. LaCount DJ, Vignali M, Chettier R, et al. A protein interaction network of the malaria parasite Plasmodium falciparum. Nature 2005; 438: 103–107.

37. Khan SM, Franke-Fayard B, Mair GR, et al. Proteome analysis of separated male and female gametocytes reveals novel sex- specific Plasmodium biology. Cell 2005; 121: 675–687.

38. Le Roch KG, Johnson JR, Florens L, et al. Global analysis of transcript and protein levels across the Plasmodium falciparum life cycle. Genome Res 2004; 14: 2308–2318.

39. Sam-Yellowe TY, Florens L, Wang T, et al. Proteome analysis of rhoptry-enriched fractions isolated from Plasmodium merozoites. J Proteome Res 2004; 3: 995–1001.

40. Lakshmanan V, Rhee KY & Daily JP. Metabolomics and malaria biology. Mol Biochem Parasitol 2011; 175: 104–111.

41. Fridman E & Pichersky E. Metabolomics, genomics, proteomics, and the identification of enzymes and their substrates and products. Curr Opin Plant Biol 2005; 8: 242–248.

42. van Brummelen AC, Olszewski KL, Wilinski D, Llinμs M, Louw AI & Birkholtz L-M. Co-inhibition of Plasmodium falciparum S-adenosylmethionine decarboxylase / ornithine decarboxylase reveals perturbation- specific compensatory mechanisms by transcriptome, proteome, and metabolome analyses. J Biol Chem 2009; 284: 4635–4646.

43. Lian L-Y, Al-Helal M, Roslaini AM, et al. Glycerol: an unexpected major metabolite of energy metabolism by the human malaria parasite. Malar J 2009; 8: 38.

44. Olszewski KL, Morrisey JM, Wilinski D, et al. Host-parasite interactions revealed by Plasmodium falciparum metabolomics. Cell Host Microbe 2009; 5: 191–199.

45. Teng R, Junankar PR, Bubb WA, Rae C, Mercier P & Kirk K. Metabolite profiling of the intraerythrocytic malaria parasite Plasmodium falciparum by (1)H NMR spectros- copy. NMR Biomed 2009; 22: 292–302.

46. Balu B, Chauhan C, Maher SP, et al. piggy- Bac is an effective tool for functional analy- sis of the Plasmodium falciparum genome. BMC Microbiol 2009; 9: 83.

47. Balu B, Shoue DA, Fraser MJ & Adams JH. High-efficiency transformation of Plasmo-

dium falciparum by the lepidopteran trans- posable element piggyBac. Proc Natl Acad Sci USA 2005; 102: 16391–16396.

48. van Ooij C, Tamez P, Bhattacharjee S, et al. The malaria secretome: from algorithms to essential function in blood stage infection. PLoS Pathog 2008; 4: e1000084.

49. Balu B, Singh N, Maher SP & Adams JH. A genetic screen for attenuated growth identifies genes crucial for intraerythrocytic development of Plasmodium falciparum. PLoS ONE 2010; 5: e13282.

50. Elemento O, Slonim N & Tavazoie S. A universal framework for regulatory element discovery across all genomes and data types. Mol Cell 2007; 28: 337–350.

51. Gonzales JM, Patel JJ, Ponmee N, et al. Regulatory hotspots in the malaria parasite genome dictate transcriptional variation. PLoS Biol 2008; 6: e238.

52. Westenberger SJ, Cui L, Dharia N, Winzeler E & Cui L. Genome-wide nucleosome map- ping of Plasmodium falciparum reveals his- tone-rich coding and histone-poor intergenic regions and chromatin remodeling of core and subtelomeric genes. BMC Genomics 2009; 10: 610.

53. Coulson RM, Hall N & Ouzounis CA. Comparative genomics of transcriptional control in the human malaria parasite Plasmodium falciparum. Genome Res 2004; 14: 1548–1554.

54. Balaji S, Babu MM, Iyer LM & Aravind L. Discovery of the principal specific transcrip- tion factors of Apicomplexa and their implication for the evolution of the AP2-integrase DNA binding domains. Nucleic Acids Res 2005; 33: 3994–4006.

55. Painter HJ, Campbell TL & Llinµs M. The Apicomplexan AP2 family: integral factors regulating Plasmodium development. Mol Biochem Parasitol 2011; 176: 1–7.

56. Templeton TJ. The varieties of gene amplification, diversification and hypervariability in the human malaria parasite, Plasmodium falciparum. Mol Biochem Parasitol 2009; 166: 109–116.

57. Aravind L, Iyer LM, Wellems TE & Miller LH. Plasmodium biology: genomic gleanings. Cell 2003; 115: 771–785.

58. Campbell TL, De Silva EK, Olszewski KL, Elemento O & Llinµs M. Identification and genome-wide prediction of DNA binding specificities for the ApiAP2 family of regulators from the malaria parasite. PLoS Pathog 2010; 6: e1001165.

59. Ganesan K, Ponmee N, Jiang L, et al. A genetically hard-wired metabolic transcriptome in Plasmodium falciparum fails to mount protective responses to lethal antifolates. PLoS Pathog 2008; 4: e1000214.

60. Le Roch KG, Johnson JR, Ahiboh H, et al. A systematic approach to understand the mechanism of action of the bisthiazolium compound T4 on the human malaria parasite, Plasmodium falciparum. BMC Genomics 2008; 9: 513.

61. Hu G, Cabrera A, Kono M, et al. Transcriptional profiling of growth perturbations of the human malaria parasite Plasmodium falciparum. Nat Biotechnol 2010; 28: 91–98.

62. Hall N, Karras M, Raine JD, et al. A comprehensive survey of the Plasmodium life cycle by genomic, transcriptomic, and proteomic analyses. Science 2005; 307: 82–86.

63. Foth BJ, Ralph SA, Tonkin CJ, et al. Dissecting apicoplast targeting in the malaria parasite Plasmodium falciparum. Science
2003; 299: 705–708.

64. Mair GR, Braks JAM, Garver LS, et al.
Regulation of sexual development of Plasmodium by translational repression. Science 2006; 313: 667–669.

65. Shock JL, Fischer KF & DeRisi JL. Whole- genome analysis of mRNA decay in Plasmodium falciparum reveals a global lengthening of mRNA half-life during the intra-erythrocytic development cycle. Genome Biol 2007; 8: R134.

66. Duraisingh MT, Voss TS, Marty AJ, et al. Heterochromatin silencing and locus repositioning linked to regulation of virulence genes in Plasmodium falciparum. Cell 2005; 121: 13–24.

67. Freitas-Junior LH, Hernandez-Rivas R, Ralph SA, et al. Telomeric heterochromatin propagation and histone acetylation control mutually exclusive expression of antigenic variation genes in malaria parasites. Cell 2005; 121: 25–36.

68. Cui L & Miao J. Chromatin-mediated epigenetic regulation in the malaria parasite Plasmodium falciparum. Eukaryot Cell 2010; 9: 1138–1149.

69. Workman JL & Kingston RE. Alteration of nucleosome structure as a mechanism of transcriptional regulation. Annu Rev Biochem 1998; 67: 545–579.

70. Chookajorn T, Dzikowski R, Frank M, et al. Epigenetic memory at malaria virulence genes. Proc Natl Acad Sci USA 2007; 104: 899–902.

71. Lopez-Rubio J-J, Mancio-Silva L & Scherf A. Genome-wide analysis of heterochromatin associates clonally variant gene regulation with perinuclear repressive centers in malaria parasites. Cell Host Microbe 2009; 5: 179–190.

72. Salcedo-Amaya AM, van Driel MA, Alako BT, et al. Dynamic histone H3 epigenome marking during the intraerythrocytic cycle of Plasmodium falciparum. Proc Natl Acad Sci USA 2009; 106: 9655–9660.

73. Flueck C, Bartfai R, Volz J, et al. Plasmodium falciparum heterochromatin protein 1 marks genomic loci linked to phenotypic variation of exported virulence factors. PLoS Pathog 2009; 5: e1000569.

74. PØrez-Toledo K, Rojas-Meza AP, Mancio- Silva L, et al. Plasmodium falciparum hetero- chromatin protein 1 binds to tri-methylated histone 3 lysine 9 and is linked to mutually exclusive expression of var genes. Nucleic Acids Res 2009; 37: 2596–2606.

75. Trelle MB, Salcedo-Amaya AM, Cohen AM, Stunnenberg HG & Jensen ON. Global his- tone analysis by mass spectrometry reveals a high content of acetylated lysine residues in the malaria parasite Plasmodium falciparum. J Proteome Res 2009; 8: 3439–3450.

76. Saha A, Wittmeyer J & Cairns BR. Chroma- tin remodelling: the industrial revolution of DNA around histones. Nat Rev Mol Cell Biol 2006; 7: 437–447.

77. Boeger H, Griesenbeck J, Strattan JS & Kornberg RD. Removal of promoter nucleosomes by disassembly rather than sliding in vivo. Mol Cell 2004; 14: 667–673.

78. Lee C-K, Shibata Y, Rao B, Strahl BD & Lieb JD. Evidence for nucleosome depletion at active regulatory regions genome-wide. Nat Genet 2004; 36: 900–905.

79. Hogan GJ, Lee C-K & Lieb JD. Cell cycle- specified fluctuation of nucleosome occupancy at gene promoters. PLoS Genet 2006; 2: e158.

80. Segal E, Fondufe-Mittendorf Y, Chen L, et al. A genomic code for nucleosome positioning. Nature 2006; 442: 772–778.

81. Segal E & Widom J. Poly(dA:dT) tracts: major determinants of nucleosome organization. Curr Opin Struct Biol 2009; 19: 65–71.

82. Kaplan N, Moore I, Fondufe-Mittendorf Y, et al. Nucleosome sequence preferences influence in vivo nucleosome organization. Nat Struct Mol Biol 2010; 17: 918–920; author reply 920–922.

83. Andersson R, Enroth S, Rada-Iglesias A, Wadelius C & Komorowski J. Nucleosomes are well positioned in exons and carry characteristic histone modifications. Genome Res 2009; 19: 1732–1741.

84. Schwartz S, Meshorer E & Ast G. Chroma- tin organization marks exon-intron structure. Nat Struct Mol Biol 2009; 16: 990–995.

85. Tilgner H, Nikolaou C, Althammer S, et al. Nucleosome positioning as a determinant of exon recognition. Nat Struct Mol Biol 2009; 16: 996–1001.

86. Sims JS, Militello KT, Sims PA, Patel VP, Kasper JM & Wirth DF. Patterns of gene-specific and total transcriptional activ- ity during the Plasmodium falciparum intraerythrocytic developmental cycle. Eukaryot Cell 2009; 8: 327–338.

87. Gopalakrishnan AM, Nyindodo LA, Ross Fergus M & López-EstraÇo C. Plasmodium falciparum: preinitiation complex occupancy of active and inactive promoters during erythrocytic stage. Exp Parasitol 2009; 121: 46–54.

88. Jin C & Felsenfeld G. Nucleosome stability mediated by histone variants H3.3 and H2A.Z. Genes Dev 2007; 21: 1519–1529. Parasite Immunology

89. Fan JY, Gordon F, Luger K, Hansen JC & Tremethick DJ. The essential histone variant H2A.Z regulates the equilibrium between different chromatin conformational states. Nat Struct Biol 2002; 9: 172–176.

90. Santisteban MS, Hang M & Smith MM. Histone variant H2A.Z and RNA polymerase II transcription elongation. Mol Cell Biol 2011; 31: 1848–1860.

91. Mok BW, Ribacke U, Rasti N, et al. Default pathway of var2csa switching and translational repression in Plasmodium falciparum. PLoS ONE 2008; 3: e1982.

92. Natalang O, Bischoff E, Deplaine G, et al. Dynamic RNA profiling in Plasmodium falciparum synchronized blood stages exposed to lethal doses of artesunate. BMC Genomics 2008; 9: 388.

93. Jiang H, Yi M, Mu J, et al. Detection of genome-wide polymorphisms in the AT-rich Plasmodium falciparum genome using a high- density microarray. BMC Genomics 2008; 9: 398.

94. Jomaa H, Wiesner J, Sanderbrand S, et al. Inhibitors of the nonmevalonate pathway of isoprenoid biosynthesis as antimalarial drugs. Science 1999; 285: 1573–1576.

95. Dharia NV, Plouffe D, Bopp SER, et al. Genome scanning of Amazonian Plasmodium falciparum shows subtelomeric instability and clindamycin-resistantparasites. Genome Res 2010; 20: 1534–1544.

96. Plouffe D, Brinker A, McNamara C, et al. *In silico* activity profiling reveals the mecha- nism of action of antimalarials discovered in a high-throughput screen. Proc Natl Acad Sci USA 2008; 105: 9059–9064.

97. Rottmann M, McNamara C, Yeung BKS, et al. Spiroindolones, a potent compound class for the treatment of malaria. Science 2010; 329: 1175–1180.

98. Mu J, Myers RA, Jiang H, et al. Plasmodium falciparum genome-wide scans for positive selection, recombination hot spots and resistance to antimalarial drugs. Nat Genet 2010; 42: 268–271.

99. Crompton PD, Kayala MA, Traore B, et al. A prospective analysis of the Ab response to Plasmodium falciparum before and after a malaria season by protein microarray. Proc Natl Acad Sci USA 2010; 107: 6958–6963.

100. Doolan DL, Mu Y, Unal B, et al. Profiling humoral immune responses to P. falciparum infection with protein microarrays. Proteomics 2008; 8: 4680–4694.

101. Sundaresh S, Doolan DL, Hirst S, et al. Identification of humoral immune responses in protein microarrays using DNA micro- array data analysis techniques. Bioinformatics 2006; 22: 1760–1766.

APPENDIX A5:

Genome-Wide Analysis of Gene Expression. *Encyclopedia of Biological Chemistry, 2nd Edition*.

Duk-Won Doug Chung and Karine G. Le Roch (2012)

Department of Cell Biology and Neurosciences, University of California at Riverside, Riverside, California, United States of America.

Key Words: Alternative splicing, Expressed sequence tag (EST), Gene expression, Genome wide analysis, Microarray, Next-Generation Sequencing Technologies, RNA-seq, SAGE, Tilling Array, Transcriptome.

Synopsis

Genome-wide expression analysis has many applications, including RNA quantification, gene discovery and mapping, intron/exon boundary detection, alternative expression analysis and organism's phenotype. The expression data gathered from a genome-wide approach can provide clues in identifying genes involved in various cellular processes by looking at relative expression levels at differing cell cycle stages or under certain induced conditions. In addition, genome-wide transcriptome analysis, particularly under varying conditions, is useful in identifying key regulatory genes and potential drug targets. Over the past two decades, several methods have been developed to analyze gene expression on a genome-wide scale. This chapter briefly discusses several of the earlier techniques employed for transcriptome analysis, followed by a description of the more recent approaches using next-generation sequencing.

Glossary Entries:

cDNA - DNA that has been created from RNA using reverse transcription.

Microarray - A device typically used for monitoring gene expression for a large number of genes. They consist of an ordered array of different elements (usually nucleic acids complementary to the genes of interest) that are usually printed or synthesized *in situ* on a solid surface. Microscopic beads can also be used.

Probe - The individual elements bound to a microarray consisting of long oligonucleotides, short oligonucleotides (25mers), fragments of genomic DNA or cDNAs.

RNA – Ribonucleic acid. Single-stranded molecules transcribed from DNA.

Read – The sequence of bases determined from a single segment of sample nucleic acid by a sequencing instrument.

Read length - The number of individual bases identified in a continuous read.

RNA-seq – Whole transcriptome analysis based on high throughput sequencing technologies.

Second Generation Sequencing (SGS) – High throughput sequencing based on short PCR amplified DNA fragments, usually based on wash-and-scan techniques.

Serial Analysis of Gene Expression (SAGE) - A method used in the genome-wide monitoring of gene expression in which short sequence tags from different cDNAs are isolated, sequenced, and counted.

Third generation sequencing (TGS)– High throughput direct sequencing of single molecules of nucleic acid without the need of amplification or halting between read steps.

Gene Expression Data for Discovery and Classification

Gene expression is the process by which gene sequences are transcribed into functional gene products such as proteins or functional RNAs (e.g, rRNA, tRNA, small RNA). The process of gene expression is used by all known organisms and can be regulated and modulated at several levels (transcription initiation, splicing, alternative splicing, mRNA stability, post-transcriptional regulation and eventually translational and post-translational regulation mechanisms). Gene expression and regulation are the basis of cell development and differentiation. They also allow the cell to adapt to different conditions. By controlling the time, location and expression level, gene transcripts can have a profound effect on the functions of genes within cells or in multicellular organisms. Genome-wide analysis of gene expression allows for the detection and the quantification of transcript levels for every known gene in the genome of the cell type and organism of interest. The primary use for whole genome expression analysis is the identification of cell or tissue phenotypes as well as the discovery of novel genes or pathways involved in particular molecular processes and disease development. For example, genome-wide analysis of gene expression can be used to distinguish two different biological states such as acute myoblastic leukemia and acute lymphoblastic leukemia, which are relatively difficult to distinguish from one another using conventional molecular and cytological approaches. New genes or pathways induced or repressed in the condition of interest, relative to the reference condition, may provide a better description of the biological processes under study, present new molecular reporters and novel drug targets, predict patient's disease classifications and also optimize treatments. To

analyze gene expression in a genome-wide manner, total RNA is collected from the condition of interest and from a control condition and analyzed by different genome-wide methodologies (e.g. microarray or sequencing technologies - see described methods below). Newly developed software are then used to examine these large genome-wide expression data projects.

Expressed sequence tag

In 1991, the primary tool employed for human gene discovery was a method using expressed sequence tags (EST). ESTs are single-pass sequence reads derived from cDNA. ESTs are typically around 100-800 base pairs (bp) in length, unedited, and randomly selected. Because the ESTs are sequenced only once, the reads are susceptible to errors. Often times, the quality of the base reads for each EST was poor at the beginning of the sequence, improved significantly in the middle, and then diminished again at the end. Phred scores provided a measure of sequence quality of ESTs and were used to extract sequences of a minimal specified quality. Today, EST is no longer widely used because of its inherent drawbacks, which are low throughput capabilities, redundancy, under-representation and over-representation.

Serial Analysis of Gene Expression.

Serial analysis of gene expression (SAGE) was developed in 1995 by Dr. Victor Velculescu (Velculescu et al., 1995) and provides a snapshot of transcript

population within a sample by analyzing small tags that correspond to the 3' fragments of mRNA (Figure A5.1 near here). First, cDNA is synthesized from mRNA using biotinylated oligo(dT) primer and then is cleaved with a restriction enzyme (anchoring enzyme) that is expected to cut each transcript at least once. The most 3' fragment of the cleaved cDNA is bound to streptavidin beads at its poly(A) tail that is complemented by biotinylated thymidines. Bound cDNA fragments are then ligated at the anchoring cleavage site with linkers that contain a type IIS restriction site. Type IIS endonucleases (tagging enzymes) cleave at a specific distance up to 20 base pairs (bp) away from the recognition site. This process is designed so that cleavage with the tagging enzymes results in the release of the linker along with a short piece of cDNA (tag) of around 9 bp. Released tags are then ligated together at their blunt ends to form a ditag, which serves as a template for polymerase chain reaction (PCR) amplification using primers specific to the linkers. Following PCR amplification, the flanking linkers are cleaved from the ditags using the anchoring enzyme. Cleaved ditags are subsequently concatenated, cloned into a plasmid vector and then sequenced.

With SAGE, up to 20,000 tags can be sequenced in a single experiment. With full-genome sequence information, the sequenced SAGE tags can be mapped back to the gene. This approach provides for a highly quantitative analysis since the number of tags for a particular gene can be counted. Since the advent of SAGE, several more robust variants, LongSAGE, RL-SAGE, and SuperSAGE, have been developed. These

newer variants allow up to 100,000 sequenced tags per trial, thus providing a more comprehensive profile of the transcriptome, and captures longer tags, up to 26 bps, enabling more confident identification of the source gene. Though there have been improvements, SAGE is no longer widely implemented because it is highly labor intensive and is relatively low throughput compared to the newly developed sequencing technologies described below.

DNA Microarray

DNA microarrays consists of an arrayed series of thousands of microscopic spots of DNA oligonucleotides, each containing specific DNA sequences, called probes or reporters. Genes or DNA fragments of interest are amplified with PCR, cleaned up and printed as a two-dimensional array using robotic microarraying devices that are able to dispense small quantities of solution at a precise location on a solid support, typically a glass slide due to its low inherent fluorescence. These glass slides are usually coated with hydrophobic poly-lysine or aminosilanes, which limit the spread of DNA on the slides. Other microarray platforms, such as Illumina, use microscopic beads instead of a solid support.

Alternatively, solid-phase *in situ* synthesis of DNA probes from genes of interest can also be implemented in constructing a microarray using a combination of light-directed oligonucleotide chemistry and photolithography. This technology

was first developed by Affymetrix but is now available from various other companies such as Agilent. Nucleic acid arrays manufactured from this method allows for high information content with up to 1,000,000 oligonucleotides selected from sequencing databases and precisely synthesized on a small platform (Figure A5.2 near here). Since these arrays are designed *in silico*, intermediates such as cDNA libraries, PCR and clones are unnecessary and thus the risk of mislabeling is reduced. In addition, this method provides more control over the sample amount deposited onto each location on the array. In order to gain confidence, different multiple probes are synthesized for a given gene.

Transcripts from the sample of interest are converted into cDNA or cRNA (target) and typically labeled with fluorophores or biotin, but chemiluminescent molecules or radioactive isotopes can also be used. Labeled cDNA or cRNA is then hybridized to the DNA microarray under high-stringency conditions. The underlying principle of gene expression level analysis using DNA microarrays is that the transcript level for a particular gene will directly correspond to the amount of hybridization detected on the probe representing that same gene. Probe-target hybridization is detected and quantified using various methods depending on the application. Usually, the background-subtracted hybridization levels from probes representing the same gene are averaged because different oligonucleotide probes have distinct hybridization properties. This approach allows for uses such as distinguishing genes that are highly expressed from those that are lowly expressed.

Tiling arrays are a subtype of microarrays that uses short probes that cover the entire length of the genome. Depending on the probe length and spacing, different resolutions can be achieved. The advantage of tiling arrays over traditional microarrays is the detection of unidentified genes, alternative splice sites, single-feature polymorphisms and transcript variants is made possible.

Microarray Analysis

High-throughput techniques such as DNA microarrays can produce a large amount of raw data. With proper analytical approaches, a significant amount of biological information can be extrapolated from microarray data. Clustering is a method of analysis that organizes data and groups genes according to their similarity in gene expression patterns. Since genes with similar biological roles will often have similar patterns of expression, genes involved in related processes are often clustered together. Pairwise average-linkage cluster analysis is a widely used tool that creates hierarchical clusters. Relationships between genes are represented by trees which branch lengths reflect the degree of similarity between genes.

Another technique for analyzing microarray data is the ontology method that implements the “guilt by association” approach by inferring possible functions to genes by using prior biological knowledge of other co-expressed genes. Often times,

known inhibitors of a biological pathway can be used to elicit the up or down-regulation of clusters of genes, and thus uncharacterized genes can be investigated by their association with other previously known genes that they are co-regulated with. Gene expression analysis by microarray is still being implemented today and is still contributing significantly to our understanding of gene expression and transcript profiles. However, with the advent of newer, cheaper, and more powerful tools such as next-generation sequencing technologies, microarray technology may be phased out in the near future.

Next-Generation Sequencing Technologies

In 1977, Sanger sequencing became available and, since then, has undergone numerous improvements that allowed for automation. Up until recent years, large sequencing projects (*e.g.* whole genome sequencing of species) have depended on Sanger sequencing. However in the past decade or so, there have been several newly developed sequencing methods, collectively called next-generation sequencing (NGS) (Figure A5.3 near here), that have both transformed the field of whole genome sequencing and also opened up new applications for sequencing-based approaches such as transcriptome analysis.

In 2005, second-generation sequencing technology (SGS) had become commercially available. Unlike Sanger sequencing, cloning the DNA template into bacterial vectors is unnecessary. On top of saving time and labor, the advantage of

bypassing the need for bacterial cloning is that cloning biases, such as cloning difficulties of AT-rich regions and genes that are toxic to bacteria, are circumvented. In addition, SGS provides for a much higher throughput than automated Sanger methods by achieving a tremendous degree of parallelization where up to billions of sequencing reactions within small volumes are taking place at the same time. In most SGS methods, template DNA is sheared into fragments, bound to adaptors and PCR amplified in order to generate clonal representations of the original template DNA. Tens of thousands of identical PCR amplified fragments are usually spatially immobilized on beads or on a given location on an array making up a PCR colony ("polony") or clusters. These clusters are read by numerous cycles of washing and scanning operations where sequencing is accomplished by employing various enzymes (*e.g.* ligases and polymerases) and reagents (*e.g.* fluorescently-labeled nucleotides) to produce a series of fluorescent signals representing the average sequence of an individual cluster. The number of cycles of *wash-and-scan* determines the length of each read and usually is limited by the amount of noise produced by the incorporation of out-of-phase nucleotides as the read becomes longer. Because of the numerous scan-and-wash cycles, the length of time required to complete a run usually lasts for several days. The massive parallel series of light signals are recorded by highly-sensitive detection systems and subsequently decoded to determine the base sequence of the DNA. During the decoding process, also known as base-calling, each called base is generally given a quality score. Timely analysis and reconstruction of the large amount of sequenced reads are

made possible by significant advancements in computational power and data storage, along with a number of software tools that are available.

Roches' 454 genome sequencer was the first commercially available SGS. Its average read length is around 400 to 500 bps, which is the longest among the SGS systems available today. It is able to produce 400-600 million bases per run or about 1 billion bases per day. The 454 sequencing method starts with the generation of a library containing single-stranded template DNA with attached adaptors. This library is then immobilized onto DNA capture beads with only one single-stranded DNA fragment per bead. By emulsion PCR, the DNA library is amplified on each bead and then placed into individual wells on a PicoTiterPlate. With the supplementation of sequencing enzymes and materials, the addition of nucleotides complementary to the template DNA result in a base-by-base chemiluminescent light that is detected and recorded by the 454 genome sequencer.

Applied Biosystems' SOLiD system is another SGS system that has a shorter read length of 35-50 bps but is able to generate a high throughput of about 100 gigabases per run (or 8 gigabases per day). Like the 454 genome sequencer, the SOLiD system uses beads to immobilize and PCR amplify template strands. Once the template strands are amplified and denatured, their 3' ends are modified to allow for covalent bonding to a glass slide. The 3' modified beads are then deposited onto glass slides which allow for high densities of beads per slide and thus a high level of

throughput. Primers are hybridized to adaptors at the ends of each template strand and a set of four fluorescently labeled di-probes are allowed to compete for the primers by interrogating every first and second base of each ligation cycle. Multiple rounds of ligation, detection and cleavage are the driving mechanism for this sequencing-by-ligation approach. After the first series of ligation cycles, the extension product is removed and the template is reset with the hybridization of primers that are offset by one nucleotide from the previous primer and a subsequent series of ligation reactions takes place. This process is repeated with a total of five primer resets for each sequence tag. For each template fragment, the multiple series of fluorescence produced by the ligation cycles are detected, recorded and threaded together in order to generate a complete read.

Illumina's (formally known as Solexa) sequencers, such as the Genome Analyser and HiSeq 2000, are capable of producing high throughput sequencing, up to 200 gigabases per run (or up to 25 gigabases a day) with a read length of 35-150 bp. Template strands are first fragmented and ligated with adaptors. Then the template fragments are hybridized to a flat, optically transparent surface, where they are bridge amplified to create a highly dense sequencing flowcell containing tens of millions of clusters each containing around 1000 copies of the same template fragment. Illumina implements a sequencing-by-synthesis approach by using reversible terminators with removable four-color fluorescent dyes to generate a base-by-base sequence. Each type of nucleotide is distinguished by a different color.

Fluorescence detection is accomplished with laser excitation and total internal reflection optics.

The Polonator is a NGS that is based upon sequencing-by-ligation that has an open-source platform where users can use off-the-shelf reagents and freely download protocols and software. The open-source approach allows for users to modify protocols, reagents and software to custom-fit their sequencing application. However, the Polonator has the shortest read lengths when compared to other NGS systems.

In addition to genome sequencing, NGS systems are applied to sequence cDNA derived from the whole transcriptomes of an organism in order to analyze expression profiles, transcript boundaries, intron/exon junctions, alternative expression, and also discover novel transcripts in a high throughput manner. The obtainment of transcriptomics data by NGS is collectively known as deep RNA sequencing or RNA-seq. In 2008, the implementation of RNA-seq was first reported in yeast, and then was quickly adopted for use of transcriptome analysis in several other organisms. The advantages of RNA-seq over previous analytical sequencing tools such as EST and SAGE are that RNA-seq has a much higher throughput and increased sequencing depth, which leads to better resolution of expression profiles. And though microarray has a genome-wide coverage, nucleic acid arrays are not available for all organisms, while RNA-seq can be used on any organism. In addition,

microarray expression data can be biased by the binding affinities of each individual probe. Another advantage is that RNA-seq is able to identify transcript boundaries and intron/exon junctions at a single-nucleotide resolution.

Third Generation Sequencing

More recently, a few new companies, such as Helicos Biosciences and Pacific Biosciences, are developing advanced sequencing methods that do not require template amplification but rather directly sequences from single template molecules. These systems are commonly referred to as third generation sequencing (TGS) platforms (Figure A5.3). The advantage of directly sequencing from single template molecules is that clonal duplication and amplification biases due to difficulties in amplifying certain regions of template strands are avoided. Also, the amplification process can introduce errors in the replicated template strands.

In the Helicos system, DNA or cDNA is sheared and turned into single-stranded molecules. Fluorescently-labeled poly-A adaptors are added to the ends of the sheared strands and then hybridized to poly-T oligonucleotide capture-strands bound to a flow-cell. A highly-sensitive camera that is able to detect single-molecule fluorescence captures the locations of each of the billions of hybridized template strands. After the fluorescent-label is removed from the adaptors, a series of wash-and-scan cycles are initiated; a single type of fluorescently-labeled nucleotides and

polymerases are washed over the flowcell, followed by fluorescence detection by imaging and removal of the fluorescent group. Then the wash-and-scan cycle is then repeated with the other three nucleotides. Multiple four-base cycles result in reads of around 25bp.

Pacific Biosciences has developed a real-time single-molecule sequencing system, PacBio RS (Figure A5.3), that is capable of 30 minute sequencing runs because it does not require the conventional wash-and-scan method, which normally can take up to several days. A thin 100nm thick metal film is perforated with an array of nanoscopic holes and bound to a glass substrate, creating wells that are tens of nanometers in diameter. A laser light is shined through the glass substrate, illuminating only the bottom 30nm of each well. Each well typically has a single DNA polymerase anchored to the glass bottom and is small enough to allow for only one template strand. Fluorescently-labeled nucleotides that have different fluorescent colors depending on the nucleotide are washed over the nanoscopic wells and are incorporated into the sequencing strand by the DNA polymerase. These nucleotides have fluorescent dyes attached to the phosphate chain instead of the base thus preventing the inactivation of the DNA polymerase. As the fluorescently-labeled nucleotides are incorporated one-by-one via the DNA polymerase, the fluorescent-dye is excited by the laser, transmits a color signal that lasts milliseconds and then is cleaved off by the DNA polymerase to allow for

additional rounds of nucleotide incorporation and fluorescence. The series of color fluorescence is detected and decoded to produce reads of over 1000 bp.

Analytical tools for NGS expression data

The high-throughput and deepness of quantitative measurements produce by NGS technology comes at the price of examining millions to billions of reads. Though the first users have created their own computer programs to study the numerous reads, a new breed of sophisticated algorithms and software tools have emerged to help analyze the massive datasets produced by RNA-seq. These analytical tools are used for transcript discovery, RNA quantification and mapping, and also identification of transcript variants and splice sites.

With a given gene model, the total number of mapped reads can be tallied for a particular gene as a measure of gene expression. The number of reads is naturally a function of the molar concentration of mRNA but must be normalized according to the length of each mRNA. A common approach is to normalize the read count by the length of the transcript and the number of millions of mappable reads to obtain reads per kilobase per million (RPKM) values. Within a given sample, RPKM values for genes are directly comparable to give relative expression values. However, RPKM values vary depending on the software package employed and can also be influenced by experimental factors such as quality of input RNA, the amount of ribosomal RNA remaining in the sample, size selection procedures and the

accuracy of the gene model used. Amplification biases can also obstruct uniform sequence coverage across the transcriptome, though TGS technologies circumvent this problem by bypassing the amplification process altogether.

In addition to quantifying gene expression, RNA-seq allows for the investigation of the correspondence between gene and transcript model. With alternative splicing, alternative promoter employment and differing 3' poly (A) tail sites, various types of transcripts can be derived by a single gene. RNA-seq is capable of using reads that map across splice junctions to identify splice sites and also discover alternatively spliced transcript isoforms. Using more advanced software, users are capable of extracting more information from their RNA-seq data such as analysis of transcript modeling, quantifying specific isoforms and investigating sequence variance due to single-nucleotide polymorphisms, private mutation or RNA editing. Bioinformatic tools are constantly being developed in order to match the quickly evolving sequencing technology.

Closing Statements

Though current analytical tools of investigating transcripts on a genome-wide level has its own distinct challenges, they offer more advantages than previous methods such as higher-throughput, wider-coverage, and increased sensitivity. Today, with RNA-seq technology, we are able to quantify differences in gene expression at a single nucleotide resolution across multiple samples and conditions.

Though we still have much to learn and discover, we draw closer to understanding how genomes are capable of encoding diverse patterns of gene expression that define each cell type/stage and also elicit a distinct response from various stimuli/conditions.

Further Reading

Adams, M. D., Kelley, J. M., Gocayne, J. D., Dubnick, M., Polymeropoulos, M. H., Xiao, H., Merril, C. R., Wu, A., Olde, B., and Moreno, R. F. (1991). Complementary DNA sequencing: expressed sequence tags and human genome project. Science *252*, 1651-1656.

DeRisi, J. L., Iyer, V. R., and Brown, P. O. (1997). Exploring the metabolic and genetic control of gene expression on a genomic scale. Science *278*, 680-686.

Nagaraj, S. H., Gasser, R. B., and Ranganathan, S. (2007). A hitchhiker's guide to expressed sequence tag (EST) analysis. Briefings in Bioinformatics *8*, 6 -21.

Nagarajan, N., and Pop, M. (2010). Sequencing and genome assembly using next-generation technologies. Methods Mol. Biol *673*, 1-17.

Pepke, S., Wold, B., and Mortazavi, A. (2009). Computation for ChIP-seq and RNA-seq studies. Nat Meth *6*, S22-S32.

Schadt, E. E., Turner, S., and Kasarskis, A. (2010). A window into third-generation sequencing. Hum. Mol. Genet *19*, R227-240.

Schena, M., Shalon, D., Davis, R. W., and Brown, P. O. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science *270*, 467-470.

Velculescu, V. E., Zhang, L., Vogelstein, B., and Kinzler, K. W. (1995). Serial analysis of gene expression. Science *270*, 484-487.

Figure Legends:

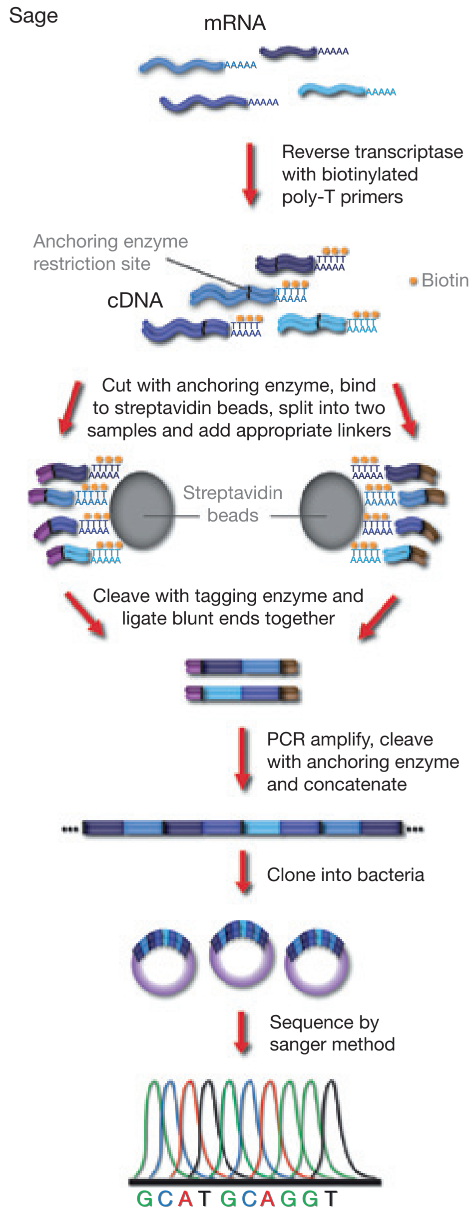


Figure A5.1: Serial Analysis of Gene Expression (SAGE).
RNA is extracted and purified from desired samples. mRNA is converted into double-stranded cDNA by reverse transcriptase with biotinylated poly-T tails as primers. The biotinylated cDNA is fragmented with anchoring enzymes that recognize and cut at their respective restriction sites that are usually naturally found every few hundred basepairs. The biotinylated ends of the fragmented cDNA are bound to streptavidin beads and split into two portions. Appropriate adaptors with endonuclease site for tagging enzymes are ligated to each cDNA fragment at the cleaved end. Tagging enzymes are added to cleave a small distance from their endonuclease sites found at each adaptor, releasing the fragmented cDNA (up to 26 basepairs) from the beads. Blunt ends from each of the two portions of fragmented cDNA are ligated together, forming a ditag, and then amplified by PCR. After PCR, ditags are cleaved at their adaptors, concatenated, and then cloned into plasmids. Once cloned, plasmids are then sequenced, mapped back to the gene and subsequently analyzed for gene expression.

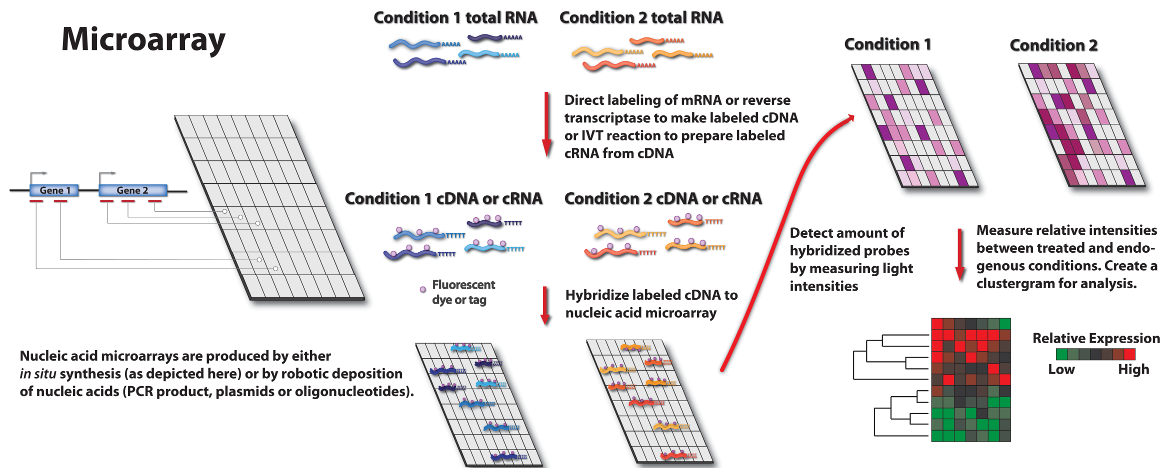
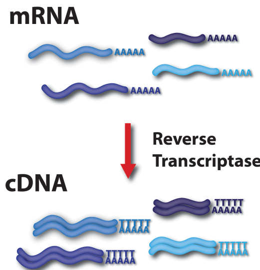


Figure A5.2: Microarray.

Nucleic acid microarrays must first be produced by either *in situ* synthesis or by robotic deposition of nucleic acids. For *in situ* synthesis, multiple probes per gene are placed on the array, while in the case of robotic deposition longer (up to 1,000 bp) double-stranded DNA probe(s) is used for each gene. RNA from various samples are extracted, purified. The RNA can be converted into cDNA by reverse transcriptase in order to produce labeled cRNA by *in vitro* transcription (IVT) or have the cDNA labeled with a tag such as biotin, or fluorescent dye. In addition, RNA can also be directly labeled. Labeled RNA/cDNA/cRNA are then hybridized to the arrays. After washing, and appropriate secondary fluorescent dyes (if necessary) are added, the intensities of light from the hybridized dyes are detected and recorded. Clustergrams and relative expression levels for each gene are extrapolated by comparing microarray data between various conditions.

RNA-Seq

mRNA

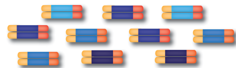


Reverse Transcriptase

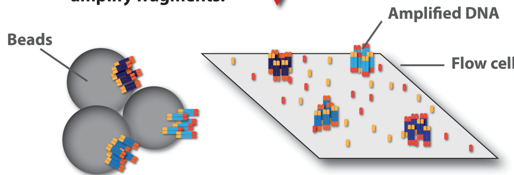
cDNA

Second Generation Sequencing

Shear up cDNA and add appropriate adaptors



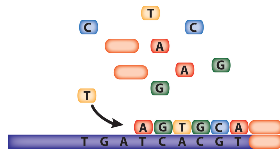
Hybridize fragments to beads or array, then amplify fragments.



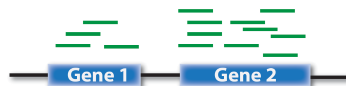
Roche 454
Applied Biosciences SOLiD

Illumina Genome Analyzer/
HiSeq 2000

Fluorescently-labeled nucleotides are incorporated while highly sensitive cameras record the sequence of lights

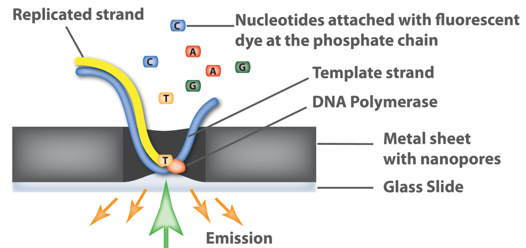


Sequence reads can be mapped back to the genome and quantitated



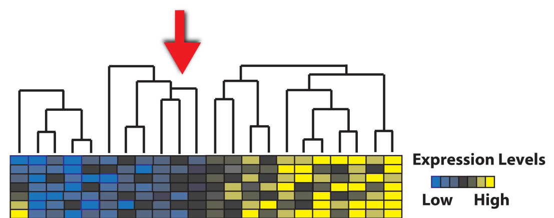
Third Generation Sequencing

Individual strands are placed into nanopores where DNA polymerases form newly synthesized strands with nucleotides attached with dyes



Pacific Biosciences PacBio RS

The real-time sequence of fluorescing lights from the incorporation of fluorescently-labeled nucleotides are detected and recorded. Reads can be used to analyse transcriptome.



Quantitative measurement of expression
Transcriptome discovery
Identification of alternative splice junctions
Transcript variant analysis

Figure A5.3: RNA-Seq with second generation and third generation sequencing technology.

Whole RNA extracts are purified and converted into cDNA. For RNA-Seq using second generation sequencing (SGS) technology, cDNA is sheared into fragments and have appropriate adaptors added. cDNA fragments are added to either DNA capture beads or flowcells and then PCR amplified creating polonies. Using a series of wash-and-scan cycles, fluorescently labeled nucleotides are incorporated or ligated according to the sequence of the template strand. Highly sensitive cameras are able to detect and record which nucleotide was incorporated by the light signal emitted. Sequence reads are then mapped to the genome for analysis. For RNA-seq using third generation sequencing (TGS) (only Pacific Biosciences technology is shown here), whole cDNA strands are individually placed into nanowells, and are replicated with fluorescently labeled nucleotides by DNA polymerases. A different color light signal is emitted for each type of nucleotide as they are incorporated into the replicated strand. The series of light signals are detected and recorded by cameras, and then converted into sequence reads, which are then mapped back to genes for analysis.

APPENDIX A6: Plasmid Maps

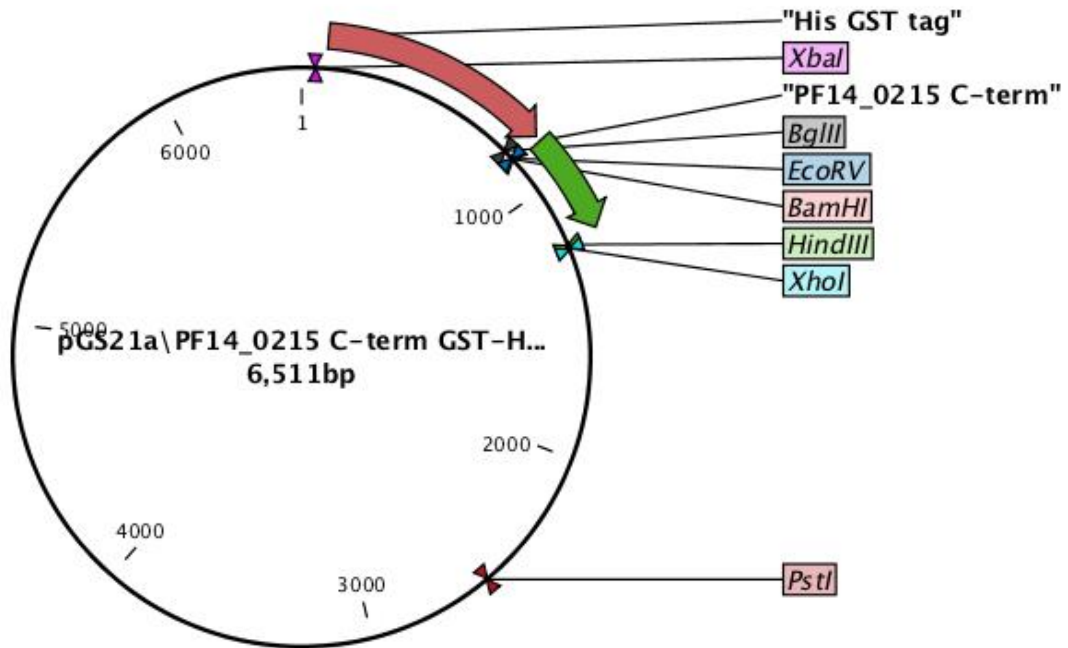


Figure A6.1 - Plasmid used for the recombinant cloning of PF14_0215 (PfHRD1) with GST and 6xHIS

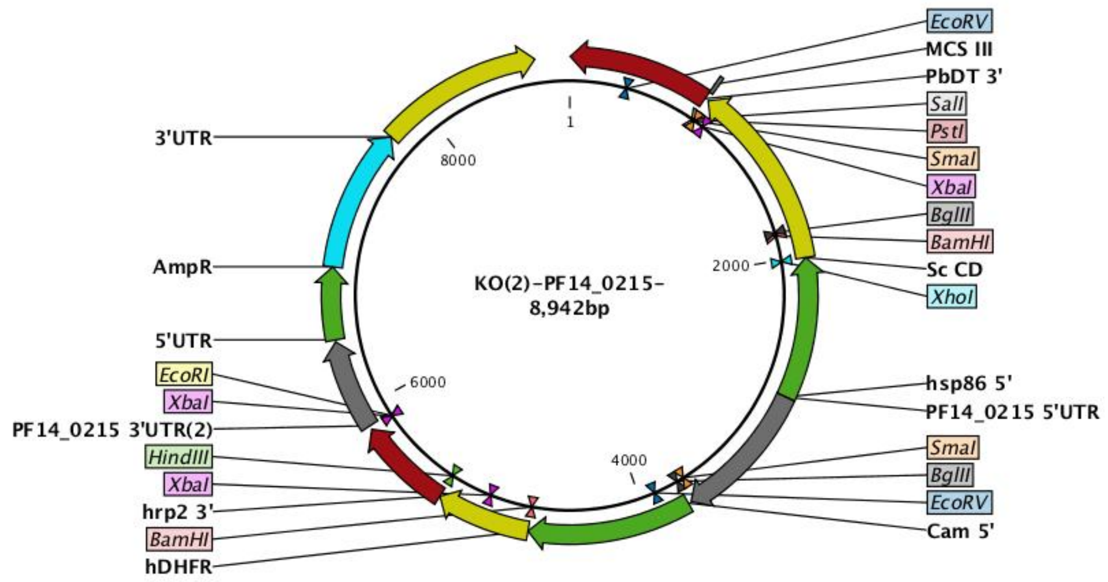


Figure A6.2 - Plasmid used to knockout the PF14_0215 (PfHRD1) gene

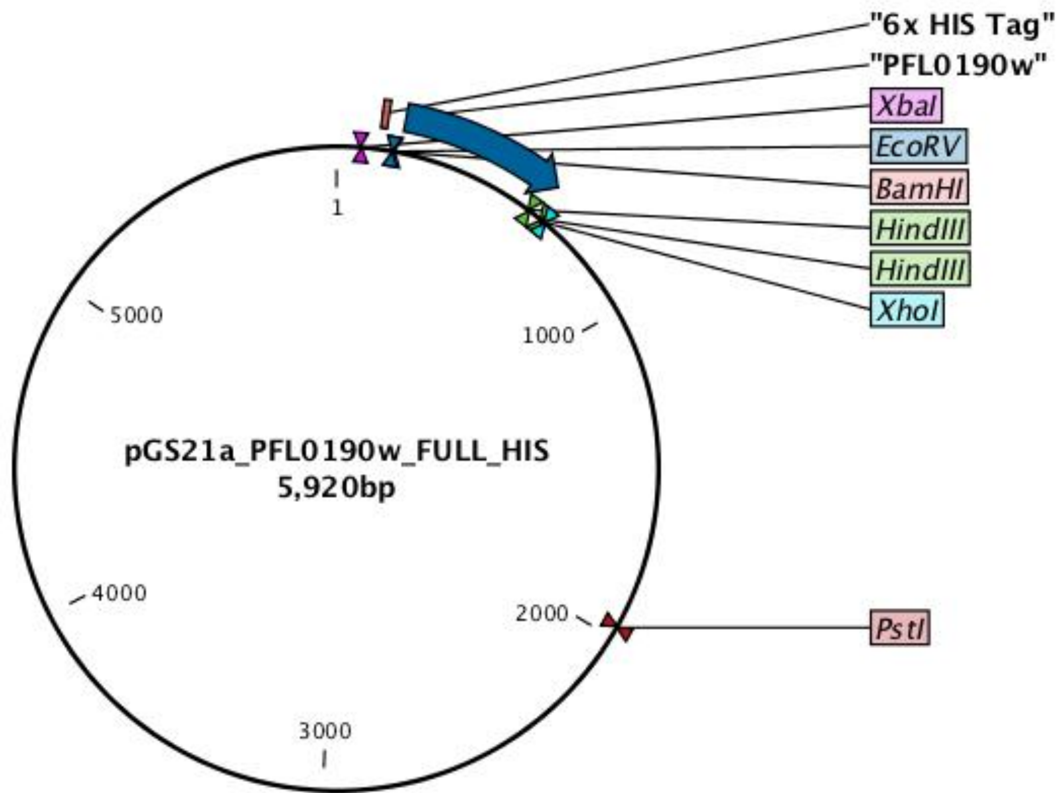


Figure A6.3 - Plasmid used for the recombinant cloning of PFL0190w (PfUBC) with 6xHIS

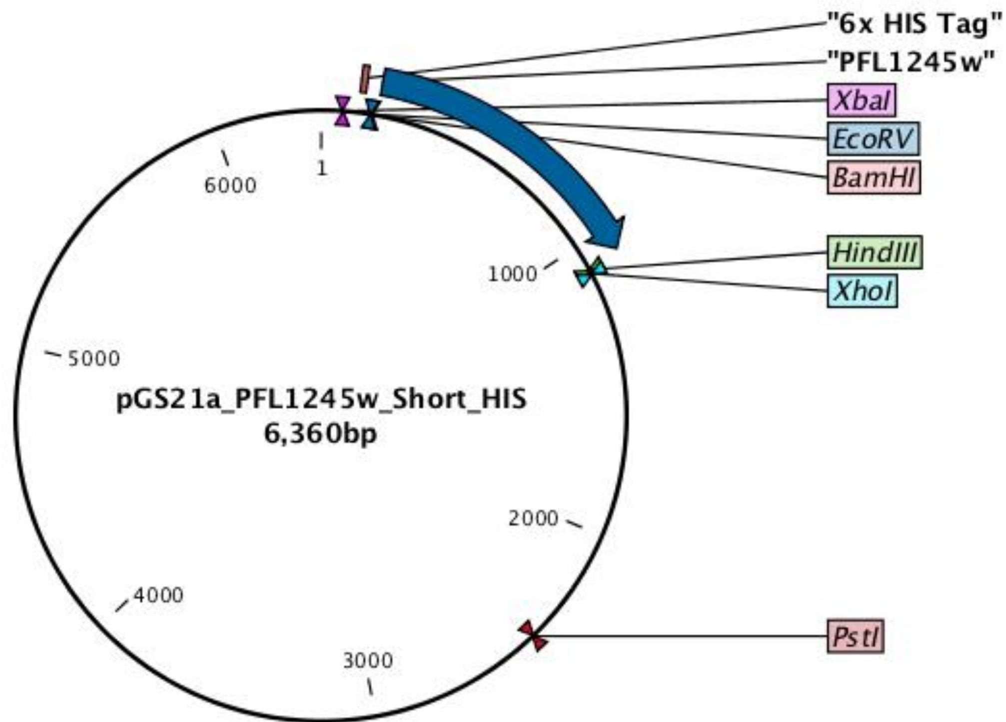


Figure A6.4 - Plasmid used for the recombinant cloning of PFL1245w (PfUBA1) with 6xHIS

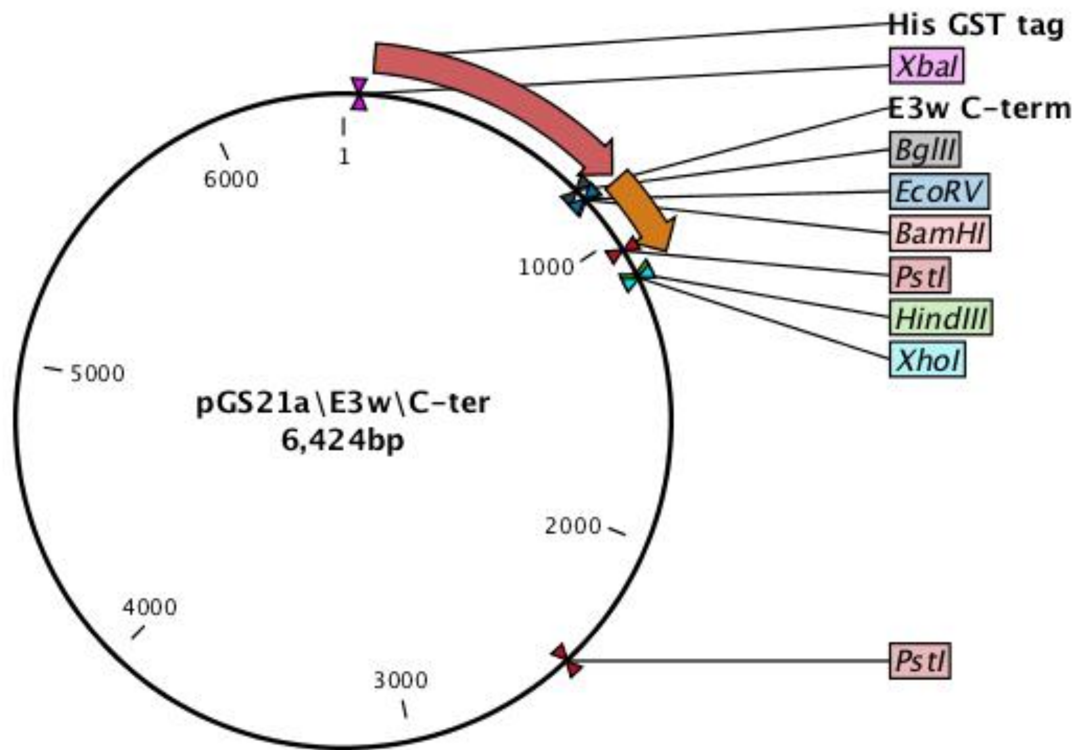


Figure A6.5 - Plasmid used for the recombinant cloning of PFC0510w (PfHRD1p) with GST and 6xHIS

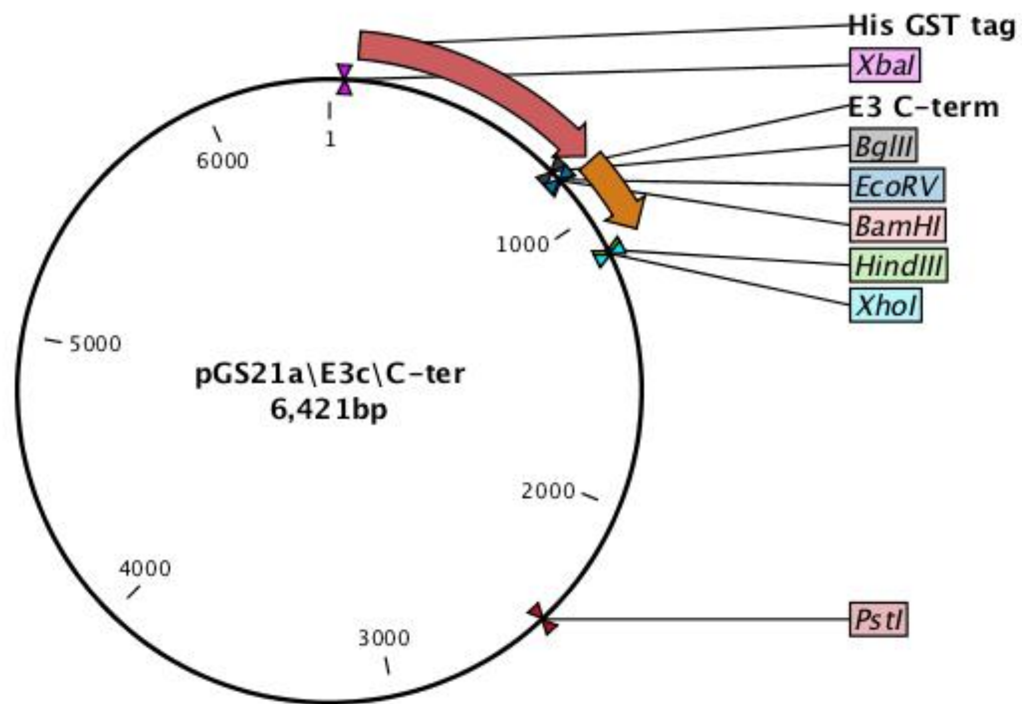


Figure A6.6 - Plasmid used for the recombinant cloning of PFC0740c (PfHRD3p) with GST and 6xHIS

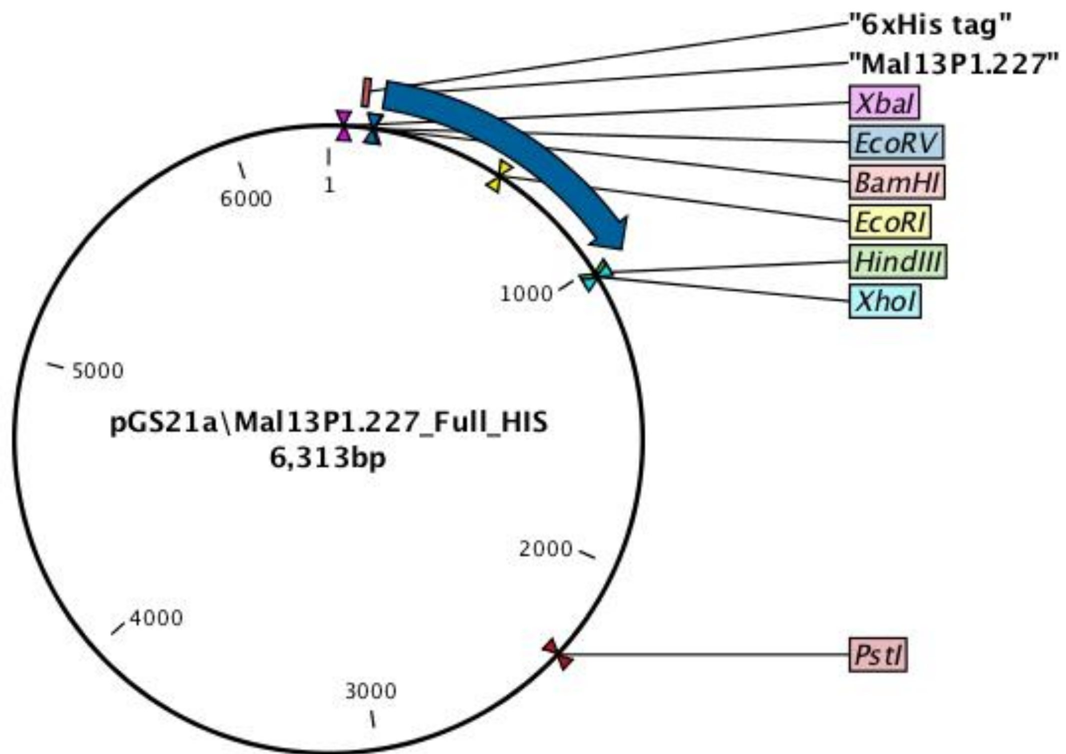


Figure A6.7 -Plasmid used for the recombinant cloning of Mal13P1.227 (PfUBCp) with 6xHIS

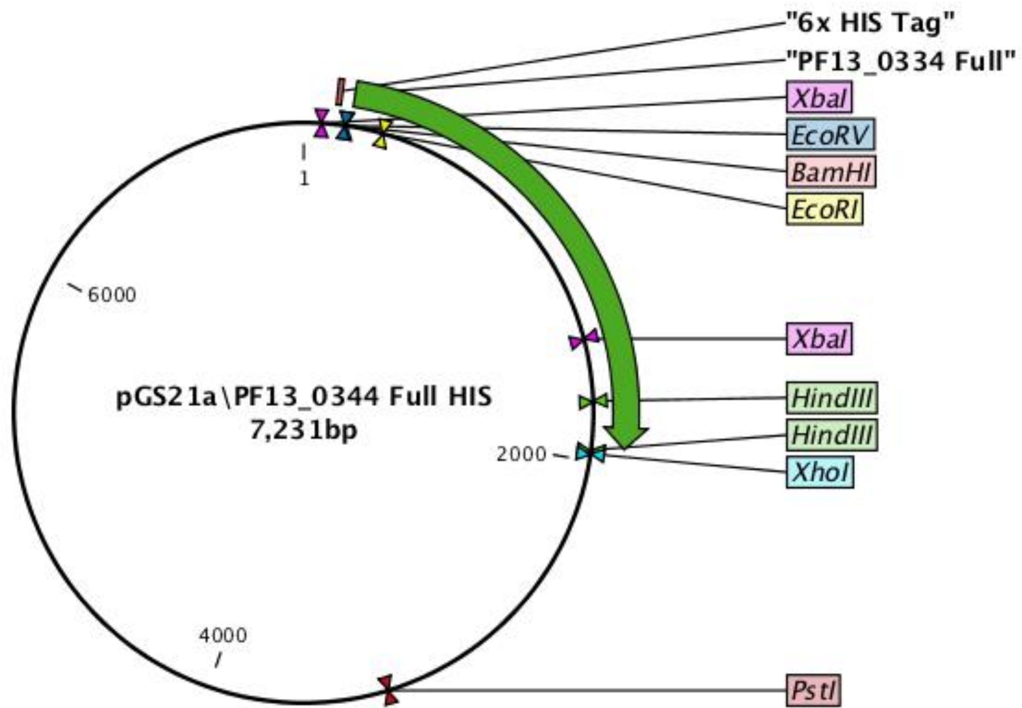


Figure A6.8 - Plasmid used for the recombinant cloning of PF13_0344 (PfUBA2p) with 6xHIS

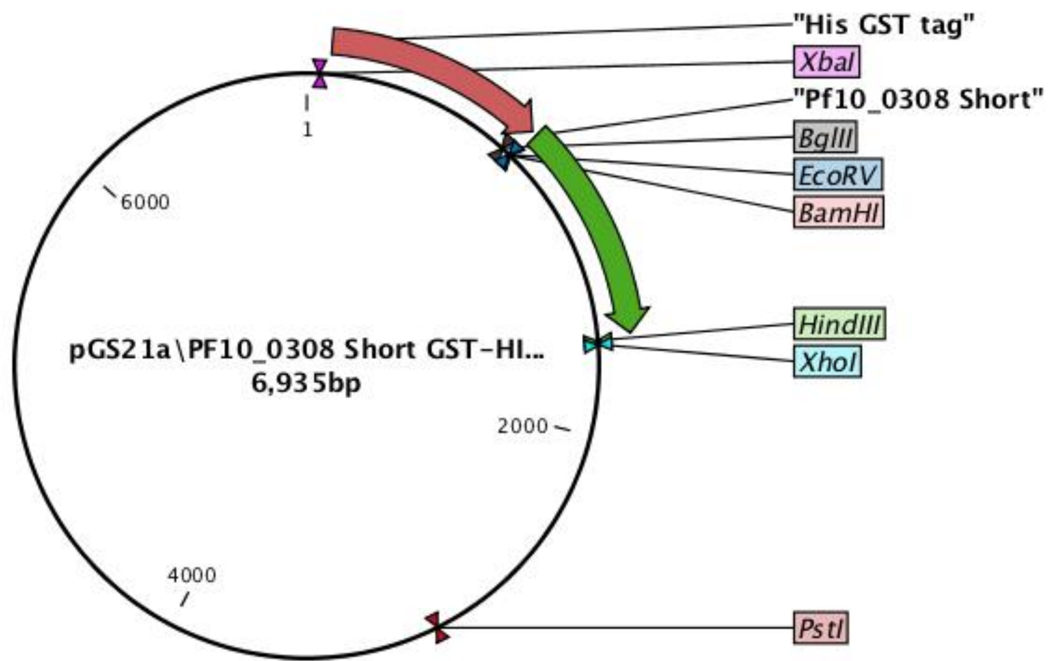


Figure A6.9 - Plasmid used for the recombinant cloning of PF10_0308 (PfOTUp) with 6xHIS

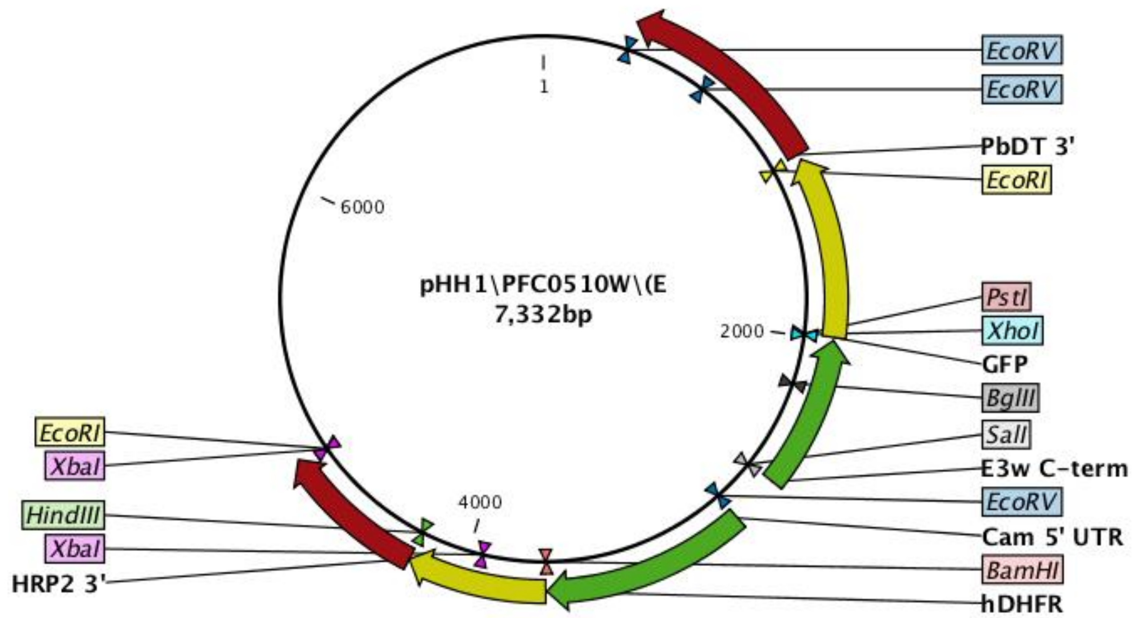


Figure A6.10 - Plasmid used to fuse GFP to the endogenous PFC0510w gene.

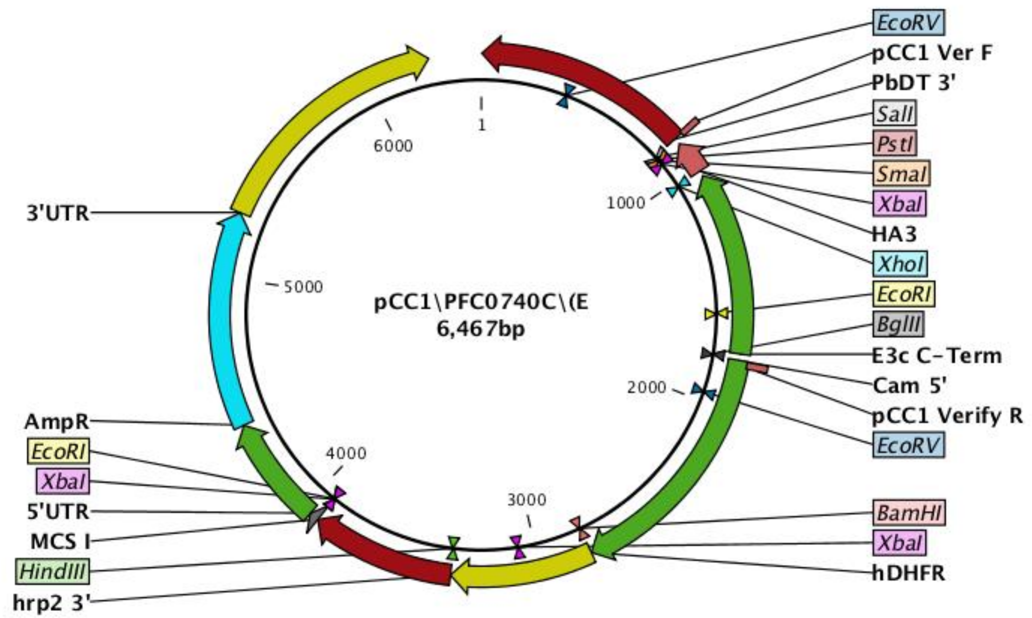


Figure A6.11 - Plasmid used to fuse a 3xHA tag to the endogenous PFC0740c gene.

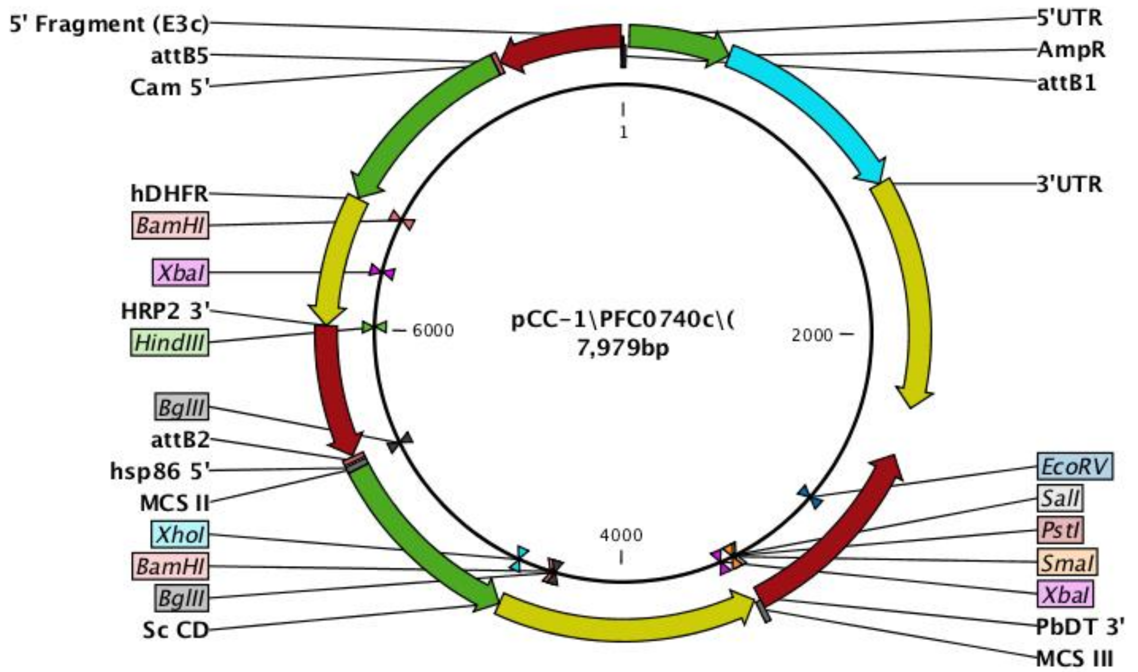


Figure A6.12 - Plasmid used to disrupt the PFC0740c gene.

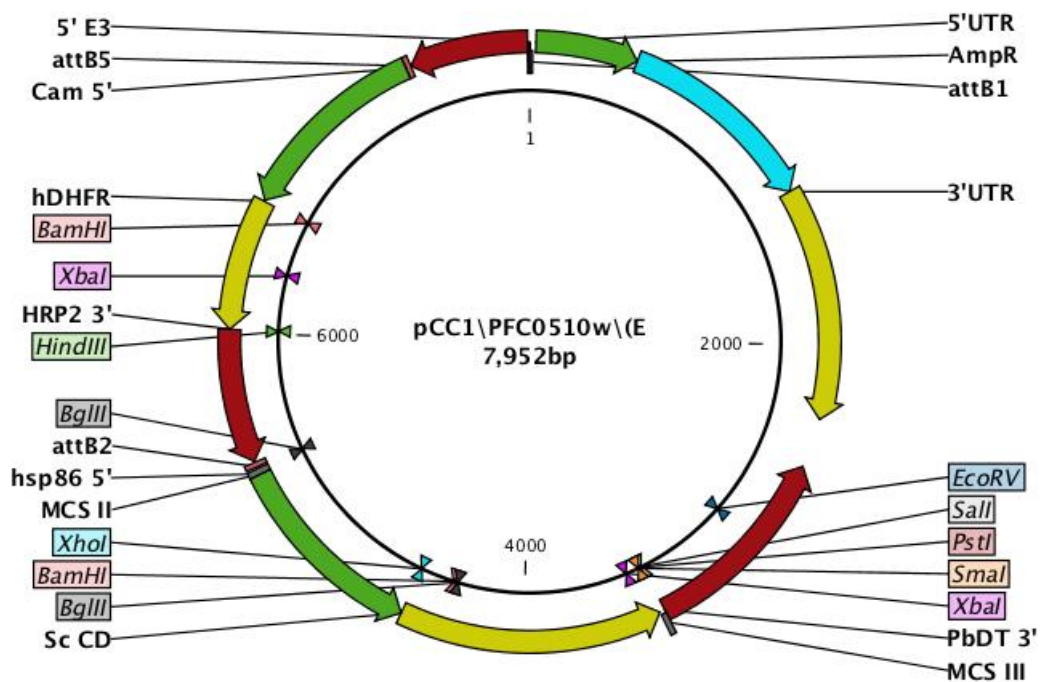


Figure A6.13 - Plasmid used to disrupt the PFC0510w gene.

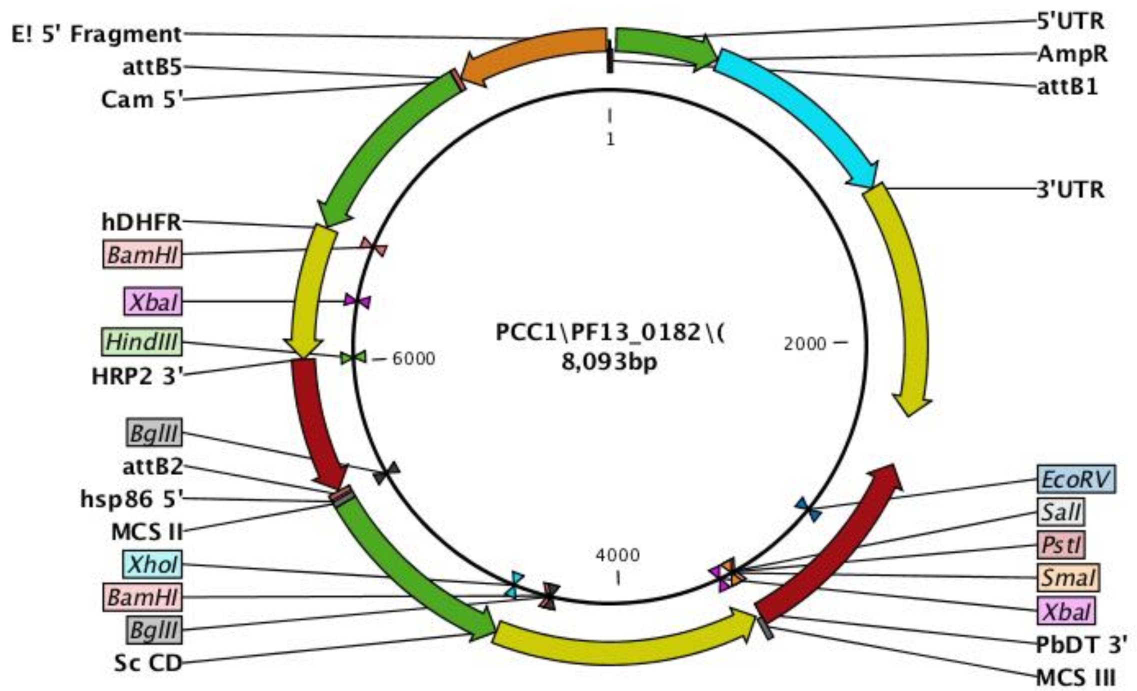


Figure A6.14 - Plasmid used to disrupt the PF13_0182 gene.

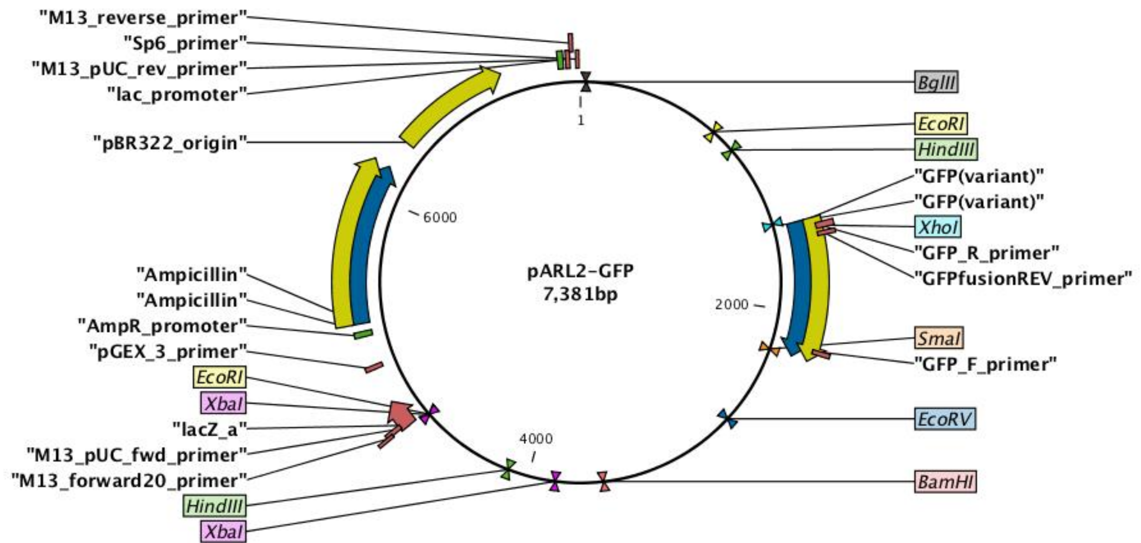


Figure A6.15 - Vector that was given to us by the Przyborski lab for episomal expression of genes tagged with GFP.

APPENDIX A7: List of published papers

Chung D-WD, Ponts N, Cervantes S & Le Roch KG (2009) Post-translational modifications in Plasmodium: more than you think! *Mol. Biochem. Parasitol.* **168:** 123–134

Chung D-WD & Le Roch KG (2010) Targeting the Plasmodium ubiquitin/proteasome system with anti-malarial compounds: promises for the future. *Infect Disord Drug Targets* **10:** 158–164

Ponts N, Saraf A, **Chung D-WD**, Harris A, Prudhomme J, Washburn MP, Florens L & Le Roch KG (2011) Unraveling the ubiquitome of the human malaria parasite. *J. Biol. Chem.* **286:** 40320–40330

Ponts N, Yang J, **Chung D-WD**, Prudhomme J, Girke T, Horrocks P & Le Roch KG (2008) Deciphering the ubiquitin-mediated pathway in apicomplexan parasites: a potential strategy to interfere with parasite virulence. *PLoS ONE* **3:** e2386

Le Roch KG, **Chung D-WD** & Ponts N (2012) Genomics and integrated systems biology in Plasmodium falciparum: a path to malaria control and eradication. *Parasite Immunology* **34:** 50–60

Le Roch KG, Johnson JR, Ahiboh H, **Chung D-WD**, Prudhomme J, Plouffe D, Henson K, Zhou Y, Witola W, Yates JR, Mamoun CB, Winzeler EA & Vial H (2008) A systematic approach to understand the mechanism of action of the bisthiazolium compound T4 on the human malaria parasite, Plasmodium falciparum. *BMC Genomics* **9:** 513

Chung D-WD & Le Roch KG (2012) Genome-Wide Analysis of Gene Expression. Encyclopedia of Biological Chemistry. 2nd Edition. Elsevier Inc. (Book Chapter)

Ponts N, **Chung D-WD** & Le Roch KG (2012) Strand-specific RNA-Seq Applied to Malaria Samples. *RNA Abundance Analysis: Methods and Protocols*. Humana Press. (Book Chapter)